A Bayesian Approach to Spatial Prediction With Flexible Variogram Models

Stefano CASTRUCCIO, Luca BONAVENTURA, and Laura M. SANGALLI

A Bayesian approach to covariance estimation and spatial prediction 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 in 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 simulation studies and by application to a well known benchmark dataset.

Key Words: Markov chain Monte Carlo; Kriging; Variogram estimation.

1. INTRODUCTION

This paper aims at developing an efficient Bayesian approach to spatial prediction based on flexible variogram models. Accurate variogram estimation 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 *a priori* 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. The choice of the distance classes as well as the variogram model to

Stefano Castruccio (✉) is a Ph.D. student, Department of Statistics, The University of Chicago, 5734 S. University Avenue, 60637 Chicago, IL, USA (E-mail: *castruccio@galton.uchicago.edu*). Luca Bonaventura and Laura M. Sangalli are Assistant Professors, MOX—Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo Da Vinci 32, 20133 Milano, Italy.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 17, Number 2, Pages 209–227
DOI: 10.1007/s13253-012-0086-x

be used in the fit is mostly done *ad hoc* and requires substantial user expertise in order to give reasonable results.

The estimation procedures can be improved in two different ways. The first way consists of modeling within a Bayesian framework the uncertainty in the variogram parameters, 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, De Oliveira, and Sanso (2001) and Handcock and Stein (1993). More recently, infill asymptotic theory for random fields has provided some important results for assessment of uncertainty of parameters for models based only on the likelihood function (see for instance Stein 1999 and Zhang 2004), but Bayesian approaches are still easier to implement and more commonly used in literature. A second direction of improvement consists of devising variogram estimation procedures that do not require too many *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, Stein, and Zhu (2007), where a semiparametric form of the spectral density was considered, consisting of 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).

Another interesting proposal toward flexible variogram models, albeit limited to fields with second order moments, is the *black-box kriging* approach introduced in Barry and Ver Hoef (1996). This approach exploits a spectral representation theorem for variogram functions that allows to characterize a subset of the piecewise polynomial functions as valid variograms with sill. Functions of this shape are then fitted to the traditional empirical variogram estimates, thus yielding a 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.

In the present work, we propose a Bayesian approach to kriging with this class of 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, Richardson, and Spiegerhalter 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 toward reducing the need for expert user intervention in the variogram model choice, which restricts in many cases the applicability of spatial analysis 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.

Another model that combines a Bayesian approach with a flexible variogram model was proposed by Ecker and Gelfand (1997). In this work, exploiting the well known Bochner's representation theorem (see, e.g., Cressie 1991, p. 84), a mixture of Bessel functions was used to model the variogram, thus allowing for a non parametric representation. An advantage of the black-box kriging approach considered in the present paper, over the model proposed by Ecker and Gelfand (1997), is that the latter is limited to isotropic fields, whereas black-box kriging handles quite naturally non isotropic random fields.

The advantages of the present method are illustrated by simulation studies with one-dimensional and two-dimensional data. In particular, 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. Also an application to the well known Wolfcamp aquifer benchmark dataset (see, e.g., Harper and Furr 1986 and Cressie 1991) is presented and compared to the results obtained by Barry and Ver Hoef (1996) with 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 implement the proposed Bayesian model is described. The results of simulation studies 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

Spatial prediction is usually formulated assuming that the data consist of a realization of a scalar random field Z, defined on a d-dimensional subset $D \subset \mathbb{R}^d$, and a valid semivariogram function $\gamma(\mathbf{h}) = \mathbb{E}[(Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x}))^2]/2$, for $\mathbf{x} \in D$ and $\mathbf{h} \in \mathbb{R}^d$ (with $\mathbf{x}+\mathbf{h} \in D$). The classical characterization of valid variograms is given in terms of conditionally negative definite functions, i.e., γ is a valid semivariogram function if and only if

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

for all $n \in \mathbb{N}$, all $\mathbf{x}_i, \mathbf{x}_j \in D$ and all $b_1, \ldots, b_n \in \mathbb{R}$ such that $\sum_{i=1}^{n} b_i = 0$. 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 $c \in \mathbb{R}$ given, the function 2γ can be represented as

$$2\gamma(h) = \int_{\mathbb{R}} \left[f(x) - f(x-h)\right]^2 dx, \tag{2.1}$$

where f is a measurable function. The main point of the flexible variogram model consists of 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 $[0, c]$ with $c > 0$ by

$$f(x; \mathbf{a}, c, k) = \sum_{j=1}^{k} a_j I\left(\frac{(j-1)c}{k} < x \leq \frac{jc}{k}\right), \tag{2.2}$$

where $I(\cdot)$ denotes the indicator function. The function f is piecewise constant. Using (2.2) in the representation theorem (2.1), 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,

$$2\gamma(h; \mathbf{a}, c, k) = (1 - V)2\gamma\left(\frac{m_l c}{k}; \mathbf{a}, c, k\right) + V2\gamma\left(\frac{m_u c}{k}; \mathbf{a}, c, k\right),$$

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

$$2\gamma(\mathbf{h}) = \int_{\mathbb{R}^d} \left[f(\mathbf{x}) - f(\mathbf{x} - \mathbf{h})\right]^2 d\mathbf{x}. \tag{2.3}$$

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

$$f\big((x_1, x_2); \mathbf{A}, \mathbf{c}, \mathbf{k}\big) = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} a_{i,j} I\left[\left(\frac{(i-1)c_1}{k_1} < x_1 \le \frac{ic_1}{k_1}\right)\left(\frac{(j-1)c_2}{k_2} < x_2 \le \frac{jc_2}{k_2}\right)\right] \tag{2.4}$$

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 (2.4) into (2.3) yields then a valid variogram function; notice that its definition is coordinate dependent. Since in general $\gamma((h_1, h_2)) = \gamma((-h_1, -h_2))$ and $\gamma((h_1, h_2)) \neq \gamma((h_1, -h_2))$, 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 $h_1 \ge c_1$ or $|h_2| \ge c_2$,

$$2\gamma\big((h_1, h_2); \mathbf{A}, \mathbf{c}, \mathbf{k}\big) = \frac{2c_1c_2}{k_1k_2} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} a_{i,j}^2;$$

- 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,

$$2\gamma\big((h_1, h_2); \mathbf{A}, \mathbf{c}, \mathbf{k}\big) = \frac{2c_1c_2}{k_1k_2} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} a_{i,j}^2 \tag{2.5}$$

$$- \frac{2c_1c_2}{k_1k_2} \sum_{i=m_1+1}^{k_1} \sum_{j=m_2+1}^{k_2} a_{i,j} a_{i-m_1, j-m_2}; \tag{2.6}$$

- 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 $m_2 < 0$,

$$2\gamma\big((h_1, h_2); \mathbf{A}, \mathbf{c}, \mathbf{k}\big) = \frac{2c_1c_2}{k_1k_2} \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} a_{i,j}^2 \tag{2.7}$$

$$- \frac{2c_1c_2}{k_1k_2} \sum_{i=m_1+1}^{k_1} \sum_{j=1}^{k_2+m_2} a_{i,j} a_{i-m_1, j-m_2}. \tag{2.8}$$

In the case of arbitrary lag values, the semivariogram is computed by bilinear interpolation between the values of the variogram at 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

$$\tilde{f}(\mathbf{x}; \mathbf{a}, c, k) = \sum_{i=1}^{k} a_i I\big((i-1)c/k < \|\mathbf{x}\| \le ic/k\big). \tag{2.9}$$

However, substitution of (2.9) into (2.3) 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. It should be mentioned that Ver Hoef, Cressie, and Barry (2004) use piecewise bilinear variogram models that are nearly isotropic. In particular, by constraining the black-box parameters a_{ij} to lie on a symmetric surface, they obtain variograms whose behavior is close to isotropic when k is large.

It should also be noticed that the flexible variogram model defined by (2.1) and (2.2) 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 (2.3) and (2.4). For example, in the one-dimensional case with $k = 2$ one has

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

that implies $2\gamma(c/2; (a_1, a_2), c, k) = 2\gamma(c/2; (a_2, a_1), c, k)$. In this very simple case, identifiability of the black-box parameters could be achieved by enforcing a constraint such as $a_1 > a_2$. However, for $k > 2$ the class of k-tuples which yield the same value of γ is not just the class of permutations of the vector $(a_1, \ldots, a_k)$, but a rather more complex symmetry class, which cannot be simply specified by some identifiability constraints. On the other hand, the real target is the estimation of γ itself (i.e., in other words, the estimation of the nodal semivariogram values $\gamma(ic/k), i = 1, \ldots, k$), rather than the estimation of the parameters a_i. 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).

Notice that the black-box variogram model is mean square continuous but is not mean square differentiable, since the second order derivative of the variogram does not exist at the origin (see, e.g., Stein 1999). Barry and Ver Hoef (1996) show that this limitation in the behavior near the origin of this piecewise linear variogram can be mitigated considering linear parts of unequal lengths.

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 spatially distributed data considered in the following shall consist of realizations of a Gaussian random field Z defined on $D \subset \mathbb{R}^d$, where $d = 1, 2$, that can be written as the sum $Z = \mu + \delta$, where

- the mean field $\mu(\mathbf{x}) = \sum_{j=1}^{p} \beta_j g_j(\mathbf{x})$ 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.

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

$$l(\mathbf{Z}|\boldsymbol{\beta}, \mathbf{A}) = \frac{1}{(2\pi)^{N/2}\sqrt{|\boldsymbol{\Sigma}_\mathbf{A}|}} \exp\left[-\frac{1}{2}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}_\mathbf{A}^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})\right]. \tag{3.1}$$

Recall that $\mathbf{k}$ influences directly the complexity of the model; this parameter can thus be chosen via a model selection criterion such as the Bayes factor.

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. It should be noticed that incorporating prior information on the black-box parameters is not easily feasible, since such priors should be consistent with the complex symmetry characterizing the equivalence classes identified by the corresponding values of $\gamma(h; \mathbf{a}, c, k)$; see Section 2. On the other hand, this is not a limitation of the black-box model, since this model has been mainly proposed to deal with cases where one has no prior information on what could be an appropriate parametric form for the variogram. We hence prefer to assume priors on the black-box parameters that in turns induce a non-informative prior on the semivariogram. The parameter specifying the prior on the black-box parameters (e.g., the mean of the exponential used in the simulation and application in Section 5) could for instance be chosen checking visually the induced prior on the semivariogram, computed via simulations; graphical comparison with an empirical semivariogram might also be useful.

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 $\mathbf{A}$. If *a priori* information on $\boldsymbol{\beta}$ is available, it can be easily incorporated in this prior. Denoting by $\pi_{\boldsymbol{\beta}}$ and $\pi_{\mathbf{A}}$ the priors of $\boldsymbol{\beta}$ and $\mathbf{A}$, respectively, we thus see that the joint posterior distribution of $\boldsymbol{\beta}$ and $\mathbf{A}$ is given by

$$\pi(\boldsymbol{\beta}, \mathbf{A}|\mathbf{Z}) = C\,\pi_{\boldsymbol{\beta}}(\boldsymbol{\beta})\,\pi_{\mathbf{A}}(\mathbf{A})\,l(\mathbf{Z}|\boldsymbol{\beta}, \mathbf{A}), \tag{3.2}$$

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 (3.2). 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:

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

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 are available, proceeding as follows. Define the vector $\hat{\boldsymbol{\gamma}}_U = [\hat{\gamma}(\mathbf{x}_0 - \mathbf{x}_1), \ldots, \hat{\gamma}(\mathbf{x}_0 - \mathbf{x}_N), 1, \ldots, 1]$ and the universal kriging matrix

$$\hat{\boldsymbol{\Gamma}}_U = \begin{bmatrix} \hat{\gamma}(\mathbf{0}) & \ldots & \hat{\gamma}(\mathbf{x}_1 - \mathbf{x}_N) & g_1(\mathbf{x}_1) & \ldots & g_p(\mathbf{x}_1) \\ \hat{\gamma}(\mathbf{x}_2 - \mathbf{x}_1) & \ldots & \hat{\gamma}(\mathbf{x}_2 - \mathbf{x}_N) & g_1(\mathbf{x}_2) & \ldots & g_p(\mathbf{x}_2) \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \hat{\gamma}(\mathbf{x}_N - \mathbf{x}_1) & \ldots & \hat{\gamma}(\mathbf{0}) & g_1(\mathbf{x}_N) & \ldots & g_p(\mathbf{x}_N) \\ g_1(\mathbf{x}_1) & \ldots & g_1(\mathbf{x}_N) & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & 0 & \ldots & 0 \\ g_p(\mathbf{x}_1) & \ldots & g_p(\mathbf{x}_N) & 0 & \ldots & 0 \end{bmatrix}. \tag{3.3}$$

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

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

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. Ver Hoef, Cressie, and Barry (2004) also avoid use of the empirical variogram by means of restricted maximum likelihood estimation of $\mathbf{A}$.

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 (3.2) alternates the following steps:

1. simulation of $\boldsymbol{\beta}$, conditional on the observations $\mathbf{Z}$ and the current value of the black-box parameters $\mathbf{A}$;
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

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

and variance matrix

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

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, 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 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. In fact, since there is no 1-1 correspondence between black-box parameters and nodal values of the variogram, there is no reason to prefer one

particular ordering of the $a_{i,j}$. Randomly permutation of the update order prevents any bias that might be originated by some specific fixed ordering of these parameters.

Notice that the Markov chains of the parameters $a_{i,j}$ will be multimodal, since as previously remarked, the black-box parameter space is partitioned in equivalence classes yielding the same variogram values. Multimodality of these chains suggests that the algorithm is correctly exploring the parameter space, without getting stuck in one of the modes but rather visiting multiple equivalence classes. In fact, this results in good mixing for the values of the variogram of the nodal points.

Though we have not considered this in the current work, a nugget effect can be included in our framework. In particular, the covariance matrix $\mathbf{\Sigma_A}$ should be modified by addition of the nugget γ_0 along the diagonal. The nugget parameter could for simplicity be assumed stochastically a priori independent of the parameters $\mathbf{A}$ and $\boldsymbol{\beta}$. The MCMC sampler should thus be modified either to update γ_0 jointly with $\mathbf{A}$ in the current step 2, or to update γ_0 separately via a further Gibbs step, preferably previous to the current step 2. Simulation studies could suggest which one of these two updating strategies was to be preferred.

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, Sections 5.1 and 5.2 deal with simulated one-dimensional and two-dimensional Gaussian random fields, while Section 5.3 shows an application to a well known benchmark in spatial data analysis, 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 SIMULATED DATA

Consider the model $Z = \mu + \delta$, where $\mu = \beta$ is a constant mean field and δ is a stationary zero-mean Gaussian random field over the real line, with semivariogram γ. We shall simulate observations from this field with semivariogram γ of the Matern class

$$\gamma\big(h \mid \sigma^2, \phi, \nu\big) = C\big(0 \mid \sigma^2, \phi, \nu\big) - C\big(h \mid \sigma^2, \phi, \nu\big),$$

$$C\big(h \mid \sigma^2, \phi, \nu\big) = \sigma^2 \frac{1}{\Gamma(\nu)\, 2^{\nu-1}} (\phi h)^{\nu} K_{\nu}(\phi h),$$

where K_ν denotes the modified Bessel function of the second kind (see, e.g., Abramowitz and Stegun 1965, pp. 374–379) the parameters (σ^2, ϕ, ν) are the scale, inverse range and smoothness parameter, respectively (with $\sigma^2, \phi, \nu > 0$). In particular, we shall simulate observations from three fields with Matern semivariograms corresponding to the following choices of the parameters (σ^2, ϕ, ν): γ_1 with $(\frac{L}{2}, 1/L, \frac{1}{2})$, γ_2 with $(\frac{L}{2}, \frac{4}{L}, \frac{1}{2})$, and γ_3 with $(\frac{L}{2}, 1/L, \frac{3}{2})$, i.e.,

$$\gamma_1(h) = \frac{L}{2}\left(1 - \exp\left\{-\frac{h}{L}\right\}\right), \tag{5.1}$$

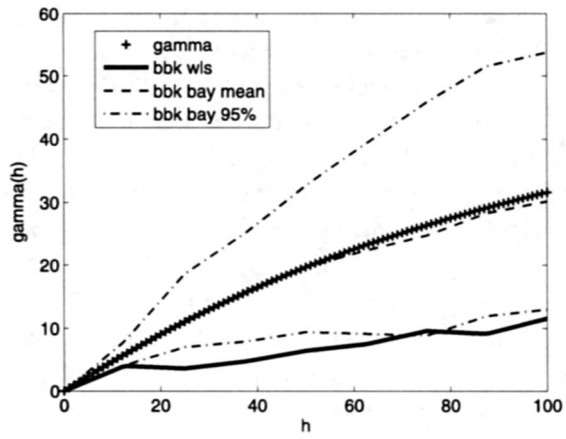

Figure 1. Bayesian and WLS black-box estimates obtained for the first simulation repetition in the case γ_1, $N = 100$ and $k = 8$; the plot also shows 95 % pointwise credibility bands for the Bayesian black-box estimate.

$$\gamma_2(h) = \frac{L}{2}\left(1 - \exp\left\{-\frac{h}{L/4}\right\}\right), \tag{5.2}$$

$$\gamma_3(h) = \frac{L}{2}\left(1 - \exp\left\{-\frac{h}{L}\right\}\left(1 + \frac{h}{L}\right)\right), \tag{5.3}$$

where we set $L = 100$. The mean β is set equal to 10.

For each semivariogram, we simulate the field at N equispaced locations on the interval $[0, N]$, as described e.g. in Cressie (1991). The sample is thus divided in a training set and a testing set, with the $N/4$ observations composing the testing set randomly selected among the N simulated observations. Using the observations in the training set, 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 Bayesian 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 priors for the black-box parameters are independent exponentials with mean 10; the prior for β is Gaussian, with mean 11 and standard deviation 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 terms of the Root Mean Squared Prediction Error (RMSPE) over the $N/4$ observations in the testing set. Moreover, to compare the prediction variances of the two methods, we consider 95 % prediction intervals for observations in the testing set, and compute their empirical coverages, counting how many of these observations fall within the corresponding prediction interval.

For each of the three semivariograms, γ_1, γ_2, γ_3, this simulation scheme is repeated for two different values of N, $N = 100$ and $N = 160$; moreover, each simulation is repeated 100 times. Also, different values of k are used for the black-box kriging estimates: when $N = 100$ we consider $k = 6$ and $k = 8$; in the case $N = 160$ we consider $k = 8$ and $k = 12$. Figure 1 shows the Bayesian and WLS black-box estimates obtained for the first simu-

lation repetition in the case γ_1, $N = 100$ and $k = 8$. Table 1, reports, for each of the 12 cases (depending on the semivariogram, the sample size N, and the value of k used for the estimates), the median and inter-quantile range of the RMSPE for the 100 semivariogram estimates obtained by Bayesian and WLS black-box kriging. For each of the 12 cases, we perform a one-sided nonparametric paired Wilcoxon test (see, e.g., Hollander and Wolfe 1999) to verify if the distribution of the RMSPE for the Bayesian estimates is stochastically lower than the distribution of the RMSPE for the WLS estimates. The p-values of these tests, reported in the eighth column, show that the RMSPE of Bayesian black-box kriging estimates are uniformly significantly lower than the ones of WLS black-box kriging estimates. A look at the inter-quantile ranges also highlights that the errors of Bayesian black-box estimates have a smaller variability, i.e., Bayesian black-box kriging estimates, besides being more precise, are also more robust than WLS black-box kriging estimates. The ninth and tenth columns show the average empirical coverages across the 100 simulations. We can see that for γ_1 and γ_2 the empirical coverage of the Bayesian black-box is very close to the 95 % nominal value, whereas the WLS black-box underestimates the variance, therefore resulting in a smaller empirical coverage. In the case of γ_3, both the Bayesian and WLS black-box overestimate the variance; this is due to the fact that the black-box variogram is less regular at the origin than γ_3, the variogram used to simulate the data. As commented in Section 3, in the proposed Bayesian framework it is possible to select the best value of k via Bayes factor. For instance, in the first replicate of the first simulation case (γ_1, $N = 100$), considering a priori equivalent the two Bayesian black-box kriging models with $k = 6$ and $k = 8$, the Bayes factor

$$\text{BF} = \frac{\text{probability density of data under BBK model with } k = 6}{\text{probability density of data under BBK model with } k = 8}$$

gives the posterior odds in favor of the model with $k = 6$ vs. $k = 8$. This quantity does not have a closed form expression and both numerator and denominator are integrals over the parameter space that can be computed separately via numerical methods. We used a Monte Carlo importance sampling algorithm (see e.g. Gilks, Richardson, and Spiegerhalter 1998) with proposal distribution being that of independent exponentially distributed parameters. The computed Bayes factor (BF = 84.75) provides strong evidence in favor of $k = 6$. As an optimal reference benchmark, Table 1 also shows the RMSPE for the estimates obtained with a Bayesian model with a Matern semivariogram $\gamma(h \mid \sigma^2, \phi, \nu)$. This model has been implemented using the function spLM in the spBayes package in R, R Development Core Team (2010). The spLM function assumes an inverse-Gamma prior for σ^2, uniform priors for ϕ and ν, and an improper uniform prior for β. We specified the following prior hyperparameters: $\sigma^2 \sim IG(2.0001, 10.001)$ (prior mean 10, prior variance 10^4), $\phi \sim \mathcal{U}(0.00001, 2)$ and $\nu \sim \mathcal{U}(0.1, 3)$. The algorithm is run for 5000 iterations and estimates are obtained using the values sampled every 50 iterations of the chain. Since this model assumes the true variogram form that was used to generate the data, it is expected to provide the best possible estimates. The seventh column of Table 1 reports the p-values of one-sided nonparametric paired Wilcoxon tests verifying that the distribution of the RMSPE of Bayesian Matern estimates are stochastically lower than those of Bayesian black-box estimates. It is interesting to note that, for data generated using the semivariograms

Table 1. One-dimensional simulated data: median (inter-quantile range) of the RMSPE for Bayesian Matern estimates (third column), Bayesian black-box kriging estimates (fifth column), WLS black-box kriging estimates (sixth column); p-values of nonparametric one-sided paired Wilcoxon tests verifying that distribution of the RMSPE of Bayesian Matern estimates is stochastically lower than the distribution of the RMSPE of Bayesian black-box estimates (seventh column) and that the distribution of the RMSPE of Bayesian black-box estimates is stochastically lower than the distribution of the RMSPE of WLS black-box estimates (eighth column). Empirical prediction coverage for a nominal 95 % confidence interval for Bayesian black-box kriging estimates (ninth column) and WLS black-box kriging estimates (tenth column).

γ	N	Bayes Matern RMSPE	k	Bayes BBK RMSPE	WLS BBK RMSPE	Bayes Matern *vs* Bayes BBK	Bayes BBK *vs* WLS BBK	Bayes BBK cover	WLS BBK cover
γ_1	100	0.79 (0.21)	6	0.78 (0.21)	0.88 (0.19)	0.9996	< 0.0001	0.958	0.855
			8	0.80 (0.20)	0.89 (0.23)	0.0003	< 0.0001	0.954	0.828
γ_1	160	0.77 (0.15)	8	0.81 (0.14)	0.98 (0.21)	< 0.0001	< 0.0001	0.949	0.730
			12	0.76 (0.15)	0.94 (0.21)	1	< 0.0001	0.963	0.782
γ_2	100	1.59 (0.38)	6	1.55 (0.42)	1.75 (0.42)	0.9986	< 0.0001	0.954	0.826
			8	1.62 (0.39)	1.74 (0.40)	0.0001	< 0.0001	0.950	0.809
γ_2	160	1.55 (0.33)	8	1.62 (0.28)	1.97 (0.41)	< 0.0001	< 0.0001	0.942	0.669
			12	1.53 (0.29)	1.83 (0.42)	1	< 0.0001	0.962	0.779
γ_3	100	0.005 (0.002)	6	0.013 (0.009)	0.014 (0.0129)	< 0.0001	< 0.0001	0.9996	1.000
			8	0.012 (0.011)	0.016 (0.010)	< 0.0001	< 0.0001	0.998	1.000
γ_3	160	0.005 (0.001)	8	0.012 (0.005)	0.016 (0.009)	< 0.0001	< 0.0001	0.999	0.997
			12	0.014 (0.005)	0.017 (0.010)	< 0.0001	< 0.0001	1.000	0.9998

Table 2. Two-dimensional simulated data: median (inter-quantile range) of the RMSPE for Bayesian Matern estimates (third column), Bayesian black-box kriging estimates (fourth column), WLS black-box kriging estimates (fifth column); p-values of nonparametric one-sided paired Wilcoxon tests verifying that distribution of the RMSPE of Bayesian Matern estimates is stochastically lower than the distribution of the RMSPE of Bayesian black-box estimates (sixth column) and that the distribution of the RMSPE of Bayesian black-box estimates is stochastically lower than the distribution of the RMSPE of WLS black-box estimates (seventh column).

θ	λ	Bayes Matern RMSPE	Bayes BBK RMSPE	WLS BBK RMSPE	Bayes Matern *vs* Bayes BBK	Bayes BBK *vs* WLS BBK
0	1	0.0126 (0.0016)	0.0185 (0.0039)	0.0222 (0.0055)	< 0.0001	< 0.0001
0	3	0.039 (0.007)	0.038 (0.007)	0.042 (0.009)	0.931	< 0.0001
0	5	0.0781 (0.0197)	0.0486 (0.0105)	0.0549 (0.0153)	1	< 0.0001
$\pi/16$	5	0.0820 (0.0178)	0.0799 (0.0196)	0.0870 (0.0220)	0.532	0.007
$\pi/4$	3	0.049 (0.007)	0.079 (0.014)	0.097 (0.022)	< 0.0001	< 0.0001
$\pi/4$	5	0.1033 (0.0177)	0.1621 (0.0221)	0.2142 (0.0517)	< 0.0001	< 0.0001

γ_1 and γ_2, the proposed Bayesian black-box kriging model, in correspondence of the best value of k ($k = 6$ when $N = 100$ for both γ_1 and γ_2, and $k = 12$ when $N = 160$ for both γ_1 and γ_2), is able to attain the optimal performances of the Bayesian Matern model, that is, assuming the true semivariogram form. Notice that the semivariogram γ_1 and γ_2 have a locally linear behavior at the origin that is well captured by the black-box model. In the case of γ_3, which has instead a quadratic behavior at the origin, both the Bayesian black-box kriging estimates and the WLS black-box kriging estimates could in fact be improved by using black-box semivariograms with linear parts of unequal length, shorter near the origin and then progressively larger (see discussion in Barry and Ver Hoef 1996).

5.2. TWO-DIMENSIONAL SIMULATED DATA

We now compare the models in a two-dimensional setting. Let $Z = \mu + \delta$, where $\mu = \beta$ is a constant mean field over the plane and δ is a stationary zero-mean Gaussian random field over the plane with semivariogram γ. We shall simulate observations from this field with a geometrically anisotropic Matern semivariogram with parameters $(\sigma^2, \phi, \nu) = (\frac{L}{2}, \frac{1}{L}, \frac{3}{2})$,

$$\gamma_4(\mathbf{h}) = \frac{L}{2}\left(1 - \exp\left\{-\frac{\|Q\mathbf{h}\|}{L}\right\}\left(1 + \frac{\|Q\mathbf{h}\|}{L}\right)\right),$$

$$Q = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix},$$

where we set $L = \sqrt{2} \cdot 20$ and consider different values of the parameters (θ, λ) in the matrix Q regulating the anisotropy, as shown in Table 2. The mean β is set equal to 10. The field is simulated on a lattice of $N \times N$ equispaced locations on $[0, N] \times [0, N]$, with $N = 20$, and the sample is then divided in a training set and a testing set, with 1/4 of the observations composing the testing set, randomly selected among the N^2 simulated

observations. Each simulation case is repeated 100 times. From observations in the training set, we obtain Bayesian and WLS black-box estimates, setting $k_1 = k_2 = 4$. Bayesian black-box estimates are obtained under the same specifications as for one-dimensional simulations. We also obtain estimates from a Bayesian model assuming an isotropic Matern semivariogram, always using the R function spLM, with the following specifications for the prior hyperparameters: $\sigma^2 \sim IG(2.1, 11)$ (prior mean 9.1 prior variance 10^4), $\phi \sim \mathcal{U}(0.00001, 2)$ and $\nu \sim \mathcal{U}(0.1, 3)$.

The third, fourth and fifth columns of Table 2 report the medians and inter-quantile ranges of the RMSPE, computed on observations of the testing set, for the 100 semivariograms estimates obtained by Bayesian Matern, Bayesian black-box and WLS black-box models, respectively. The sixth and seventh columns show the p-values of one-sided nonparametric paired Wilcoxon tests verifying, respectively, if the distribution of the RMSPE of Bayesian Matern estimates is stochastically lower than the distribution of the RMSPE of Bayesian black-box estimates (sixth column), and if the distribution of the RMSPE of Bayesian black-box estimates is stochastically lower than the one of WLS black-box estimates (seventh column). These results show that Bayesian black-box kriging uniformly outperforms WLS black-box kriging, as the distribution of RMSPE of the former is significantly lower than that of WLS black-box estimates in all simulation cases; also, likewise for the simulations in the one-dimensional setting, a look at the inter-quantile ranges of RMSPE highlights that the errors of Bayesian black-box estimates have a smaller variability, i.e., Bayesian black-box kriging estimates, besides being more precise, are also more robust than WLS black-box kriging estimates.

The comparison with the Bayesian Matern model is also interesting. The latter model assumes the correct semivariogram form, but does not take into account the possible anisotropy. Notice also that the Matern semivariogram used to generate the data is locally quadratic at the origin, likewise γ_3 in Section 5.1; among the simulation cases considered in the previous section, this was the less favorable to the use of black-box kriging, which instead assumes a less regular semivariogram. Table 2 shows that in the isotropic case ($\theta = 0, \lambda = 1$) the Bayesian isotropic Mater model correctly provides better estimates than the black-box model, even if the significance is not particulary strong. Instead, in the anisotropic cases corresponding to ($\theta = 0, \lambda = 3$), ($\theta = 0, \lambda = 5$) and ($\theta = \pi/16, \lambda = 5$), the Bayesian Matern model does not provide better predictions than Bayesian black-box kriging, even if the latter model assumes a semivariogram that is only piecewise linear; Bayesian black-box estimates are in fact significantly better than Bayesian Matern estimates (in two cases) or not significantly different from Bayesian Matern estimates (in one case). Finally, the anisotropic cases ($\theta = \pi/4, \lambda = 3$) and ($\theta = \pi/4, \lambda = 5$) illustrate that the type of anisotropy that the black-box kriging is best at capturing are those deployed mostly along one of the axes (while in the last two simulation cases the direction along which the anisotropy is deployed is at 45° with respect to the two axes). This was expected, as the definition of the black-box model over a two-dimensional domain, recalled here in Equations (2.3) and (2.4), is coordinate dependent, as commented in Section 2.

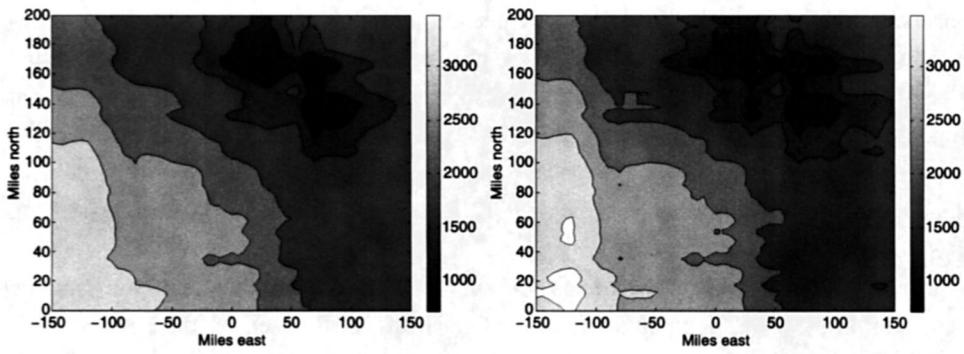

Figure 2. Kriging reconstruction of the Wolfcamp Aquifer data: predicted values with Bayesian (left) and WLS (right) black-box variogram estimates.

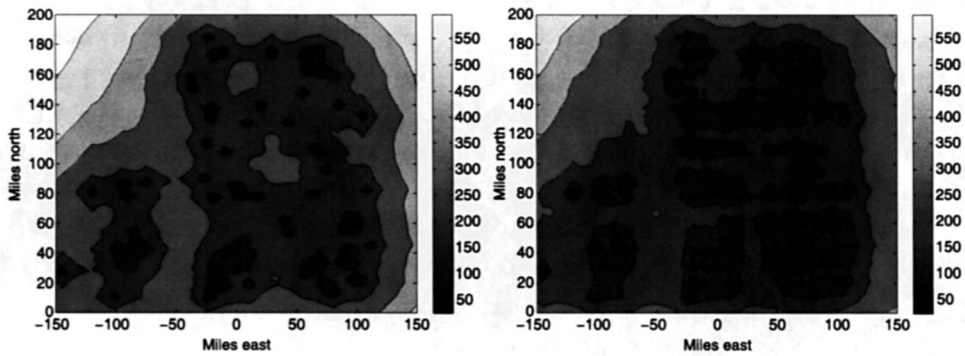

Figure 3. Kriging reconstruction of the Wolfcamp Aquifer data: standard deviation with Bayesian (left) and WLS (right) black-box variogram estimates.

5.3. THE WOLFCAMP AQUIFER DATASET

We now show an application to a well known benchmark in spatial data analysis, 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 measurements of piezometric head scattered over a rectangular domain of approximately 300 miles times 200 miles 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.

For the application of the flexible black-box variogram model to the Wolfcamp dataset, we set $k_1 = 4$, $k_2 = 5$, as in Barry and Ver Hoef (1996). The WLS here presented is the one computed by Barry and Ver Hoef (1996). For the Bayesian approach, we use the model described in Section 3, with a constant mean field $\mu(\mathbf{x}) = \beta$. We run the Hastings-within-Gibbs algorithm under the following specifications: the priors for the black-box parameters are independent exponentials with mean 10; the prior for β is Gaussian with mean 1500 and standard deviation 1000; the algorithm is run for 7000 iterations and estimates are obtained using the values sampled every 50 iterations of the chain.

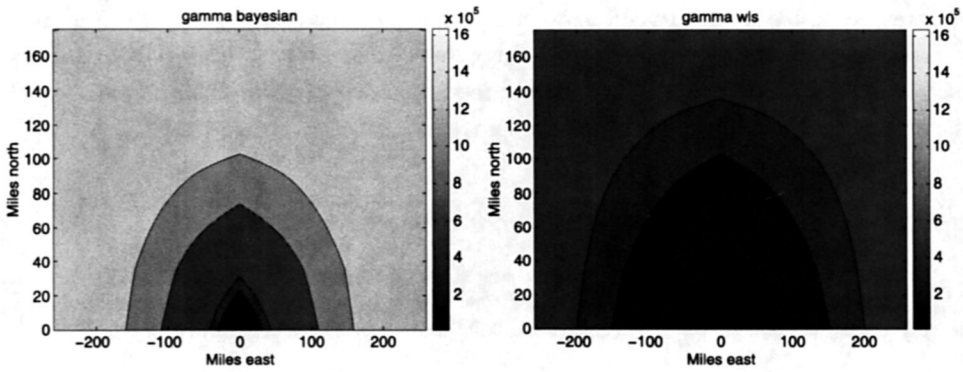

Figure 4. Comparison between the Bayesian (left) and WLS (right) estimated variograms.

The leave-one-out Root Mean Square Error (RMSE) of the Bayesian and WLS black-box estimates are 175.90 and 173.22, 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 0.24 and 0.06, respectively. Figures 2 and 3 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. Figure 4 shows the estimated Bayesian and WLS estimated semivariograms; the behavior is very similar, although the Bayesian variogram tends to have larger values for large lags. 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 predicted distributions of a new realization at any location of the field.

Since Figure 2 seems to display a linear trend, we also tested a Bayesian black-box model with mean field $\mu(\mathbf{x}) = \sum_{j=1}^{4} \beta_j g_j(\mathbf{x})$ where $g_1 = 1$, $g_2 = x_1$, $g_3 = x_2$, $g_4 = x_1 x_2$; this model, as well as models nested within this one, has higher leave-one-out RMSE than the model with constant mean field, and 95 % credibility intervals for the parameters $\beta_2, \beta_3, \beta_4$ crossing zero. We thus concluded that the simpler model with constant mean field was better from a prediction point of view.

6. DISCUSSION AND CONCLUSIONS

We have developed a Bayesian approach to covariance estimation and spatial prediction based on flexible black-box variograms. These variogram models, originally introduced by Barry and Ver Hoef (1996), 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. By means of simulation studies with both one-dimensional and two-dimensional data, we have shown that Bayesian black-box kriging yields superior estimates

with respect to the standard WLS variogram estimation described in Barry and Ver Hoef (1996). Bayesian black-box estimates are more precise and robust, having stochastically lower RMSPE, and with a smaller variance; moreover, they provide a better estimation of the field variance, as illustrated by their empirical predicted coverages.

ACKNOWLEDGEMENTS

We would like to thank Professor Piercesare Secchi for his useful comments and thorough review of this work. We are also grateful to the associate editor and referees for thoughtful suggestions, which led to a much improved paper. This study has been supported by Politecnico di Milano within the Lecco Campus Point initiative.

[Received January 2011. Accepted February 2012. Published Online March 2012.]

REFERENCES

Abramowitz, M., and Stegun, I. (1965), *Handbook of Mathematical Functions*, New York: Dover.

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: Wiley.

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

Cressie, N. (1991), *Statistics for Spatial Data*, New York: 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 Practice*, London: Chapman & Hall.

Gorsich, D., and Genton, M. (2000), "Variogram Model Selection via Nonparametric Derivative Estimation," *Mathematical Geology*, 32, 249–270.

Ecker, M. D., and Gelfand, A. E. (1997), "Bayesian Variogram Modeling for an Isotropic Spatial Process," *Journal of Agricultural, Biological, and Environmental Statistics*, 2 (4), 347–369.

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*, 2nd ed. *Wiley Series in Probability and Statistics: Texts and References Section*. New York: Wiley.

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.

R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. ISBN 3-900051-07-0, URL *http://www.R-project.org*.

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

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.

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

Ver Hoef, J. M., Cressie, N., and Barry, R. P. (2004), "Flexible Spatial Models for Kriging and Cokriging Using Moving Averages and the Fast Fourier Transform (FFT)," *Journal of Computational and Graphical Statistics*, 13, 265–282.

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

Walden, A. T., and Guttorp, P. (1992), *Statistics in the Environmental and Earth Sciences: New Developments in Theory and Practice*, London: Hodder Arnold.

Zhang, H. (2004), "Inconsistent Estimation and Asymptotically Equal Interpolations in Model Based Geostatistics," *Journal of the American Statistical Association*, 99, 250–261.