Statistics and Computing **13**: 127–136, 2003

Flexible spatio-temporal stationary variogram models

RUBÉN FERNÁNDEZ-CASAL*, WENCESLAO GONZÁLEZ-MANTEIGA† and MANUEL FEBRERO-BANDE†

**Department of Statistics and Operational Research, University of Vigo, Facultade de Ciencias Empresariais, Campus de Ourense, 32004 Ourense, Spain*
rubenfc@uvigo.es
†*Department of Statistics and Operational Research, University of Santiago de Compostela, Facultad de Matemáticas, Campus Sur, 15782 Santiago de Compostela, Spain*
wences@zmat.usc.es, febrero@zmat.usc.es

Received November 2001 and accepted August 2002

In this paper we propose a generalization of the Shapiro and Botha (1991) approach that allows one to obtain flexible spatio-temporal stationary variogram models. It is shown that if the weighted least squares criterion is chosen, the fitting of such models to pilot estimations of the variogram can be easily carried out by solving a quadratic programming problem. The work also includes an application to real data and a simulation study in order to illustrate the performance of the proposed space-time dependency modeling.

Keywords: geostatistics, anisotropy, spectral representation, ordinary kriging

1. Introduction

The modeling of spatio-temporal processes is a crucial problem in many disciplines (for example, in environmental science, climatology, image handling, hydrology, geology, etc.) in which, in recent years, the use of geostatistical spatio-temporal models has been applied (see, for example, Kyriakidis and Journel 1999). Nonetheless, many of these approaches are based on models initially developed for cases that are either exclusively spatial or temporal, and spatio-temporal dependencies are thus handled separately or inadequately exploited.

In traditional geostatistical approaches (see, for example, Cressie 1993) certain hypotheses are usually assumed, such as isotropy or geometric anisotropy, which facilitate the modeling of spatial dependency considerably. These hypotheses are usually inappropriate, however, for the spatio-temporal case (see, for example, Journel 1986). To avoid this problem, other approaches also based on classical covariogram (or variogram) models are considered, such as the assumption of separability (e.g. Guttorp, Sampson and Newman 1992) or variations on this (e.g. Haas 1995).

More recently some non-separable spatio-temporal stationary variogram models have been proposed (Jones and Zhang 1997, Cressie and Huang 1999), but apart from the reduced number of spatio-temporal variogram models known at present, their fit to the empirical variogram could be unsatisfactory in practice. It would therefore be highly desirable to have available models with sufficient flexibility to adapt to any situation.

Our paper describes a generalization of the Shapiro and Botha (1991) approach for the case of anisotropy in two components, with a view to obtaining flexible stationary variogram models suitable for modeling spatio-temporal dependency. In Section 2 we describe the notation used together with a review of some basic theoretical results. In Section 3 various families of valid anisotropic two component stationary semivariogram models are defined (also applicable to the spatio-temporal case). Section 4 describes an example of the application of some of those models to real data. In section 5 different models are compared via a simulation study. Finally, in Section 6 some comments and conclusions of the approach described are included.

2. Notation and preliminary remarks

We can obtain valid spatio-temporal semivariogram models as particular cases of anisotropic two component models, therefore

we will focus on anisotropic two component stationary spatial processes in following sections.

Let us suppose that $Z(\mathbf{s})$ is a stationary spatial process in $\mathbb{R}^d$, in other words:

$$E(Z(\mathbf{s})) = \mu(\mathbf{s}) = \mu,$$

$$Cov(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = C(\mathbf{s}_1, \mathbf{s}_2) = C(\mathbf{h}),$$

where $\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$, and assume that, even though the covariogram is not necessarily isotropic, there is isotropy within two components of the lag vector:

$$C(\mathbf{h}) = C(\|\mathbf{h}_1\|, \|\mathbf{h}_2\|) = C(r, u),$$

where $\mathbf{h}_1 \in \mathbb{R}^{d_1}$ is a vector formed by d_1 components of the lag $\mathbf{h}$ and $\mathbf{h}_2 \in \mathbb{R}^{d_2}$ is the vector formed by the $d_2 = d - d_1$ remaining components (e.g. the case of a spatio-temporal process with space-isotropic/time-isotropic covariogram $C(\mathbf{h}, t) \equiv C(\|\mathbf{h}\|, |t|)$).

Let us also assume for the moment that the covariogram is continuous at the origin. From Bochner's theorem (Bochner 1959), since the function is nonnegative definite, a representation of the covariogram in the following form is possible:

$$C(\mathbf{h}_1, \mathbf{h}_2) = \int_{\mathbb{R}^{d_1}} \int_{\mathbb{R}^{d_2}} e^{i(\boldsymbol{\omega}\cdot\mathbf{h}_1 + \boldsymbol{\tau}\cdot\mathbf{h}_2)}\, dF(\boldsymbol{\omega}, \boldsymbol{\tau}), \qquad (1)$$

where dF is a finite positive measure. If the function F (called the spectral distribution function) is differentiable, we can express the above equation in terms of the spectral density function $f(\boldsymbol{\omega}, \boldsymbol{\tau}) \geq 0$.

Analogously to the case of spatial isotropy (see for example, Stein 1999, pp. 42–43), we can obtain the following expression for the previous equation (1) corresponding to an anisotropic two component covariogram:

$$C(r, u) = \int_0^\infty \int_0^\infty \kappa_{d_1}(\lambda r)\kappa_{d_2}(\upsilon u)\, dG(\lambda, \upsilon), \qquad (2)$$

where

$$\kappa_d(x) = \left(\frac{2}{x}\right)^{(d-2)/2} \Gamma\left(\frac{d}{2}\right) J_{(d-2)/2}(x),$$

J_p is the Bessel function of order p, and G is a bounded positive function on $[0, \infty) \times [0, \infty)$ with properties similar to those of a bidimensional distribution function. This function takes the form:

$$G(\lambda, \upsilon) = \int_{\|\boldsymbol{\omega}\| < \lambda} \int_{\|\boldsymbol{\tau}\| < \upsilon} dF(\boldsymbol{\omega}, \boldsymbol{\tau}),$$

where dF is a symmetric positive measure. Moreover, from the expressions of the Bessel function of order p (see Abramovitz and Stegun 1965, 9.1.20 and 9.1.10), it can be seen that $\kappa_d(0) = 1$, and therefore:

$$C(0, 0) = \int_0^\infty \int_0^\infty dG(\lambda, \upsilon).$$

A special case of (2) frequently used in practice and which simplifies the problem of space-time dependency modeling considerably, is the assumption of separable covariograms $C(\mathbf{h}, t) \equiv C_1(\|\mathbf{h}\|)C_2(|t|)$ (which corresponds to $f(\boldsymbol{\omega}, \tau) \equiv f_1(\|\boldsymbol{\omega}\|) f_2(|\tau|)$). The non-separable models proposed by Cressie and Huang (1999) are also particular cases of (2).

In general the expressions of equation (2) will include trigonometric functions when d_i is odd and Bessel functions of integer order when d_i is even. For example, for $d_1 = 2, d_2 = 1$ and $d_1 = 3, d_2 = 1$, we would obtain expressions as follows:

$$d_1 = 2, d_2 = 1 : C(r, u) = \int_0^\infty \int_0^\infty J_0(\lambda r)\cos(\upsilon u)\, dG(\lambda, \upsilon),$$

$$d_1 = 3, d_2 = 1 : C(r, u) = \int_0^\infty \int_0^\infty \frac{\sin(\lambda r)}{\lambda r}\cos(\upsilon u)\, dG(\lambda, \upsilon),$$

It is well-known that a spatial covariogram valid on $\mathbb{R}^d$ is also valid on $\mathbb{R}^{d_0}, \forall\, d_0 \leq d$. Moreover, it can be deduced analogously to the results obtained for isotropic spatial covariograms (see, for example, Stein 1999, pp. 44–45), that a function is a (continuous) two component anisotropic covariogram valid in any spatial dimension d_1, if and only if it can be represented as:

$$C(r, u) = \int_0^\infty \int_0^\infty e^{-\lambda^2 r^2} \kappa_{d_2}(\upsilon u)\, dG(\lambda, \upsilon),$$

where G is a bounded positive function on $[0, \infty) \times [0, \infty)$. We can thus denote $\kappa_\infty(x) \equiv e^{-x^2}$.

It is also important to note that:

$$\begin{aligned} C(r, u) &= \int_0^\infty \int_0^\infty \kappa_\infty(\lambda r)\kappa_\infty(\upsilon u)\, dG(\lambda, \upsilon) \\ &= \int_0^\infty \int_0^\infty e^{-\lambda^2 r^2 - \upsilon^2 u^2} dG(\lambda, \upsilon), \end{aligned}$$

is a valid covariogram for any d_1 and d_2.

3. Families of spatio-temporal variogram models

In general, the covariogram of a spatial process $Z(\mathbf{s})$ may not be continuous at the origin, in which case it can be expressed as $C(\mathbf{h}) = c_0 \cdot \delta(\mathbf{h}) + C_0(\mathbf{h})$, where $c_0 \geq 0$ (nugget effect), $\delta(\cdot)$ is the indicator function of the origin ($\delta(\mathbf{0}) = 1$ and $\delta(\mathbf{h}) = 0$ if $\mathbf{h} \neq \mathbf{0}$) and $C_0(\cdot)$ is a covariogram continuous at the origin. Moreover, since its covariogram and semivariogram are related by $\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$, the semivariogram of an anisotropic two component stationary process takes the form:

$$\gamma(r, u) \begin{cases} \nu_0 - \nu(r, u) & \text{if } r > 0 \text{ or } u > 0 \\ 0 & \text{if } r = u = 0 \end{cases}, \qquad (3)$$

where ν_0 is a positive constant and $\nu(r, u)$ is a nonnegative definite function continuous at the origin and which can be represented in the form of (2).

Reciprocally, any function that verifies (3) as well as $\nu_0 - \nu(r, u) \geq 0\ \forall\, r, u$ is an anisotropic two component semivariogram valid in $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Moreover, since a nonnegative definite function has its absolute value bounded by its value at the

origin, the above restriction is equivalent to $\nu_0 - \nu(0, 0) \geq 0$. The case of $c_0 = \nu_0 - \nu(0, 0) > 0$, will correspond to the presence of a nugget effect.

The anisotropic semivariogram models considered will take the form of (3), where ν_0 is a positive constant and where $\nu(r, u)$ is a nonnegative definite function, continuous at the origin. It can be represented as in (2) and such that:

$$\nu_0 - \nu(0, 0) = \nu_0 - \int_0^\infty \int_0^\infty dG(\lambda, \upsilon) \geq 0. \tag{4}$$

We can consider an analogous simplification to that of Shapiro and Botha (1991) which facilitates the fit of models of this kind to pilot semivariogram estimates, assuming that dG, in the equations above, is an atomic measure. In other words, we can assume that G is a step function with a finite number of positive jumps z_{kl} at points (x_k, y_l), taking the form:

$$G(x, y) = \sum_{x_k \leq x} \sum_{y_l \leq y} z_{kl}.$$

For convenience sake, let us also assume that the points (x_k, y_l) are regularly spaced (the results will be identical for positions that are irregularly spaced), i.e. $x_k = \phi_1 \cdot k$, $k = 1, \ldots, m$, $y_l = \phi_2 \cdot l$, $l = 1, \ldots, n$, where ϕ_1 and ϕ_2 are two positive numbers.

The (bounded) semivariogram models obtained, which we will designate as extended Shapiro-Botha models, are of the form (3), where $\nu(r, u)$ is reduced to:

$$\nu(r, u) = \sum_{k=1}^{m} \sum_{l=1}^{n} \kappa_{d_1}(x_k r) \kappa_{d_2}(y_l u) z_{kl}, \tag{5}$$

and the restriction (4) is converted into the linear restriction:

$$\nu_0 - \sum_{k=1}^{m} \sum_{l=1}^{n} z_{kl} \geq 0.$$

In practice, if $\{\hat{\gamma}_{ij} = \hat{\gamma}(r_i, u_j) : (i, j) \in I\}$ are pilot semivariogram estimates (where I represents the set of indices corresponding to the lags considered), then the problem of the fit of a valid model is reduced to finding the vector $\boldsymbol{\theta} = (z_{11}, \ldots, z_{mn}, \nu_0)'$ of dimensions $m \times n + 1$ that minimizes the function:

$$Q(\boldsymbol{\theta}) = \sum_{(i,j) \in I} w_{ij} \left(\hat{\gamma}_{ij} - \nu_0 + \sum_{k=1}^{m} \sum_{l=1}^{n} \kappa_{d_1}(x_k r_i) \kappa_{d_2}(y_l u_j) z_{kl} \right)^2,$$

subject to the linear restrictions:

$$z_{kl} \geq 0, \quad k = 1, \ldots, m, l = 1, \ldots, n; \quad \text{and}$$

$$\nu_0 - \sum_{k=1}^{m} \sum_{l=1}^{n} z_{kl} \geq 0.$$

Just as in the Shapiro and Botha (1991) isotropic case, we are dealing with a quadratic programming problem that will be resolved similarly. Naturally, for this minimization problem to have a single solution it has to be verified that $|I| \geq m \times n + 1$, where $|I|$ is the number of lags used in the fit. If, moreover, we apply the criterion for the choice of weights proposed by Cressie (1985) of taking $w_{ij} = n_{ij} / \gamma(r_i, u_j)^2$, where n_{ij} is the number of pairs used in the estimation $\hat{\gamma}_{ij} = \hat{\gamma}(r_i, u_j)$, then we must proceed iteratively, taking $w_{ij} = 1$ for the first step and recalculating the weights for each iteration until convergence is obtained. If $\bar{\boldsymbol{\theta}} = (\bar{z}_{11}, \ldots, \bar{z}_{mn}, \bar{\nu}_0)'$ is the optimum solution obtained by resolving this problem, then the fitted semivariogram model will be given by:

$$\bar{\gamma}(r, u) = \bar{\nu}_0 - \sum_{k=1}^{m} \sum_{l=1}^{n} \kappa_{d_1}(x_k r) \kappa_{d_2}(y_l u) \bar{z}_{kl}, \tag{6}$$

for $(r, u) \neq (0, 0)$; and the corresponding value of weighted least squares (*WLS*) will be:

$$WLS = \sum_{(i,j) \in I} n_{ij} \left(\frac{\hat{\gamma}(r_i, u_j)}{\bar{\gamma}(r_i, u_j)} - 1 \right)^2. \tag{7}$$

If we use the empirical estimator of the semivariogram, then the $\hat{\gamma}_{ij}$ estimates may be highly variable. Since the family of semivariograms obtained in this way is very flexible, the fitted semivariogram $\bar{\gamma}(r, u)$ may have an irregular shape (above all if $|I|$ is large). In order to avoid an "overfit" of this kind, the inclusion of additional smoothness, monotony and convexity (linear) restrictions may be considered (see Shapiro and Botha 1991, pp. 91–94, for the spatial case). A different approach could be the use of nonparametric kernel estimators of the semivariogram (commented briefly in next section). In addition to obtaining more efficient estimates, these being smoother, it is not necessary to add further restrictions to the fit.

This approach could equally be extended to the case of anisotropy in more than two components, for example, to the spatio-temporal case with anisotropy in the spatial coordinates.

4. Application to a real case

In this section an example of the application of these spatio-temporal semivariogram models to a set of real data is given, the same data as those used in the paper by Cressie and Huang (1999). The observations are measurements taken every six hours (time unit) of the east-west component of wind speed on a grid of 17 × 17 spatial positions (regularly spaced, approx. 210 kms) located in the tropical western Pacific Ocean. Rather than use all the time measurements as in the Cressie and Huang paper (480 time intervals corresponding to the period from November 1992 to February 1993), it was decided to take into consideration only the first 20 time points in order to ensure that the number of observations was manageable. The calculations were thus based on a spatio-temporal sample of 5780 observations, corresponding to 289 spatial locations and 20 time points.

Following an exploratory analysis of the data, it could be seen (Cressie and Huang 1999, p. 1336) that the assumption of spatio-temporal stationarity is reasonable (which permits us to use the ordinary kriging model and directly apply the classical estimator of the semivariogram to the data).

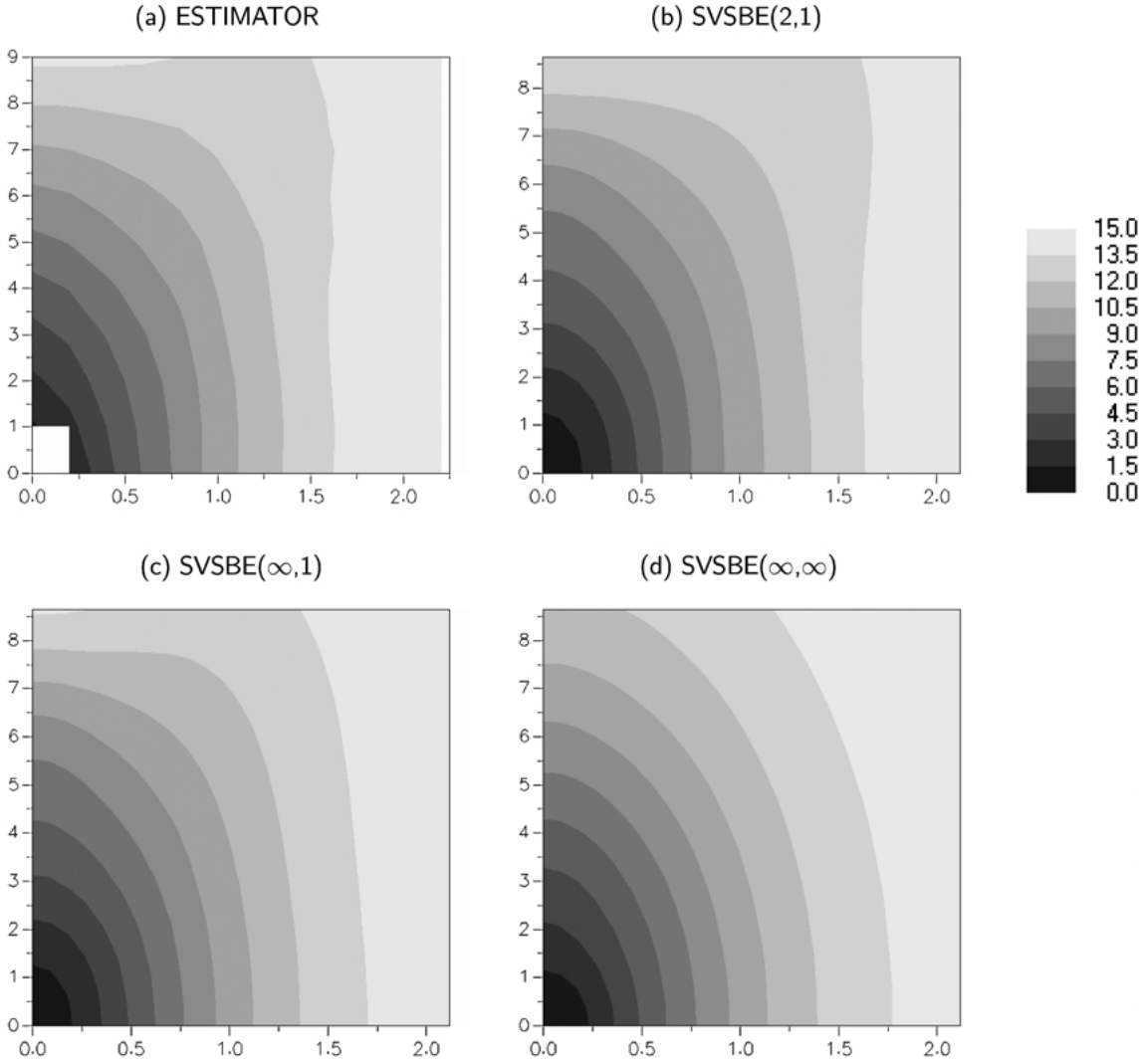

Fig. 1. *Contour surfaces corresponding to the empirical estimator of the semivariogram (evaluated in the set of lags used for the fit); and the fitted extended Shapiro-Botha models. The horizontal axis represents the spatial lag (in thousands of kilometers) and the vertical axis the temporal lag (1 unit = 6 hours)*

If we denote the observed values of the E-W component of wind speed as $Z(\mathbf{s}_1, t_1), \ldots, Z(\mathbf{s}_N, t_N)$, then using the classical estimator, the pilot semivariogram estimates $\{\hat{\gamma}_{lu} = \hat{\gamma}(h(l), u) : 0 \leq l \leq L,\ 0 \leq u \leq U,\ (l, u) \neq (0, 0)\}$ will be given by:

$$\hat{\gamma}(h(l), u) = \frac{1}{2|N(h(l), u)|} \sum_{N(h(l),u)} (Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2,$$

where $N(h(l), u) = \{(i, j) : \|\mathbf{s}_i - \mathbf{s}_j\| \in Tol(h(l)), |t_i - t_j| = u\}$, and $Tol(h(l))$ is a tolerance region around $h(l)$. Moreover, following the recommendations of Journel and Huijbregts (1978, p. 194), the considered models were fitted using half of the possible lags and in such a way that the number of contributions to the estimation in each lag was at least 30. Figure 1(a) shows the estimates obtained with the classical estimator of the semivariogram in the lags used for the fit: $h(i) = 0.20046 \cdot i$ (thousands of kilometers), $i = 0, \ldots, 11$, $u = 0, \ldots, 9$ and $(i, u) \neq (0, 0)$ (note that the value of the semivariogram at the origin is always cero). The tolerance regions were taken so that $r \in Tol(h(l))$ if $-0.10023 \leq r - h(l) < 0.10023$.

In order to avoid the "overfit" problem with the use of the empirical semivariogram estimator (commented in Section 3), we could, for example, obtain a pilot estimate of $\gamma(r, u)$ by multivariate local linear least squares regression, minimizing:

$$\sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left((Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2 - (\beta_0, \beta_{10}, \beta_{01}) \times \begin{pmatrix} 1 \\ \|\mathbf{s}_i - \mathbf{s}_j\| - r \\ |t_i - t_j| - u \end{pmatrix} \right)^2 \cdot K_H \left(\begin{pmatrix} \|\mathbf{s}_i - \mathbf{s}_j\| - r \\ |t_i - t_j| - u \end{pmatrix} \right), \tag{8}$$

Table 1. *Values for ϕ_1 and ϕ_2, nugget effect estimates and WLS obtained for the fit of the extended Shapiro-Botha models: [SVSBE(2,1)] model (6) with $d_1 = 2$ and $d_2 = 1$, [SVSBE(∞,1)] model (6) with $d_1 = \infty$ and $d_2 = 1$ and [SVSBE(∞,∞)] model (6) with $d_1 = \infty$ and $d_2 = \infty$*

	ϕ_1	ϕ_2	c_0	*WLS*
SVSBE(2,1)	0.85	0.1	0.5896	1337.98
SVSBE(∞,1)	0.25	0.1	0.5549	3387.291
SVSBE(∞,∞)	0.55	0.085	0.0000	7933.438

with $K_H(\mathbf{v}) = \frac{1}{|H|}K(H^{-1}\mathbf{v})$, where $K(\cdot)$ is a bidimensional kernel and H a bandwidth matrix. One of the main reasons for choosing this estimator among the local kernel estimators is the absence of boundary effects (this is especially important near the origin; see e.g. Stein 1999, chapter 3). If $(\hat{\beta}_0, \hat{\beta}_{10}, \hat{\beta}_{01})$ is the solution of the straightforward weighted linear least squares problem (8), the pilot estimate of $\gamma(r, u)$ will be $\hat{\gamma}(r, u) = \hat{\beta}_0$. However, this approach will be considered in a future study.

For the fit of the extended Shapiro-Botha models the routine QPROG from the IMSL library was used to resolve the quadratic programming problem. The procedure was iterative, with $w_{ij} = 1$ (OLS) used for the first step and with the weights recalculated for each iteration until convergence. The values of m and n in (5) were set as equal to the number of lags used in the fit minus one, i.e. $m = 11$ and $n = 9$. The values for ϕ_1 and ϕ_2 were taken so as to minimize the sum of weighted squares of the fit (these were considered as two further parameters; in order to estimate them, the objective function was evaluated on a bidimensional grid and values that minimized *WLS* were selected). Table 1 displays the values obtained for ϕ_1 and ϕ_2, the nugget effect estimates and the *WLS* values (equation (7)) for the fit to the empirical semivariogram. It can be observed that the model with the best fit is that which is specific to the dimensions of the data (using more general models would mean a loss of flexibility). Figure 1(b)–(d) illustrate the adjusted semivariogram models.

In order to verify if the fitted variogram models adequately described the spatio-temporal dependence of the data, the cross-validation technique was used (see, for example, Cressie 1993, pp. 101–104). For each spatio-temporal location $(\mathbf{s}_i, t_i)$, using the ordinary kriging method, a predictor $\hat{Z}_{-i}(\mathbf{s}_i, t_i)$ of $Z(\mathbf{s}_i, t_i)$ and the corresponding mean-squared prediction error $\sigma^2_{-i}(\mathbf{s}_i, t_i)$ were calculated (based on the set of observations without $Z(\mathbf{s}_i, t_i)$). Using these values the following measurements were obtained:

- Cross-validation averaged squared error:

$$CVASE = \frac{1}{N}\sum_{i=1}^{N}(\hat{Z}_{-i}(\mathbf{s}_i, t_i) - Z(\mathbf{s}_i, t_i))^2,$$

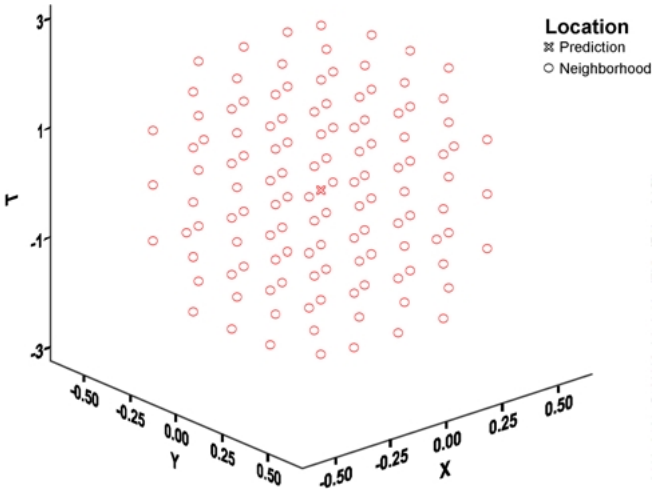

Fig. 2. *Neighborhood around the prediction location used in the calculation of the cross-validation measurements*

- Cross-validation dimensionless averaged squared error:

$$CVDASE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}((\hat{Z}_{-i}(\mathbf{s}_i, t_i) - Z(\mathbf{s}_i, t_i))/\sigma_{-i}(\mathbf{s}_i, t_i))^2},$$

  This should be close to 1 if there is agreement between the kriging variances and the observed variances.
- Coverage (*CVCOV*) of the 95% prediction interval (assuming normality).

In the calculation of the cross-validation measurements for each location $(\mathbf{s}_i, t_i)$, instead of using all the remaining observations $\{Z(\mathbf{s}_j, t_j) : j \neq i\}$ (which in principle would require inversion or factorization of a 5780×5780 matrix), only 122 of the data closest to the prediction location were used (although the nugget effect estimates would suggest that smaller neighborhoods could also have been appropriate, but there being no computational problems a larger neighborhood was selected). The configuration of the relative spatio-temporal locations of the neighborhood data for the area of the prediction position is shown in Fig. 2. Once the neighborhood was fixed, the procedure was analogous to that of Cressie (1993, pp. 158–161). The kriging weights corresponding to each position in the neighborhood were obtained (resolving a system of dimension 123). For each prediction location in which there was the totality of the neighborhood observations (1694 locations in total), the kriging prediction and the kriging variance were calculated using the weights obtained in the first stage. The total computational time for this procedure, on a PIII at 750 Mhz, was less than two minutes.

The results obtained for the fitted extended Shapiro-Botha models can be seen in Table 2. It is observed from the averaged squared error that the most flexible models with the best fit to the empirical estimator (Table 1) produce poorer results. Nevertheless, a comparison of the values of the dimensionless averaged squared error indicates that the SVSBE(∞,∞) model

Table 2. *Local cross-validation measurements corresponding to the fitted extended Shapiro-Botha models*

		SVSBE(2,1)	SVSBE(∞,1)	SVSBE(∞,∞)
CVASE	Mean	.5569	.6128	.0138
	Median	.2316	.2452	.0046
	Std. Dev.	.8626	.9702	.0470
CVDASE	Mean	.8075	.8955	3.0857
	Median	.3358	.3584	1.0279
	Std. Dev.	1.2508	1.4178	10.5326
CVCOV	Mean	.9705	.9569	.8146
	Std. Dev.	.1693	.2031	.3887

obtains prediction variance estimates (kriging variance) that are excessively optimistic, and therefore coverage of the prediction intervals is reduced.

5. Simulation study

In the simulation study the extended Shapiro-Botha models described in Section 3 were compared with the following spatio-temporal semivariogram models:

- Semivariogram model (SVCH2) corresponding to Example 2 in Cressie and Huang (1999, p. 1333):

$$\gamma(\mathbf{h}, u \mid \boldsymbol{\theta}) = \begin{cases} c_0 + \sigma^2\left(1 - \dfrac{1}{(a|u|+1)^{d/2}} \exp\left\{-\dfrac{b^2\|\mathbf{h}\|^2}{a|u|+1}\right\}\right) & \text{if } \mathbf{h} \neq \mathbf{0} \text{ or } u \neq 0 \\ 0 & \text{if } \mathbf{h} = \mathbf{0} \text{ and } u = 0 \end{cases}$$

  where d is the spatial dimension and $\boldsymbol{\theta} = (c_0, a, b, \sigma^2)'$, $c_0 \geq 0$ (nugget effect), $a \geq 0$ (time scale parameter), $b \geq 0$ (space scale parameter) and $\sigma^2 > 0$ (partial sill).
- Anisotropic spherical semivariogram model (SVESFA):

$$\gamma(\mathbf{h}, u \mid \boldsymbol{\theta}) = g(\sqrt{\|\mathbf{h}\|^2 + bu^2} \mid \tilde{\boldsymbol{\theta}}),$$

$$g(r \mid \tilde{\boldsymbol{\theta}}) = \begin{cases} 0 & \text{if } r = 0 \\ c_0 + \sigma^2\left\{\dfrac{3}{2}\dfrac{r}{a} - \dfrac{1}{2}\left(\dfrac{r}{a}\right)^3\right\} & \text{if } 0 < r \leq a \\ c_0 + \sigma^2 & \text{if } r > a \end{cases}$$

  where d is the spatial dimension and $\boldsymbol{\theta} = (c_0, a, b, \sigma^2)'$, $c_0 \geq 0$ (nugget effect), $a \geq 0$ (spherical semivariogram range), $b \geq 0$ (spatio-temporal interaction parameter) and $\sigma^2 > 0$ (partial sill).
- Semivariogram model corresponding to a separable exponential covariogram (SVSEXP):

$$\gamma(\mathbf{h}, u \mid \boldsymbol{\theta}) = \begin{cases} c_0 + \sigma^2\left(1 - \exp\left\{-\dfrac{u}{a} - \dfrac{\|\mathbf{h}\|}{b}\right\}\right) & \text{if } \mathbf{h} \neq \mathbf{0} \text{ or } u \neq 0 \\ 0 & \text{if } \mathbf{h} = \mathbf{0} \text{ and } u = 0 \end{cases}$$

  where d is the spatial dimension and $\boldsymbol{\theta} = (c_0, a, b, \sigma^2)'$, $c_0 \geq 0$ (nugget effect), $a \geq 0$ (time scale parameter), $b \geq 0$ (space scale parameter) and $\sigma^2 > 0$ (partial sill).

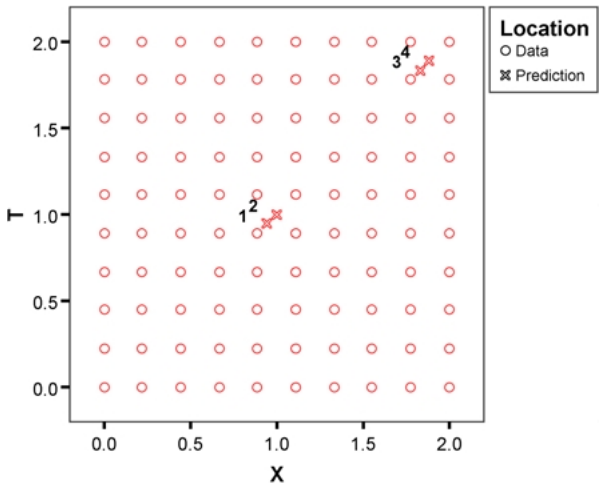

Fig. 3. *Spatial locations of the data and the predictions*

A square 10×10 lattice of ($N = 100$) points in the domain $D = [0, 2] \times [0, 2] \subset \mathbb{R} \times \mathbb{R}$ (spatial dimension $d = 1$) was chosen as the spatio-temporal configuration (fixed design). Using the ordinary kriging model, 1000 simulations were generated for these spatio-temporal locations and in 4 prediction locations (Fig. 3). The semivariogram models described above were used to generate the values, with parameters as follows: $\boldsymbol{\theta}_0 = (c_0 = 1, a = 1, b = 1, \sigma^2 = 1)'$ in the Cressie and Huang model; $\boldsymbol{\theta}_0 = (c_0 = 1, a = 0.5, b = 0.25, \sigma^2 = 1)'$ in the anisotropic spherical model; and finally, $\boldsymbol{\theta}_0 = (c_0 = 1, a = 0.5, b = 1, \sigma^2 = 1)'$ in the separable exponential model.

For the estimation of the variogram and the fit of the extended Shapiro-Botha models, an analogous procedure to that described in the previous section was followed. Alternatively, weighted least squares estimators were obtained for the models given above following an iterative algorithm. For each iteration:

1. The weights were recalculated: $w_{ij} = n_{ij}/\bar{\gamma}(r_i, t_j)^2$ (in the first iteration the weights were set equal to 1).
2. The nugget effect and the partial sill were estimated using a weighted least squares linear regression (if the other parameters are fixed then the models are a linear function of these); A negative value became zero and the other parameter was estimated again.
3. The rest of the parameters were estimated (fixing the nugget effect and the partial sill) using a modified Levenberg-Marquardt algorithm with restrictions in the parameters (routine BCLSF of the IMSL).

5.1. *Fitting the models*

In order to compare the different models, in addition to the *WLS* values, relative squared errors were also calculated to measure

Table 3. *Fit measurements obtained using the different models*

MODEL		WLS			RSE			RSET		
Theoretical	Fitted	Mean	Median	S. dev.	Mean	Median	S. dev.	Mean	Median	S. dev.
SVCH2	SVSBE(1,1)	19.48	18.28	7.75	.232	.216	.093	1.998	.951	3.43
	SVSBE(∞,1)	24.14	23.00	8.61	.298	.284	.110	1.974	.933	3.44
	SVSBE(∞,∞)	26.99	25.68	9.58	.334	.314	.124	1.899	.856	3.39
	SVCH2	30.54	29.40	10.92	.371	.350	.136	1.831	.819	3.19
	SVESFA	34.24	31.13	15.39	.424	.381	.195	1.755	.846	2.90
	SVEXPS	51.97	39.61	44.46	.635	.482	.544	1.702	.845	2.61
	TEOR	186.9	98.18	280.5	2.27	1.18	3.56	.000	.000	.00
SVESFA	SVSBE(1,1)	22.73	19.91	12.30	.268	.237	.142	1.470	.804	2.16
	SVSBE(∞,1)	27.53	25.14	12.92	.339	.310	.158	1.349	.698	2.13
	SVSBE(∞,∞)	31.33	28.98	13.73	.390	.360	.170	1.310	.658	2.13
	SVCH2	51.82	46.02	24.74	.618	.549	.289	1.551	.824	2.42
	SVESFA	35.59	32.82	15.01	.442	.407	.184	1.368	.689	2.26
	SVEXPS	39.31	35.40	17.83	.474	.431	.206	1.274	.613	2.07
	TEOR	148.5	92.59	188.6	1.70	1.06	2.16	.000	.000	.00
SVEXPS	SVSBE(1,1)	22.48	19.70	12.44	.264	.232	.142	1.882	.967	3.22
	SVSBE(∞,1)	24.91	22.22	12.29	.303	.271	.149	1.775	.866	3.22
	SVSBE(∞,∞)	28.01	25.50	13.38	.340	.308	.162	1.725	.812	3.21
	SVCH2	48.75	41.15	28.39	.585	.499	.339	1.987	.929	3.68
	SVESFA	34.83	31.30	15.92	.427	.380	.198	1.762	.811	3.37
	SVEXPS	43.96	37.03	25.89	.522	.449	.300	1.590	.779	2.82
	TEOR	180.4	105.1	272.0	2.12	1.23	3.28	.000	.000	.00

the fit to the theoretical and empirical variograms:

$$RSET = \sum_{(i,j)\in I} \left(\frac{\bar{\gamma}(r_i, t_j)}{\gamma(r_i, t_j)} - 1 \right)^2,$$

$$RSE = \sum_{(i,j)\in I} \left(\frac{\hat{\gamma}(r_i, t_j)}{\bar{\gamma}(r_i, t_j)} - 1 \right)^2,$$

where $\bar{\gamma}(\cdot,\cdot)$ is the fitted semivariogram, $\gamma(\cdot,\cdot)$ the theoretical semivariogram and I the set of indices corresponding to the lags used for the fit.

The values obtained are shown in Table 3. With a view to facilitating a comparison of the results obtained with the adjusted models, the values corresponding to the theoretical variogram (TEOR) used to generate the observations were also calculated. If we compare the empirical semivariogram fits we can see that the extended Shapiro-Botha models are much more flexible and adapt better to the semivariogram estimates in all the cases (and among these, the more specific models are those which produce the best fits). If we compare the *RSET* values, as a general rule, the opposite occurs and the more specific extended Shapiro-Botha models are poorer fits. This might lead one to think that the classical semivariogram estimator is not very efficient (see the *WLS* and *RSE* values obtained using the theoretical semivariograms), and that other pilot estimates of the semivariogram might be more suitable; see the comments given at the end of Section 3.

Figure 4 compares the nugget effect estimates obtained using the different models. The estimates obtained with the extended Shapiro-Botha models are those nearest to the true value, and so we could say that the behavior of these models near the origin is quite satisfactory.

In order to diagnose the fit of the different models the cross-validation statistics described in the previous section were also calculated. In this case the calculations were not made locally and all remaining observations $\{Z(\mathbf{s}_j, t_j) : j \neq i\}$ were taken into account in order to obtain predictions and kriging variances for each location $(\mathbf{s}_i, t_i)$. Nonetheless, instead of calculating these values in the usual way (resolving N systems of order N), they were obtained on the basis of the kriging matrix constructed using all the data (bearing in mind the expressions of the kriging equations leaving out one datum and the fact that the kriging methods are exact interpolators, the expressions from which we can obtain the weights for the cross-validation can be easily deduced; see Dubrule 1983). Thus, by calculating the predictions, the cross-validation values were also calculated at no additional computational cost. Table 4 shows the values for the cross-validation statistics obtained. It can be seen that for the averaged squared error the extended Shapiro-Botha models are those that obtain the better values and, as with the example shown in the previous section, the most general model (and that with the poorest fit to the empirical semivariogram) is that with the lowest *CVASE*. If we compare the agreement values for the kriging variances and the observed variances (*CVDASE*), as well as the coverage of the prediction intervals, it can be also observed that SVSBE(∞, ∞) produces the most satisfactory behavior.

Table 4. *Cross-validation values*

MODEL		CVASE			CVDASE			CVCOV	
Theoretical	Fitted	Mean	Median	S. dev.	Mean	Median	S. dev.	Mean	S. dev.
SVCH2	SVSBE(1,1)	1.107	1.095	.172	1.044	1.037	.042	.940	.019
	SVSBE(∞,1)	1.119	1.107	.175	1.028	1.021	.039	.944	.019
	SVSBE(∞,∞)	1.119	1.107	.170	1.007	1.006	.020	.949	.015
	SVCH2	1.142	1.130	.173	1.016	1.013	.030	.948	.017
	SVESFA	1.150	1.133	.174	1.022	1.015	.036	.945	.018
	SVEXPS	1.175	1.159	.182	1.049	1.029	.069	.938	.025
	TEOR	1.137	1.128	.165	.993	.992	.073	.951	.022
SVESFA	SVSBE(1,1)	1.596	1.584	.264	1.050	1.045	.038	.939	.018
	SVSBE(∞,1)	1.516	1.506	.243	1.034	1.025	.057	.942	.022
	SVSBE(∞,∞)	1.519	1.505	.239	1.002	1.000	.020	.950	.016
	SVCH2	1.869	1.837	.327	1.017	1.012	.034	.948	.017
	SVESFA	1.581	1.570	.256	1.001	1.002	.025	.950	.016
	SVEXPS	1.583	1.575	.253	1.019	1.012	.039	.946	.018
	TEOR	1.563	1.550	.235	.994	.992	.075	.951	.022
SVEXPS	SVSBE(1,1)	1.393	1.383	.228	1.050	1.044	.042	.938	.019
	SVSBE(∞,1)	1.334	1.329	.211	1.034	1.024	.050	.942	.020
	SVSBE(∞,∞)	1.323	1.313	.207	1.002	1.001	.020	.951	.015
	SVCH2	1.571	1.536	.291	1.022	1.018	.035	.946	.017
	SVESFA	1.385	1.373	.211	1.007	1.007	.028	.949	.016
	SVEXPS	1.373	1.363	.213	1.026	1.014	.057	.945	.021
	TEOR	1.360	1.343	.203	.994	.990	.075	.952	.022

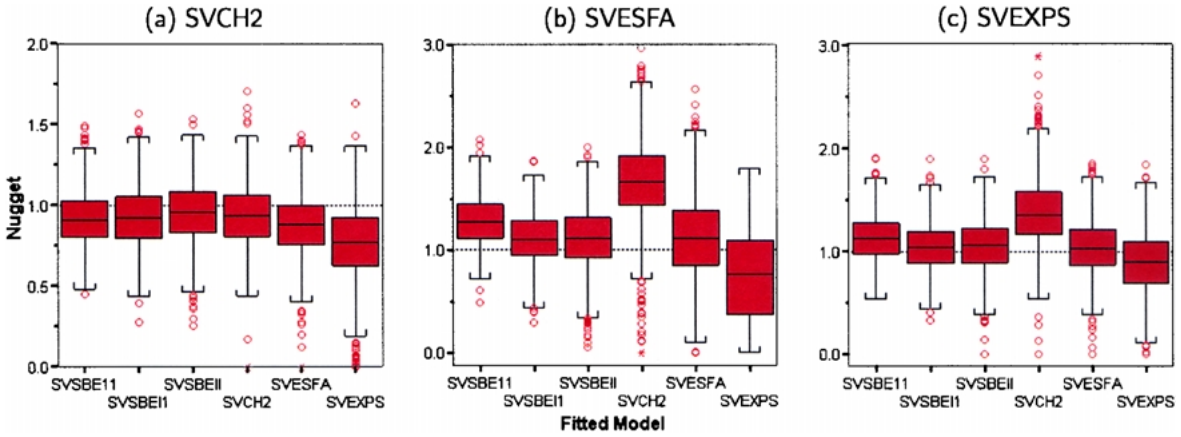

Fig. 4. *Nugget effect estimates obtained with the different models (the dotted lines represent the theoretical values)*

5.2. *Predictions*

In order to analyze the predictions obtained using the different models, the following measurements were calculated for each prediction location $(\mathbf{s}_0, t_0)$ (Fig. 3):

$$PASE = (\hat{Z}(\mathbf{s}_0, t_0) - Z(\mathbf{s}_0, t_0))^2,$$
$$PDMSE2 = ((\hat{Z}(\mathbf{s}_0, t_0) - Z(\mathbf{s}_0, t_0))/\sigma(\mathbf{s}_0, t_0))^2,$$
$$PCOV = 95\% \text{ coverage of the prediction interval (assuming normality).}$$

Table 5 shows a summary of the values obtained for these statistics in the simulations. It can be seen that, with respect to the averaged squared prediction error (*PASE*), the SVSBE(∞, ∞) model always behaves satisfactorily and is occasionally even the best performer, whereas the SVSBE(1,1) model is in general the poorest performer (along with the SVSBE(∞,1) model, obtaining intermediate values between the above two models). With regard to the kriging variance estimates, and bearing in mind that the values for the dimensionless averaged squared error (*PDMSE2*) corresponding to these models are large, it can be observed that the extended Shapiro-Botha models obtain

Table 5. *Efficiency measurements for the predictions (for all the prediction locations)*

MODEL		PASE			PDMSE2			PCOV	
Theoretical	Fitted	Mean	Median	S. dev.	Mean	Median	S. dev.	Mean	S. dev.
SVCH2	SVSBE(1,1)	1.207	.569	1.65	1.230	.572	1.70	.925	.264
	SVSBE(∞,1)	1.203	.570	1.65	1.193	.559	1.67	.927	.260
	SVSBE(∞,∞)	1.170	.534	1.60	1.109	.501	1.53	.936	.245
	SVCH2	1.163	.538	1.59	1.058	.488	1.46	.942	.233
	SVESFA	1.169	.535	1.60	1.109	.510	1.53	.937	.243
	SVEXPS	1.184	.544	1.62	1.067	.486	1.48	.941	.236
	TEOR	1.128	.522	1.53	.987	.456	1.34	.955	.208
SVESFA	SVSBE(1,1)	1.586	.732	2.19	1.143	.527	1.60	.931	.253
	SVSBE(∞,1)	1.528	.713	2.11	1.164	.528	1.65	.929	.256
	SVSBE(∞,∞)	1.512	.695	2.06	1.124	.510	1.57	.934	.249
	SVCH2	1.776	.855	2.44	.991	.478	1.39	.953	.212
	SVESFA	1.529	.699	2.08	1.045	.476	1.46	.942	.233
	SVEXPS	1.549	.730	2.12	.945	.441	1.30	.956	.205
	TEOR	1.450	.659	1.97	.988	.450	1.34	.955	.207
SVEXPS	SVSBE(1,1)	1.487	.692	2.04	1.227	.560	1.71	.924	.265
	SVSBE(∞,1)	1.476	.684	2.05	1.260	.562	1.82	.918	.274
	SVSBE(∞,∞)	1.452	.656	2.00	1.188	.532	1.67	.924	.265
	SVCH2	1.545	.705	2.12	1.036	.481	1.45	.944	.230
	SVESFA	1.442	.663	1.97	1.104	.508	1.53	.939	.239
	SVEXPS	1.495	.697	2.06	1.031	.484	1.43	.945	.227
	TEOR	1.392	.638	1.89	.988	.455	1.34	.956	.206

excessively optimistic prediction variances estimates. Consequently, and in contrast to the cross-validation values, the coverage of the prediction intervals by these models was reduced (possibly due to an underestimation of the semivariogram for large lags).

Finally, the possible existence in the fitted models of different behaviors dependent on prediction locations (interior, exterior, near to or far from the data) was also studied, but there were apparently no differences.

6. Conclusions

- A generalization of the Shapiro and Botha approach, as described in this paper, makes a wide range of stationary semivariogram models valid for the spatio-temporal case available (and can be easily extended to the case of two or more anisotropic components).
- The fit of the pilot semivariogram estimates, being linear functions of the parameters, can be easily carried out using quadratic programming (thus avoiding the problems of nonlinear multidimensional minimization that occur when using parametric models).
- The models obtained using this approach are extremely flexible and adaptable, thus the use of efficient pilot estimates being very important.
- The comparisons made in the simulation study indicate that the results obtained using the models proposed for kriging in this paper are similar to, and sometimes better than, those obtained using traditional parametric models.

Acknowledgments

The authors would like to thank Noel Cressie and Hsin-Cheng Huang for helping us with the wind speed data used in Section 4. This work was supported in part by grant PGIDT99MA 20701 of the Xunta de Galicia, Spain. In addition, research of Wenceslao González-Manteiga and Manuel Febrero-Bande was partially supported by grant PB98-0182-C02-02 of the DGESIC, Ministerio de Ciencia y Tecnología, Spain.

References

Abramowitz M. and Stegun I.A. 1965. Handbook of Mathematical Functions. Dover, New York.

Bochner S. 1959. Lectures on Fourier Integrals. Princeton University Press, New Jersey.

Cressie N. 1985. Fitting variogram models by weighted least squares. Mathematical Geology 17: 563–586.

Cressie N. 1993. Statistics for Spatial Data (rev. ed.). Wiley, New York.

Cressie N. and Huang H.-L. 1999. Classes of non-separable, spatio-temporal stationary covariance functions. *JASA* 94: 1330–1340.

Dubrule O. 1983. Cross validation of kriging in a unique neighborhood. Mathematical Geology 15: 687–699.

Guttorp P., Sampson P.D., and Newman K. 1992. Non-parametric estimation of spatial covariance with application to monitoring network evaluation. Statistics in the Environmental and Earth Sciences. Edward Arnold, London, pp. 39–51.

Haas T.C. 1995. Local prediction of a spatio-temporal process with an application to wet sulphate deposition. JASA 90: 1189–1199.

Jones R.H. and Zhang Y. 1997. Models for continuous stationary space-time processes. Modeling Longitudinal and Spatially Correlated Data. Lecture Notes in Statistics 122. Springer, New York, pp. 289–298.

Journel A.G. 1986. Geostatistics: Models and tools for the earth sciences. Mathematical Geology 18: 119–140.

Journel A.G. and Huijbregts C.J. 1978. Mining Geostatistics. Academic Press, London.

Kyriakidis P. and Journel A. 1999. Geostatistical space-time models: A review. Mathematical Geology 31: 651–684.

Shapiro A. and Botha J.D. 1991. Variogram fitting with a general class of conditionally non-negative definite functions. Computational Statistics and Data Analysis 11: 87–96.

Stein M. 1999. Interpolation of Spatial Data: Some Theory for Kriging. Springer, New York.