MOX–Report No. 21/2009

A Bayesian approach to geostatistical interpolation with flexible variogram models

luca Bonaventura, Stefano Castruccio,
Laura M. Sangalli

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

A Bayesian approach to geostatistical interpolation with flexible variogram models

Luca Bonaventura, Stefano Castruccio*,
Laura M. Sangalli

July 21, 2009

MOX – Modelling and Scientific Computing,
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
luca.bonaventura@polimi.it, laura.sangalli@polimi.it

Keywords: Markov Chain Monte Carlo methods, kriging, variogram estimation.

AMS Subject Classification: 62-07, 62N99, 62P12, 62P30, 62F03.

Abstract

A Bayesian approach to covariance estimation and geostatistical interpolation based on flexible variogram models is introduced. In particular, we consider black-box kriging models. These variogram models do not require restrictive assumptions on the functional shape of the variogram; furthermore, they can handle quite naturally non isotropic random fields. The proposed Bayesian approach does not require the computation of an empirical variogram estimator, thus avoiding the arbitrariness implied by the construction of the empirical variogram itself. Moreover, it provides a complete assessment of the uncertainty in the variogram estimation. The advantages of this approach are illustrated via an extensive simulation study and by application to a well known benchmark dataset.

*Present address: Department of Statistics, The University of Chicago, 5734 S. University Avenue, 60637 Chicago, IL, castruccio@uchicago.edu

1 Introduction

Accurate variogram estimation is at the core of geostatistical interpolation procedures and is crucial for making reliable predictions on the basis of spatially correlated data. Traditionally, isotropy assumptions are required and variogram estimation is achieved by generalized least square fitting procedures (see e.g. Christensen (1991), Cressie (1991), Wackernagel (1995), Walden and Guttorp (1992) for a complete discussion). These fitting approaches require the construction of an empirical variogram estimator and an assumption concerning the variogram shape. However, each of these steps allows for remarkable arbitrariness. Empirical variogram estimators are built clustering data pairs into classes according to the distance between the two points, but the selection of such distance classes is not uniquely dictated by the data. Their choice and that of the variogram model to be used in the fit are mostly done *ad hoc* and require substantial user expertise in order to give reasonable results.

These estimation procedures can be improved in two different ways. On the one hand, the uncertainty in the variogram parameters can be modeled in a Bayesian framework, in order to estimate their posterior distribution with respect to the available data (the reader may refer e.g. to Bernardo and Smith (1994) for an introduction to the basic concepts and definitions of Bayesian statistics). This leads to the so called Bayesian kriging approaches, proposed e.g. in Berger et al. (2001) and Handcock and Stein (1993). On the other hand, variogram estimation procedures can be devised that do not require so many restrictive *ad hoc* assumptions. For example, more flexible variogram models have been proposed in Shapiro and Botha (1991), where a cosine series variogram model was introduced, and in Im et al. (2007), where a semiparametric form of the spectral density was considered, consisting in a combination of cubic splines for low frequencies and of a polynomially decreasing tail for high frequencies. Other non parametric approaches were proposed e.g. in Gorsich and Genton (2000).

An interesting proposal in the latter direction, albeit limited to fields with second order moments, is the flexible variogram model introduced in Barry and Ver Hoef (1996) and also known as *black-box kriging* approach. This approach exploits a special representation theorem for variogram functions that allows to characterize a subset of piecewise polynomial functions as valid variograms with sill. Functions of this shape are then fitted to the traditional empirical variogram estimates, thus yielding a very flexible procedure that does not require to specify a variogram model of fixed functional form. Indeed, it is proven in Barry and Ver Hoef (1996) that, in the one-dimensional case, any generic variogram with sill can be approximated by a valid piecewise linear one. Another advantage of this approach is that it handles quite naturally non isotropic random fields.

In the present work, we propose a Bayesian approach to kriging with flexible variogram models. The parameters characterizing a generic piecewise linear valid variogram, according to the representation theorem of Barry and Ver Hoef (1996), are assumed to be random variables with a chosen *a priori* distribution. The *a posteriori* distribution of these parameters given the available data can be computed by an appropriate Markov Chain

Monte Carlo (MCMC) scheme, and a complete assessment of the uncertainty in the variogram estimation is achieved (see e.g. Gilks et al. (1998) and Robert and Casella (2004) for an introduction to MCMC methods). The proposed technique addresses both previously reviewed shortcomings of traditional techniques and constitutes a further step towards reducing the need for expert user intervention in the variogram model choice, that restricts in many cases the applicability of geostatistical interpolation techniques. With respect to the standard weighted least square variogram estimation used in Barry and Ver Hoef (1996), the present Bayesian approach does not require the computation of an empirical variogram estimator, thus avoiding the potential arbitrariness implied in the construction of the empirical variogram itself. Furthermore, as in Barry and Ver Hoef (1996), anisotropy can be accounted for in a very straightforward way.

The advantages of the present method are illustrated by an extensive simulation study with synthetic one-dimensional data, as well as by an application to the well known Wolfcamp aquifer benchmark dataset (see e.g. Harper and Furr (1986), Cressie (1991) and Barry and Ver Hoef (1996)). Kriging reconstructions obtained with the estimated variograms are also presented and validated. It is shown that the estimates obtained by the present approach yield significant improvements over the fitting technique used in the original black-box kriging approach.

In section 2, the flexible black-box variogram models are reviewed. In section 3, the Bayesian approach to flexible black-box variogram estimation is introduced, while in section 4 the MCMC algorithm used to deal with the proposed model is described. The results of simulation studies on synthetic data and of applications to real data are presented in section 5, while in section 6 we draw some conclusive remarks.

2 Flexible black-box variogram models

Geostatistical interpolation is usually formulated assuming that the data consist in a realization of a random field $Z : D \times \Omega \to \mathbb{R}, D \subset \mathbb{R}^d$ with a deterministic mean $m(\mathbf{x})$ and a valid semivariogram function $\gamma(\mathbf{h}) = \mathbb{E}[(Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x}))^2]/2$. The classical characterizazion of valid variograms is given in terms of conditionally negative definite functions, i.e., $\gamma(\mathbf{h})$ is a valid semivariogram function if and only if

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \gamma(\mathbf{x}_i - \mathbf{x}_j) \leq 0,$$

for all $n \in \mathbb{N}$, all $\mathbf{x}_i, \mathbf{x}_j \in D$ and all $a_1, \ldots, a_n \in \mathbb{R}$ such that $\sum_{i=1}^{n} a_i = 0$. In general, a piecewise linear function (in more than one dimension, a piecewise multilinear one) is not conditionally negative definite, so that simple interpolation of the values of an empirical variogram estimator does not yield a valid variogram function.

It was proven in Barry and Ver Hoef (1996) that, for $d = 1$, under the assumption that the semivariogram is constant for $h > c$, with $c > 0$ given,

the function 2γ can be represented as

$$2\gamma(h) = \int_{\mathbb{R}} [f(x) - f(x-h)]^2 dx, \tag{1}$$

where f is a measurable function. The main point of the flexible variogram model consists in choosing a piecewise constant function f, to yield as a consequence piecewise linear valid variogram. More specifically, for any positive integer k and vector of positive real numbers $\mathbf{a} = (a_1, \ldots, a_k)$, define the function f with support $[0, c]$ by

$$f(x; \mathbf{a}, c, k) = \sum_{j=1}^{k} a_j I\left(\frac{(j-1)c}{k} < x \leq \frac{jc}{k}\right), \tag{2}$$

where $I(\cdot)$ denotes the indicator function. The function f is piecewise constant. Using (2) in the representation theorem (1), after some algebra we have an explicit expression of the semivariogram. For convenience, values at the nodal points $h = mc/k$, for $m = 1, \ldots, k$, are first computed, and the remaining values are then recovered by linear interpolation, which is justified since the resulting function is indeed piecewise linear. The resulting semivariogram can then be described as follows for $h > 0$:

- if $h \geq c$

$$2\gamma(h; \mathbf{a}, c, k) = \frac{2c}{k} \sum_{i=1}^{k} a_i^2 \,;$$

- if $h < c$ and there exists an integer m such that $h = mc/k$,

$$2\gamma(h; \mathbf{a}, c, k) = \frac{2c}{k} \sum_{i=1}^{k} a_i^2 - \frac{2c}{k} \sum_{i=m+1}^{k} a_i a_{i-m} \,;$$

- if $h < c$, but h is not an integer multiple of c/k,

$$2\gamma(h; \mathbf{a}, c, k) = (1 - V) 2\gamma\left(\frac{m_l c}{k}; \mathbf{a}, c, k\right) + V 2\gamma\left(\frac{m_u c}{k}; \mathbf{a}, c, k\right)$$

  where $m_l = \lfloor hk/c \rfloor$ and $m_u = \lceil hk/c \rceil$ and $V = (h - m_l c/k)/(c/k)$, that is, the value of the semivariogram is given by linear interpolation of the two values at the nearest multiple integers of c/k enclosing h.

In Barry and Ver Hoef (1996), specific variograms were then obtained by fixing k and c and estimating the a_i from the data, starting from standard empirical estimators such as those proposed in Cressie and Hawkins (1980) and Hawkins and Cressie (1984), and applying a weighted least square (WLS) algorithm. The integer k represents the number of equal size intervals in which $[0, c]$ is divided and over which the variogram is represented by a different linear function; hence, k influences directly the complexity of the variogram model. In general, k has to be smaller than the number of different lags used in an empirical variogram estimator.

The representation theorem introduced above also holds in the multidimensional case, so that for $d > 1$ one has

$$2\gamma(\mathbf{h}) = \int_{\mathbb{R}^d} [f(\mathbf{x}) - f(\mathbf{x} - \mathbf{h})]^2 d\mathbf{x}. \tag{3}$$

In the following, only the two-dimensional case shall be considered. More specifically, along the lines of Barry and Ver Hoef (1996), we can define piecewise constant functions on the two-dimensional rectangular domain $[0, c_1] \times [0, c_2]$ as

$$f\big((x_1, x_2); \mathbf{A}, \mathbf{c}, \mathbf{k}\big) = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} a_{i,j} I\left[\left(\frac{(i-1)c_1}{k_1} < x_1 \leq \frac{ic_1}{k_1}\right)\left(\frac{(j-1)c_2}{k_2} < x_2 \leq \frac{jc_2}{k_2}\right)\right] \tag{4}$$

where $\mathbf{c} = (c_1, c_2)$, $\mathbf{k} = (k_1, k_2)$ with k_1, k_2 positive integers, and $\mathbf{A}$ is a matrix of positive real numbers with entries $\{\mathbf{A}\}_{ij} = a_{i,j}$. Substituting (4) into (3) yields then a valid variogram function. Since in general $2\gamma(h_1, h_2) = 2\gamma(-h_1, -h_2)$ and $2\gamma(h_1, h_2) \neq 2\gamma(h_1, -h_2)$, it is sufficient to compute the variogram for $h_1 > 0$ only. As for the one-dimensional case, values at the nodal points $(h_1, h_2) = (m_1 c_1/k_1, m_2 c_2/k_2)$, for $m_1 = 1, \ldots, k_1$ and $m_2 = -k_2, \ldots, -1, 1, \ldots, k_2$, are first computed, and the remaining values are recovered by bilinear interpolation. The resulting anisotropic piecewise bilinear semivariogram can be described as follows:

- if $h_1 \geq c_1$ or $|h_2| \geq c_2$,

$$2\gamma(h_1, h_2; \mathbf{A}, \mathbf{c}, \mathbf{k}) = \frac{2cd}{k_1 k_2} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} a_{i,j}^2 \,;$$

- if $0 < h_1 < c_1$ and $0 < h_2 < c_2$, with $h_1 = m_1 c_1/k_1$ and $h_2 = m_2 c_2/k_2$ for some positive integers m_1 and m_2,

$$2\gamma(h_1, h_2; \mathbf{A}, \mathbf{c}, \mathbf{k}) = \frac{2cd}{k_1 k_2} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} a_{i,j}^2 - \frac{2cd}{k_1 k_2} \sum_{i=m_1+1}^{k_1} \sum_{j=m_2+1}^{k_2} a_{i,j} a_{i-m_1, j-m_2} \,;$$

- if $0 < h_1 < c_1$ and $-c_2 < h_2 < 0$, with $h_1 = m_1 c_1/k_1$ and $h_2 = m_2 c_2/k_2$ for some integers $m_1 > 0$ and $m_2 < 0$,

$$2\gamma(h_1, h_2; \mathbf{A}, \mathbf{c}, \mathbf{k}) = \frac{2cd}{k_1 k_2} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} a_{i,j}^2 - \frac{2cd}{k_1 k_2} \sum_{i=m_1+1}^{k_1} \sum_{j=1}^{k_2+m_2} a_{i,j} a_{i-m_1, j-m_2} \,.$$

In the case of arbitrary lag values, the semivariogram is computed by bilinear interpolation between the values of the variogram on the corners of the rectangle containing (h_1, h_2) whose vertices are the nearest integer multiples of c_1/k_1 and c_2/k_2.

The variograms obtained by this procedure are clearly anisotropic. An interesting point, apparently not dealt with in the original paper, is whether

the isotropic piecewise linear one-dimensional model could be extended to define an isotropic multidimensional model. Following the same approach as in the one-dimensional case, one could define a function that is piecewise constant in the radial direction

$$\tilde{f}(\mathbf{x};\mathbf{a},c,k)=\sum_{i=1}^{k}a_i I\left((i-1)c/k<\|\mathbf{x}\|\leq ic/k\right). \tag{5}$$

However, substitution of (5) into (3) does not yield a piecewise linear variogram, because when using polar coordinates to carry out the integration analytically, the jacobian factor $\rho\, d\rho\, d\theta$ leads to a piecewise quadratic function. More generally, for a generic $d-$dimensional field the jacobian of the coordinate transformation to hyperspherical coordinates with angles $\phi_1,\ldots,\phi_d$ would be given by

$$\rho^{d-1}\sin^{d-2}(\phi_1)\sin^{d-3}(\phi_2)\ldots\sin(\phi_{d-2})d\rho d\phi_1\ldots d\phi_{d-1},$$

so that piecewise polynomials of increasing order would arise. Thus, it appears to be impossible to have a multidimensional variogram that is both isotropic and piecewise linear using this representation theorem. Although the piecewise quadratic form could possibly turn out to be useful, we have not pursued its application within this work.

Finally, it should be noticed that the flexible variogram model defined by (1)-(2) is not identifiable in the black-box parameters $a_i, i=1,\ldots,k$, and the same can be said for the two-dimensional flexible variogram model defined by (3)-(4). For example, in the one-dimensional case with $k=2$ one has

$$2\gamma\left(\frac{c}{2};(a_1,a_2),c,k\right)=\frac{c}{2}a_1^2+\frac{c}{2}a_2^2-\frac{c}{2}a_1a_2,$$

that implies $2\gamma(c/2;(a_1,a_2),c,k)=2\gamma(c/2;(a_2,a_1),c,k)$. Moreover, for $k>2$ the class of k-uples which yield the same value of γ is not just a permutation of the vector $(a_1,\ldots,a_k)$, but a rather more complex symmetry class. On the other hand, the real interest lies on the estimation of γ itself, rather than on the black-box parameters $a_i, i=1,\ldots,k$. Thus, we shall regard the parameter space of the $a_i, i=1,\ldots,k$ as a quotient space with equivalence classes identified by the corresponding values of $\gamma(h;\mathbf{a},c,k)$ (and similarly for the two-dimensional case).

3 A Bayesian approach for the estimation of flexible variogram models

A novel Bayesian approach for the estimation of flexible black-box variogram models is now introduced. The model for geostatistical data considered in the following shall consist of realizations of a Gaussian random field $Z: D\times\Omega\to\mathbb{R}, D\subset\mathbb{R}^d$ (where $d=1,2$), that can be written as the sum $Z=\mu+\delta$, where

- the mean field $\mu(\mathbf{x})=\sum_{j=0}^{p}\beta_j g_j(\mathbf{x})$ is given by a linear combination of known functions $g_j(\mathbf{x})$ with random coefficients $\boldsymbol{\beta}=[\beta_1,\ldots,\beta_p]^T$;

- the field δ is a second order stationary zero-mean Gaussian random field with flexible semivariogram $\gamma(\mathbf{h}; \mathbf{A}, \mathbf{c}, \mathbf{k})$, where the parameters $\mathbf{A}$ are random and the parameters $\mathbf{c}$ and $\mathbf{k}$ are fixed (recall in particular that $\mathbf{k}$ influences directly the complexity of the model).

As a result, the data $\mathbf{Z} = [Z(\mathbf{x}_1), \ldots, Z(\mathbf{x}_N)]^T$, at N distinct locations $\mathbf{x}_i, \ldots, \mathbf{x}_N$ in space, can be represented as

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_{\mathbf{A}})$$

where $\{\mathbf{X}\}_{ij} = g_i(\mathbf{x}_j)$ is the design matrix and the covariance matrix has entries

$$\{\boldsymbol{\Sigma}_{\mathbf{A}}\}_{ij} = \gamma(\infty; \mathbf{A}, \mathbf{c}, \mathbf{k}) - \gamma(\mathbf{x}_i - \mathbf{x}_j; \mathbf{A}, \mathbf{c}, \mathbf{k}).$$

The likelihood function is thus given by

$$l(\mathbf{Z}|\boldsymbol{\beta}, \mathbf{A}) = \frac{1}{(2\pi)^{N/2}\sqrt{|\boldsymbol{\Sigma}_{\mathbf{A}}|}} \exp\left[-\frac{1}{2}\left(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}\right)^T \boldsymbol{\Sigma}_{\mathbf{A}}^{-1}\left(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}\right)\right].$$

We assume that γ and $\boldsymbol{\beta}$ are stochastically independent *a priori*, i.e. that the parameters $\mathbf{A}$ and $\boldsymbol{\beta}$ are *a priori* independent. For simplicity, we also assume that the black-box parameters in $\mathbf{A}$ are *a priori* independent among themselves. For these black-box parameters we shall always assume non-informative priors. In the simulations and applications shown in Section 5 we will use exponential priors with large mean parameters; we will also show that the results obtained are robust with respect to the choice of non-informative priors with different forms, such us Gamma or Chi-squared non-informative distributions. For $\boldsymbol{\beta}$ we choose a Gaussian prior distribution with mean $\mathbf{m}$ and covariance matrix $\mathbf{G}$ (with $\mathbf{m}$ and $\mathbf{G}$ fixed). This prior distribution has the advantage of being conjugate with the likelihood, so that it is possible to derive analytically the conditional distribution of $\boldsymbol{\beta}$ given $\mathbf{Z}$ and $\mathbf{A}$. Denoting by $\pi_{\boldsymbol{\beta}}$ and $\pi_{\mathbf{A}}$ the priors of $\boldsymbol{\beta}$ and $\mathbf{A}$ respectively, we thus have that the joint posterior distribution of $\boldsymbol{\beta}$ and $\mathbf{A}$ is given by

$$\pi(\boldsymbol{\beta}, \mathbf{A}|\mathbf{Z}) = C\, \pi_{\boldsymbol{\beta}}(\boldsymbol{\beta})\, \pi_{\mathbf{A}}(\mathbf{A})\, l(\mathbf{Z}|\boldsymbol{\beta}, \mathbf{A}) \tag{6}$$

where C is a normalizing constant. This posterior distribution cannot be easily computed analytically, so that an appropriate MCMC sampler must be employed. In the next section we shall describe in detail a MCMC scheme that can be used for sampling from (6). As a result of this numerical sampling, information about the *a posteriori* variability of γ can be recovered. In particular, the posterior estimate of the semivariogram $\hat{\gamma}$ is computed by averaging the semivariograms determined along the Markov chain:

$$\hat{\gamma}(\cdot) = \frac{1}{V - W} \sum_{v=W+1}^{V} \gamma(\cdot; \mathbf{A}^{[v]}, \mathbf{c}, \mathbf{k})$$

where $\mathbf{A}^{[v]}$ are the black-box parameters sampled at the v-th iteration, V is the total number of iterations and W is the number of initial iterations discarded as burn-in of the Markov chain (to reduce the correlation among adjoining values of the chain, the estimate $\hat{\gamma}$ may also be computed averaging only the semivariograms computed each M iterations, for M large enough).

Furthermore, the predictive distribution and the posterior kriging variance distribution can be recovered by solving the standard universal kriging equations. More specifically, we can compute the optimal linear estimate $\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{N} \lambda_i Z(\mathbf{x}_i)$ at some point $\mathbf{x}_0$ where no data is available, proceeding as follows. Define the vector $\hat{\boldsymbol{\gamma}}_U = [\hat{\gamma}(\mathbf{x}_0, \mathbf{x}_1), \cdots, \hat{\gamma}(\mathbf{x}_0, \mathbf{x}_N), 1, \cdots, 1]$ and the universal kriging matrix

$$\hat{\boldsymbol{\Gamma}}_U = \begin{bmatrix} \hat{\gamma}(\mathbf{x}_1, \mathbf{x}_1) & \dots & \hat{\gamma}(\mathbf{x}_1, \mathbf{x}_N) & g_1(\mathbf{x}_1) & \dots & g_p(\mathbf{x}_1) \\ \hat{\gamma}(\mathbf{x}_2, \mathbf{x}_1) & \dots & \hat{\gamma}(\mathbf{x}_2, \mathbf{x}_N) & g_1(\mathbf{x}_2) & \dots & g_p(\mathbf{x}_2) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \hat{\gamma}(\mathbf{x}_N, \mathbf{x}_1) & \dots & \hat{\gamma}(\mathbf{x}_N, \mathbf{x}_N) & g_1(\mathbf{x}_N) & \dots & g_p(\mathbf{x}_N) \\ g_1(\mathbf{x}_1) & \dots & g_1(\mathbf{x}_N) & 0 & \dots & 0 \\ \dots & \dots & \dots & 0 & \dots & 0 \\ g_p(\mathbf{x}_1) & \dots & g_p(\mathbf{x}_N) & 0 & \dots & 0 \end{bmatrix}.$$

The kriging coefficients $\lambda_1, \cdots, \lambda_N$ can then be computed by solving the linear system $\hat{\boldsymbol{\Gamma}}_U \boldsymbol{\lambda}_U = \hat{\boldsymbol{\gamma}}_U$, where $\boldsymbol{\lambda}_U = [\lambda_1, \cdots, \lambda_N, \beta_0, \cdots, \beta_p]^T$; furthermore, the kriging variance can be computed as

$$\hat{\sigma}^2_U(\mathbf{x}_0) = \mathbb{E}\left(Z(\mathbf{x}_0) - \hat{Z}(\mathbf{x}_0)\right)^2 = \boldsymbol{\lambda}_U^T \hat{\boldsymbol{\gamma}}_U = \hat{\boldsymbol{\gamma}}_U^T \hat{\boldsymbol{\Gamma}}_U^{-1} \hat{\boldsymbol{\gamma}}_U.$$

This is one of the distinctive advantages of Bayesian kriging, since in this way the complete one-dimensional distribution of the reconstructed field is estimated, which is exactly the type of information required in many applications. It should be remarked that, in this estimation process, no empirical variogram estimator is employed, thus avoiding the potential arbitrariness implied in the construction of the empirical variogram itself.

4 The MCMC sampler

We now give the details of a MCMC algorithm that samples from the posterior distribution of $\boldsymbol{\beta}$ and $\mathbf{A}$. A Gibbs sampling algorithm for sampling from (6) alternates the following steps:

1. simulation of $\boldsymbol{\beta}$, conditional on the observations $\mathbf{Z}$ and the current value of the parameter $\mathbf{A}$;
2. simulation of $\mathbf{A}$, conditional on the observations $\mathbf{Z}$ and the current value of the parameter $\boldsymbol{\beta}$.

The update of $\boldsymbol{\beta}$ is straightforward. Having chosen a conjugate prior for $\boldsymbol{\beta}$, it is possible to derive analytically the conditional distribution of $\boldsymbol{\beta}$ given $\mathbf{Z}$ and $\mathbf{A}$, which is still Gaussian with mean vector

$$(\mathbf{G}^{-1} + \mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{A}}^{-1} \mathbf{X})^{-1} (\mathbf{G}^{-1} \mathbf{m} + \mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{A}}^{-1} \mathbf{Z})$$

and variance matrix

$$(\mathbf{G}^{-1} + \mathbf{X}^T \boldsymbol{\Sigma}_{\mathbf{A}}^{-1} \mathbf{X})^{-1}.$$

The first step is thus performed by sampling directly from this conditional distribution.

The update of the semivariogram $\gamma(\cdot; \mathbf{A}, \mathbf{c}, \mathbf{k})$, which is performed via the update of the black-box parameters $\mathbf{A}$, is computationally more demanding. If the parameters $a_{i,j}$ in $\mathbf{A}$ are chosen to be *a priori* independent among themselves, i.e. $\pi_{\mathbf{A}}(\mathbf{A}) = \prod_{i=1}^{k_1} \prod_{j=1}^{k_2} \pi_a(a_{i,j})$, where π_a is the common prior distribution of the parameters $a_{i,j}$, then step 2 can be carried out as follows:

2. for $i \in \{1, \ldots, k_1\}$ and $j \in \{1, \ldots, k_2\}$, simulation of $a_{i,j}$ conditional on the observations $\mathbf{Z}$ and the current values of the parameters $\boldsymbol{\beta}$ and $\mathbf{A}_{\sim(i,j)}$, where $\mathbf{A}_{\sim(i,j)}$ is the set of parameters in $\mathbf{A}$ with $a_{i,j}$ removed.

The conditional distribution distribution of $a_{i,j}$ given $\mathbf{Z}$, $\boldsymbol{\beta}$ and $\mathbf{A}_{\sim(i,j)}$ is proportional to $\pi(a_{i,j})\, l(\mathbf{Z}|\mathbf{A}, \boldsymbol{\beta})$, and cannot be sampled directly. An appropriate Metropolis-Hastings step is thus required. We can for example use a multiplicative random walk sampler. Setting $w_{i,j} = \log(a_{i,j})$, from the current state $a_{i,j} = \exp\{w_{i,j}\}$ we propose a move to $a^*_{i,j} = \exp\{w_{i,j} + N\}$ where $N \sim \mathcal{N}(0, \sigma^2)$, for some fixed σ. With the change of variable from $a_{i,j}$ to $w_{i,j}$ the invariant distribution becomes $a_{i,j}\pi_a(a_{i,j})l(\mathbf{Z}|\mathbf{A}, \boldsymbol{\beta})$, so that the move is accepted with probability

$$\min\left\{1,\ \frac{a^*_{i,j}\, \pi_a(a^*_{i,j})\, l(\mathbf{Z}|\mathbf{A}^*, \boldsymbol{\beta})}{a_{i,j}\, \pi_a(a_{i,j})\, l(\mathbf{Z}|\mathbf{A}, \boldsymbol{\beta})}\right\}$$

where $\mathbf{A}^*$ coincides with $\mathbf{A}$ apart for the (i,j)-entry which is replaced by $a^*_{i,j}$. In particular, σ can be chosen by tuning its value over short chains, in order to achieve a sufficient acceptance ratio. Note that instead of updating the parameters $a_{i,j}$ in some fixed order, these can also be updated in a random order, by sampling, at each MCMC iteration, a random permutation of the indices.

5 Simulation studies and applications to real data

We now illustrate the Bayesian flexible black-box variogram estimation via simulation studies and an application to a real data set. In particular, Section 5.1 deals with simulated one-dimensional Gaussian random fields, while Section 5.2 shows an application to a well known benchmark in geostatistical interpolation, the Wolfcamp aquifer dataset, that was originally described in Harper and Furr (1986).

We thoroughly compare the proposed technique to the original flexible variogram estimation procedure used in Barry and Ver Hoef (1996), highlighting the advantages of the Bayesian approach.

5.1 One-dimensional synthetic data

We consider the model $Z : [0, L] \times \Omega \to \mathbb{R}$, $Z = \delta$, where δ is a stationary zero-mean Gaussian random field having semivariogram γ. In particular,

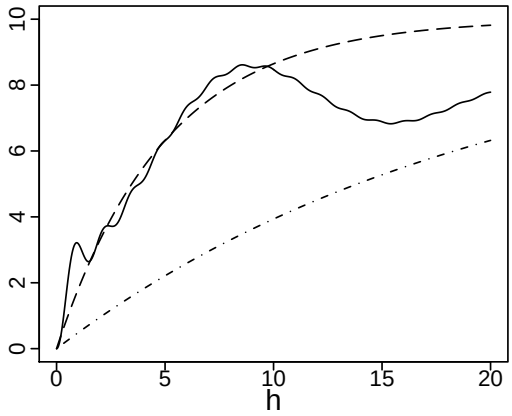


Figure 1: Three considered semivariograms: γ_1 (dash and dot line), γ_2 (dashed line) and γ_3 (solid line).

we shall consider different semivariogram shapes:

$$\gamma_1(h) = \frac{L}{2}\left(1 - \exp\left\{-\frac{h}{L}\right\}\right),$$

$$\gamma_2(h) = \frac{L}{2}\left(1 - \exp\left\{-\frac{h}{L/4}\right\}\right),$$

$$\gamma_3(h) = \frac{L}{8}\left(1 - \frac{L}{100}\frac{\sin(100h/L)}{h}\right) + \frac{L}{4}\left(1 - \frac{L}{10}\frac{\sin(10h/L)}{h}\right),$$

for $0 < h < L$, $L = 20$. The three above semivariograms are shown in Figure 1.

For either semivariogram, we simulate the field (at N randomly chosen locations) as described e.g. in Cressie (1991). From the sampled field, we estimate the semivariogram by a black-box semivariogram with k black-box parameters, using the WLS technique described in Barry and Ver Hoef (1996), and using the Bayesian approach proposed here. In particular, the Bayesain black-box kriging estimates are obtained by running the Hastings-within-Gibbs algorithm, that has been described in Section 4, under the following specifications: the prior for the black-box parameters are independent exponentials with mean 10; the algorithm is run for 5000 iterations and estimates are obtained using the values sampled every 50 iterations of the chain.

The estimates obtained by Bayesian black-box kriging and by WLS black-box kriging are thus compared in term of the following errors:

- the $L1$ relative error $\|\hat{\gamma} - \gamma\|_1 / \|\gamma\|_1$, where γ is the true variogram, $\hat{\gamma}$ is the variogram estimate, and $\|\cdot\|_1$ is the L^1 norm over the interval $[0, L]$;
- the leave-one-out root mean squared error (RMSE).

In particular, we proceed along the following simulation scheme. For 100 times,

- we sample N independent locations from a uniform distribution over the interval $[0, L]$;
- we generate the random field at the N sampled locations;
- we compute the Bayesian black-box kriging estimate, with k black-box parameters, and compute its $L1$ relative error and RMSE;
- we compute the WLS black-box kriging estimate, with k black-box parameters, and compute its $L1$ relative error and RMSE.

This simulation scheme is repeated for two different values of N, $N = 20, 30$, and for each of the three semivariograms, $\gamma_1, \gamma_2, \gamma_3$; moreover, the semivariogram estimates are computed for two different values of k, $k = 4, 8$. We thus have 12 different cases (depending on the semivariogram, the value of N, and the value of k), each composed of 100 simulations.

γ	N	k	Bayes bbk $L1$ rel. err.	WLS bbk $L1$ rel. err.	p-value	Bayes bbk RMSE	WLS bbk RMSE	p-value
γ_1	20	4	0.345 (0.302)	0.590 (0.486)	<0.001	0.607 (0.206)	0.665 (0.210)	<0.001
γ_1	20	8	0.324 (0.371)	0.597 (0.592)	<0.001	0.563 (0.159)	0.662 (0.242)	<0.001
γ_1	30	4	0.407 (0.374)	0.492 (0.487)	0.04	0.512 (0.123)	0.536 (0.160)	<0.001
γ_1	30	8	0.564 (0.557)	0.527 (0.563)	0.02	0.489 (0.113)	0.587 (0.210)	<0.001
γ_2	20	4	0.337 (0.209)	0.518 (0.659)	<0.001	1.165 (0.368)	1.307 (0.497)	<0.001
γ_2	20	8	0.471 (0.190)	0.573 (0.874)	<0.001	1.042 (0.332)	1.230 (0.405)	<0.001
γ_2	30	4	0.187 (0.128)	0.466 (0.498)	<0.001	0.988 (0.243)	1.063 (0.352)	<0.001
γ_2	30	8	0.295 (0.189)	0.494 (0.454)	<0.001	0.987 (0.225)	1.199 (0.417)	<0.001
γ_3	20	4	0.333 (0.167)	0.519 (0.758)	<0.001	1.510 (0.503)	1.585 (0.682)	<0.001
γ_3	20	8	0.444 (0.195)	0.556 (0.682)	<0.001	1.315 (0.332)	1.623 (0.523)	<0.001
γ_3	30	4	0.235 (0.121)	0.447 (0.571)	<0.001	1.305 (0.404)	1.393 (0.436)	<0.001
γ_3	30	8	0.267 (0.140)	0.473 (0.527)	<0.001	1.189 (0.252)	1.374 (0.376)	<0.001

Table 1: Median (inter-quantile range) of $L1$ relative error and RMSE for Bayesian black-box kriging estimates and WLS black-box kriging estimates; results of nonparametric paired Wilcoxon tests.

Table 1, reports, for each of the 12 cases, the median and inter-quantile range of the $L1$ relative error and RMSE for the 100 semivariograms estimates obtained by Bayesian black-box kriging, and for the 100 semivariograms estimates obtained by WLS black-box kriging. For each of the 12 cases, we perform a nonparametric paired Wilcoxon test (see, e.g., Hollander and Wolfe, 1999), to verify if the distribution of the $L1$ relative error of Bayesian estimates is stochastically lower than the distribution of the $L1$ relative error of WLS estimates; the same is done for the distributions of RMSE. The p-values of these tests, reported in the sixth and ninth column of the table, show that the $L1$ relative errors and RMSEs of Bayesian black-box kriging estimates are significantly lower than the ones of WLS black-box kriging estimates. Figure 2 illustrates graphically the superiority of Bayesian black-box kriging estimates over WLS black-box kriging estimates, displaying the boxplots of the distributions of $L1$ relative error and RMSE, for the case $N = 20$, $k = 4$. The analysis of the boxplots shows that the errors of Bayesian estimates not only are lower, but also have a smaller

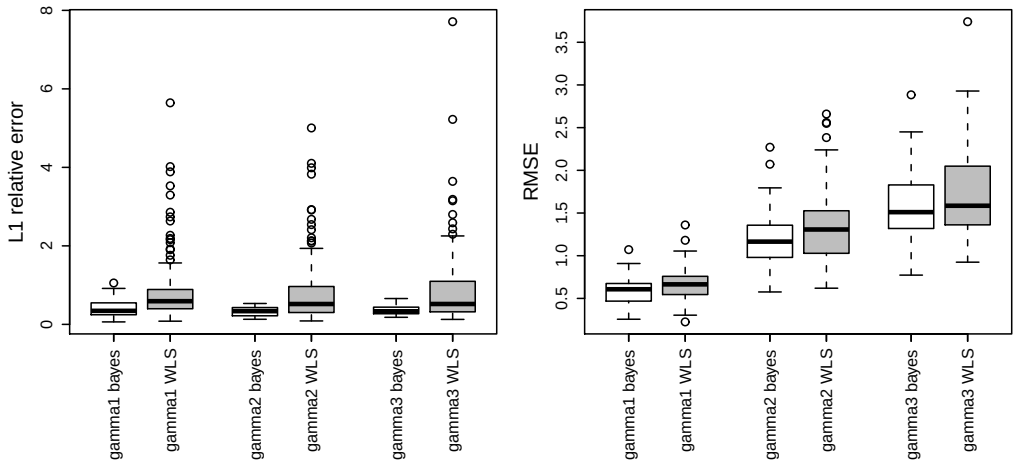


Figure 2: Boxplots of the distributions of $L1$ relative error and RMSE, for the case $N = 20$, $k = 4$.

variability, i.e., Bayesian black-box kriging estimates, besides being more accurate, are also more robust than WLS black-box kriging estimates.

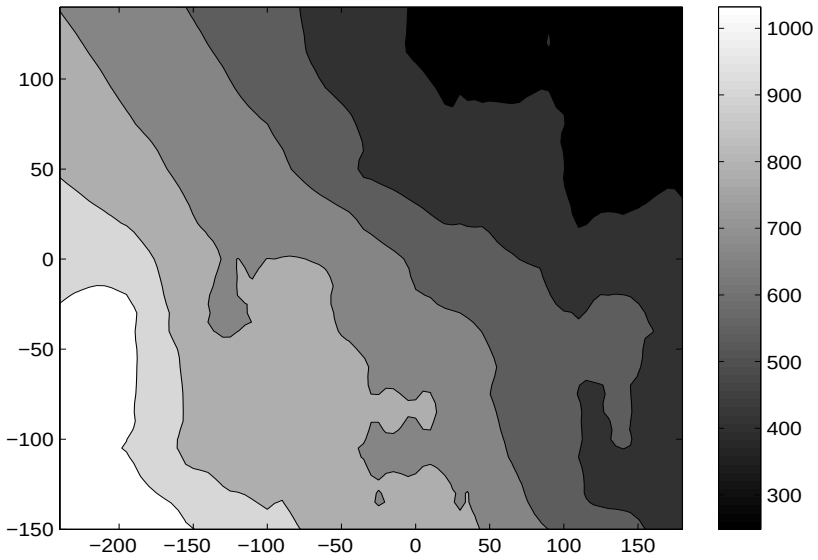
5.2 The Wolfcamp Aquifer dataset

We now show an application to a well known benchmark in geostatistical interpolation, the Wolfcamp aquifer dataset (see, e.g., Harper and Furr, 1986; Cressie, 1991; Barry and Ver Hoef, 1996). This dataset consists of a set of 85 measurement of piezometric head scattered over a rectangular domain of approximately 350km times 250km centered over Amarillo, Texas. A thorough description of classical kriging reconstructions for these data is presented in Cressie (1991), highlighting the need for rather complex analyses to achieve an appropriate data detrending and to account for data anisotropy. On the contrary, as remarked in section 2, flexible black-box variogram models can adapt naturally to the characteristics of anisotropic data.

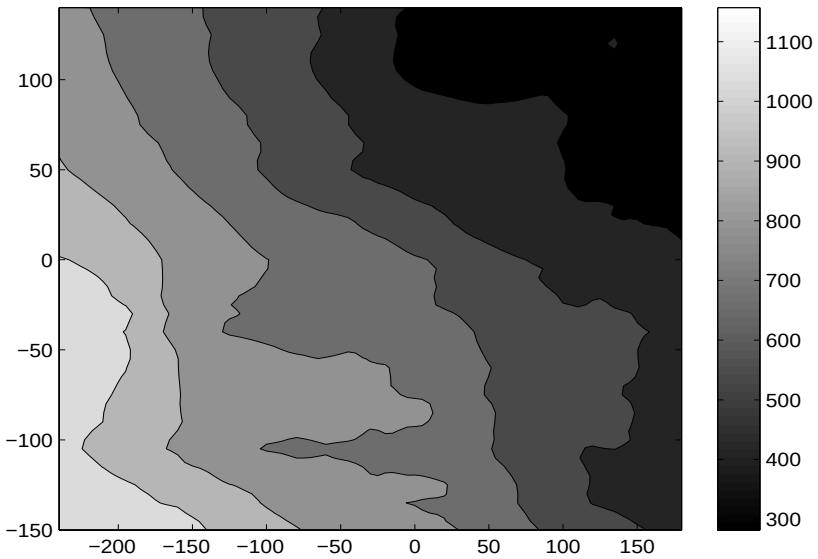
For the application of the flexible black-box variogram model to the Wolfcamp dataset, we set $m_1 = 4$, $m_2 = 5$. In the WLS approach, the deterministic trend has to be removed from the dataset to perform variogram estimation. We thus carry out a recursive generalized least square detrending procedure (see, e.g., Cressie, 1991). The anisotropic analog of the estimator proposed by Cressie-Hawkins in Hawkins and Cressie (1984) is used to compute an empirical variogram.

In the Bayesian approach, we use the model described in Section 3, with $g_0(\mathbf{x}) = 1$, and $\boldsymbol{\beta} = \beta_0$. We run the Hastings-within-Gibbs algorithm under the following specifications: the prior for the black-box parameters are independent exponentials with mean 10; the algorithm is run for 7000 iterations and estimates are obtained using the values sampled every 50 iterations of the chain.

The leave-one-out RMSE of the Bayesian and WLS estimates are 5.35

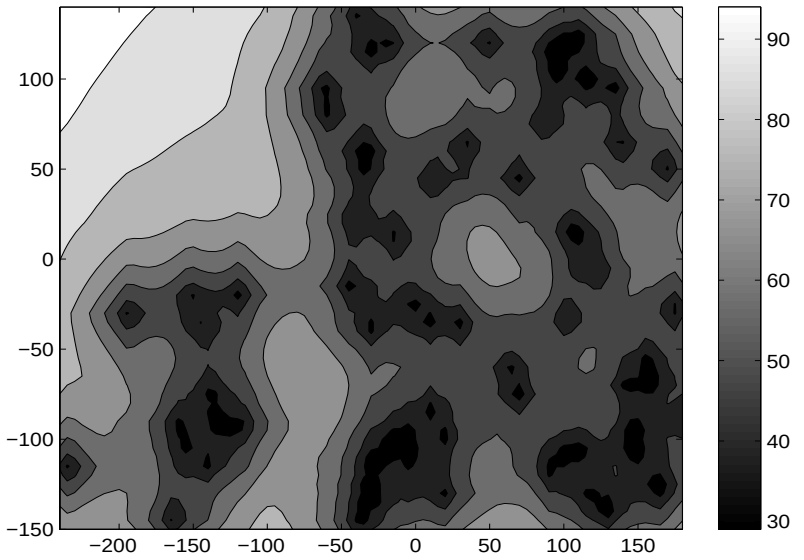


(a)

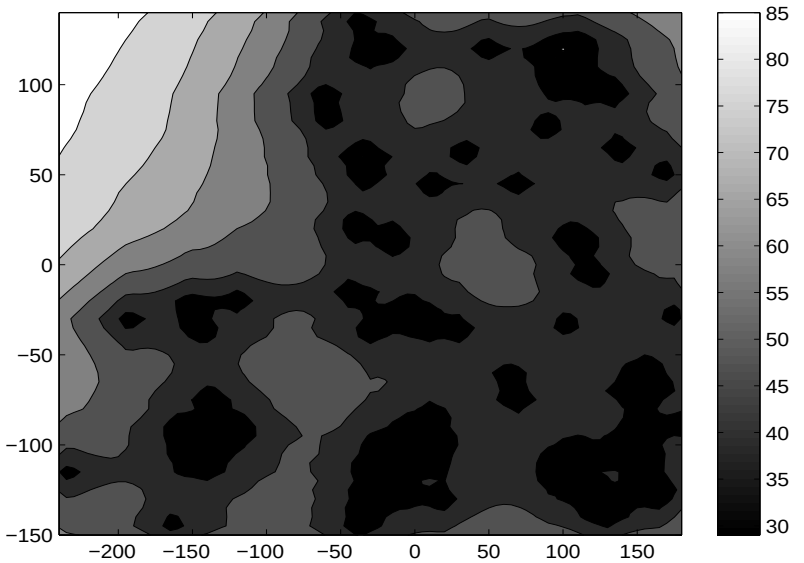


(b)

Figure 3: Kriging reconstruction of the Wolfcamp Aquifer data: predicted values with Bayesian (a) and WLS (b) black-box variogram estimates.



(a)



(b)

Figure 4: Kriging reconstruction of the Wolfcamp Aquifer data: standard deviation with Bayesian (a) and WLS (b) black-box variogram estimates.

and 5.46, respectively; the p-values of the Shapiro-Wilks normality test (see, e.g., Shapiro and Wilk, 1965) on the rescaled residuals of Bayesian and WLS estimates are 0.21 and 0.14, respectively. Figures 3 and 4 display the predicted values obtained with the Bayesian and WLS black-box variogram estimates, and the corresponding standard deviations, showing consistency of the results provided by the two approaches. On the other hand, the Bayesian approach, with respect to the WLS approach, yields a complete assessment of the uncertainty in the variogram estimation, providing the complete posterior distributions of the variogram at any location of the field, and the complete predicted distributions of a new realization at any location of the field.

6 Conclusions

We have developed a Bayesian approach to covariance estimation and geostatistical interpolation based on flexible black-box variograms. These variogram models, originally introduced by Barry and Ver Hoef (1996), do not require restrictive assumptions on the functional shape of the variogram; moreover, they can handle quite naturally non isotropic random fields. These variogram models exploit a special representation theorem for variogram functions that allows to characterize a subset of piecewise polynomial functions as valid variograms with sill. In the Bayesian framework, the parameters characterizing a generic piecewise linear valid variogram, according to this representation theorem, are assumed to be random variables with a chosen *a priori* distribution. The *a posteriori* distribution of these parameters given the available data can be computed by an appropriate Markov Chain Monte Carlo (MCMC) scheme, yielding a complete assessment of the uncertainty in the variogram estimation. The proposed Bayesian approach, with respect to the standard WLS variogram estimation described in Barry and Ver Hoef (1996), does not require the computation of an empirical variogram estimator, thus avoiding the arbitrariness implied by the construction of the empirical variogram itself. Moreover, an extensive simulation study has shown that the estimates provided by Bayesian black-box kriging are significantly more accurate and more robust than the one provided by WLS black-box kriging.

Acknowledgements

We would like to thank Professor Piercesare Secchi for his useful comments and thorough review of this work. This study has been supported by Politecnico di Milano within the Lecco Campus Point initiative.

References

Barry, R. and Ver Hoef, J. (1996), "Blackbox Kriging: Spatial Prediction without Specifying Variogram Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 1, 297–322.

Berger, J., De Oliveira, V., and Sanso, B. (2001), "Objective Bayesian analysis of spatially correlated data," *Journal of the American Statistical Association*, 96, 1361–1374.

Bernardo, J.-M. and Smith, A. F. M. (1994), *Bayesian theory*, Chichester: John Wiley & Sons Ltd.

Christensen, R. (1991), *Linear Models for Multivariate, Time Series and Spatial data*, Springer Verlag.

Cressie, N. (1991), *Statistics for spatial data*, Wiley.

Cressie, N. and Hawkins, D. (1980), "Robust estimation of the variogram," *Journal of the International Association for Mathematical Geology*, 12, 115–125.

Gilks, W., Richardson, S., and Spiegerhalter, D. (1998), *Markov chain Monte Carlo in pratice*, Chapman & Hall.

Gorsich, D. and Genton, M. (2000), "Variogram model selection via nonparametric derivative estimation," *Mathematical Geology*, 32, 249–270.

Handcock, M. and Stein, M. (1993), "A Bayesian analysis of kriging," *Tecnometrics*, 35, 403–410.

Harper, W. and Furr, J. (1986), "Geostatistical analysis of potentiometric data in the Wolfcamp Aquifer of the Palo Duro Basin, Texas," Technical report BMI/ONWI-587, Bettelle Memorial Institute, Columbus, OH.

Hawkins, D. and Cressie, N. (1984), "Robust kriging-a proposal," *Journal of the International Association for Mathematical Geology*, 16, 3–18.

Hollander, M. and Wolfe, D. A. (1999), *Nonparametric statistical methods*, Wiley Series in Probability and Statistics: Texts and References Section, New York: John Wiley & Sons Inc., 2nd ed., a Wiley-Interscience Publication.

Im, H., Stein, M., and Zhu, Z. (2007), "Semiparametric estimation of spectral density with irregular observation," *Journal of the American Statistical Association*, 102, 726–735.

Robert, C. P. and Casella, G. (2004), *Monte Carlo statistical methods*, Springer Texts in Statistics, New York: Springer-Verlag, 2nd ed.

Shapiro, A. and Botha, J. (1991), "Variogram Fitting With a General Class of Conditionally Nonnegative Definite Functions," *Computational statistics and data analysis*, 11, 87–96.

Shapiro, S. S. and Wilk, M. B. (1965), "An analysis of variance test for normality: Complete samples," *Biometrika*, 52, 591–611.

Wackernagel, H. (1995), *Multivariate Geostatistics*, Berlin: Springer Verlag.

Walden, A. T. and Guttorp, P. (1992), *Statistics in the environmental and earth sciences*, Arnold.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

21/2009 L. Bonaventura, S. Castruccio, L.M. Sangalli:
A Bayesian approach to geostatistical interpolation with flexible variogram models

20/2009 N.Accoto, T. Rydén, P. Secchi:
Bayesian hidden Markov models for performance-based regulation of continuity of electricity supply

19/2009 L. Formaggia, A.Mola, N. Parolini, M. Pischiutta:
A three-dimensional model for the dynamics and hydrodynamics of rowing boats

18/2009 J. Silva Soares, P. Zunino:
A mathematical model for water uptake, degradation, erosion, and drug release from degradable polydisperse polymeric networks

17/2009 M. Longoni, C. Magistroni, P. Ruffo, G. Scrofani:
3D Inverse and Direct Structural Modeling Workflow

16/2009 G. Aletti, C. May, P. Secchi:
A functional equation whose unknown is P([0; 1]) valued

15/2009 S. Perotto, A. Ern, A. Veneziani:
Hierarchical local model reduction for elliptic problems I: a domain decomposition approach

14/2009 L. Beirão da Veiga, M. Verani:
A posteriori boundary control for FEM approximation of elliptic eigenvalue problems

13/2009 E.Miglio, A. Villa:
A mathematical derivation of compaction and basin models

12/2009 S. Badia, A. Quaini, A. Quarteroni:
Coupling Biot and Navier-Stokes equations for modelling fluid-poroelastic media interaction