KRIGING WITH MIXED EFFECTS MODELS (*)

A. Pollice, M. Bilancia

1. INTRODUCTION

From its origins Geostatistics was developed largely as an applied field, and somehow independently of general (spatial) statistic methodology. Spatial interpolation and prediction for continuos processes are rarely performed with explicit reference to distributive assumptions: as a consequence Geostatistical practice is seldom based on modern likelihood-based statistical methods. Standard second order assumptions, though regarded as essential for the applicability of the methodologies, prove to be insufficient to guarantee their optimality. It is well know that least mean squared error predictors such as Kriging are not optimal in the absence of Gaussian assumptions (Ripley, 1981); indeed they are optimal under Gaussianity only if mean and variance/covariance parameters are known or considered so, the optimal predictor generally being nonlinear when parameters are unknown (Cressie, 1993). These considerations push towards the adoption of widely spread model-based statistical techniques in the Geostatistical context, explicit distributional assumptions and the use of likelihood-based inferential procedures being central to this purpose. Given that the methodological bulk of theoretical and applied Geostatistics is still to be found in variogram modelling and the use of minimum variance predictors, since the late 80's (Vecchia, 1988) until recently (Stein, 1999) more and more contributions have been proposed, involving the explicit adoption of spatial stochastic models.

A convenient representation of the spatial process is obtained by splitting the total variability in a systematic term or mean effect, a spatially correlated component and a random noise, the latter being quite regardful in Geostatistical series due to measurement erraticity. Such a representation is easily obtainable by the use of linear mixed models (Christensen, 1991 p. 273), specifically developed for the analysis of correlated observations, typically in the context of repeated measu-

(*) Both authors were supported by the MURST grant "Metodi e Tecniche Statistiche per l'Analisi dei Dati a Struttura Spaziale e Spazio-Temporale" - COFIN 99. Address for correspondence: Alessio Pollice, Dipartimento di Scienze Statistiche - Università degli Studi di Bari, Via C. Rosalba n.53, 70124 Bari, Italia. Tel.: +39/0805049243 - Fax: +39/0805049147 - Email: apollice@dss.uniba.it

rement data. These models have recently experienced meaningful advances in estimation and selection procedures (Searle *et al.*, 1992; Davidian and Giltinan, 1995; Pinheiro and Bates, 2000) and consequently a growing popularity among applied statisticians. A further advantage in the use of linear mixed models is associated to their hierarchical representation, that enables the formulation of the joint distribution of the spatial process as a combination of simpler conditional component models (Gelman *et al.*, 1995). The construction of models in complex settings is considerably simplified by the hierarchical approach, which enables to highlight the relationships among different causes of variability (Berliner, 2000).

The hierarchical representation of stochastic models is intrinsically connected to Bayesian inference, where probability statements on observable quantities are (hierarchically) specified conditionally on the values of random unknown parameters, provided with their own prior distributions. Recent advances in computational methods made hierarchical methods and the Bayesian approach widely used, causing an explosion in the number and the variety of applications (Gilks *et al.*, 1996). Approaching Geostatistics within the Bayesian framework enables to define spatial predictors explicitly accounting for uncertainty in the unknown spatial correlation structure. Moreover the use linear mixed models in this context, allows obtaining classical Geostatistical results as special cases (Kitanidis, 1986; Ribeiro and Diggle, 1999*a*) and the treatment of complex spatial situations including anisotropic (Ecker and Gelfand, 1997) and multivariate processes (Royle and Berliner, 2000). Finally it is to notice how the assumption of Gaussianity of the spatial process can also be relaxed by the use of generalised linear mixed models (Gelfand *et al.* 2000; Diggle *et al.*, 1998).

In the following sections some results concerning the interaction between Geostatistics and the theory of linear mixed effects models are reviewed in a unified framework. An illustrative case study is also proposed concerning the soil structure of an agricultural area in the Foggia district.

2. GAUSSIAN MIXED MODELS

Consider a finite set of spatial locations $\mathrm{t} = (t_1, \ldots, t_n)$, and only one measurement of a one-dimensional spatially dependent random variable Y at each location. Assume that the data vector $y(\mathrm{t}) = (y(t_1), \ldots, y(t_n)) = (y_1, \ldots, y_n) = \mathrm{y}$ is a finite realization of a second order stationary single-valued spatial (stochastic) process $\{Y(t); t \in D\}$ where $D \subset \Re^2$.

The stochastic behavior of the random vector $Y(\mathrm{t}) = (Y(t_1), \ldots, Y(t_n))$ is assumed to depend on observed covariates and a latent spatial process $S(\mathrm{t})$, according to a *spatial Gaussian* or *linear mixed model*

$$Y(\mathrm{t}) = X(\mathrm{t})\beta + S(\mathrm{t}) + \varepsilon(\mathrm{t}) \tag{1}$$

This decomposition corresponding to classical Kriging with measurement error (Cressie, 1993 and many others) is essential to the definition of a common framework to embed maximum likelihood and Bayesian inference for continuos spatial processes. It also proves to be useful in the correct definition of spatial predictions in the presence of measurement error (§ 3.1 and 3.2). In the latter expression $\varepsilon(\mathrm{t})$ is a spatially independent Gaussian process with zero mean and variance τ^2 (nugget), modeling the measurement error. Random effects $S(\mathrm{t})$ are assumed to account for spatial variability being a spatially correlated stationary Gaussian process

$$S(\mathrm{t}) \sim \mathrm{N_n}\left(0, \sigma^2 H_{11}(\phi)\right) \tag{2}$$

where the scalar σ^2 is the variance (partial sill) and the correlation matrix $H_{11}(\phi)$ is built according to a valid correlation function $h(\nu;\phi)$, where ν is the distance between locations and ϕ is a correlation parameter (range). The fixed effect part of the model $X(\mathrm{t})\beta$ is given by the product of an $n \times p$ matrix of spatially referenced non random variables (coordinates and/or covariates) and a *p*-dimensional parameter vector (spatial trend parameter). Independence of $S(\mathrm{t})$ and $\varepsilon(\mathrm{t})$ implies that elements of $Y(\mathrm{t})$ are independent and normally distributed conditionally on $X(\mathrm{t})$ and $S(\mathrm{t})$ (model (1) is also referred to as *conditional independence model*). Integrating random effects out of model (1) a *marginal model* is obtained in which regression parameters retain their meaning and random effects contribute in a simple way to the variance-covariance structure. The marginal distribution of the spatial process $Y(\mathrm{t})$ is

$$Y(\mathrm{t}) \sim \mathrm{N_n}\left(X(\mathrm{t})\beta, \Sigma_{11}\right) \tag{3}$$

with $\Sigma_{11} = \sigma^2 H_{11}(\phi) + \tau^2 I_n$.

3. MAXIMUM LIKELIHOOD ESTIMATION

Being able to specify the marginal distribution of the spatial process $Y(\mathrm{t})$ in closed form is crucial in likelihood-based inference: such ability depends on the assumptions of Gaussianity and that of linear dependence of the response on fixed and random effects. Expression (3) implies that the marginal likelihood for the mean parameter β and covariance structure parameters $\varpi = \left(\sigma^2, \phi, \tau^2\right)$ is simply that of a general linear model with correlated errors, given by

$$L(\varpi, \beta; \mathrm{y}) = \left[(2\pi)^n |\Sigma_{11}|\right]^{-1/2} \exp\left[-\frac{1}{2}(\mathrm{y} - X\beta)' \Sigma_{11}^{-1} (\mathrm{y} - X\beta)\right] \tag{4}$$

where vector y contains a finite realization of the spatial process and X stands for $X(\mathrm{t})$ from now on. Assuming the covariance matrix Σ_{11}, or related parameters ϖ are known, the maximum likelihood estimator of the trend parameter vector β (fixed effects) is given by the Aitken (GLS or WLS) estimator

$$\hat{\beta} = \left(X^{'}\Sigma_{11}^{-1}X\right)^{-1} X^{'}\Sigma_{11}^{-1}\mathrm{y} \tag{5}$$

Conditionally on the value of Σ_{11} expression (5) is a Best Lineat Unbiased Estimator (BLUE) with

$$\mathrm{Cov}\left(\hat{\beta}\right) = \left(X^{'}\Sigma_{11}^{-1}X\right)^{-1} \tag{6}$$

The main disadvantage due to considering Σ_{11} as fixed, neglecting the uncertainty in the unknown covariance parameter vector ϖ, is to be found in the value of covariance (6) being underestimated.

Notice that also the joint maximization of (4) with respect to ϖ and β produces the GLS estimator for the mean parameter β, so maximum likelihood estimates of covariance parameters $\varpi = \left(\sigma^2, \phi, \tau^2\right)$ are obtained by maximizing (4) conditionally on estimated values of β. Derivatives of the likelihood function are often nonlinear with respect to σ, ϕ and τ and lead to non-explicit solutions to the estimation problem, obtainable by numeric iterative methods (Newton Raphson or EM). ML estimates of covariance parameters tend to be (negatively) biased as a consequence of considering β as fixed (see Davidian and Giltinan, 1995 and also Ripley, 1988 for a full discussion of ML covariance estimation in the spatial context).

Rather than conditioning on β, an unbiased alternative to ML estimation of spatial covariance parameters ϖ is obtained by the maximization of a likelihood function based on residuals $\mathrm{y} - X\hat{\beta}$. It can be shown (Diggle *et al.*, 1994) that $\mathrm{y} - X\hat{\beta}$ and $\hat{\beta}$ are independent, so that the full likelihood (4) can be factorized as

$$L(\varpi, \beta; \mathrm{y}) = L\left(\varpi; \mathrm{y} - X\hat{\beta}\right) L\left(\varpi, \beta; \hat{\beta}\right)$$

The so called REML likelihood can then be obtained as

$$L\left(\varpi; \mathrm{y} - X\hat{\beta}\right) = \frac{L(\varpi, \beta; \mathrm{y})}{L\left(\varpi, \beta; \hat{\beta}\right)} \propto \left|\Sigma_{11}\right|^{-1/2} \left|X^{'}\Sigma_{11}^{-1}X\right|^{-1/2} \exp\left[-\frac{1}{2}\left(\mathrm{y} - X\hat{\beta}\right)^{'} \Sigma_{11}^{-1} \left(\mathrm{y} - X\hat{\beta}\right)\right] \tag{7}$$

Fixed effects estimates (5) are substituted into REML likelihood (7) which is then maximized by suitable numerical methods.

Estimates of β are then updated to reflect the current estimates of ϖ and conversely, according to an iterative estimation process. Considering the REML likelihood is equivalent to correcting the likelihood function for the loss of degrees of freedom due to estimating β , thus covariance parameter estimates obtained by the iterative estimation process result to be unbiased (Cressie, 1993).

3.1. *Predictions*

Say t_0 is a vector of m unsampled locations and $Y(\mathrm{t}_0)=Y_0$ the corresponding m-dimensional vector of unobserved realizations of the spatial process with

$$Y_0 = X_0\beta + S_0 + \varepsilon_0 \tag{8}$$

where $S_0 \sim \mathrm{N}_m(0,\sigma^2 H_{00}(\phi))$ with $\mathrm{Cov}(\mathrm{S},\mathrm{S}_0)=\sigma^2 H_{10}(\phi)$. Covariance matrices $H_{00}(\phi)$ and $H_{10}(\phi)$ are also built according to the correlation function $h(v;\phi)$. The random noise component $\varepsilon_0 \sim \mathrm{N}_m(0,\tau^2 I_m)$ is assumed to be independent of S_0 and ε, so that the covariance between observed and unobserved locations is not supposed to be influenced by the measurement error. The latter condition can only be expressed in terms of spatial linear mixed effects models (1) and (8); their adoption becomes crucial in the definition of spatial predictions in the presence of measurement error (Christensen *et al.*, 1992; Cressie, 1993 pp. 127-130). Observed and unobserved realizations can then be given the following joint distribution

$$\begin{pmatrix} Y \\ Y_0 \end{pmatrix} \sim \mathrm{N}_{n+m}\left(\begin{pmatrix} X\beta \\ X\beta_0 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{10} \\ \Sigma_{10}^{'} & \Sigma_{11} \end{pmatrix}\right) \tag{9}$$

where X_0 is an $m \times p$ matrix containing coordinates/covariates for unsampled locations t_0 and covariance matrices Σ_{00} and Σ_{10} are respectively given by

$$\begin{aligned} \Sigma_{00} &= \sigma^2 H_{00}(\phi) + \tau^2 I_m \\ \Sigma_{10} &= \sigma^2 H_{10}(\phi) \end{aligned} \tag{10}$$

As is well known from Normal distribution theory, the minimum variance predictor of Y_0 conditional on values of β and ϖ , is given by

$$\mathrm{E}(Y_0 \mid Y = \mathrm{y}) = X_0\beta + \Sigma_{10}^{'}\Sigma_{11}^{-1}(\mathrm{y} - X\beta) \tag{11}$$

When substituting a BLUE for β , as the GLS estimator $\hat{\beta}$ in (5), one obtains a so called *Best Linear Unbiased Predictor* (BLUP, Robinson, 1991, Davidian and Giltinan, 1995)

$$\hat{\mathrm{Y}}_0 = X_0\hat{\beta} + \Sigma_{10}^{'}\Sigma_{11}^{-1}(\mathrm{y} - X\hat{\beta}) \tag{12}$$

Expression (12) has (least) prediction MSE given by

$$\mathrm{MSE}\left(\hat{Y}_0\right)=\Sigma_{00}-\Sigma_{10}^{'}\Sigma_{11}^{-1}\Sigma_{10}+\left(X_0-\Sigma_{10}^{'}\Sigma_{11}^{-1}X\right)\left(X^{'}\Sigma_{11}^{-1}X\right)^{-1}\left(X_0-\Sigma_{10}^{'}\Sigma_{11}^{-1}X\right)^{'} \tag{13}$$

Also the latter expression contains a negative bias due to considering ϖ as known.

Expressions (11) and (12) are referred to as *simple Kriging* and *universal Kriging* predictor in Geostatistical terms, when the covariance matrix Σ_{10} becomes Σ_{11} for $\mathrm{t}=\mathrm{t}_0$ (which is not our case), leading to the so-called exactness of such predictors, *i.e.* to the feature of predicting observed points by the points themselves. This is only reasonable when no random noise is assumed for the spatial process and predictions must coincide with observed measurements at sampled locations, but can lead to overfitting of the predicted surface in the presence of measurement error or small scale variability (Cressie, 1993 p. 129), producing seriously dangerous extrapolations and confusing interpolations. In these cases the use of the mixed effect spatial model (9) with covariance structure (10) is suggested, where the error variance τ^2 is not contained in the expression of the predictor but for observed locations. Notice that one can easily obtain an exact predictor in the mixed effect context, by modeling Σ_{10} as in Ecker (1999). Robinson (1991) and Christensen (1991) first recognized the Kriging predictor as a BLUP in the context of linear mixed models. The latter author advocates the introduction of a measurement error term as in model (1) when there's evidence of erraticity in the data, to avoid the aforementioned exactness of the predictor.

3.2. *Measurement error and residuals*

In classical Geostatistics, when no measurement error is considered for the data at hand and the spatial covariance structure does not include any nugget effect, the kriging predictor is said to be exact in that it exactly reproduces observed values when calculated at sampled locations. In order to preserve this exactness when a nugget effect is included in the covariance structure, the predictor has to be discontinuous at $\nu\to 0$. In other words in this case the predicted surface is indeed smooth with useless spikes corresponding to observed values at sampled locations. As a matter of fact one could be interested in noiseless predictions of the spatial process at data points, obtainable by a spatial covariance function $\sigma^2 h(\nu;\phi)$ more plausibly continuous at $\nu\to 0$, relegating the nugget effect τ^2 to the additive random noise term ε (Christensen *et al.*, 1992). As (10) shows, spatial linear mixed effects models enable to specify the covariance Σ_{10} between observations Y and unobserved realizations of the process Y_0 as totally due to spatially correlated random effects and not to the random noise term ε. These considerations lead to the definition of the following *noiseless predictor for sampled locations*

$$\hat{Y} = X\hat{\beta} + \sigma^2 H_{11}\left(\tau^2 I_n + \sigma^2 H_{11}\right)^{-1}\left(\mathrm{y} - X\hat{\beta}\right) \tag{14}$$

which can also be written in the form of a James-Stein (shrinkage) estimator

$$\hat{Y} = \mathrm{y} + \left(I_n - \sigma^2 H_{11}\Sigma_{11}^{-1}\right)\left(\mathrm{y} - X\hat{\beta}\right) \tag{15}$$

with the appealing feature that the more $\sigma^2 H_{11}$ approaches Σ_{11}, *i.e.* the smaller the nugget effect τ^2, the more predictions $\hat{Y}$ are shrunken towards observations y.

Conversely, when the nugget effect is remarkable, more weight is attributed to the average mean profile $X\hat{\beta}$. Notice that the *mean squared error* of predictor $\hat{Y}$ in (14) can be obtained by suitably adapting expression (13).

Expression (14) enables to define residuals as the difference between observed and fitted (predicted) values

$$\mathrm{e} = \mathrm{y} - \hat{Y} = -A\mathrm{y} \tag{16}$$

with $A = (I_n - B)(I_n - C)$, $B = \sigma^2 H_{11}\Sigma_{11}^{-1}$ and $C = X\left(X'\Sigma_{11}^{-1}X\right)^{-1}X'\Sigma_{11}^{-1}$. Unbiasedness of $\hat{Y}$ implies that if the model is correct, then

$$\mathrm{e} \mid \varpi \sim N_n\left(0, A\Sigma_{11}A'\right) \tag{17}$$

4. BAYESIAN INFERENCE

As the main disadvantage with the former approach to prediction is still to be found in the bias induced by considering estimated values of covariance parameters as if they were known, this point is going to be explicitly discussed in this section. The Bayesian approach to Geostatistics has a long history: seminal theoretical papers include those of Kitanidis (1986), Omre (1987) and Le and Zidek (1992) who first proposed the use of hierarchical prior models. Handcock and Stein (1993) compare the results of a Bayesian analysis with those obtained by kriging and perform some sensitivity analysis with respect to the choice of the priors. De Oliveira *et al.* (1997) extend the work of the latter authors to non-Gaussian random fields by a family of Box-Cox transformations. More recent contributions (Diggle *et al.*, 1998; Gaudard *et al.*, 1999) are more and more suitable to extensive applications making use of powerful MCMC estimation techniques.

Both the response variable Y and parameters (β, ϖ) are considered as random quantities within the Bayesian approach. External information on model parameters (*i.e.* expert beliefs, physical constraints, etc.) is specified in terms of the *joint prior distribution* $\pi(\beta, \varpi) = \pi(\beta)\pi(\tau^2)\pi(\sigma^2, \phi)$, possibly depending on some *hyperparameters*. The random variable Y has marginal probability distribution (likelihood) given by (3) or equivalently

$$p(\mathrm{y} \mid \beta, \varpi) = p(\mathrm{y} \mid \beta, \tau^2, \sigma^2, \phi) = \int p(\mathrm{y} \mid S, \beta, \tau^2) p(S \mid \sigma^2, \phi) dS \tag{18}$$

where the conditional dependence on covariance parameters is made explicit. External information regarding parameters β and ϖ is updated after data collection by combining the prior distribution with the likelihood function in the *posterior distribution* obtainable by Bayes' theorem

$$\begin{aligned} &\pi(\beta, \varpi \mid \mathrm{y}) = \pi(\beta, \tau^2, \sigma^2, \phi \mid \mathrm{y}) = \\ &= \frac{\int p(\mathrm{y} \mid S, \beta, \tau^2) p(S \mid \sigma^2, \phi) \pi(\beta) \pi(\tau^2) \pi(\sigma^2, \phi) dS}{\int \cdots \int p(\mathrm{y} \mid S, \beta, \tau^2) p(S \mid \sigma^2, \phi) \pi(\beta) \pi(\tau^2) \pi(\sigma^2, \phi) d\beta dS d\tau^2 d\sigma^2 d\phi} \end{aligned} \tag{19}$$

Bayesian inference on mean and covariance parameters β and ϖ is based on numerical summaries of (19), whereas predictions of the spatial process at unsampled locations involve the consideration of the predictive distribution given by the following posterior expectation

$$\begin{aligned} &p(\mathrm{y}_0 \mid \mathrm{y}) = \iint p(\mathrm{y}_0 \mid \mathrm{y}, \beta, \varpi) \pi(\beta, \varpi \mid \mathrm{y}) d\beta d\varpi = d\beta d\varpi \\ &= \iint \frac{p(\mathrm{y}_0, \mathrm{y} \mid \beta, \varpi) \pi(\beta, \varpi)}{\iint p(\mathrm{y} \mid \beta, \varpi) \pi(\beta, \varpi) d\beta d\varpi} d\beta d\varpi \end{aligned} \tag{20}$$

where the joint density $p(\mathrm{y}_0, \mathrm{y} \mid \beta, \varpi)$ is given by (9). Whether maximum likelihood and traditional geostatistical prediction methods consider the expected value of the conditional distribution $p(\mathrm{y}_0 \mid \mathrm{y}, \beta, \varpi)$ for estimated values of mean and covariance parameters, the Bayesian predictive approach is essentially based on averaging such a distribution over the whole parameter space using the posterior $\pi(\beta, \varpi \mid \mathrm{y})$ as a weighting function. In other words the Bayesian predictive distribution, rather than focusing on the maximum likelihood estimates of the mean and covariance parameters, takes into account the whole likelihood surface.

Calculation of integrals in (19) and (20) can be extremely difficult to carry out either analytically and numerically, the choice of prior distributions being obviously crucial. The use of *conjugate* priors, defined as those inducing posterior densities that belong to the same functional class of the prior, is often motivated by computational convenience and easy interpretability. In practice, for complicated likelihood models, conjugate prior distributions may not even be found. The lack of prior information is sometimes represented by distributions guaranteed to play a minimal role in the posterior, referred to as *noninformative* or flat priors (Gelman *et al.*, 1995).

4.1. *A simplified model*

Following Searle *et al.* (1992) we begin by considering the simplest Bayesian *hierarchical* form of the normal linear mixed effects spatial model, avoiding prior specification for covariance parameters ϖ (*i.e.* assuming their values are known with probability one, as it is often done in standard geostatistical practice)

$$Y \mid \beta, S \sim \mathrm{N_n}\left(X(\mathrm{t})\beta + S, \tau^2 I_n\right)$$

$$\beta \sim \mathrm{N}_p(\beta_0, B_0) S \sim \mathrm{N}_n\left(0, \sigma^2 H_{11}(\phi)\right) \tag{21}$$

β, S independent

Notice that the same model can equivalently be set by replacing the first condition in (21) by $\varepsilon \sim \mathrm{N}_n\left(0, \tau^2 I_n\right)$ independently of β and S, where ε is the random term in (1). It is easily shown that the joint distribution of the random vector $(Y, \beta, S)'$ is given by

$$\begin{pmatrix} Y \\ \beta \\ S \end{pmatrix} \sim \mathrm{N}_{2n+p}\left(\begin{pmatrix} X\beta_0 \\ \beta_0 \\ 0 \end{pmatrix}, \begin{pmatrix} XB_0X' + \Sigma_{11} & XB_0 & \sigma^2 H_{11}(\phi) \\ B_0X' & B_0 & \mathrm{O} \\ \sigma^2 H_{11}(\phi) & \mathrm{O} & \sigma^2 H_{11}(\phi) \end{pmatrix}\right) \tag{22}$$

This oversimplified model has little to do with a full Bayesian treatment of the Geostatistical estimation problem, where priors for ϖ should be specified involving a third level of the hierarchy. Nevertheless it proves to be useful in elucidating connections between ML and REML estimation techniques and the Bayesian estimation framework.

The mean parameter β and the spatial signal S are treated similarly in model (21), in that they both have their prior distributions and no distinction is made between fixed and random effects. The posterior distribution (19) reduces to

$$p(\beta \mid \mathrm{y}) = \frac{\int p(\mathrm{y} \mid \beta, S)\pi(\beta)p(S)dS}{\iint p(\mathrm{y} \mid \beta, S)\pi(\beta)p(S)d\beta dS}$$

Standard normal theory arguments lead to the conclusion that $p(\beta \mid \mathrm{y})$ is Normal, with the following expressions for the posterior expectation (the Bayesian estimator of the trend parameter β)

$$\mathrm{E}(\beta \mid \mathrm{y}) = \left(X'\Sigma_{11}^{-1}X + B_0^{-1}\right)^{-1}\left(X'\Sigma_{11}^{-1}\mathrm{y} + B_0^{-1}\beta_0\right) \tag{23}$$

and the posterior variance

$$\mathrm{Cov}(\beta \mid \mathrm{y}) = \left(X'\Sigma_{11}^{-1}X + B_0^{-1}\right)^{-1} \tag{24}$$

Estimates based on the posterior distribution require specification of β_0 and B_0. If one assumes that no prior information is available on β and chooses a non-informative flat prior of the form $B_0^{-1} = \mathrm{O}$ (*i.e.* null precision for prior $\pi(\beta)$) it is easy to notice that expressions (23) and (24) coincide with those of the BLUE (5) and its covariance (6).

Until now we have been considering covariance parameters as known, or equivalently assuming a degenerate prior giving probability one to a single value of ϖ. If attention is restricted to model (21), assuming $\beta = \beta_0$ is fixed, then obviously the posterior distribution for ϖ is proportional to the likelihood function (3.1) calculated at $\beta = \beta_0$. Therefore ML estimation of covariance parameters can be derived within model (21) as a Bayesian estimator with the value of β taken as a fixed unknown constant β_0.

Suppose instead that the locally uniform prior $\pi(\beta) \propto 1$ is specified, then after some algebra (Davidian and Giltinan, 1995)

$$\pi(\varpi \mid \mathrm{y}) \propto L(\varpi; \mathrm{y}) = \iint p(\mathrm{y} \mid \beta, S)\pi(\beta)p(S)\,d\beta dS = \int p(\mathrm{y} \mid \beta)\pi(\beta)\,d\beta =$$
$$= \int \left[(2\pi)^n |\Sigma_{11}|\right]^{-1/2} \exp\left[-\frac{1}{2}(\mathrm{y} - X\beta)'\Sigma_{11}^{-1}(\mathrm{y} - X\beta)\right] d\beta = \ldots = L\left(\varpi; \mathrm{y} - X\hat{\beta}\right)$$

Thus REML estimation of variance components corresponds to a Bayesian estimation procedure where both spatial random effects S and the trend parameter β are averaged out, the latter using a flat weighting function.

In line with what was done in section 3.2 we consider an extension of model (22) as in (9), to include unobserved spatial locations: as a consequence the random vector $(Y, Y_0, \beta, S, S_0)'$ results to be jointly distributed according to a $(2n + 2m + p)$-variate normal with mean vector $(X\beta_0, X_0\beta_0, \beta_0, \mathbf{0}, \mathbf{0})'$ and covariance matrix

$$\begin{pmatrix} XB_0X' + \Sigma_{11} & XB_0X_0' + \Sigma_{10} & XB_0 & \sigma^2 H_{11}(\phi) & \Sigma_{10} \\ \left(XB_0X_0' + \Sigma_{10}\right)' & X_0B_0X_0' + \Sigma_{00} & X_0B_0 & \Sigma_{10}' & \sigma^2 H_{00}(\phi) \\ B_0X' & B_0X_0' & B_0 & \mathrm{O} & \mathrm{O} \\ \sigma^2 H_{11}(\phi) & \Sigma_{10} & \mathrm{O} & \sigma^2 H_{11}(\phi) & \Sigma_{10} \\ \Sigma_{10}' & \sigma^2 H_{00}(\phi) & \mathrm{O} & \Sigma_{10}' & \sigma^2 H_{00}(\phi) \end{pmatrix} \tag{25}$$

Again it is easily shown that this setting leads to a Gaussian predictive distribution with mean and covariance parameters respectively given by following expressions

$$\mathrm{E}(Y_0 \mid \mathrm{y}) = X_0\beta_0 + \left(XB_0X_0' + \Sigma_{10}\right)'\left(XB_0X' + \Sigma_{11}\right)^{-1}(\mathrm{y} - X\beta_0) \tag{26}$$

$$\mathrm{Cov}(Y_0 \mid \mathrm{y}) = X_0 B_0 X_0^{'} + \Sigma_{00} - \left(XB_0 X_0^{'} + \Sigma_{10}\right)^{'} \left(XB_0 X^{'} + \Sigma_{11}\right)^{-1} \left(XB_0 X_0^{'} + \Sigma_{10}\right) \tag{27}$$

Notice that the Bayesian predictor (26) reduces to the least MSE predictor (11) when $B_0 = \mathrm{O}$, *i.e.* when a degenerate prior for β is assumed, assigning probability one to a specific value β_0. If a non informative flat prior ($B_0^{-1} = \mathrm{O}$) is used, then it can be shown (Davidian and Giltinan, 1995) that parameters of the predictive distribution coincide with (12) and (13), so the Bayesian predictor is equivalent to the BLUP.

Neglecting to predict the measurement error part of model (1) one can also obtain a so called *noiseless predictive distribution* (Ecker and Gelfand, 1997) given by $p(X_0\beta + S_0 \mid \mathrm{y})$ which is also Gaussian under model (25). This approach enables to predict the mean of the response at unsampled spatial locations, rather than the response itself, and leads to the same Bayesian predictor $\mathrm{E}(X_0\beta + S_0 \mid \mathrm{y}) = \mathrm{E}(Y_0 \mid \mathrm{y})$ but to a smaller posterior predictive variance

$$\begin{aligned} &\mathrm{Cov}(X_0\beta + S_0 \mid \mathrm{y}) = \\ &= X_0 B_0 X_0^{'} + \sigma^2 H_{00}(\phi) - \left(XB_0 X_0^{'} + \Sigma_{10}\right)^{'} \left(XB_0 X^{'} + \Sigma_{11}\right)^{-1} \left(XB_0 X_0^{'} + \Sigma_{10}\right) \end{aligned} \tag{28}$$

The difference between (27) and (28) is that the latter expression does not take the measurement error into account when measuring the variance of unobserved locations. Notice that expressions (27) and (28) are estimates of the dispersion of the predictive distributions and do not correspond to frequentist variances of the two predictors, the latter being only conceivable under repeated sampling assumptions. Decision theory considerations, and the specification of a loss function, enable the evaluation of the quality of estimators in the Bayesian framework.

Expressions for the spatial trend parameter estimator (23) and the predictor (26) involve unknown covariance parameters ϖ. Taking into account uncertainty on covariance parameters is essential for a full Bayesian treatment of Geostatistical data. Formal Bayesian estimation of the spatial correlation would imply extending model (21) to include prior specification for ϖ, deriving new expressions for the fixed effects estimator and the predictor, by averaging out these parameters according to their priors. As this can be quite cumbersome to carry out in practice, an intuitively appealing alternative would be to derive a point estimate of ϖ from a marginal distribution and substitute it into expressions (23) and (26). The general strategy of substituting point estimates of parameters into posterior distribution descriptors is known as *empirical Bayes estimation*. For the problem at hand it can be easily shown (Searle *et al.*, 1992) that the marginal likelihood for ϖ is Gaussian with parameters

$$\begin{aligned} &\mathrm{Cov}(\mathrm{y} \mid \varpi) = A = \left(\Sigma_{11}^{-1} - \Sigma_{11}^{-1} XCX^{'} \Sigma_{11}^{-1}\right)^{-1} \\ &\mathrm{E}(\mathrm{y} \mid \varpi) = A \Sigma_{11}^{-1} XCB_0^{-1} \beta_0 \end{aligned} \tag{29}$$

with $C = \left(B_0^{-1} + X' \Sigma_{11}^{-1} X\right)^{-1}$. Notice that when hyperparameters β_0 and B_0 are considered as specified values, marginal maximum likelihood estimates of elements of ϖ can be calculated and substituted into (23) and (26) to obtain the empirical Bayes estimate of the mean parameter β and the empirical Bayes predictor. Variances of such estimators are not easily obtainable, substitution of estimates of ϖ into (24), (27) and (28) rather leading to possibly negatively biased estimated variances.

4.2. *Priors on covariance structure parameters*

As Geostatistics is concerned with the estimation of the spatial variability and correlation structure, full Bayesian Geostatistical modeling involves prior specification for covariance parameters ϖ. In this section we review some recent attempts to extend model (21) by adding a third level to the hierarchy, assuming ϖ is a priori distributed as $p(\varpi)$. Formal Bayesian estimation of β and ϖ consists in obtaining summaries of the location of the posterior distribution

$$\pi(\beta, \varpi \mid \mathrm{y}) = \frac{p(\mathrm{y} \mid \beta, \varpi)\pi(\beta, \varpi)}{\int p(\mathrm{y} \mid \beta, \varpi)\pi(\beta, \varpi) d\beta d\varpi} \tag{30}$$

If this strategy is adopted considering the posterior mode with a conjugate Gaussian prior for β and a non-informative (flat) prior for ϖ, one obviously obtains the same results as in section 4.1.

Now let's first assume that the range or any correlation structure parameter ϕ and the relative nugget $\tau_R^2 = \tau^2 / \sigma^2$ are known or fixed, so that the spatial covariance matrix Σ_{11} is known up to the scale factor σ^2 (say $\Sigma_{11} = \sigma^2 D_{11}$, where D_{11} is known). Uncertainty on the trend parameter β and the partial sill σ^2 can be taken into account by considering the jointly conjugate prior (Kitanidis, 1986; Ribeiro and Diggle, 1999*a*) for $\left(\beta, \sigma^2\right)$, *i.e.* the Normal-Inverse Gamma distribution given by

$$\beta, \sigma^2 \sim \mathrm{N}_p\left(\beta_0, \sigma^2 B\right) \mathrm{IG}(a, b) \tag{31}$$

where B_0 was reparametrized as $\sigma^2 B$, leading to a joint posterior distribution of the same form for $\left(\beta, \sigma^2\right)$

$$\beta, \sigma^2 \mid \mathrm{y} \sim \pi\left(\beta \mid \sigma^2, \mathrm{y}\right) \pi\left(\sigma^2 \mid \mathrm{y}\right) = \mathrm{N}_p\left(\beta_{\mid \sigma^2, \mathrm{y}}, \sigma^2 B_{\mid \sigma^2, \mathrm{y}}\right) \mathrm{IG}\left(a_{\mid \mathrm{y}}, b_{\mid \mathrm{y}}\right) \tag{32}$$

with updated parameters

$$\beta_{|\sigma^2,\mathrm{y}} = \left(X^{'}D_{11}^{-1}X + B^{-1}\right)^{-1}\left(X^{'}D_{11}^{-1}\mathrm{y} + B^{-1}\beta_0\right) = FE$$
$$B_{|\sigma^2,\mathrm{y}} = \left(X^{'}D_{11}^{-1}X + B^{-1}\right)^{-1} = F \tag{33}$$

where the latter expressions are equivalent to setting $B_0 = \sigma^2 B$ in (23) and (24), and

$$a_{|\mathrm{y}} = a + \frac{n}{2}$$
$$b_{|\mathrm{y}} = b + \frac{1}{2}\left[\mathrm{y}^{'}D_{11}^{-1}\mathrm{y} + \beta_0^{'}B^{-1}\beta_0 - E^{'}FE\right] \tag{34}$$

With respect to the simplified model of section 5.1, the introduction of the jointly conjugate prior (31) allows the simultaneous estimation of fixed effects β and the scale parameter σ^2. The Bayesian estimator of the partial sill σ^2 is given by the mode of the posterior $\pi\left(\sigma^2 \mid \mathrm{y}\right)$, *i.e.* by $b_{|\mathrm{y}} / \left(a_{|\mathrm{y}} + 1\right)$. Integration of the joint posterior (32) with respect to σ^2 leads to the conclusion that the marginal posterior for β is multivariate Student-*t* with $2a_{|\mathrm{y}}$ degrees of freedom, mean $\beta_{|\sigma^2,\mathrm{y}}$ and scale parameter $b_{|\mathrm{y}} B_{|\sigma^2,\mathrm{y}} / a_{|\mathrm{y}}$

$$\beta \mid \mathrm{y} \sim \mathrm{t}_{2\mathrm{a}_{|\mathrm{y}}}\left(\beta_{|\sigma^2,\mathrm{y}}, \frac{b_{|\mathrm{y}}}{a_{|\mathrm{y}}} B_{|\sigma^2,\mathrm{y}}\right) \tag{35}$$

Within the same setting, the analytic computation of the predictive distribution $p\left(\mathrm{y}_0 \mid \mathrm{y}\right)$ requires obtaining the following conditional predictive distribution first

$$\mathrm{y}_0 \mid \sigma^2, \mathrm{y} \sim \mathrm{N_m}\left(\mu_{0|\sigma^2,\mathrm{y}}, \sigma^2 H_{0|\sigma^2,\mathrm{y}}\right) \tag{36}$$

with parameters given by

$$\mu_{0|\sigma^2,\mathrm{y}} = GFB^{-1}\beta_0 + \left(GFX^{'}D_{11}^{-1} + H_{10}^{'}D_{11}^{-1}\right)\mathrm{y}$$
$$H_{0|\sigma^2,\mathrm{y}} = D_{00} - H_{10}^{'}D_{11}^{-1}H_{10} + GFG^{'} \tag{37}$$

where F was introduced in (33), $G = X_0 - H_{10}^{'}D_{11}^{-1}X$ and $D_{00} = \tau_R I_m + H_{00}(\phi)$.

Then the predictive distribution is given by

$$p\left(\mathrm{y}_0 \mid \mathrm{y}\right) = \int p\left(\mathrm{y}_0 \mid \sigma^2, \mathrm{y}\right)\pi\left(\sigma^2 \mid \mathrm{y}\right) = \mathrm{t}_{2\mathrm{a}_{|\mathrm{y}}}\left(\mu_{0|\sigma^2,\mathrm{y}}, \frac{b_{|\mathrm{y}}}{a_{|\mathrm{y}}} H_{0|\sigma^2,\mathrm{y}}\right) \tag{38}$$

The non informative improper prior for the same situation is given by

$$\pi\left(\beta,\sigma^2\right)\propto\frac{1}{\sigma^2} \tag{39}$$

and corresponds to assuming $a=0$ and $B^{-1}=\mathrm{O}$ in (31). Substitution of these values into (34), (35) and (38) leads to expressions of posterior and predictive distributions for this special case. The Bayesian estimator of the spatial trend parameter and the predictor turn out to be respectively equivalent to the BLUE (5) and the BLUP (12).

No general methods are available for the analytical derivation of the posterior distribution $\pi(\beta,\varpi\,|\,\mathrm{y})$ when the spatial correlation structure is unknown in all its parameters $\varpi=\left(\sigma^2,\phi,\tau^2\right)$. As a matter of fact the posterior $\pi(\varpi\,|\,\mathrm{y})$ is not obtainable in closed form for most currently used covariance functions and numerical simulation-based integration methods must be used instead.

Ecker and Gelfand (1997, 1999) choose to simulate the joint posterior $\pi(\beta,\varpi\,|\,\mathrm{y})$ by importance sampling (Geweke, 1989), assuming non-informative prior distributions and obtaining the importance sampling density by West's adaptive mixture method (West, 1993). Posterior parameter estimation is performed by summarizing sampled values, which are also attached to $p(\mathrm{y}_0\,|\,\mathrm{y},\varpi,\beta)$ in order to obtain random draws of $\mathrm{y}_0\,|\,\mathrm{y}$, *i.e.* simulations of the predictive distribution.

Ribeiro and Diggle (1999b) take full advantage of the former analytic results by turning to posterior simulation only for parameters ϕ and τ_R^2. They notice that $\pi(\beta,\varpi\,|\,\mathrm{y})=\pi\left(\beta,\sigma\,|\,\phi,\tau_R^2,\mathrm{y}\right)\pi\left(\phi,\tau_R^2\,|\,\mathrm{y}\right)$ implies

$$\pi\left(\phi,\tau_R^2\,|\,\mathrm{y}\right)\propto\frac{p(\mathrm{y}\,|\,\beta,\varpi)\pi\left(\beta,\sigma^2\,|\,\phi,\tau_R^2\right)}{\pi\left(\beta,\sigma^2\,|\,\phi,\tau_R^2,\mathrm{y}\right)}\pi\left(\phi,\tau_R^2\right) \tag{40}$$

where the numerator is the product of the likelihood (4) and prior (31) and the denominator is given by (32). A discrete prior for ϕ and τ_R^2 on a pre-specified reference grid of values is assumed and random draws from (40) are obtained and attached to (32) and (38) which are sampled again to give draws from the posterior and the predictive distribution respectively.

5. CASE STUDY: SPATIAL PREDICTIONS OF TAXONOMICAL SOIL PROPERTIES

In this section simple geostatistical predictions are going to be obtained within the mixed models framework. Data concerning the soil structure of an agricultural area in the Foggia district were supplied by “Istituto Sperimentale Agronomico

– Bari” and originally collected by the “Consorzio di bonifica della Capitanata”, in order to inform the water supply management about water needs in this area, traditionally cropped with durum wheat.

The composition of the soil texture in sand, silt and clay was measured at 175 sample locations on an area of about 75000*ha*, as part of the complete characterization of the soil properties of the area. The area under investigation is known to be quite rich in silt and clay, exploratory data analysis suggesting the presence of a drift towards northwest of observed sand measurements and an isotropic process underlying their spatial distribution. Assumptions underlying the correct use of linear mixed effects models are exactly the same as those necessary to rely on the optimality of the classical linear (kriging) predictor, *i.e.* Gaussianity, stationarity and isotropy of the data-generating process. Compositional data, as those considered in this case study, are clearly non-Gaussian by definition, nevertheless the marginal distribution of sand rates in figure 1 looks symmetric and bell shaped (still Gaussianity of the marginal distribution isn't but a necessary condition for the joint normality of the process). The common behavior of the four directional variograms, covering the whole range of directions with a tolerance of 22.5 degrees, shows the reasonableness of the isotropy assumption. Finally the contourplot of the spatial distribution of sand rates, obtained by smoothing the original noisy surface by triangulation of data points and linear interpolation on an even 40×40 grid (as the ones in the following figures 5.2-5.6), pushes towards the consideration of a trend within the spatial model. Evidence for the presence of spatial outliers was also excluded on the basis of exploratory tools such as variogram clouds and graphs described in Haining (1990, p. 214).

Though only approximately providing a probabilistic description of the data generating process in this illustrative example, linear mixed effects models were used in the light of the result by Verbeke and Lesaffre (1997) showing that MLEs consistency and asymptotic normality still apply when the random effects distribution is not normal. The composition in sand of the soil samples was then considered as the response variable in a model including a spatial trend or fixed effects term. Simultaneous estimates of fixed effects and correlation structure parameters together with some overall fit likelihood-based statistics were obtained by the proposed parametric modelling approach. After rescaling the spatial coordinates nine different models for three possible trend surfaces including one or both spatial coordinates and three standard covariance functions (exponential, spherical and Gaussian) were estimated by minimizing the REML likelihood using the ridge-stabilized Newton-Raphson algorithm implemented in the SAS-MIXED procedure (Littel *et al.*, 1996). Incidentally notice that values of the estimates comparable to those shown in table 1 were obtained using Splus-lme and Splus-gls, the latter including a nugget effect within the covariance structure. The three models with trend in both spatial coordinates (I, IV and VII) had fixed effects estimates close to zero and associated to high standard errors. On the other hand the remaining six models had significant fixed effects and, having the same number of parameters, could be compared in terms of attained residual likelihood.

The selected model VI corresponds to the lowest value of overall fit statistics and shows a trend in the south-north direction ($\hat{\beta}_{\text{lat}} = 0.4$) and a spherical correlation structure. Notice that the estimated value of the range parameter ($\hat{\phi} = 11.42$ km) corresponds to a long-range spatial correlation structure typically associated to taxonomical features of soils as the composition in sand (Castrignanò *et al.*, 1999).

When compared to the empirical one the spherical variogram estimated by model VI shows a reasonable fit to observed data (figure 2); a further check of the model fit was done considering the residuals from the fitted mixed effect model defined in §3.2, as they ought to be randomly distributed over the area under investigation: the contourplot in figure 2 shows no drift in any direction and some residual spatial pattern neglected by the fitted model.

Still the empirical variogram of residuals looks almost constant around the value 15: the covariance between residuals at pairs of spatial locations doesn't seem to depend on the distance, *i.e.* all the relevant spatial correlation in the data is reproduced by model VI.

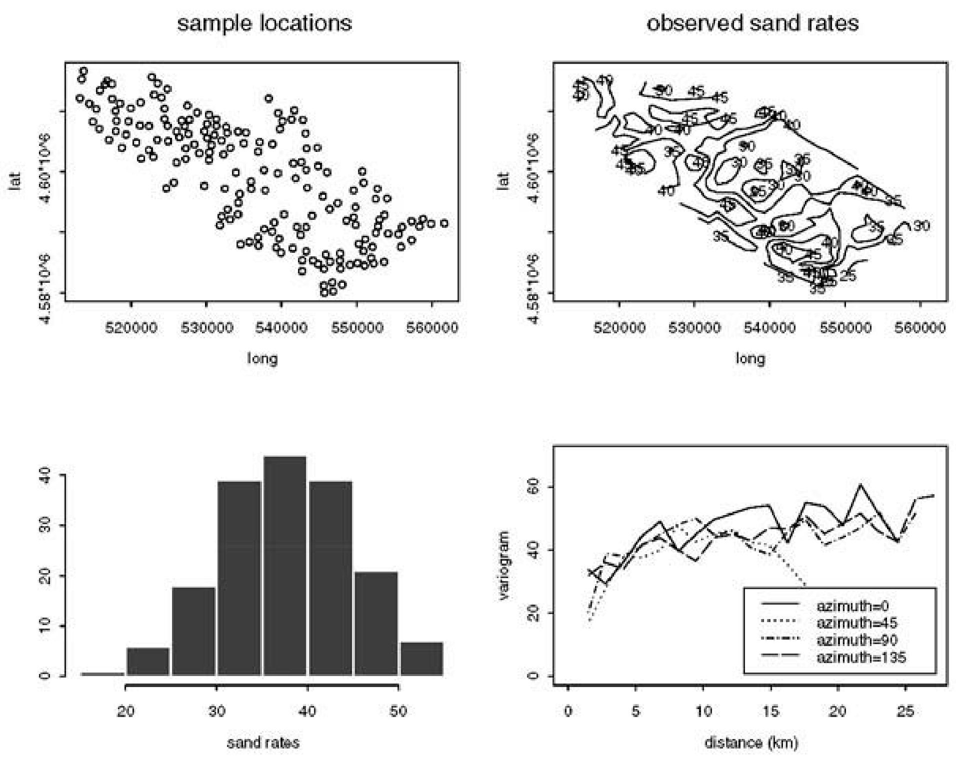

Figure 1 – Scatterplot of the 175 sampled locations, contourplot (based on a linear interpolation), histogram and directional variograms of the rates of composition in sand of the soil samples.

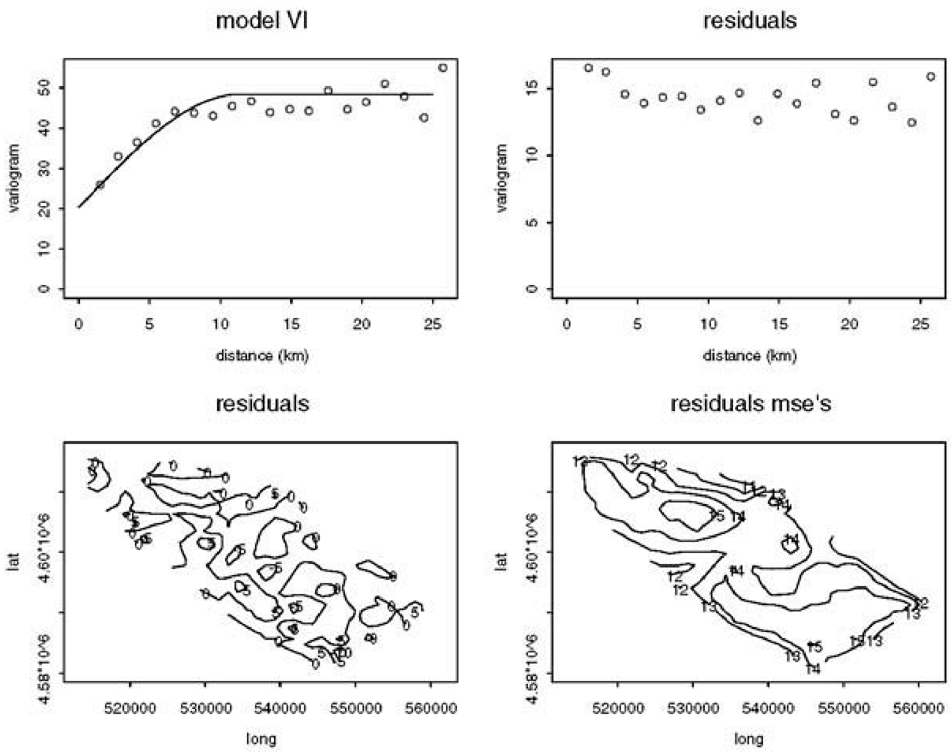

Figure 2 – Empirical and estimated spherical variogram for model VI, empirical variogram of residuals, contourplot (based on a linear interpolation) of residuals and their mse's.

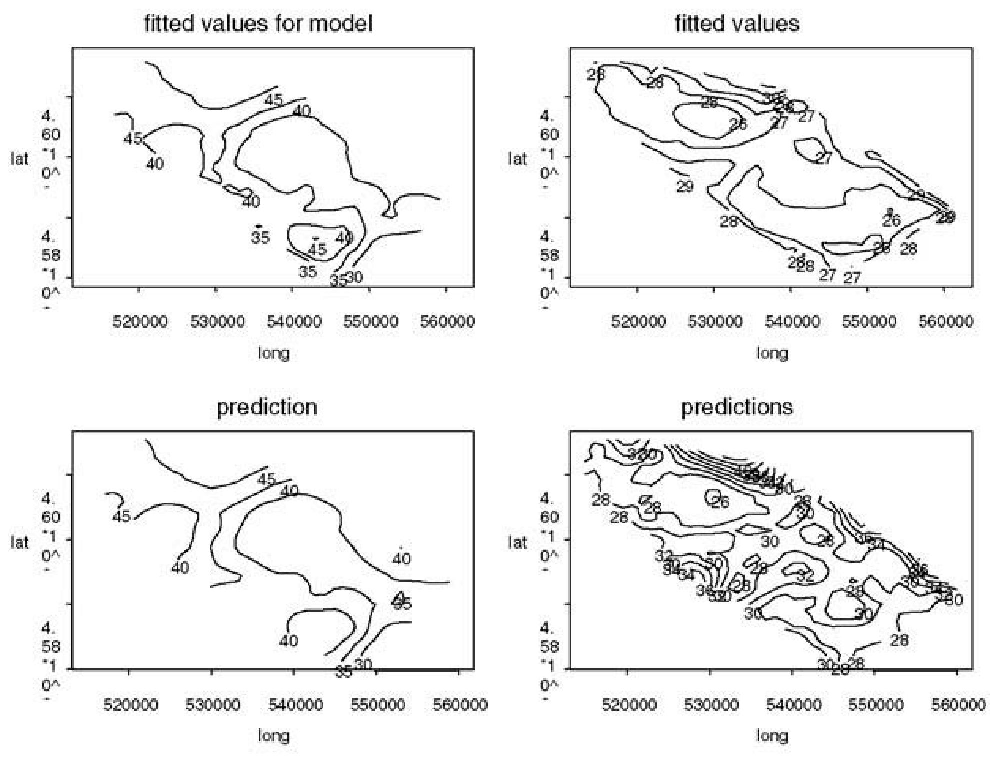

Figure 3 – Contourplots (based on a linear interpolation) of fitted values and predictions on a square grid for model VI and their mse's.

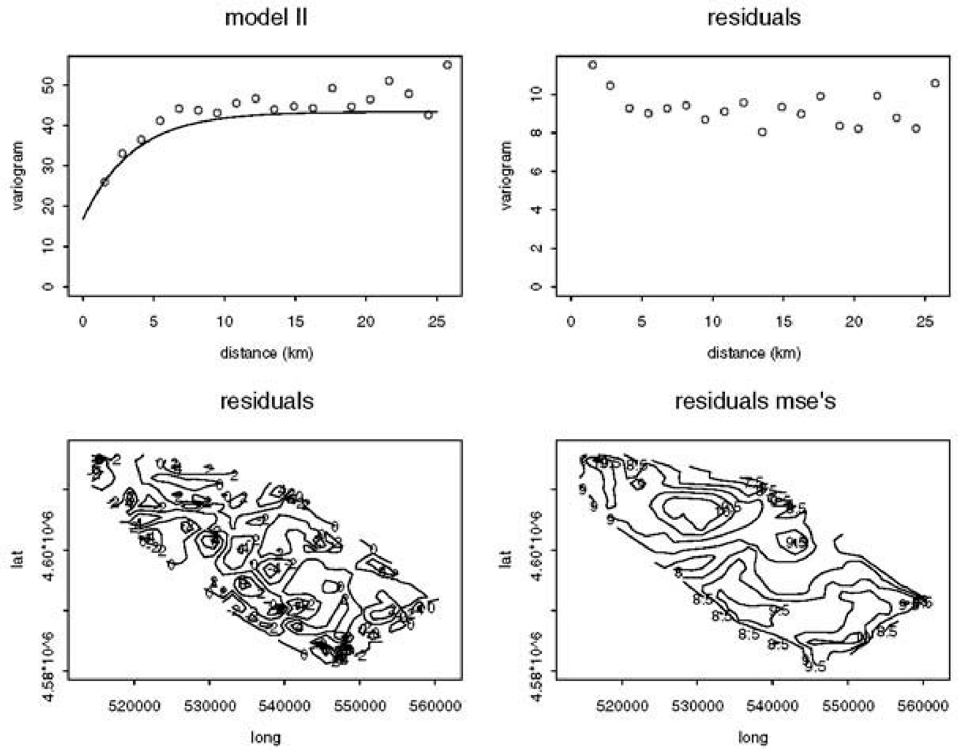

Figure 4 – Empirical and estimated spherical variogram for model II, empirical variogram of residuals, contourplot (based on a linear interpolation) of residuals and their mse's.

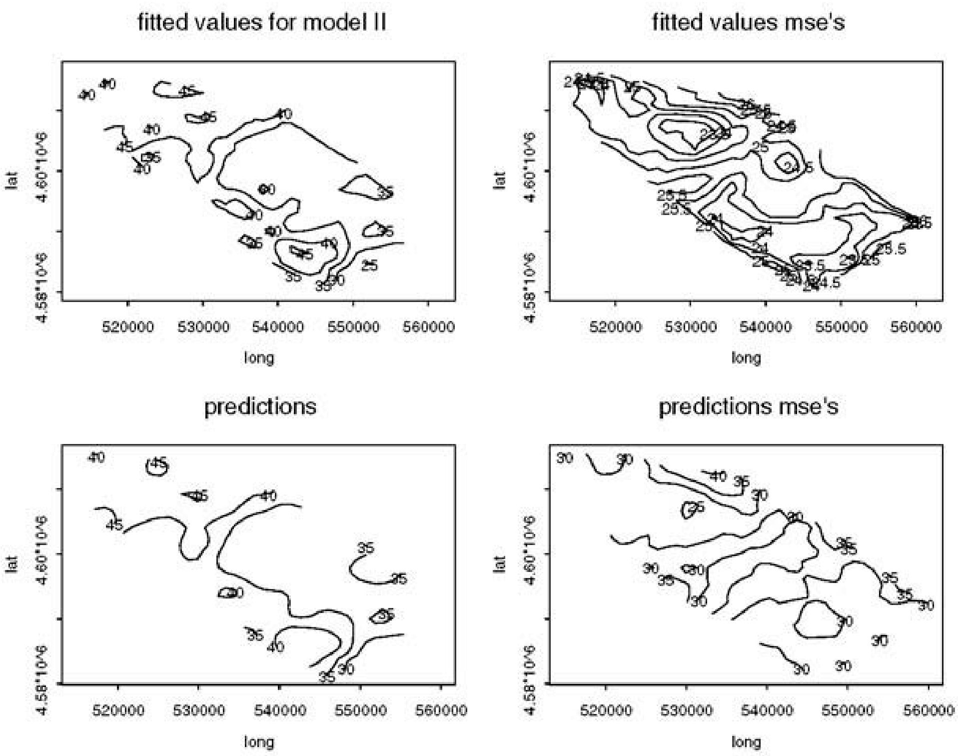

Figure 5 – Contourplots (based on a linear interpolation) of fitted values and predictions on a square grid for model II and their mse's.

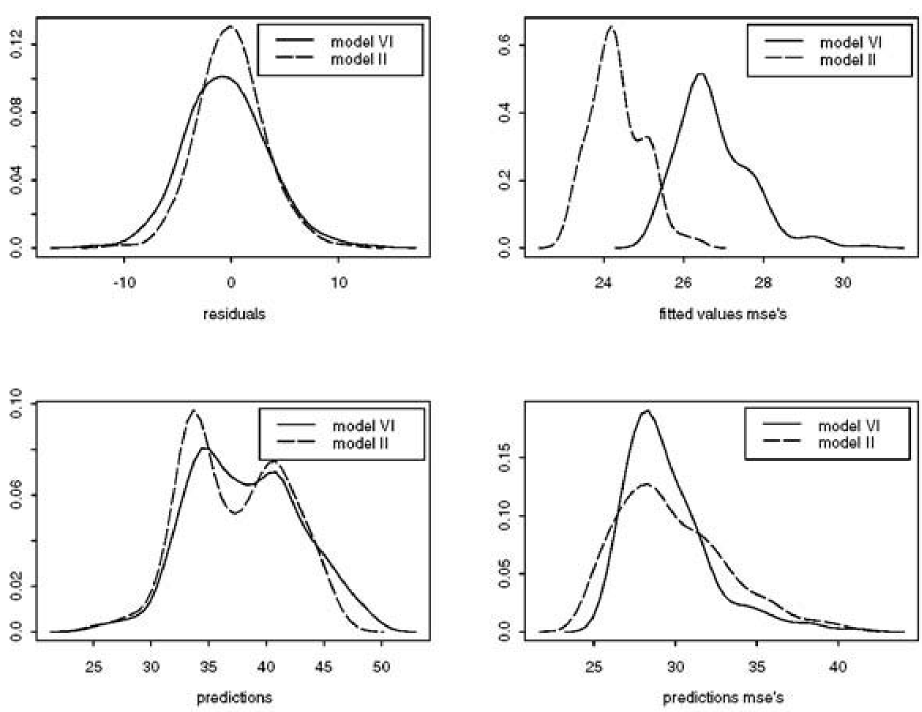

Figure 6 – Marginal density plots (based on a kernel smoothing) of residuals, fitted values mse's, predictions and predictions mse's for models VI and II.

Prediction locations were then identified as 409 points obtained by cutting a square 30×30 grid along the edges of the sampled area. Fitted values for model VI, together with spatial predictions obtained by computing BLUP's as in (12) and shown in figure 3, reproduce the observable relevant spatial pattern quite well. Estimated fitted values and prediction mean squared errors (13) look almost constant and slightly higher where sampled locations are more sparsely distributed.

TABLE 1

Nine different spatial models: covariance structure (COV), REML estimates (EST) of fixed effects (EFF) and their standard errors (SE) and estimated values of parameters (NUG, SIL, RAN), overall fit statistics (-2RLL, AIC, BIC).

	COV	EFF	EST	SE	NUG	SIL	RAN	-2RLL	AIC	BIC
I	exp	long lat	-0.17 0.23	0.135 0.172	17.60	28.16	4.35	1112.9	1118.9	1128.4
II	exp	long	-0.28	0.094	16.81	26.58	3.54	1112.9	1118,9	1128.4
III	exp	lat	0.37	0.149	17.99	30.42	5.08	1112.2	1118.2	1127.6
IV	sph	Long lat	-0.18 0.20	0.115 0.149	20.46	22.81	9.25	1111.9	1117.9	1127.3
V	sph	long	-0.28	0.082	22.05	20.58	8.84	1111.7	1117.7	1127.2
VI	sph	lat	0.40	0.134	20.36	28.07	11.42	1111.3	1117.3	1126.8
VII	gau	Long lat	-0.18 0.21	0.120 0.154	16.81	26.58	3.54	1112.9	1118.9	1128.4
VIII	gau	long	-0.28	0.080	24.34	17.56	4.24	1112.2	1118.2	1127.7
IX	gau	lat	0.38	0.127	24.96	21.62	5.38	1112.0	1118.0	1127.4

If we alternatively select the "short range" model II ($\hat{\phi} = 3.54\,\text{km}$) we get a weaker fit of the empirical to the estimated exponential variogram (figure 4) with some evidence for the underestimation of the sill and nugget parameters: as a consequence resulting residuals look less spatially structured (narrower contours), their empirical variogram being constant around the value 10. Fitted values and spatial predictions do not differ that much from the previous ones but rather their precisions do: in figures 5 and 6 fitted values mse's show a weaker spatial correlation than those for model VI, due to the mentioned underestimation of both the nugget and the sill. On the other hand predictions from model II result to be less stable in that their mse's are more variable over the sampled area (figure 5). Finally figure 6 highlights the compromise between residuals and predictions precision obtained by REML estimation: the more peaked marginal distribution of residuals from model II corresponds to sharper but more unstable predictions: their contours look narrower and the empirical variogram is again almost constant around the value 10.

We now turn to the analysis of the same data-set with the same mixed effects spatial model (though only the exponential and spherical correlation structures were considered for illustrative purposes) within the Bayesian framework, by the freely available R package geoR (Ribeiro and Diggle, 1999b). Trying to restrict the role of prior information as much as possible, the non informative prior (39) was chosen for the trend parameter β and the partial sill σ^2. A 201×31 regular grid of values was defined for ϕ and τ_R^2 in the range 2 to 13 and 0 to 1 respectively. Notice how this specification adds little information to the whole setting, the range chosen for both parameters being as wide as to include all plausible values of ϕ and τ_R^2. While values of τ_R^2 were assigned uniform probabilities of being selected, four different prior choices were considered for the range parameter ϕ. The uniform, reciprocal, squared reciprocal and exponential prior put more and more weight on smaller values of the range parameter ϕ. Results of 10000 random draws from posterior distributions (32) and (40) are summarized in table 2 where the modes of the samples are given.

At first we notice that the uniform prior on ϕ gives results quite close to those obtained by REML, due to the very weak prior information added to the data. When additional prior information concerning smaller values of the range is added, all the estimates change accordingly. Posterior predictions obtained by sampling the predictive distribution (38) were not reported as they tend to be similar to those obtained previously. It is to notice that predictions show more variability due to properly considering covariance parameters as estimated rather than known. This feature makes the Bayesian approach the formally correct solution to prediction in that it does not imply conditioning on estimated values of parameters. Moreover, though in the previous example the level of prior information was kept to a minimum in order to compare results to those formerly obtained, this is not always the case: the possibility of properly introducing external information (historical data, physical knowledge of the phenomenon) in the

TABLE 2

Six different spatial models: Bayesian posterior estimates of fixed effects (LON and LAT) and covariance parameters (NUG, SIL, RAN) for four different choices of the prior distribution of the range parameter (PRI) and two alternative specifications of the covariance structure (COV).

	COV	PRI	LON	LAT	NUG	SIL	RAN
I	exp	uni	-0.17	0.23	14.91	27.96	4.26
		rec	-0.18	0.20	14.23	26.69	3.16
		squ	-0.19	0.17	15.24	26.90	2.61
		exp	-0.19	0.17	15.38	27.14	2.44
II	exp	uni	-0.28		14.17	26.57	3.43
		rec	-0.28		15.00	26.47	2.72
		squ	-0.27		15.25	26.92	2.28
		exp	-0.27		15.54	27.42	2.22
III	exp	uni		0.37	15.08	30.16	5.03
		rec		0.36	15.50	29.05	3.87
		squ		0.35	15.24	28.57	3.27
		exp		0.34	15.24	28.57	2.88
IV	sph	uni	-0.18	0.20	19.72	22.76	11.46
		rec	-0.18	0.20	19.58	22.59	9.10
		exp	-0.18	0.19	19.40	22.39	8.82
		squ	-0.22	0.10	19.58	23.49	4.59
V	sph	uni	-0.28		19.83	22.03	8.93
		rec	-0.28		19.65	21.84	8.60
		exp	-0.28		18.96	21.07	8.22
		squ	-0.27		19.51	23.41	4.53
VI	sph	uni		0.40	21.35	27.84	11.52
		rec		0.40	21.29	27.77	11.40
		squ		0.36	19.00	24.79	9.21
		exp		0.20	18.20	23.74	7.45

estimation process increases the flexibility of mixed effects models and can be a key issue for applied scientists. On the other hand, general principles that may be helpful in the elicitation of the prior distribution are rarely available. Responsible data analysis is thus strongly suggested to rely upon suitable Bayesian model choice criteria for comparing competing prior choices. A popular approach due to Gelfand and Ghosh (1998) consists in considering a vector y_{rep}, randomly generated by likelihood (4), and an appropriate distance measure $L(\mathrm{y}_{rep},\mathrm{y})$ as the Euclidean distance $L(\mathrm{y}_{rep},\mathrm{y})=\sum_i (y_{i,rep}-y_i)^2$ in the case of Gaussian data. Model choice can be based on minimizing the posterior expected loss obtained by averaging L with respect to the posterior predictive distribution (20)

$$G(m)=\mathrm{E}(L(\mathrm{y}_{rep},\mathrm{y})\mid \mathrm{y}) \tag{41}$$

Unfortunately no general software that can monitor and summarize (41) by means of the MCMC output is currently available, a suitable modification of the geoR package being easily implementable and warmly desirable.

6. DISCUSSION

Likelihood and Bayesian inferences were implemented assuming a stationary Gaussian process with mean and covariance function of known shape; nevertheless the last few years were marked by a growth of theoretical and applied studies aimed to increase the flexibility of linear models with respect to initial distributional assumptions. For the case of measurements which are not comfortably approximated as Gaussian, Generalized linear mixed effects models are considered introducing latent spatial random effects suitably transformed by a link function to explain the mean of the spatial process. A Gaussian process model for spatial random effects is generally assumed to be sensible (Diggle *et al.*,1998; Gelfand *et al.*, 2000; Christensen *et al.*, 2000). Some important developments of generalized linear mixed models have to do with their extension to bidimensional responses (Banerjee *et al.*, 2000) and the definition of spatially correlated Gaussian random effects obtained through the convolution of a white noise process and a smoothing kernel function. In the latter case (Higdon, 2001) the covariance structure of the model is implicitly determined by (the latent process and) the smoothing kernel instead of being parametrically a priori specified.

In a more general perspective mixed models have recently been the subject of many investigations aimed at diminishing the parametric assumptions in view of their possible extensions to the semiparametric framework (Mallick *et al.*, 2000). The typical exponential family distributional assumption implies the unimodality and an implicit relation between the first two moments of the assumed stochastic process (though on a transformed scale). In the case of continuous data, as those deriving from georeferenced environmental measurements, the possibility of avoiding the assumption of Gaussianity of random effects becomes crucial. This assumption, while convenient from a mathematical point of view turns out to be scarcely realistic in many applied fields. Erroneous probabilistic assumptions may reduce the efficiency of fixed effects and covariance structure parameter estimates and invalidate the random effects predictions, which are highly relevant in the geostatistical context. Finite mixture distribution models (Magder and Zeger, 1996) are today a more flexible and easily applicable alternative thanks to the great developments in simulation-based inferential techniques (Richardson and Green, 1997). The nonparametric alternative consists in the lack of any specific distributional assumption for random effects that can be accomplished by the use nonparametric density estimation (e.g. the predictive recursion algorithm, Follman and Lambert, 1989; Tao *et al.*, 1999). In the Bayesian framework the nonparametric extension of mixed effects models consists in assuming a Dirichlet process prior on random effects, such prior is nothing but a probability distribution on the space of probability distributions (Gelfand and Kottas, 1999; Kleinman and Ibrahim, 1998; Gelfand and Mukhopadhyay, 1995).

Dipartimento di Scienze Statistiche
Università degli Studi di Bari

ALESSIO POLLICE
MASSIMO BILANCIA

RIFERIMENTI BIBLIOGRAFICI

S. BANERJEE, A.E. GELFAND, W. POLASEK, (2000), *Geostatistical modelling for spatial interaction data with application to postal service performance*, "Journal of Statistical Planning and Inference", 90, pp. 87-105.

L. M. BERLINER, (2000), *Hierarchical Bayesian modeling in the environmental sciences*, Technical Report, Ohio State University.

A. CASTRIGNANÒ, R. COLUCCI, D. FERRI, P. LA CAVA, N. MARTINELLI, M. STELLUTI, (1999), *Valutazione e descrizione della fertilità di terreni meridionali mediante la geostatistica multivariata*, in "Atti del convegno della Società Italiana della Scienza del Suolo", Aosta.

G. CHRISTAKOS, (1984), *On the problem of permissible covariance and variogram models*, "Water Resources Research", 20, pp. 251-265.

O.F. CHRISTENSEN, J. MØLLER , R. WAAGEPETERSEN, (2000), *Analysis of spatial data using Generalized Linear Mixed Models and Langevin-type Markov Chain Monte Carlo*, Technical Report, Department of mathematical sciences, Aalborg University, DK.

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

R. CHRISTENSEN, W. JOHNSON, L.M. PEARSON, (1992), *Prediction diagnostics for spatial linear models*, "Biometrika", 79, pp. 583-591.

N.A.C. CRESSIE, (1993), *Statistics for Spatial Data*, John Wiley & Sons.

M. DAVIDIAN, D.M. GILTINAN, (1995), *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall.

V.D. DE OLIVEIRA, B. KEDEM, D.A. SHORT, (1997), *Bayesian prediction of transformed Gaussian random fields*, "Journal of the American Statistical Association", 92, pp. 1422-1433.

P.J. DIGGLE, K.Y. LIANG, S.L. ZEGER, (1994), *Analysis of Longitudinal Data*, Oxford University Press.

P.J. DIGGLE, R.A. MOYEED, J.A. TAWN, (1998), *Model-based geostatistics*, "Applied Statistics", 47, pp. 299-350.

M.D. ECKER, A.E. GELFAND, (1997), *Bayesian variogram modeling for an isotropic spatial process*, "Journal of Agricultural, Biological and Environmental Statistics", 2, pp. 347-368.

M.D. ECKER, A.E. GELFAND, (1999), *Bayesian modeling and inference for geometrically anisotropic spatial data*, "Mathematical Geology", 31, pp. 67-83.

D.A. FOLLMAN, D. LAMBERT, (1989), *Generalising logistic regression by nonparametric mixing*, "Journal of the American Statistical Association", 84, pp. 295-300.

M. GAUDARD, M. KARSON, E. LINDER, D. SINHA, (1999), *Bayesian spatial prediction*, "Environmental and Ecological Statistics", 6, pp. 147-171.

A.E. GELFAND, S.K. GHOSH, (1998), *Model choice: a minimum posterior predictive loss approach*, "Biometrika", 85, pp. 1-11.

A.E. GELFAND, A. KOTTAS, (1999), *A Computational approach for full nonparametric Bayesian inference under Dirichlet process mixture model*, Technical Report 99-08, University of Connecticut, Department of Statistics.

A.E GELFAND, S. MUKHOPADHYAY, (1995), *On nonparametric Bayesian inference for the distribution of a random sample*, "The Canadian Journal of Statistics", 23, pp. 411-420.

A.E. GELFAND, N. RAVISHANKER, M.D. ECKER, (2000), *Modeling and inference for point-referenced binary spatial sata*, in D. DEY, S. GHOSH, B. MALLICK, (eds.), *Generalized Linear Models: A Bayesian Perspective*, Marcel Dekker, pp. 381-394.

A. GELMAN, J.B. CARLIN, H.S. STERN, D.B. RUBIN, (1995), *Bayesian Data Analysis*, Chapman and Hall.

J. GEWEKE, (1989), *Bayesian inference in econometric models using Monte Carlo integration*, "Econometrica", 57, pp. 1317-1339.

W.R. GILKS, S. RICHARDSON, D.J. SPIEGELHALTER, (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hall.

R. HAINING, (1990), *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press.

M.S. HANDCOCK, M.L. STEIN, (1993), *A Bayesian analysis of Kriging*, "Technometrics", 35, pp. 403-410.

D. HIGDON, (2001), *Space and space-time modelling using process convolutions*, Technical Report 01-03, Institute of Statistics and Decision Sciences, Duke University, USA.

P.K. KITANIDIS, (1986), *Parameter uncertainty in estimation of spatial functions: Bayesian analysis*, "Water Resources Research", 22, pp. 499-507.

P.K. KITANIDIS, (1997), *Introduction to Geostatistics - Applications in Hydrogeology*, Cambridge University Press.

K.P. KLEINMAN, J.C.IBRAHIM, (1998), *A semiparametric Bayesian approach to the random effects model*, "Biometrics", 54, pp. 921-938.

N.D. LE, J.V. ZIDEK, (1992), *Interpolation with uncertain spatial covariances: a Bayesian alternative to Kriging*, "Journal of Multivariate Analysis", 43, pp. 351-374.

R.C. LITTEL, G.A. MILLIKEN, W.W. STROUP, R.D. WOLFINGER, (1996), *SAS System for Mixed Models*, SAS Institute.

L.S. MAGDER, S.L. ZEGER, (1996), *A Smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians*, "Journal of the American Statistical Association", 91, pp. 1141-1152.

B.K. MALLICK, D.G.T. DENISON, A.F.M. SMITH, (2000), *Semiparametric Generalized Linear Models: Bayesian approaches*, in D. DEY, S. GHOSH, B. MALLICK, (eds.), *Generalized Linear Models: A Bayesian Perspective*, Marcel Dekker, pp. 217-230.

H. OMRE, (1987), *Bayesian Kriging - merging observations and qualified guesses in Kriging*, "Mathematical Geology", 19(1), pp. 25-39.

J.C. PINHEIRO, D.M. BATES, (2000), *Mixed-Effects Models in S and S-Plus*, Springer Verlag.

P.J. RIBEIRO, P.J. DIGGLE, (1999*a*), *Bayesian inference in Gaussian model-based Geostatistics*, Technical Report, ST-99-08, Department of Mathematics and Statistics, Lancaster University.

P.J. RIBEIRO, P.J. DIGGLE, (1999b), *geoR/geoS: functions for geostatistical analysis using R or S-plus*, Technical Report, ST-99-09, Department of Mathematics and Statistics, Lancaster University.

S. RICHARDSON, P.J. GREEN, (1997), *On Bayesian analysis of mixtures with an unknown number of components*, "Journal of the Royal Statistical Society B", 59, pp. 731-792.

B.D. RIPLEY, (1981), *Spatial Statistics*, John Wiley & Sons.

B.D. RIPLEY, (1988), *Statistical Inference for Spatial Processes*, Cambridge University Press.

G.K. ROBINSON, (1991), *That BLUP is a good thing: the estimation of random effects*, "Statistical Science", 6, pp. 15-51.

J.A. ROYLE, L.M. BERLINER, (2000), *A hierarchical approach to multivariate spatial modeling and prediction*, Technical Report, Ohio State University.

S.R. SEARLE, G. CASELLA, C.E. MC CULLOCH, (1992), *Variance Components*, John Wiley & Sons.

M.L. STEIN, (1999), *Interpolation of Spatial Data - Some Theory for Kriging*, Springer Verlag.

H. TAO, M. PALTA, B.S. YANDELL, M.A. NEWTON, (1999), *An estimation method for the semiparametric mixed effect model*, "Biometrics", 55, pp. 102-110.

A.V. VECCHIA, (1988), *Estimation and model identification for continuous spatial processes*, "Journal of the Royal Statistical Society B", 50, pp. 297-312.

G. VERBEKE, E. LESAFFRE, (1997), *The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data*, "Computational Statistics and Data Analysis", 23, pp. 541-556.

M. WEST, (1993), *Approximating posterior distributions by mixtures*, "Journal of the Royal Statistical Society B", 55, pp. 563-586.

RIASSUNTO

Kriging e modelli a effetti misti

In questo lavoro viene illustrata l'efficacia dell'uso dei modelli a effetti misti per la stima e la previsione riferite a fenomeni spaziali continui negli ambiti classico e bayesiano. Le metodologie esposte vengono successivamente applicate ad un caso di studio riferito a dati agricoli.

SUMMARY

Kriging with mixed effects models

In this paper the effectiveness of the use of mixed effects models for estimation and prediction purposes in spatial statistics for continuous data is reviewed in the classical and Bayesian frameworks. A case study on agricultural data is also provided.