Environ Earth Sci (2012) 66:615–624
DOI 10.1007/s12665-011-1269-4

ORIGINAL ARTICLE

An approach for valid covariance estimation via the Fourier series

Pilar García-Soidán · Raquel Menezes · Óscar Rubiños-López

Received: 12 May 2011 / Accepted: 20 July 2011 / Published online: 20 August 2011
© Springer-Verlag 2011

Abstract The use of kriging for construction of prediction or risk maps requires estimating the dependence structure of the random process, which can be addressed through the approximation of the covariance function. The nonparametric estimators used for the latter aim are not necessarily valid to solve the kriging system, since the positive-definiteness condition of the covariance estimator typically fails. The usage of a parametric covariance instead may be attractive at first because of its simplicity, although it may be affected by misspecification. An alternative is suggested in this paper to obtain a valid covariance from a nonparametric estimator through the Fourier series tool, which involves two issues: estimation of the Fourier coefficients and selection of the truncation point to determine the number of terms in the Fourier expansion. Numerical studies for simulated data have been conducted to illustrate the performance of this approach. In addition, an application to a real environmental data set is included, related to the presence of nitrate in groundwater in Beja District (Portugal), so that pollution maps of the region are generated by solving the kriging equations with the use of the Fourier series estimates of the covariance.

P. García-Soidán (✉)
Department of Statistics and Operations Research,
University of Vigo, Campus de A Xunqueira,
36005 Pontevedra, Spain
e-mail: pgarcia@uvigo.es

R. Menezes
Department of Mathematics and Applications,
University of Minho, Campus de Azurém,
4800-058 Guimarães, Portugal

Ó. Rubiños-López
Department of Signal Theory and Communications,
University of Vigo, Campus de Lagoas-Marcosende,
36310 Vigo, Spain

Keywords Covariance function · Fourier series · Kriging · Truncation point

Introduction

The need to construct a prediction map over the whole observation region from a finite set of data can be found in a broad spectrum of areas, such as geostatistics, hydrology, atmospheric science, etc. The use of kriging for this purpose (Cressie 1993) needs estimation of the dependence structure of the random process, namely, the semivariogram or the covariance function, depending on the type of stationarity assumed from the random process.

Several procedures have been suggested in the literature for estimation of the semivariogram or the covariance function and used for comparison in numerical studies covering different spatial settings (Menezes et al. 2005). In a first step, the nonparametric methods may be applied to approximate the dependence structure, such as the experimental estimator (Matheron 1963), the kernel-type estimator (Hall and Patil 1994) or more robust estimators (Cressie and Hawkins 1980; Genton 1998). Nevertheless, they are not necessarily valid to solve the kriging system and, consequently, give rise to a negative mean-squared prediction error. It is noteworthy that validity requires the semivariogram or the covariance function to respectively satisfy the conditionally negative-definiteness property or the positive-definiteness condition.

The aforementioned problem is typically addressed in practice by choosing a valid parametric family and selecting the parameter estimates which best fit the data. For the latter aim, possible approaches are based on the minimum variance, the maximum likelihood and the least squares criteria (Christakos 1984). In addition, a multi-objective bilevel

programming method has been developed for better approximation of the parameters (Huang and Hu 2009), through a cross-validation procedure.

The parametric approach has been applied to describe the correlation structure of different spatial variables, such as soil moisture (Romshoo 2003) or the hydraulic conductivities (Chen 2005). The use of the parametric estimator and the linear kriging methods, together with other techniques (regression, sequential simulation, etc.), allow the prediction of variables at unsampled locations, which have been employed to quantify the trace metal content (Satapathy et al. 2009) or the water reservoir capacity (Rakhmatullaev et al. 2011), among other variables. Furthermore, assessment of contamination risk can be addressed by using the probability kriging or the indicator kriging (LaMotte and Green 2007; Lin et al. 2011), enabling the researcher to determine those locations with low, medium or high probability of exceeding a predetermined value of the variable involved. In the referred applications, the dependence structure was approximated through the parametric approach by departing from the experimental estimator. However, when outliers are present in data, an appropriate characterization of the spatial correlation and also an accurate prediction require employing robust geostatistical approaches (Lark 2000; Zhao et al. 2007).

The simplicity and validity of the parametric estimator, widely used in practice as presented above, make it attractive at first, although one of its main drawbacks is related to the criteria followed to select the parametric model. The latter is usually chosen by eye, from the basis of a nonparametric estimator previously obtained, and the graphical diagnostic employed for checking its validity is often difficult to assess, since the shape of some parametric models is very similar.

To avoid selection of a parametric model, an approach was proposed (Hall et al. 1994), based on first truncating a nonparametric estimator and then Fourier-inverting it to produce a valid one, although selection of the truncation term is an open issue.

This research is focused on the covariance function estimation, which directly provides a semivariogram estimator; in addition, a similar approach could be derived for the semivariogram. In the current paper, a valid covariance is obtained from a nonparametric estimator, through the Fourier series tool, without requiring the choice of a parametric model. This technique has been applied to other statistical problems, such as those concerning the density or the regression estimation (Tarter and Lock 1993; Efromovich 1999; Eubank 1988). The underlying idea is based on selecting a nonparametric estimator and approximating it by a finite expansion, which provides a valid estimator of the unknown covariance function. This approach involves two issues that will be addressed in this paper: specification of the truncation point and estimation of the Fourier coefficients in the expansion. To ensure that the resulting estimator is positive-definite, only those terms in the approximation corresponding to positive Fourier coefficients will be included.

The approximation of the Fourier coefficients in the expansion requires a prior nonparametric covariance, which will be called the pilot estimator. This may be either supplied to carry out a specific study or selected from the different alternatives existing in the geostatistics literature. In this respect, the Fourier series approach may be viewed as a procedure for transformation of a given covariance estimator into a valid one.

The choice of a smoothing parameter is necessary to specify the number of terms to be used in the expansion, which will be referred to as the cutoff or the truncation point. Different methods have been proposed for the latter selection with the aim of overcoming inconsistency of the resulting estimators (Hart 1985; Diggle and Hall 1986). In the current study, an explicit procedure for choice of the truncation point is provided, based on the minimization of the corresponding mean integrated squared error (MISE).

This paper is organized as follows. The main results are developed in Sect. 2 and their implementation in practice is given in Sect. 3, where a kernel-type estimator is taken as the pilot covariance. Section 4 describes the numerical studies conducted for simulated data to illustrate the performance of this approach. In addition, an application to the construction of pollution maps is included in Sect. 5, from a real data set measuring the nitrate in groundwater in Beja District (Portugal). Finally, the main conclusions of the current work are summarized in Sect. 6.

Main results

Let $\{Z(s) : s \in D \subset \mathbb{R}^2\}$ be a random process, where D is a bounded observation region. As discussed in Sect. 1, the random process will be assumed to satisfy second-order stationarity, with covariance function C, so that:

(i) $E[Z(s)] = \mu$, for all $s \in D$ and some $\mu \in \mathbb{R}$.
(ii) $\mathrm{Cov}[Z(s), Z(s')] = C(s - s')$, for all $s, s' \in D$.

Suppose that n data, $Z(s_1), \ldots, Z(s_n)$, are collected, at known spatial locations, $s_1, \ldots, s_n \in D$. From condition (ii), estimation of $C(t)$ will be addressed, with $t \in E = \{s - s' : s, s' \in D\}$ rather that $t \in D$. For the sake of simplicity, E will be assumed to be a bounded rectangle and, more precisely, $E = [0, e_1] \times [0, e_2]$. The latter is not restrictive in practice, since D is a bounded region and, therefore, either E is a bounded rectangle or it is strictly contained

within it. Consequently, the estimation of the covariance function in a bounded rectangle includes that in E.

The idea behind the Fourier series approach will be summarized in Appendix 1 and applies for the covariance function C. This theory yields that there exists a set of functions $\{\psi_i : E \longrightarrow I\!R : i \in I\!N\}$, defined in (10), so that $C(t)$ can be approximated by the Fourier expansion:

$$C_m(t) = \sum_{i=0}^{m} \theta_{C,i}\psi_i(t), \quad \text{for all } t \in E \tag{1}$$

where m is referred to as the cutoff or the truncation point and $\theta_{C,i} = \int_E C(t)\psi_i(t)\,\mathrm{d}t$ is the ith Fourier coefficient of C.

From the foregoing definition of C_m, it is clear that estimating the covariance function through the Fourier series approach requires:

- Approximating coefficients $\theta_{C,i}$, dependent on the theoretical covariance function.
- Selecting the cutoff m, which will specify the number of terms in the expansion.

The first issue can be addressed by appropriately choosing a pilot estimator $\hat{C}$ of the theoretical covariance C. This estimator must satisfy the following constraints:

$$\sup_{t\in E} |\mathrm{Bias}[\hat{C}(t)]| \overset{n\to\infty}{\longrightarrow} 0 \quad \text{and} \quad \sup_{t\in E} \mathrm{Var}[\hat{C}(t)] \overset{n\to\infty}{\longrightarrow} 0 \tag{2}$$

The latter will guarantee that $\theta_{\hat{C},i} = \int_E \hat{C}(t)\psi_i(t)\,\mathrm{d}t$ provides a reliable approximation of the theoretical coefficient $\theta_{C,i}$, as proved in Appendix 2, where its bias and variance are derived.

In addition, replacing $\theta_{C,i}$ by $\theta_{\hat{C},i}$ in (1), one obtains:

$$\hat{C}_{1,m}(t) = \sum_{i\le m} \theta_{\hat{C},i}\psi_i(t) \tag{3}$$

which gives a consistent approximation of $C(t)$, whose MISE tends to zero and is developed in Appendix 3.

A further step in this research is the specification of the cutoff m. With this aim, observe from relation (13) that $\mathrm{MISE}[\hat{C}_{1,m}, C] = M_1(m) + \int_B C(t)^2\,\mathrm{d}t$, with:

$$M_1(m) = \sum_{i\le m} (\mathrm{Var}[\theta_{\hat{C},i}] + \mathrm{Bias}[\theta_{\hat{C},i}]^2 - \theta_{C,i}^2) \tag{4}$$

To choose the truncation point m, an idea may be that of minimizing $\mathrm{MISE}[\hat{C}_{1,m}, C]$ or, equivalently, $M_1(m)$, since $\int_E C(t)^2\,\mathrm{d}t$ is a constant value. Nevertheless, all the terms appearing in $M_1(m)$ are unknown and must be estimated. For this purpose, the pilot covariance $\hat{C}$ can be again used, yielding an estimator $\hat{M}_1(m)$ of $M_1(m)$. In particular, $\theta_{C,i}^2$ can be estimated by $\theta_{\hat{C},i}^2$. Hence the cutoff could be taken as the minimizer of $\hat{M}_1$, namely:

$$\hat{m} = \underset{0\le m}{\mathrm{argmin}}\, \hat{M}_1(m)$$

The resulting covariance estimator (3) has some of the properties of the Fourier basis, such as the degree of smoothness. However, it is not necessarily positive-definite, conveying that it cannot be directly used for spatial prediction, as remarked in Sect. 1. This problem can be solved by imposing an additional requirement on the proposed approach, namely, by selecting those terms in the expansion with positive coefficients $\theta_{\hat{C},i}$.

This leads to the following approximating function:

$$\hat{C}_{2,m}(t) = \sum_{i\le m} w_i\theta_{\hat{C},i}\psi_i(t) \tag{5}$$

with $w_i = I_{\{\theta_{\hat{C},i} > 0\}}$. The positive-definiteness of (5) can be easily derived, as shown in Appendix 4.

The truncation point m, necessary for implementation of estimator (5), must be recalculated to produce an accurate approximation, as some terms originally appearing in (3) may now not be included. The optimal cutoff can be again obtained by minimizing the MISE of $\hat{C}_{2,m}$, which equals $\mathrm{MISE}[\hat{C}_{2,m}, C] = M_2(m) + \int_B C(t)^2\,\mathrm{d}t$, as given in (15), with:

$$M_2(m) = \sum_{i\le m} w_i(\mathrm{Var}[\theta_{\hat{C},i}] + \mathrm{Bias}[\theta_{\hat{C},i}]^2 - \theta_{C,i}^2) \tag{6}$$

Minimization of $M_2(m)$, equivalent to that of $\mathrm{MISE}[\hat{C}_{2,m}, C]$, again provides a key idea for selection of the cutoff, where the unknown terms can be appropriately estimated through the pilot covariance. This leads to a consistent estimator $\hat{M}_2(m)$, so that the truncation point m would be given by:

$$\hat{m} = \underset{0\le m}{\mathrm{argmin}}\, \hat{M}_2(m)$$

Remark 1 A pilot covariance estimator $\hat{C}$ is needed to approximate the Fourier coefficients as well as the truncation point; in the latter case, it is used for estimation of the unknown terms in the objective function. Therefore, this approach may be viewed as a tool for transformation of a covariance estimator $\hat{C}$ into a valid one, where simple requirements to guarantee consistency of the resulting estimator are imposed, such as those established in (2).

Remark 2 An estimator of the semivariogram γ could be easily derived from the relationship between γ and C, namely, $\gamma(t) = C(0) - C(t)$. In fact, the use of (3) or (5) for approximation of C in the latter relation would provide a semivariogram estimator, which would satisfy the conditionally negative-definiteness property in the second case, on account of the validity of (5).

On the other hand, the approach proposed in this section for estimation of the covariance function, through the

Fourier series tool, could be adapted for direct approximation of the semivariogram. In this respect, the analog of expansions (3) or (5) could be obtained for the semivariogram, by previously selecting a pilot estimator; furthermore, validity in the second case would be achieved with the same Fourier basis, provided that the finite expansion solely involved negative Fourier coefficients instead of positive ones.

Practical implementation

The approach introduced in Sect. 2 requires the selection of a pilot covariance estimator $\hat{C}$ satisfying (2), which is necessary for:

- Approximation of the Fourier coefficients, used in the construction of $\hat{C}_{1,m}$ and $\hat{C}_{2,m}$ given in (3) and (5), respectively.
- Estimation of the unknown terms in the objective functions M_1 and M_2, defined in (4) and (6). Minimization of each estimated objective function $\hat{M}_i$ leads to the optimal cutoff of $\hat{C}_{i,m}$, for $i = 1, 2$.

With the above purpose, the kernel covariance estimator can be taken as the pilot covariance, which is given by:

$$\hat{C}_h(t) = \frac{\sum_{j,k} K\left(\frac{t-(s_j-s_k)}{h}\right)(Z(s_j)-\bar{Z})(Z(s_k)-\bar{Z})}{\sum_{j,k} K\left(\frac{t-(s_j-s_k)}{h}\right)} (t \in \mathbb{R}^2) \tag{7}$$

where $\bar{Z} = n^{-1}\sum_{j=1}^{n} Z(s_j)$, K denotes a d-variate kernel function and $h = h_n$ is the bandwidth parameter, where h is assumed to be smaller as the sample size n increases.

The asymptotic properties of $\hat{C}_h(t)$ have been established, as well as the dominant terms of the bias and the variance of the kernel covariance estimator (Hall and Patil 1994; Hall et al. 1994). The latter results allow concluding that condition (2) is satisfied, under several assumptions related to the increasing rate of the observation region, the decreasing rate of the bandwidth h and some properties concerning the moments of the random process. In addition, one has that $\hat{B}_h(t)$ and $I_{\{t=t'\}}\hat{V}_h(t)$ provide adequate estimators of $\mathrm{Bias}[\hat{C}_h(t)]$ and $\mathrm{Cov}[\hat{C}_h(t), \hat{C}_h(t')]$, where:

$$\hat{B}_h(t) = \frac{\sum_{j,k=1}^{n} K\left(\frac{t-(s_j-s_k)}{h}\right)X(s_j,s_k)}{\sum_{j,k=1}^{n} K\left(\frac{t-(s_j-s_k)}{h}\right)}$$

$$\hat{V}_h(t) = \frac{\sum_{j,k=1}^{n} K\left(\frac{t-(s_j-s_k)}{h}\right)^2 X(s_j,s_k)^2}{\left(\sum_{j,k=1}^{n} K\left(\frac{t-(s_j-s_k)}{h}\right)\right)^2}$$

with $X(s,s') = \hat{C}_h(s-s') - (Z(s)-\bar{Z})(Z(s')-\bar{Z})$.

Remark 3 If the aim were the semivariogram estimation, similar properties as described above for the bias and variance for the kernel covariance would hold for the kernel-type semivariogram (García-Soidán 2007). Furthermore, an adaptation of the kernel estimator has been proposed for clustered data (Menezes et al. 2008).

Using the kernel-type covariance $\hat{C}_h$, parameter $\theta_{C,i}$ can be estimated by its counterpart $\theta_{\hat{C}_h,i} = \int_E \hat{C}_h(t)\psi_i(t)\,\mathrm{d}t$. From the properties of $\hat{C}_h$, the bias and variance of $\theta_{\hat{C}_h,i}$, respectively, derived from (11) and (12), may be approximated by $B_{n,i} = \int_E \hat{B}_h(t)\psi_i(t)\,\mathrm{d}t$ and $V_{n,i} = \int_E \hat{V}_h(t)\psi_i(t)^2\,\mathrm{d}t$. These estimators enable the approximation of the unknown terms in functions M_1 and M_2 yielding:

$$\hat{M}_1(m) = \sum_{i\le m}(V_{n,i} + B_{n,i}^2 - \theta_{\hat{C}_h,i}^2)$$

$$\hat{M}_2(m) = \sum_{i\le m} w_i(V_{n,i} + B_{n,i}^2 - \theta_{\hat{C}_h,i}^2)$$

with $w_i = I_{\{\theta_{\hat{C}_h,i} > 0\}}$, which can be used for an optimal choice of the cutoff m in the Fourier expansions $\hat{C}_{1,m}$ and $\hat{C}_{2,m}$, respectively.

Remark 4 The kernel-type estimator is a weighted average, which conveys the use of all the data. The weight associated to each term $(Z(s_j)-\bar{Z})(Z(s_k)-\bar{Z})$ in $\hat{C}_h(t)$ does not depend on the values of the variable involved, but on the locations and, more specifically, on the difference $t-(s_j-s_k)$. The latter means that the effect of the outliers on the kernel covariance can be substantial and, therefore, it should be replaced by a more robust estimator in such a situation (Zhao et al. 2007). An alternative way to proceed, when outliers are present in data, could be that of diminishing their impact on $\hat{C}_h$, which can be accomplished in the following manner. On one hand, the detected outliers could be removed from the data to compute the kernel covariance, although the whole observed values should be considered for prediction to allow the outliers to have an appropriate influence on the nearby locations. A second option would be that of truncating $\hat{C}_h(t)$, so that a percentage of pairs would be included in the weighted average, instead of all of them, by eliminating those (j,k) corresponding to the extreme values of $(Z(s_j)-\bar{Z})(Z(s_k)-\bar{Z})$. Proceeding in either of these ways with the kernel covariance, the practical implementation developed in this section would follow for the truncated kernel estimator by also restricting $\hat{B}_h$ and $\hat{V}_h$ to the same set of pairs as $\hat{C}_h$.

Numerical studies with simulated data

This section summarizes the results of several studies conducted with simulated data to illustrate the performance of the Fourier series estimator, where the kernel covariance is the pilot estimator, as described in the previous section.

Firstly, stationary Gaussian processes on $D \subset \mathbb{R}^2$ were simulated, with $D = [0, 5] \times [0, 5]$. The uniform density on D was used for random generation of the spatial locations s_i, for $i = 1, \ldots, n$ and $n = 50, 100$. With these locations, the data $Z(s_i)$ were obtained from Gaussian processes with zero mean and anisotropic exponential covariance:

$$C_{d_1,d_2,r}(x_1, x_2) = C_{d_1,d_2}\left(\sqrt{x_1^2 + r x_2^2}\right) \tag{8}$$

where r denotes the anisotropy ratio and C_{d_1,d_2} is the isotropic exponential model with variance d_1 and effective range d_2 defined as follows:

$$C_{d_1,d_2}(x) = d_1 \exp\left(-\frac{3x}{d_2}\right)$$

More specifically, the data were simulated from the anisotropic covariance function above with $(r, d_1, d_2) = (0.25,\ 1,\ 3)$.

The aim of this numerical study was the estimation of the covariance function at t, with $t \in E = [0, 3] \times [0, 3]$, by considering different alternatives. To start, the kernel covariance $\hat{C}_h$ in (7) was computed by using $h = 0.4$ and the kernel $K_{2,\mathrm{ep}}(x_1, x_2) = \prod_{i=1}^{2} K_{\mathrm{ep}}(x_i)$, with K_{ep} the univariate Epanechnikov kernel, namely $K_{\mathrm{ep}}(x) = 0.75\,(1 - x^2) I_{\{|x|<1\}}$. Then, estimates of $\hat{C}_{2,\mathrm{m}}$ were obtained by using a bidimensional orthonormal basis on E, constructed from the unidimensional cosine system on $[0, 3]$, as described in Appendix 1.

Parametric estimators of the covariance function were also obtained by selecting valid covariance families and deriving maximum likelihood estimates to approximate the unknown parameters. The exponential and the spherical models were used for the latter purpose, thus enabling a comparison between the orthonormal series covariance and two parametric estimators: one provided by the theoretical model, which is an advantageous candidate, and the other affected by misspecification. The isotropic spherical model is defined as:

$$C_{d_1,d_2}(x) = \begin{cases} d_1\left(1 - \frac{3x}{2d_2} + \frac{x^3}{2d_2^3}\right), & \text{if } x \le d_2 \\ 0, & \text{if } x > d_2 \end{cases}$$

and the anisotropic spherical covariance was obtained by taking (8) with C_{d_1,d_2} above.

For each data set, the integrated quadratic error (ISE), defined as $\int_E (C(t) - \bar{C}(t))^2\,\mathrm{d}t$, has been approximated numerically for each of the estimators $\bar{C}$ implemented, including the nonvalid kernel covariance, as the Fourier series estimator requires it. Proceeding in this way with 100 samples of size n, the mean and the standard deviation values of the ISE were computed, whose results are summarized in Table 1.

Table 1 Mean and standard deviation (St) values of the ISE

Sample size	Orthonormal		Maximum likelihood				Kernel	
	Series		Exponential		Spherical		$\hat{C}_{h_n}$	
	Mean	St	Mean	St	Mean	St	Mean	St
$n = 50$	0.654	0.046	0.553	0.525	1.026	1.020	0.989	0.393
$n = 100$	0.651	0.097	0.451	0.202	2.655	1.880	0.856	0.420

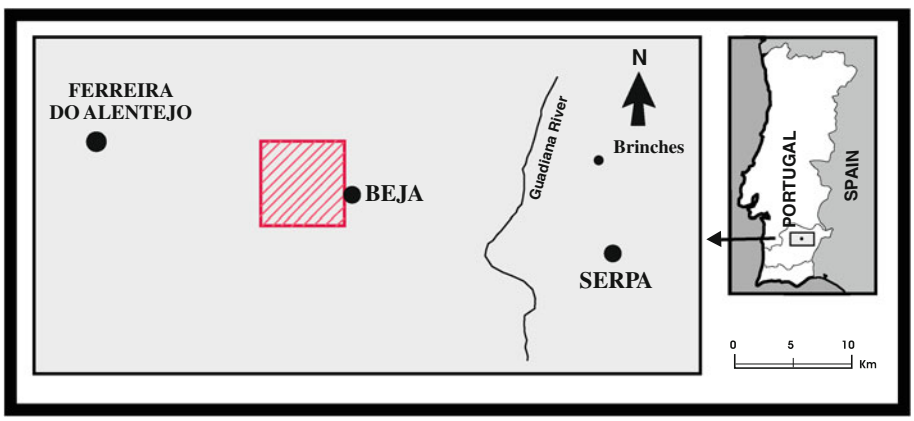

Fig. 1 Geographic context associated with the aquifer system of Gabros de Beja (350 km^2) and the area studied (50 km^2)

In view of the mean estimates appearing in Table 1, performance of estimator (5) improves the one shown by both the parametric estimator obtained from a wrong model and the kernel covariance itself, used as the pilot estimator. The same conclusion, related to the better behavior of the Fourier series approach, remains valid when considering the standard deviation values and can be even extended to the parametric covariance following the theoretical model, which does not produce the expected smaller dispersion.

Table 2 Summary statistics for nitrate concentration levels in groundwater measured in 1998 and 2000 in Beja District

Statistic	Untransformed		Log-transformed	
	1998	2000	1998	2000
Number of locations	50	69	50	69
Mean	72.74	86.43	4.15	4.36
Standard deviation	37.00	31.86	0.56	0.53
Minimum	14	10	2.64	2.30
Maximum	190	162	5.25	5.09

Application to groundwater quality data

The following study describes an application to real data of the approach suggested in Sect. 2, which will exemplify its practical use. It involves groundwater quality data collected in Beja District (in the south of Portugal) in 1998 and 2000. The observation region $D \subset I\!R^2$ forms an approximated 50 km^2 area, as illustrated in Fig. 1.

Measurements of nitrate were taken, as this chemical element is quite related to the agricultural activity, which is of great importance in this area. On the other hand, Beja District is part of one of the driest regions of Portugal, thus making the quality of water a very important issue. The nitrate concentration levels depends on the season (Paralta and Ribeiro 2003), and the current study was restricted to the same month (July) in these 2 years. The data analysis was carried out to construct and compare maps of nitrate concentration in 1998 and 2000, under distinct methods of covariance estimation.

The measured nitrate concentration included three gross outliers in 1998 and also in 2000, which were replaced by the averages of the remaining values of the corresponding year's survey. Table 2 gives the summary statistics for the

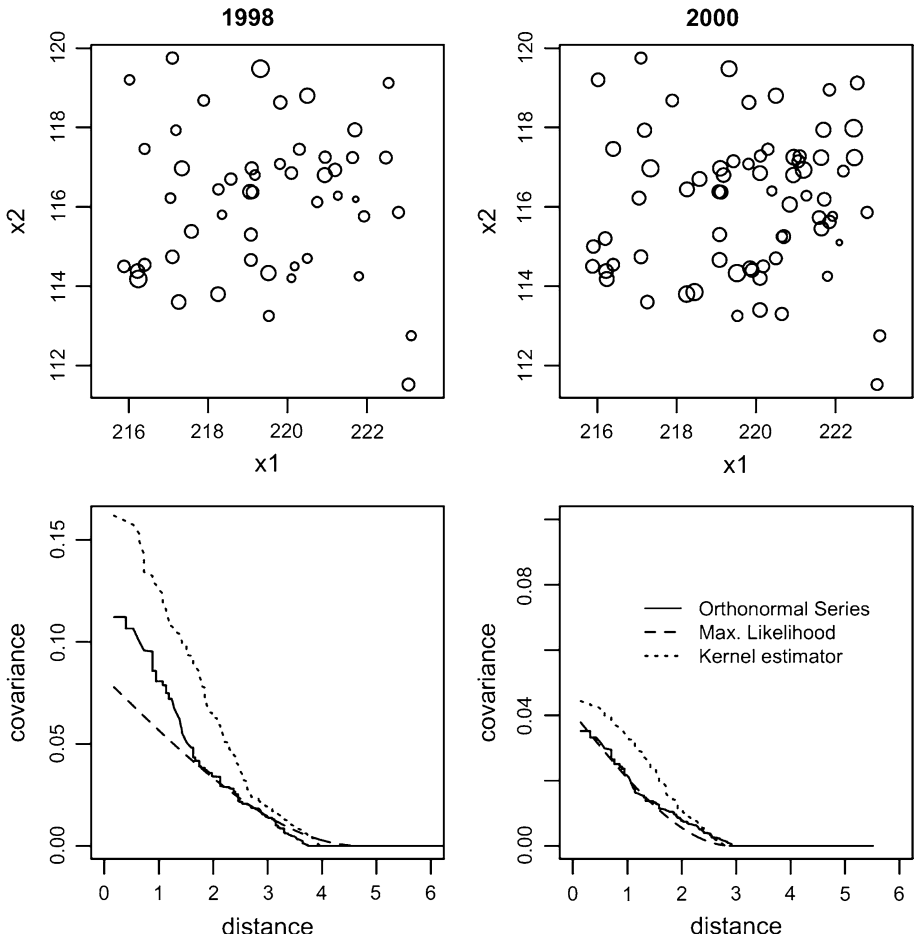

Fig. 2 *Top panels* give the sample locations for 1998 and 2000 in the Beja data. Note that the size of the bullets is proportional to the value measured of nitrate. *Bottom panels* present the likelihood estimator (*dash line*), the kernel covariance (*dot line*) and its valid Fourier series estimator (*full line*), obtained for nitrate concentration data in Beja District in the northeast–southeast direction. The unit of distance is 1 km

data collected in 1998 and 2000. Note that the mean response is higher for the 2000 than for 1998 data, which would be consistent with an overall increment in levels of pollution over the 2 years between both surveys. Also, the log-transformation eliminates an apparent variance–mean relationship in the data and leads to more symmetric distributions of measured values.

Next, as the construction of pollution maps of the region requires the estimation of the spatial dependency, covariance estimates were obtained for the 2 years. Firstly, the kernel estimator $\hat{C}_h$ in (7) was computed with a bandwidth h equal to 1.2 and 1.0 for 1998 and 2000, respectively. Note that the smaller number of sample points in the first survey (see top panels of Fig. 2), more spread over the observation region, leads to a larger value of the bandwidth.

Secondly, the Fourier series approach was applied by using the bidimensional system on $[0, 5] \times [0, 5]$ and $[0, 4] \times [0, 4]$, constructed from the unidimensional cosine basis on [0,5] and [0,4], in 1998 and 2000, respectively. To obtain the cutoff m, specifying the number of terms in the expansion, the bias and variance of Fourier coefficients were estimated, as described in Sect. 3. The corresponding integrals were approximated by using the trapezoid rule for areas in the plane, and the optimal cutoff was acquired for $m = 17$. In addition, the maximum likelihood estimator was obtained when adopting the spherical model (Paralta and Ribeiro 2003).

The top panels of Fig. 2 show the locations of the monitoring stations for each year, where the distance unit is 1 km. The size of each bullet is proportional to the value found in each location, so that larger bullets identify larger values of nitrate concentration. The bottom panels compare the resulting covariance estimates, when applied to the groundwater quality data. Only the values achieved in the northeast–southeast direction are shown, since similar results were obtained in other distinct directions. The largest discrepancy between the likelihood approach and the Fourier series estimator is found in 1998, showing larger estimates for the nugget effect (possibly some measurement error) and for the total variance of the spatial process. Anyway, in both years, the results of the orthonormal series estimator are closer to those from the likelihood one, when compared with the kernel estimator.

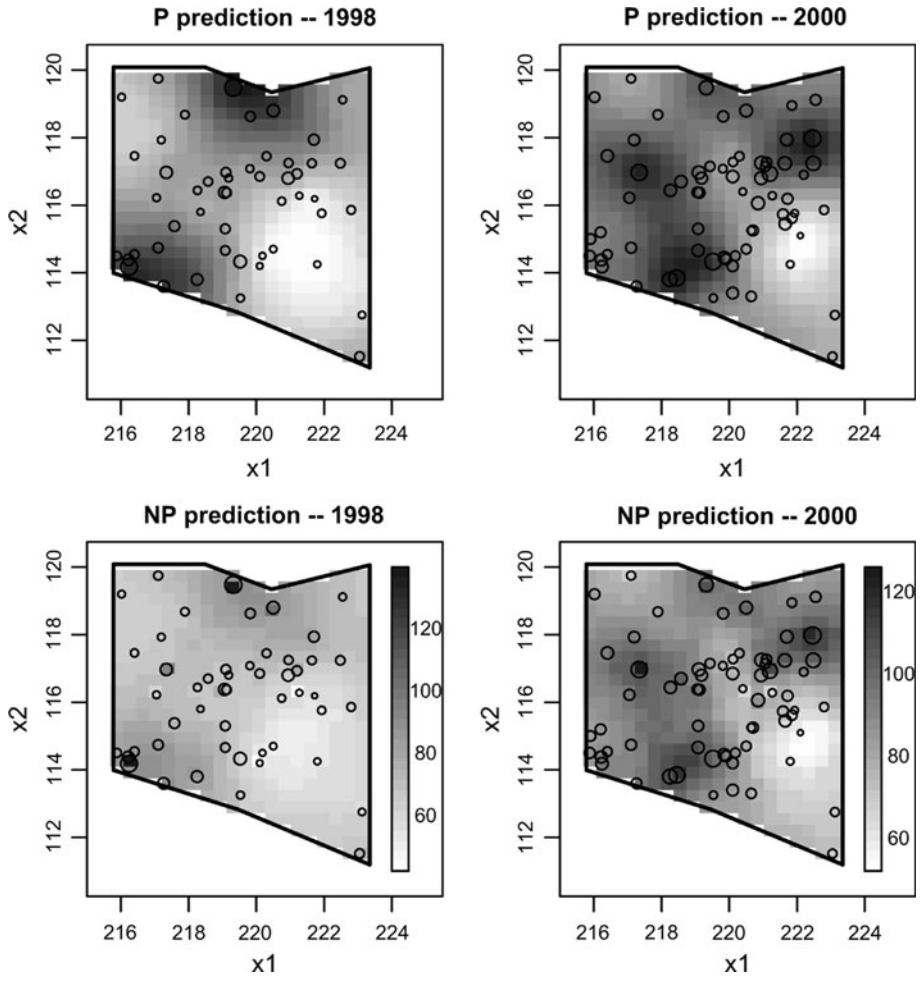

Fig. 3 Predicted surfaces for the original nitrates data by using the likelihood estimator (P) and the proposed valid covariance estimator (P). *Bullets* represent sample data

The next step was the application of the ordinary kriging technique to proceed with prediction over the observation region. The predicted surface was obtained for the groundwater quality data in two distinct ways, with the proposed valid covariance estimator and with the maximum likelihood one. Figure 3 displays the resulting prediction maps obtained over a grid of a total of 500 points, allowing us to identify the location of areas more polluted by the presence of nitrates.

The results of both approaches, the parametric (P) and the nonparametric (NP), confirm an overall increment in levels of pollution over the 2 years between the surveys. Surprisingly, the highest predicted values are found in 1998, which might be explained by the existence of stronger outliers in this year than in 2000. Therefore, the use of the Fourier series covariance seems to offer a good alternative to a classic approach, such as the maximum likelihood one, which can be directly used for prediction.

Conclusions

In the current paper, the Fourier series tool is applied for estimation of the covariance function. Furthermore, a valid covariance estimator may be obtained to be directly used for spatial prediction under a simple requirement, which is the choice of those terms in the expansion corresponding to positive Fourier coefficients. The proposed approach includes a criterion for approximation of the unknown characteristic, the truncation point, which specifies the number of terms to be used in the expansion.

The main advantages of this proposal are that it avoids the model misspecification problem and provides a covariance estimator that inherits several attractive properties from the Fourier basis, such as the degree of smoothness, the simplicity for implementation or, even, the positive-definiteness.

In practice, selection of a pilot covariance estimator is needed, which may be supplied to carry out a specific study or may be selected from the different alternatives existing in the literature, such as the kernel covariance estimator considered in Sect. 3. Therefore, an additional profit of the suggested approach is that it may be used for transformation of a given covariance estimator into a valid one.

Acknowledgments The authors thank the helpful suggestions from the reviewers, which have been reflected in the current paper. This work has been supported in part by grant INCITE-08-PXIB-322219-PR from Consellería de Innovación e Industria (Xunta de Galicia, Spain). R. Menezes acknowledges financial support from the projects PTDC/MAT/104879/2008 and PTDC/MAT/112338/2009 (FEDER support included) of the Portuguese Ministry of Science, Technology and Higher Education. Ó. Rubiños-López's research has also been supported by FEDER through Xunta de Galicia Researching programs (Grupos de referencia competitiva). P. García-Soidán and Ó. Rubiños-López acknowledge financial support from the project CONSOLIDER-INGENIO CSD2008-00068.

Appendix 1: A Fourier series approach

Let $f : E \longrightarrow \mathbb{R}$ be a bounded function, with $E = [0, e_1] \times [0, e_2]$, for some $e_1, e_2 > 0$. Then, a complete and countable orthonormal basis $\{\psi_i : E \longrightarrow \mathbb{R} : i \in \mathbb{N}\}$, can be constructed, satisfying that:

$$f(t) = \sum_{i \in \mathbb{N}} \theta_{f,i} \psi_i(t), \quad \text{for all } t \in E \tag{9}$$

where $\theta_{f,i} = \int_E f(t)\psi_i(t)\,\mathrm{d}(t)$ is the ith Fourier coefficient of f. The linear expansion above is called a Fourier expansion of f.

The existence of such a basis can be derived from the following properties:

- The unidimensional cosine system $\psi_{i,e}(x) = a_i \cos(i\pi x e^{-1})$ is a complete orthonormal basis on $[0, e]$, where a_i equals $e^{-1/2}$ or $(0.5\ e)^{-1/2}$, for $i = 0$ or $i > 0$, respectively.
- The set $\{\psi_{i_1,i_2} : E \longrightarrow \mathbb{R} : i_1, i_2 \in \mathbb{N}\}$, with $\psi_{i_1,i_2}(t) = \psi_{i_1,e_1}(t_1)\psi_{i_1,e_2}(t_2)$ and $t = (t_1, t_2)$, is a complete orthonormal basis on E (Zygmund 2002).
- A bijection $g : \mathbb{N}^2 \to \mathbb{N}$ can be established, referred to as Cantor's diagonal function, with $g(i_1, i_2) = 0.5\ (i_1 + i_2)(i_1 + i_2 + 1) + i_1$, whose values are displayed in Table 3.

Then, a complete orthonormal basis on E would be given by:

$$\psi_i(t) = \psi_{i_1,e_1}(t_1)\psi_{i_2,e_2}(t_2), \quad \text{with } (i_1, i_2) = g^{-1}(i) \tag{10}$$

For example, the first values of function g^{-1} are $g^{-1}(0) = (0,0), g^{-1}(1) = (0,1), g^{-1}(2) = (1,0), \ldots$

Furthermore, for every $\varepsilon > 0$ there exists a number $M \in \mathbb{N}$ such that $|f(t) - f_m(t)| < \varepsilon$ for all $m > M$, with:

Table 3 Values of Cantor's diagonal function

	0	1	2	3	4	⋯
0	0	1	3	6	10	⋯
1	2	4	7	11	16	⋯
2	5	8	12	17	23	⋯
3	9	13	18	24	31	⋯
4	14	19	25	32	40	⋯
⋮	⋮	⋮	⋮	⋮	⋮	⋱

The entry for row i_1 and column i_2 represents $g(i_1, i_2)$, where g is the Cantor diagonal function

$$f_m(t) = \sum_{i \le m} \theta_{f,i}\psi_i(t), \quad \text{for all } t \in E$$

For practical reasons, it is customary to use the truncated partial sum $f_m(t)$, instead of expansion (9), for approximation of f.

Appendix 2: Consistency of $\theta_{\hat{C},i}$

From relation (2) and the fact that the basis is orthonormal, one has:

$$\begin{aligned} |\text{Bias}[\theta_{\hat{C},i}]| &= \left| \int_B \text{Bias}[\hat{C}(t)]\psi_i(t)\,\mathrm{d}t \right| \\ &\le \sup_{t\in E} |\text{Bias}[\hat{C}(t)]| \int_E |\psi_i(t)|\,\mathrm{d}t \\ &\le b_i \sup_{t\in E} |\text{Bias}[\hat{C}(t)]| \overset{n\to\infty}{\longrightarrow} 0 \end{aligned} \tag{11}$$

for some positive constant b_i, together with:

$$\begin{aligned} \text{Var}[\theta_{\hat{C},i}] &= \int_E \int_E \text{Cov}[\hat{C}(t), \hat{C}(t')]\psi_i(t)\psi_i(t')\,\mathrm{d}t\,\mathrm{d}t' \\ &\le \left(\int_E \text{Var}[\hat{C}(t)]^{1/2} |\psi_i(t)|\,\mathrm{d}t \right)^2 \\ &\le \sup_{t\in E} \text{Var}[\hat{C}(t)] \left(\int_E |\psi_i(t)|\,\mathrm{d}t \right)^2 \\ &\le b_i^2 \sup_{t\in E} \text{Var}[\hat{C}(t)] \overset{n\to\infty}{\longrightarrow} 0 \end{aligned} \tag{12}$$

The fact that the bias and variance of $\theta_{\hat{C},i}$ tend to zero implies that $\theta_{\hat{C},i} \xrightarrow{P} \theta_{C,i}$ and, therefore, the consistency of $\theta_{\hat{C},i}$.

Appendix 3: MISE of $\hat{C}_{1,m}$

Observe that:

$$\begin{aligned} \text{MISE}[\hat{C}_{1,m}, C] &= \text{E}\left[\int (\hat{C}_{1,m}(t) - C(t))^2\,\mathrm{d}t \right] \\ &= \sum_{i \le m} \text{E}[(\theta_{\hat{C},i} - \theta_{C,i})^2] + \sum_{i > m} \theta_{C,i}^2 \\ &= \sum_{i \le m} (\text{Var}[\theta_{\hat{C},i}] + \text{Bias}[\theta_{\hat{C},i}]^2) + \sum_{i > m} \theta_{C,i}^2 \end{aligned}$$

By Parseval's identity (Efromovich 1999), one has:

$$\sum_{i > m} \theta_{C,i}^2 = \int_E C(t)^2\,\mathrm{d}t - \sum_{i \le m} \theta_{C,i}^2$$

Then:

$$\begin{aligned} \text{MISE}[\hat{C}_{1,m}, C] = \int_B C(t)^2\,\mathrm{d}t + \sum_{i \le m} (&\text{Var}[\theta_{\hat{C},i}] + \text{Bias}[\theta_{\hat{C},i}]^2 \\ &- \theta_{C,i}^2) \end{aligned} \tag{13}$$

Appendix 4: Properties of $\hat{C}_{2,m}$

Firstly, the positive-definiteness of $\hat{C}_{2,m}$ will be proved. For the latter aim, take into account that each function in the basis $\{\psi_i : i \in \mathbb{N}\}$ is obtained from the unidimensional cosine system, as given in (10). Then, ψ_i is positive-definite on account of the fact that $\cos(x - x') = \cos(x)\cos(x') + \sin(x)\sin(x')$. This means that for each $i \in \mathbb{N}$:

$$c_i = \sum_{j=1}^{n} \sum_{k=1}^{n} d_j d_k \psi_i(s_j - s_k) \ge 0 \tag{14}$$

for any set of locations $\{ s_j \}_{j=1}^n$ and real numbers $\{ d_j \}_{j=1}^n$.

By (14) and the definition of the weights w_i, one has:

$$\sum_{j=1}^{n} \sum_{k=1}^{n} d_j d_k \hat{C}_{2,m}(s_j - s_k) = \sum_{i \le m} c_i w_i \theta_{\hat{C},i} \ge 0$$

to conclude the positive-definiteness of $\hat{C}_{2,m}$.

By proceeding similarly as in Appendix 3, the MISE of $\hat{C}_{2,m}$ could be developed to yield:

$$\begin{aligned} \text{MISE}[\hat{C}_{2,m}, C] &= E\left[\int (\hat{C}_{2,m}(t) - C(t))^2\,\mathrm{d}t \right] \\ &= \sum_{i \le m} \text{E}[(w_i\theta_{\hat{C},i} - \theta_{C,i})^2] + \sum_{i > m} \theta_{C,i}^2 \\ &= \sum_{i \le m} w_i E[(\theta_{\hat{C},i} - \theta_{C,i})^2] \\ &\quad + \sum_{i \le m} (1 - w_i)\theta_{C,i}^2 + \sum_{i > m} \theta_{C,i}^2 \\ &= \sum_{i \le m} w_i (\text{Var}[\theta_{\hat{C},i}] + \text{Bias}[\theta_{\hat{C},i}]^2 - \theta_{C,i}^2) \\ &\quad + \int_E C(t)^2\,\mathrm{d}t \end{aligned} \tag{15}$$

References

Chen X (2005) Statistical and geostatistical features of streambed hydraulic conductivities in the Platte River, Nebraska. Environ Geol 48:693–701. doi:10.1007/s00254-005-0007-1

Christakos G (1984) On the problem of permissible covariance and variogram models. Water Resour Res 20:251–265. doi:10.1029/WR020i002p00251

Cressie N (1993) Statistics for spatial data. Wiley, New York

Cressie N, Hawkins D (1980) Robust estimation of the variogram. Math Geol 12:115–125. doi:10.1007/BF01035243

Diggle P, Hall P (1986) The selection of terms in an orthogonal series density estimator. J Am Stat Assoc 81:230–233

Efromovich S (1999) Nonparametric curve estimation: methods, theory and applications. Springer, New York

Eubank RL (1988) Spline smoothing and nonparametric regression. Marcel and Dekker, New York

García-Soidán P (2007) Asymptotic normality of the Nadaraya-Watson semivariogram estimators. TEST 16:479–503. doi:10.1080/10485250902878655

Genton MG (1998) Highly robust variogram estimation. Math Geol 30:213–221. doi:10.1023/A:1021728614555

Hall P, Patil P (1994) Properties of nonparametric estimators of autocovariance for stationary random fields. Probab Theory Relat Fields 99:399–424. doi:10.1007/BF01199899

Hall P, Fisher NI, Hoffman B (1994) On the nonparametric estimation of covariance functions. Ann Stat 22:2115–2134. doi:10.1214/aos/1176325774

Hart JD (1985) On the choice of a truncation point in Fourier series density estimation. J Stat Comput Simul 21:95–116. doi:10.1080/00949658508810808

Huang B, Hu T (2009) BLP approach for estimation of variogram parameters. Environ Earth Sci 59:421–428. doi:10.1007/s12665-009-0040-6

LaMotte AE, Green EA (2007) Spatial analysis of land use and shallow groundwater vulnerability in the watershed adjacent to Assateague Island National Seashore, Maryland and Virginia, USA. Environ Geol 52:1413–1421. doi:10.1007/s00254-006-0583-8

Lark RM (2000) A comparison of some robust estimators of the variogram for use in soil survey. Eur J Soil Sci 51:137–157. doi:10.1046/j.1365-2389.2000.00280.x

Lin YP, Chu HJ, Huang YL, Cheng BY, Chang TK (2011) Modeling spatial uncertainty of heavy metal content in soil by conditional Latin hypercube sampling and geostatistical simulation. Environ Earth Sci 62:299–311. doi:10.1007/s12665-010-0523-5

Matheron G (1963) Principles of geostatistics. Econ Geol 58:1246–1266. doi:10.2113/gsecongeo.58.8.1246

Menezes R, García-Soidán P, Febrero M (2005) A comparison of approaches for valid variogram achievement. Comput Stat 20(4):623–642. doi:10.1007/BF02741319

Menezes R, García-Soidán P, Febrero M (2008) A kernel variogram estimator for clustered data. Scand J Stat 35:18–37. doi:10.1111/j.1467-9469.2007.00566.x

Paralta E, Ribeiro L (2003) Monitorização e modelação estocástica da contaminação por nitratos no Aquífero Gabro-diorítico na Região de Beja. Resultados, conclusões e recomendações. Seminário sobre Águas Subterráneas, APRH/LNEC, Lisboa. http://repositorio.lneg.pt/bitstream/10400.9/478/1/33618.pdf Accessed 2 December 2009

Rakhmatullaev S, Marache A, Huneau F, LeCoustumer P, Bakiev M (2011) Geostatistical approach for the assessment of the water reservoir capacity in arid regions: a case study of the Akdarya reservoir, Uzbekistan. Environ Earth Sci 63:447–460. doi:1007/s12665-010-0711-3

Romshoo S (2003) Geostatistical analysis of soil moisture measurements and remotely sensed data at different spatial scales. Environ Geol 45:339–349. doi:10.1007/s00254-003-0891-1

Satapathy DR, Salve PR, Katpatal YB (2009) Spatial distribution of metals in ground/surface waters in the Chandrapur district (Central India) and their plausible sources. Environ Geol 56:1323–1352. doi:10.1007/s00254-008-1230-3

Tarter ME, Lock MD (1993) Model-free curve estimation. Chapman and Hall, Los Angeles

Zhao Y, Xu X, Huang B, Sun W, Shao X, Shi X, Ruan X (2007) Using robust kriging and sequential Gaussian simulation to delineate the copper- and lead-contaminated areas of a rapidly industrialized city in Yangtze River Delta, China. Environ Geol 52:1423–1433. doi:10.1007/s00254-007-0667-0

Zygmund A (2002) Trigonometric series. Cambridge University Press, Cambridge

Copyright of Environmental Earth Sciences is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.