Anisotropic Matérn correlation and spatial prediction using REML

Article *in* Journal of Agricultural Biological and Environmental Statistics · June 2007

DOI: 10.1198/108571107X196004

3 authors, including:

Kathryn Haskard
University of Adelaide
17 PUBLICATIONS **428** CITATIONS
SEE PROFILE

Arunas Petras Verbyla
97 PUBLICATIONS **3,973** CITATIONS
SEE PROFILE

Anisotropic Matérn Correlation and Spatial Prediction Using REML

Kathryn A. HASKARD, Brian R. CULLIS, and Arūnas P. VERBYLA

The Matérn correlation function provides great flexibility for modeling spatially correlated random processes in two dimensions, in particular via a smoothness parameter, whose estimation allows data to determine the degree of smoothness of a spatial process. The extension to include anisotropy provides a very general and flexible class of spatial covariance functions that can be used in a model-based approach to geostatistics, in which parameter estimation is achieved via REML and prediction is within the E-BLUP framework. In this article we develop a general class of linear mixed models using an anisotropic Matérn class with an extended metric. The approach is illustrated by application to soil salinity data in a rice-growing field in Australia, and to fine-scale soil pH data. It is found that anisotropy is an important aspect of both datasets, emphasizing the value of a straightforward and accessible approach to modeling anisotropy.

Key Words: Geometric anisotropy; Kriging; Model-based geostatistics; Residual maximum likelihood; Spatial correlation.

1. INTRODUCTION

Rapid and cost-effective measurement of soil salinity via apparent electrical conductivity (ECa) of soil profiles is becoming an important management tool for determining the suitability of soils for growing rice in parts of New South Wales, Australia. The current protocol involves measurements of ECa from a ground-based electromagnetic induction instrument, EM31, which is linked to a differential global positioning system and towed behind a four-wheel motorbike, providing a large number of geographically referenced observations. From one rice field, 2000 observations were gathered in a serpentine fashion throughout an irregularly shaped field, as displayed in Figure 1. There were 1995 distinct locations, with five locations having two observations. The aim was to produce a fine-scale map to determine where ECa is at least 150 mS/m (milli-Siemens per meter), fulfilling one requirement for suitability for growing rice (Beecher, Hume, and Dunn 2002).

Kathryn A. Haskard was Senior Biometrician, BiometricsSA, SARDI, GPO Box 397, Adelaide, SA 5001, Australia. (Current address: Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom (E-mail: *kathy.haskard@bbsrc.ac.uk*)). Brian R. Cullis is Principal Research Scientist, New South Wales Department of Primary Industries, Wagga Wagga, NSW 2650, Australia. Arūnas P. Verbyla is Professor, School of Agriculture, Food and Wine, The University of Adelaide, PMB 1, Glen Osmond, SA 5064 Australia.

Journal of Agricultural, Biological, and Environmental Statistics, Volume 12, Number 2, Pages 147–160
DOI: 10.1198/108571107X196004

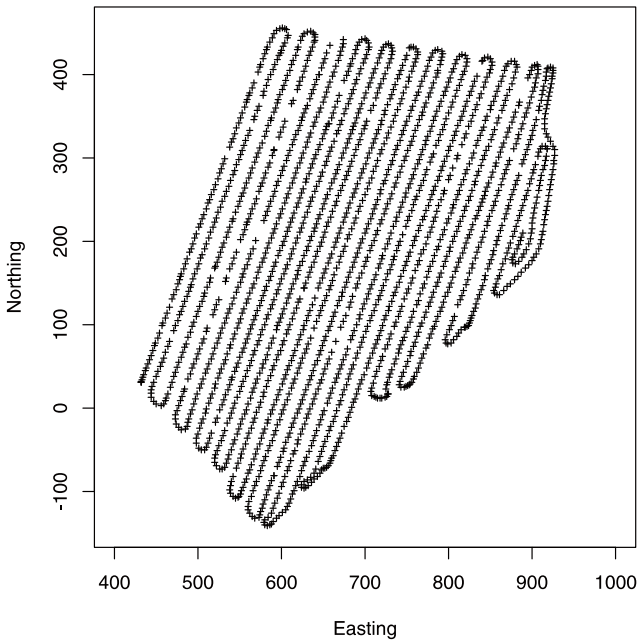

Figure 1. Rice bay ECa sampling pattern.

A second set of spatial data arose as part of a larger study conducted at Wagga Wagga by Dr. Mark Conyers of the New South Wales Department of Primary Industries. Soil pH was determined for five replicates of 100 one cm^3 cubes in a 10 cm $\times$ 10 cm square at the soil surface in a field that had been cropped and grazed for several years. The five blocks were individually oriented for sampling convenience, with orientation not recorded. The aim of this study was to identify the covariance structure of pH on a fine spatial scale.

To perform fine scale mapping, a common approach is to use geostatistical methods based on kriging using an estimated semivariogram or covariance structure (see, for example, Webster and Oliver 2001). Thus, both examples require estimation of the spatial covariance structure. In this article the focus is on a model-based approach rather than the classical approach. In model-based geostatistics, a spatial process is described by a formal statistical model, often a Gaussian linear mixed model. See, for example, Diggle, Ribeiro, and Christensen (2003). Residual maximum likelihood (REML) in the linear mixed models framework has been recognized for some years as an objective approach to providing a semivariogram estimate for kriging in geostatistics. Kitanidis (1983) was the first to apply REML to the estimation of spatial covariance parameters, while Stein (1999) suggested that REML and best linear unbiased prediction (BLUP) form a coherent conceptual package for the estimation and prediction for these models, including extensions to where the spatially correlated random component is an intrinsic random function (IRF). This coupled with best linear unbiased prediction provides kriging estimates and kriging variance for spatial processes.

As with the classical geostatistics approach of estimating a semivariogram based on

an empirical semivariogram, the procedure is limited by the class of models available for consideration. The Matérn class has been promoted by various authors, including Stein (1999), as a natural and wide class of spatial correlation functions. Its mathematical form has made it difficult to implement, however, and it has not been readily available in standard statistical software.

Isotropic covariance models are often used in practice. However, spatial processes may be anisotropic. To date, the capability to fit general anisotropic processes has not been readily available. Combining geometric anisotropy with the Matérn correlation function provides a powerful class of covariance models. A preliminary implementation of software to estimate the anisotropic Matérn parameters using REML has now been made in the computer package ASReml (Gilmour, Gogel, Cullis, and Thompson 2005), and its application is investigated in this article by analysis of the two datasets, and by Haskard (2007) with simulations.

Sections 2 and 3 describe the formulation of the Gaussian geostatistical model as a linear mixed model, and kriging as prediction, with particular reference to the general anisotropic Matérn model. Estimation of the variance parameters using REML is discussed in Section 4. The two motivating datasets are analyzed in Sections 5 and 6, and the article concludes with a short discussion and summary.

2. GEOSTATISTICAL MIXED MODELS

Our development of the model follows the aims of analysis of the datasets. Geostatistics is usually concerned with the problem of producing a map or interpolation of a quantity of interest over a particular area (in $\mathbb{R}^2$). We assume that we have observed data at a set of n locations (of which b, possibly less than n, are distinct), with the ith observation y_i taken at the location identified by a vector $\mathbf{s}_i$, $i = 1 \ldots n$, and that a model for y_i is

$$y_i = f(\mathbf{s}_i) + e_i, \tag{2.1}$$

where $f(\mathbf{s}_i)$ is some function of the spatial location $\mathbf{s}_i$ and the e_i are mutually independent $\mathrm{N}(0, \sigma^2)$ random variables. If $\mathbf{s}$ represents the set of b distinct observed locations, and $\mathbf{f}(\mathbf{s}) = (f(\mathbf{s}_1), \ldots, f(\mathbf{s}_n))^T$, then we assume

$$\mathbf{f}(\mathbf{s}) = \mathbf{X}_s \boldsymbol{\tau}_s + \mathbf{Z}_s \mathbf{u}_s(\mathbf{s}), \tag{2.2}$$

where $\mathbf{X}_s^{n\times p}$ is a matrix of polynomials in $\mathbf{s}$, often of degree 1, $\boldsymbol{\tau}_s^{p\times 1}$ is a vector of polynomial regression coefficients, $\mathbf{Z}_s^{n\times b}$ is an indicator matrix for random effects at distinct locations, accommodating duplicated locations (typically $\mathbf{Z}_s = \mathbf{I}_n$ if all the locations are distinct, otherwise $b < n$), and $\mathbf{u}_s(\mathbf{s})$ is a realization of a stationary Gaussian process distributed independently of $\mathbf{e} = (e_1, \ldots, e_n)^T$, with zero mean and variance matrix $\gamma_s \sigma^2 \mathbf{G}_s$, so that γ_s is the ratio of the variance of the spatially correlated process to the nugget variance. The elements of $\mathbf{G}_s$ are given by $\rho(\mathbf{s}_i - \mathbf{s}_j, \boldsymbol{\phi})$, $\rho(\cdot)$ being a correlation function with a parameter vector $\boldsymbol{\phi}$ and dependent on the spatial separation vector $\mathbf{h}_{ij} = \mathbf{s}_i - \mathbf{s}_j$. The matrix $\mathbf{G}_s$ is assumed positive definite. Subscripts s indicate elements specifically related to spatial effects.

Combining Equations (2.1) and (2.2) we have, in matrix notation,

$$\mathbf{y}(\mathbf{s}) = \mathbf{X}_s\boldsymbol{\tau}_s + \mathbf{Z}_s\mathbf{u}_s(\mathbf{s}) + \mathbf{e}, \tag{2.3}$$

and

$$\mathbf{y}(\mathbf{s}) \sim \mathrm{N}\left(\mathbf{X}_s\boldsymbol{\tau}_s, \sigma^2(\gamma_s\mathbf{Z}_s\mathbf{G}_s\mathbf{Z}_s^T + \mathbf{I}_n)\right), \tag{2.4}$$

which is a Gaussian linear mixed model with spatially correlated random effects $\mathbf{u}_s$ and identically and independently distributed residual errors $\mathbf{e}$. In geostatistical terms, the identity matrix in (2.4) models a nugget effect.

Alternatively, provided the n locations are distinct ($b = n$) and $\mathbf{Z}_s = \mathbf{I}_n$, the roles of $\mathbf{e}$ and $\mathbf{u}_s$ can be switched, formulating the model with spatially correlated residual errors $\mathbf{u}_s \sim N(\mathbf{0}, \sigma_s^2\mathbf{G}_s)$, with the nugget effect modeled as an independent random $n \times 1$ effect $\mathbf{e} \sim N(\mathbf{0}, \sigma_s^2\gamma\mathbf{I}_n)$, with design matrix $\mathbf{Z} = \mathbf{I}_n$,

$$\mathbf{y}(\mathbf{s}) = \mathbf{X}_s\boldsymbol{\tau}_s + \mathbf{e} + \mathbf{u}_s, \tag{2.5}$$

and

$$\mathbf{y}(\mathbf{s}) \sim \mathrm{N}(\mathbf{X}_s\boldsymbol{\tau}_s, \sigma_s^2(\gamma\mathbf{I}_n + \mathbf{G}_s)). \tag{2.6}$$

Here $\sigma_s^2 = \sigma^2\gamma_s$, and $\gamma = 1/\gamma_s$ is the ratio of nugget variance to variance of the spatially correlated process; the nugget effect can be excluded by setting $\gamma = 0$, equivalent to dropping $\mathbf{e}$ from (2.5). The first form (2.3) and (2.4) is necessary when there are locations with multiple observations, and the second form (2.5) and (2.6) is required if it desired to fit a model with no nugget effect. This is because a residual error term must always be present in the variance ratio formulation we are using (involving a residual error variance which must be positive, and a variance ratio γ_s or γ), in contrast to the variance components formulation (using only σ^2 and σ_s^2, of which either may be zero).

Many forms for the spatial correlation function $\rho(\cdot)$, which determines the elements of $\mathbf{G}_s$, have been suggested in the geostatistical literature. Following the recommendations of Stein (1999) we base our correlation model on the Matérn family of correlation functions and, further, incorporate geometric anisotropy and a choice of distance metrics, by defining

$$\rho(\mathbf{h}; \boldsymbol{\phi}) = \rho_M(d(\mathbf{h}; \delta, \alpha, \lambda); \phi, \nu),$$

where $\mathbf{h} = (x, y)^T$ is the spatial separation vector between two locations, and $\rho_M(\cdot)$ is the Matérn family of (isotropic) correlation functions for a specified metric $d(\cdot)$. Our metric is the Minkowski metric, with parameter $\lambda > 0$, applied to linearly transformed spatial separation vectors. The transformation corresponds to rotating the original coordinate axes through α radians then stretching or shrinking them relative to each other by a factor $\delta^{2/\lambda}$, achieved by multiplying the rotated x-coordinates by $\delta^{1/\lambda}$ and dividing the rotated y-coordinates by $\delta^{1/\lambda}$. This leads to the (non-Euclidean) metric

$$d(\mathbf{h}; \delta, \alpha, \lambda) = \left(\delta|x'|^\lambda + \frac{1}{\delta}|y'|^\lambda\right)^{1/\lambda}$$

for

$$\mathbf{h}' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{T}\mathbf{h}.$$

When $\lambda = 2$, we see that

$$d^2 = \mathbf{h}^T \mathbf{T}^T \mathbf{S}^2 \mathbf{T} \mathbf{h} = \mathbf{h}''^T \mathbf{h}'',$$

where

$$\mathbf{S}^2 = \begin{bmatrix} \delta & 0 \\ 0 & 1/\delta \end{bmatrix}$$

and $\mathbf{h}'' = \mathbf{STh}$, so that the metric corresponds to the usual Euclidean distance applied to spatial separation $\mathbf{h}''$ in the rotated and stretched/shrunken axis coordinates. With appropriate choice of α and δ, elliptical contours of constant correlation on the original axes are transformed into circles. Typically, $\lambda = 2$ will be used.

When $\lambda = 1$, our metric corresponds to the city-block metric on the transformed axes, widely used in the analysis of field trials (with $\nu = \frac{1}{2}$ in the Matérn correlation function, giving exponential correlation, and stretching/shrinking but no rotation of axes; see Gilmour, Cullis, and Verbyla 1997).

We note that there is nonuniqueness in this metric $d(\cdot)$, since inverting δ and adding $\frac{\pi}{2}$ to α gives the same distance. This nonuniqueness can be removed by constraining $0 \leqslant \alpha < \frac{\pi}{2}$ and $\delta > 0$, or by constraining $0 \leqslant \alpha < \pi$ and either $0 < \delta \leqslant 1$ or $\delta \geqslant 1$. For $\lambda = 2$, isotropy corresponds to $\delta = 1$, and then the rotation angle α is irrelevant: correlation contours are circles, compared with ellipses in general. For $\lambda = 1$, correlation contours are diamonds, with squares if $\delta = 1$. True isotropy is not possible with λ other than 2.

The isotropic Matérn correlation function is given by

$$\rho_M(d; \phi, \nu) = \left\{2^{\nu-1}\Gamma(\nu)\right\}^{-1} \left(\frac{d}{\phi}\right)^{\nu} K_\nu\left(\frac{d}{\phi}\right), \tag{2.7}$$

where $\phi > 0$ is a range parameter, $\nu > 0$ is a smoothness parameter, $\Gamma(\cdot)$ is the gamma function, and $K_\nu(.)$ is the modified Bessel function of the third kind of order ν (Abramowitz and Stegun 1965, sec. 9.6). For a given ν, the range parameter ϕ affects the rate of decay of $\rho(\cdot)$ with increasing d. The parameter $\nu > 0$ controls the analytic smoothness of the underlying process $\mathbf{u}_s$, the process being $\lceil \nu \rceil - 1$ times mean-square differentiable, where $\lceil \nu \rceil$ is the smallest integer greater than or equal to ν (Stein 1999, p. 31). Larger ν correspond to smoother processes.

We note that $\nu = \frac{1}{2}$ yields the exponential correlation function, $\rho_M(d; \phi, \frac{1}{2}) = \exp(-d/\phi)$, while $\nu = 1$ yields Whittle's elementary correlation function, $\rho_M(d; \phi, 1) = (d/\phi)K_1(d/\phi)$ (Webster and Oliver 2001, p. 119). When $\nu = m + \frac{1}{2}$ for any positive integer m, $\rho_M(\cdot)$ is of the form $\exp(-d/\phi)$ times a polynomial in d of degree m. In the limit as ν approaches infinity, with $\psi = 2\nu^{1/2}\phi$ remaining constant, the Matérn correlation approaches the Gaussian correlation function, $\exp(-d^2/\psi^2)$.

This model can be extended to intrinsic random functions (IRFs) of Matheron (1973), thereby accommodating a range of nonstationarity, but we do not consider this here.

The anisotropic Matérn correlation model corresponds to Matérn correlation (2.7) with the same smoothness parameter ν in all directions, but with the range parameter varying between $\phi/\sqrt{\delta}$ in direction α and $\phi\sqrt{\delta}$ in direction $\alpha \pm \frac{\pi}{2}$.

3. PREDICTION (KRIGING)

Recalling that one aim of our analysis is to produce a map, we seek to predict $f(\cdot)$ at a new location $\mathbf{s}_0 \in \mathbb{R}^2$, say, namely $f(\mathbf{s}_0) = \mathbf{x}_0^T\boldsymbol{\tau} + u_s(\mathbf{s}_0)$, where $\mathbf{x}_0$ is the vector of polynomial functions of $\mathbf{s}_0$, for many different $\mathbf{s}_0$. It can be shown (Stein 1999) that for known $(\gamma_s, \boldsymbol{\theta})$, where here $\boldsymbol{\theta} = (\phi, \nu, \delta, \alpha)^T$, the best linear unbiased predictor (BLUP) of $f(\mathbf{s}_0)$ is

$$\tilde{f}(\mathbf{s}_0) = \mathbf{x}_0^T\hat{\boldsymbol{\tau}}_s + \mathbf{g}_0^T\mathbf{G}_s^{-1}\tilde{\mathbf{u}}_s, \tag{3.1}$$

where $\mathbf{g}_0 = \text{cor}\,(\mathbf{u}_s,\ u_s(\mathbf{s}_0))$, and $\hat{\boldsymbol{\tau}}_s$ and $\tilde{\mathbf{u}}_s$ are solutions to the mixed model equations,

$$\begin{bmatrix} \mathbf{X}_s^T\mathbf{X}_s & \mathbf{X}_s^T\mathbf{Z}_s \\ \mathbf{Z}_s^T\mathbf{X}_s & \mathbf{Z}_s^T\mathbf{Z}_s + \frac{1}{\gamma_s}\mathbf{G}_s^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\tau}}_s \\ \tilde{\mathbf{u}}_s \end{bmatrix} = \begin{bmatrix} \mathbf{X}_s^T\mathbf{y} \\ \mathbf{Z}_s^T\mathbf{y} \end{bmatrix}. \tag{3.2}$$

Equation (3.2) can be written as

$$\mathbf{C}\tilde{\boldsymbol{\beta}} = \mathbf{W}^T\mathbf{y},$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}_s^T\mathbf{X}_s & \mathbf{X}_s^T\mathbf{Z}_s \\ \mathbf{Z}_s^T\mathbf{X}_s & \mathbf{Z}_s^T\mathbf{Z}_s + \frac{1}{\gamma_s}\mathbf{G}_s^{-1} \end{bmatrix},$$

$\tilde{\boldsymbol{\beta}} = [\hat{\boldsymbol{\tau}}_s^T \quad \tilde{\mathbf{u}}_s^T]^T$, and $\mathbf{W} = [\mathbf{X}\ \mathbf{Z}]$. Writing

$$\tilde{f}(\mathbf{s}_0) - f(\mathbf{s}_0) = \mathbf{w}_0^T(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) + \mathbf{g}_0^T\mathbf{G}_s^{-1}\,\mathbf{u}_s - u_s(\mathbf{s}_0),$$

where $\mathbf{w}_0^T = \begin{bmatrix}\mathbf{x}_0^T & \mathbf{g}_0^T\mathbf{G}_s^{-1}\end{bmatrix}$ and $\boldsymbol{\beta} = [\boldsymbol{\tau}_s^T \quad \mathbf{u}_s^T]^T$, and noting that

$$\text{var}\begin{pmatrix} \tilde{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \mathbf{u}_s \\ u_s(\mathbf{s}_0) \end{pmatrix} = \sigma^2 \begin{bmatrix} \mathbf{C}^{-1} & -\mathbf{C}^{WZ} & -\mathbf{C}^{WZ}\mathbf{G}_s^{-1}\mathbf{g}_0 \\ -\mathbf{C}^{ZW} & \gamma_s\mathbf{G}_s & \gamma_s\mathbf{g}_0 \\ -\mathbf{g}_0^T\mathbf{G}_s^{-1}\mathbf{C}^{ZW} & \gamma_s\mathbf{g}_0^T & \gamma_s \end{bmatrix},$$

where $\mathbf{C}^{-1}$ is partitioned as

$$\mathbf{C}^{-1} = \begin{bmatrix} \mathbf{C}^{XX}_{(p\times p)} & \mathbf{C}^{XZ}_{(p\times b)} \\ \mathbf{C}^{ZX}_{(b\times p)} & \mathbf{C}^{ZZ}_{(b\times b)} \end{bmatrix},$$

and $\mathbf{C}^{ZW} = [\mathbf{C}^{ZX} \quad \mathbf{C}^{ZZ}] = (\mathbf{C}^{WZ})^T$, the prediction error variance (PEV, or mean square error of prediction, MSEP) is

$$\text{var}\left(\tilde{f}(\mathbf{s}_0) - f(\mathbf{s}_0)\right) = \sigma^2\left(\mathbf{w}_0^T\mathbf{C}^{-1}\mathbf{w}_0 - \gamma_s\mathbf{g}_0^T\mathbf{G}_s^{-1}\mathbf{g}_0 + \gamma_s\right). \tag{3.3}$$

Gilmour et al. (2004) presented a general approach for efficiently computing (3.1) and (3.3).

We note that best linear unbiased prediction is called kriging in the geostatistical literature (see Stein 1999, p. 8 for example). If the mean $(\mathbf{X}_s\boldsymbol{\tau}_s)$ is simply an unknown constant, then this BLUP is known as ordinary kriging. BLUP for our model is known as universal kriging, while BLUP for a known mean is known as simple kriging.

4. ESTIMATION

The practical interpretation of (3.1) as well as the direct aim of describing the data via the model (2.3), as in the soil pH example, both require estimates of the variance parameters $\boldsymbol{\rho} = (\sigma^2, \gamma_s, \boldsymbol{\theta}^T)^T$ for fixed $\lambda = 1$ or 2. Since Equation (2.3) is a linear mixed model under the Gaussian assumption (2.4), residual maximum likelihood (REML) provides the natural approach for estimation of $\boldsymbol{\rho}$. The residual log-likelihood is given by

$$\ell_R(\boldsymbol{\rho}; \mathbf{y}) = -\tfrac{1}{2}\left\{\log|\mathbf{H}| + \log|\mathbf{X}_s^T\mathbf{H}^{-1}\mathbf{X}_s| + (n-p)\log\sigma^2 + \frac{1}{\sigma^2}\mathbf{y}^T\mathbf{P}\mathbf{y}\right\}, \qquad (4.1)$$

where $\mathbf{H} = \gamma_s\mathbf{Z}_s\mathbf{G}_s\mathbf{Z}_s^T + \mathbf{I}$ and $\mathbf{P} = \mathbf{I} - \mathbf{W}\mathbf{C}^{-1}\mathbf{W}^T$ (Verbyla 1990).

Maximization of (4.1) usually requires an iterative approach which we briefly outline in the next section.

4.1 COMPUTATIONAL DETAILS AND MODEL FITTING

Haskard (2007) presents details for the anisotropic Matérn class of an implementation of the average information (AI) algorithm due to Gilmour, Thompson, and Cullis (1995). This requires evaluation of the derivatives of the residual log-likelihood with respect to the components of $\boldsymbol{\rho}$. These are relatively straightforward to compute analytically for ϕ, δ, and α, though more complex for ν. In practice, to avoid occasional numerical problems, we recommend the use of numerical methods to obtain $\partial\mathbf{G}_s/\partial\nu$ (see Haskard 2007), though this is often avoided through our modeling strategy.

4.2 MODEL BUILDING, INFERENCE, AND DIAGNOSTICS

An advantage of using the mixed model framework is that tests of hypotheses for both fixed effects and variance models are widely available. For fixed effects, Wald tests or the improved Wald-type tests proposed by Kenward and Roger (1997) can be conducted, while tests of hypotheses concerning variance models are available within the likelihood ratio paradigm.

REML likelihood ratio tests can be used to test an hypothesis H_0 nested within hypothesis H_1, where H_1 contains an additional k parameters. The residual likelihood ratio test statistic is given by

$$D = -2\left\{\ell_R(\hat{\boldsymbol{\rho}}_0; \mathbf{y}) - \ell_R(\hat{\boldsymbol{\rho}}_1; \mathbf{y})\right\}$$

where $\hat{\boldsymbol{\rho}}_0$ and $\hat{\boldsymbol{\rho}}_1$ are the REML estimates of $\boldsymbol{\rho}$ for the models under H_0 and H_1, respectively.

Under the usual regularity conditions, the statistic D is asymptotically a chi-square variable with k degrees of freedom. However, the distribution theory for D is complicated when a parameter is on the boundary of the parameter space under H_0 (see Stram and Lee 1994, for further details), for example in a test of the need for the inclusion of a nugget effect. See also Haskard and Verbyla (2007) on a test for (geometric) anisotropy which appears nonstandard but is not.

In deciding on appropriate models we are generally guided by contour plots and empirical semivariograms of both raw data and residuals adjusted for global trend (from an

ordinary least squares model). We emphasize however that, as noted by Stein (1999, p. 176), "using empirical semivariograms for model selection can work disastrously for smooth processes," and so we use these diagnostics for guidance but not as definitive. With a suitably rich class of correlation models available in the one framework, such as that considered here, reliance on such diagnostics is reduced. This richness includes anisotropy. It makes good sense to fit the anisotropic model even if anisotropy is not suspected, as the isotropic model is a special case.

Full REML estimation of ν is possible but sometimes difficult, and it may be sufficient to choose a value of ν by profiling the REML log-likelihood from a range of values, say 0.5, 1.0, 1.5, and 2. A starting value for ν can be chosen from the profile likelihood approach or based on previous experience. Fitting the simpler isotropic model can provide a good starting value for ϕ for the anisotropic model. Alternatively, since Matérn correlation functions all have correlation approximately 0.14 at a separation distance of $2\phi\sqrt{2\nu}$, at least for ν larger than about 0.5, a starting value for ϕ is close to the spatial separation distance at which the semivariogram (or a rough "average" of directional semivariograms) reaches 86% of its maximum, divided by $2\sqrt{2\nu}$. Starting values for the geometric anisotropy parameters can be obtained from empirical directional semivariograms: for δ, the ratio (> 1) of the separation distances at which the slowest-rising and fastest-rising directional semivariograms reach some chosen height (such as half their maximum), when $\lambda = 2$, and the square root of this ratio when $\lambda = 1$; for α, the angle corresponding to the fastest-rising semivariogram, or the angle corresponding to the slowest-rising semivariogram $\pm\frac{\pi}{2}$.

The two example datasets were analyzed using ASReml (Gilmour, Gogel, Cullis, and Thompson 2005). The Minkowski metric parameter λ was set to its typical value of 2 for both examples.

5. ANALYSIS OF THE ELECTROMAGNETIC SALINITY DATA

Scatterplots of ECa against northing and easting suggested possible nonstationarity which may be accounted for by inclusion of second degree polynomial effects in northing and easting. Figure 2 presents empirical semivariograms in four directions for the raw data and for residuals after an ordinary least squares (OLS) fit of a quadratic surface in northing and easting, which provide added support of the need to allow for a global trend. However, there remains evidence of anisotropy which we investigate more formally in the following.

The anisotropic Matérn model formulation (2.3) and (2.4) was used, to accommodate the five duplicated locations. Full REML analysis gave a maximized residual log-likelihood (ignoring constants) of -4851.13, with parameter estimates (and standard errors, although these are of dubious value for variance-model parameters) $\hat{\phi} = 40.98$ (SE 8.36), $\hat{\nu} = 0.770$ (0.060), $\hat{\delta} = 2.627$ (0.171), $\hat{\alpha} = -0.4100$ (0.0366), $\hat{\sigma}_s^2 = 614.2$ (123.8), and $\hat{\sigma}^2 = 6.555$ (1.182). Full REML analysis assuming isotropy gave a maximized residual log-likelihood (ignoring constants) of -5042.69, with parameter estimates $\hat{\phi} = 18.60$ (SE 2.09), $\hat{\nu} = 1.122$ (0.099), $\hat{\sigma}_s^2 = 415.7$ (44.6), and $\hat{\sigma}^2 = 6.563$ (1.108). Twice the difference in log-likelihood is 383.12, which is highly significant ($p \ll 0.001$) when compared with a

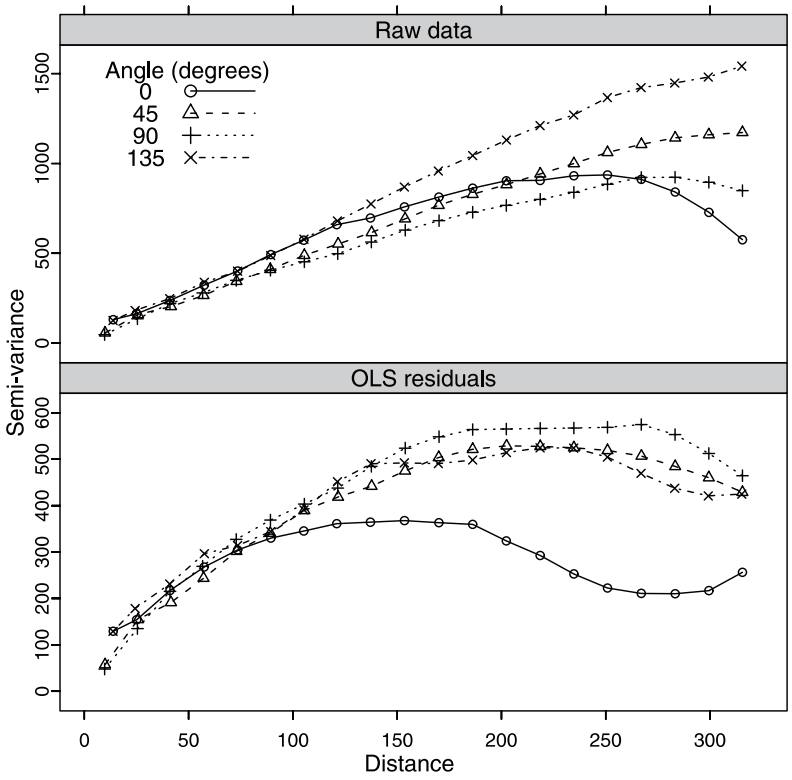

Figure 2. ECa salinity data empirical directional semivariograms for raw data and OLS residuals from a quadratic surface.

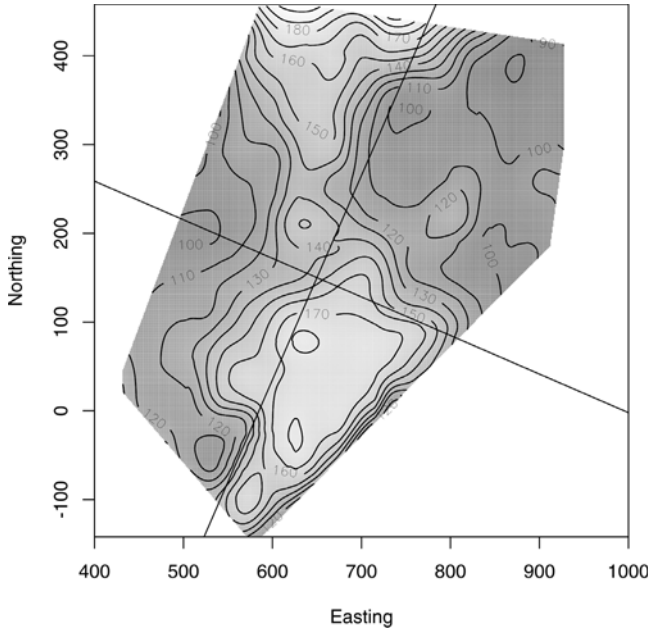

Figure 3. ECa salinity data predictions and rotated axes.

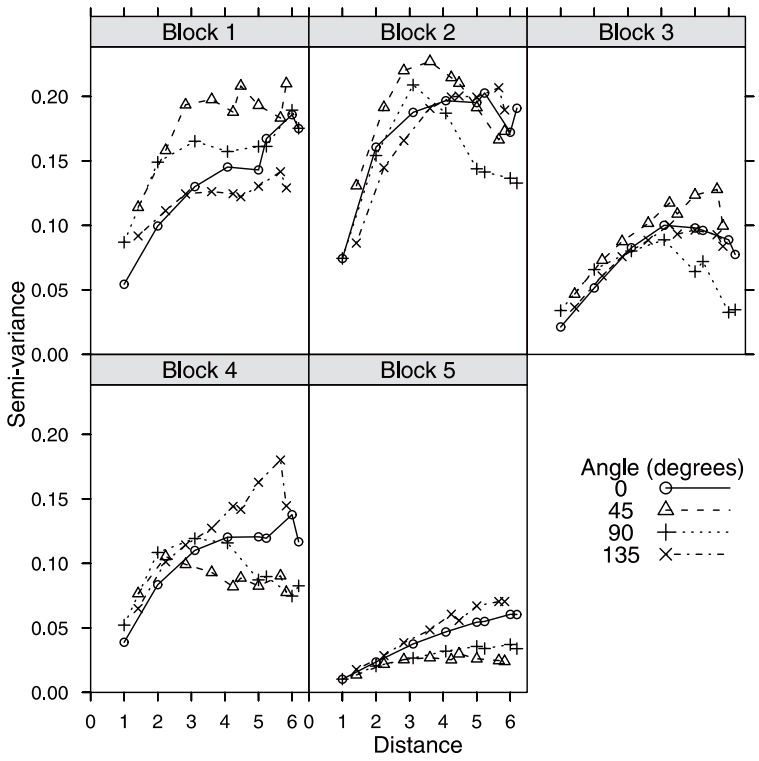

Figure 4. Soil pH residuals from OLS fit of linear trend surface for each block, empirical directional semivariograms on the same scale.

χ^2_2 distribution to test for geometric anisotropy, indicating that anisotropy is important in the covariance structure for this dataset.

The fitted quadratic surfaces for the two models (standard errors in parentheses) are

$$\begin{aligned}
\text{anisotropic: } & -0.000660\,x^2 + 0.000282\,y^2 - 0.000242\,xy + 0.900\,x + 0.113\,y - 170 \\
& \quad (0.000354) \quad (0.000195) \quad (0.000376) \quad (0.440) \quad (0.213) \quad (136) \\
\text{isotropic: } & -0.000763\,x^2 + 0.000223\,y^2 - 0.000163\,xy + 1.044\,x + 0.029\,y - 211 \\
& \quad (0.000193) \quad (0.000129) \quad (0.000221) \quad (0.241) \quad (0.128) \quad (76)
\end{aligned}$$

Figure 3 is a map of predicted values in the region observed. Overlaid on this map are the estimated rotated axes. Correlation remains highest (the range parameter is largest) along the nearer-to-vertical axis, and drops most quickly along the more horizontal axis ($-23.5°$ to the horizontal). Two regions above 150 mS/m can be identified (lighter shading), aligned north-south.

6. ANALYSIS OF THE FINE-SCALE SOIL pH DATA

Empirical directional semivariograms were inspected for both the raw fine-scale soil pH data and the ordinary least squares (OLS) residuals after fitting a linear global trend surface separately to each block, to allow for broad trends. The latter are displayed in Figure 4. The

Table 1. Models fitted to soil pH data, with parameters either separate for the five soil blocks or common to all, maximized residual log-likelihoods, and some likelihood ratio tests comparing models. Models 1 and 2 are isotropic; all others are anisotropic. The parameters are: spatial variance σ_s^2, Matérn parameters ϕ and ν, anisotropy ratio δ, anisotropy angle α, and nugget variance σ^2.

Model	Separate	Common	logL	c.f. Model	LR test p
1	σ_s^2	$\phi, \nu, \sigma^2,\ \delta = 1$	497.058	4	0.006
2	ϕ	$\sigma_s^2, \nu, \sigma^2,\ \delta = 1$	499.237	6	0.005
3	α	$\sigma_s^2, \phi, \nu, \delta, \sigma^2$	< 450*	6	< 0.001
4	α, σ_s^2	$\phi, \nu, \delta, \sigma^2$	506.103	5	0.129
5	α, σ_s^2, ϕ	ν, δ, σ^2	509.670	6	0.664
6	$\boldsymbol{\alpha, \phi}$	$\boldsymbol{\sigma_s^2, \nu, \delta, \sigma^2}$	**508.473**		**preferred**
7	α, ϕ, ν	$\sigma_s^2, \delta, \sigma^2$	510.814	6	0.332
8	α, ϕ, δ	$\sigma_s^2, \nu, \sigma^2$	509.920	6	0.576

* convergence not achieved

semivariograms suggest anisotropy may be present. Using formulation (2.5) and (2.6), the nugget effect variance was found to be very small ($< 1\%$ of the spatial variance) and in some models went to the zero bound, which is not unrealistic since pH was determined for complete adjoining 1 cm^3 cubes, effecting a complete survey of the five small blocks.

It is reasonable to expect the five blocks to have similar spatial covariance structure, as it can be presumed that the same random process generated the pH for these five blocks from the same field. Since the orientations of the five blocks were different and unknown, the anisotropy angle was estimated separately for each block. Table 1 displays several of the fitted models. It was found that common ν and δ were acceptable but it was not possible to force both σ_s^2 and ϕ to be common across all five blocks. Model 6 with common σ_s^2 but separate ϕ was chosen over Model 4 with separate σ_s^2 and common ϕ (with the nugget variance estimated at the zero boundary), based on the Akaike information criterion (AIC, Akaike 1973). A block effect and linear terms in location coordinates x and y were fitted separately for each block in all models. Table 2 displays variance parameter estimates for the final model and the corresponding isotropic model, with standard errors, although these must be interpreted cautiously for variance-model parameters, whose estimates are likely to be far from normally distributed.

An extension of the χ_2^2 test of isotropy (involving an anisotropy ratio and one angle) to the situation with five separate anisotropy angles yields an asymptotic χ_6^2 likelihood-ratio test statistic with $p = 0.005$, comparing Models 2 and 6. Models 1 and 4 with separate σ_s^2 and common ϕ also yield a statistically significant test of isotropy ($p = 0.006$). Like the ECa data, this example illustrates that anisotropy is present in this soil variable, and to ignore it, as is most often done using commonly available models, is to use an inferior and less appropriate model.

7. DISCUSSION AND CONCLUSIONS

Analyses of the two example datasets illustrate that simultaneous estimation via REML of all four Matérn and geometric anisotropy parameters, together with a nugget variance

Table 2. Parameter estimates and their estimated standard errors in the preferred model (anisotropic) and the corresponding isotropic model for the soil pH data, models 6 and 2, respectively, in Table 1.

Parameter	Anisotropic estimates	(SEs)	Isotropic estimates	(SEs)
σ_s^2	0.1805	(0.0399)	0.1739	(0.0350)
ϕ	0.982, 1.040, 1.975, 1.461, 3.705	(0.314, 0.334, 0.728, 0.495, 1.414)	0.883, 0.917, 1.659, 1.300, 3.314	(0.262,0.275, 0.571,0.411, 1.213)
ν	1.078	(0.270)	1.183	(0.298)
δ	1.360	(0.102)		
α	1.574, 0.854, 1.399, 1.277, 2.218	(0.397,0.259, 0.276,0.257, 0.252)		
σ^2	0.001370	(0.001902)	0.001957	(0.001633)

ratio if appropriate, is feasible, although it can be time-consuming with the large dataset. The time taken per iteration on a 750 MHz Pentium III with 256 Mb of random access memory for the analysis of the ECa example (with 1995 locations) can range between 30 minutes for isotropy with ν fixed and 100 minutes when estimating all parameters, including ν and anisotropy parameters. The major source of computational load is the lack of sparsity in $\mathbf{G}_s^{-1}$. The burden in computing the likelihood, score and average information matrix for large irregularly spaced datasets is a potential obstacle to routine use of REML. However approximate methods are available, involving conditional or simultaneous models.

For example, building on the work of Vecchia (1998), Stein, Chi, and Welty (2004) described a conditional approach to efficiently obtain REML estimates in large spatial datasets. Essentially their approach considers the residual likelihood as a product of a marginal density for a small subset of the data and conditional densities conditioned only on "previous" observations in some ordering, then judiciously pruning the conditioning sets. This is equivalent to zeroing many elements of $\mathbf{G}_s^{-1}$, thus increasing sparsity. This approach has much promise and we plan to implement it in the future.

Anisotropy was found to be important in both examples. Previously the capability has not been available to easily model anisotropy in general cases, and common practice has been to ignore anisotropy, which can lead to poor predictions and inappropriate estimates of prediction variance if anisotropy is present. For example, prediction standard errors derived from the better-fitting anisotropic model for the ECa data are usually a little smaller, but sometimes larger, than those derived by assuming the inferior isotropic model.

The use of so-called "plug-in" estimates of prediction error variance has been criticized in the Bayesian literature—for example, by Diggle, Ribeiro, and Christensen (2003)—for underestimating prediction error variance. However, simulations in Haskard (2007) illustrate good agreement between the model-based and empirical mean square errors of prediction when the true model was fitted. With a poor-fitting model the prediction error variances can be far from correct. The ability to choose a good-fitting model is far more critical than the approximation involved in plugging in estimates of variance parameters

without allowing for their uncertainty. This emphasizes the importance of an approach which facilitates selection from a wide range of models, such as provided by the anisotropic Matérn model.

Given that isotropy is available as a special case, it makes good sense to routinely model anisotropy. Lark and Cullis (2004, p. 804) ignored anisotropy in their two examples of 126 and 100 observations, claiming "there were two few data to do otherwise". However, simulations by Haskard, Cullis, and Verbyla (2007) demonstrate that simultaneous estimation of Matérn and geometric anisotropy parameters can be feasible with as few as 100 observations.

The anisotropic Matérn model described here encompasses within a single general model a range of correlation models currently in common use for geostatistics, with an optional nugget effect, as well as the extension for anisotropy. Further, by formulation in the linear mixed models framework, it caters for simple, ordinary, and universal kriging within the same general model, and enables other effects, spatial or otherwise (such as assigned treatment effects), to be easily incorporated. Together, all these attributes make this a very flexible model with the potential for wide and valuable applicability in geostatistical spatial analysis.

ACKNOWLEDGMENTS

The authors wish to thank the Grains Research and Development Corporation (GRDC), Australia, for funding for K. A. Haskard to pursue this research, Mr. Brian Dunn and Dr. Mark Conyers, both of New South Wales Department of Primary Industries, Australia, for kindly providing the data for the examples, and Dr. Arthur Gilmour, also of New South Wales Department of Primary Industries, Australia, for helpful comments regarding the geometric anisotropy formulation and for implementing the model in ASReml.

[Received November 2005. Revised February 2006.]

REFERENCES

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

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Proceedings 2nd International Symposium on Information Science, Budapest*, eds. B. N. Petrov and F. Csaki, Akademai Kiado, pp. 267–281.

Beecher, H. G., Hume, I. H., and Dunn, B. W. (2002), "Improved Method for Assessing Rice Soil Suitability to Restrict Recharge," *Australian Journal or Experimental Agriculture*, 42, 297–307.

Diggle, P. J., Ribeiro, P. J. J., and Christensen, O. F. (2003), "An Introduction to Model-Based Geostatistics," in *Spatial Statistics and Computational Methods*, ed. J. Møller, New York: Springer, pp. 43–86.

Gilmour, A. R., Cullis, B. R., and Verbyla A. P. (1997), "Accounting for Natural and Extraneous Variation in the Analysis of Field Experiments," *Journal of Agricultural, Biological and Environmental Statistics*, 2, 269–293.

Gilmour, A. R., Cullis, B. R., Welham, S. J., Gogel, B. J., and Thompson, R. (2004), "An Efficient Computing Strategy for Prediction in Mixed Linear Models," *Computational Statistics and Data Analysis*, 44, 571–586.

Gilmour, A. R., Gogel, B. J., Cullis, B. R., and Thompson, R. (2005), *ASReml User Guide Release 2.0*, VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Gilmour, A. R., Thompson, R., and Cullis, B. R. (1995), "Average Information REML: An Efficient Algorithm for Variance Parameter Estimation in Linear Mixed Models," *Biometrics*, 51, 1440–1450.

Haskard, K. A. (2007), "Anisotropic Matérn Correlation and Other Issues in Model-Based Geostatistics," unpublished Ph. D. thesis, BiometricsSA, University of Adelaide. in preparation.

Haskard, K. A., Cullis, B. R., and Verbyla, A. P. (2007), "Estimation and Sampling Issues for an Anisotropic Matérn Correlation Model," submitted.

Haskard, K. A., and Verbyla, A. P. (2007), "A Note on an Apparent Non-regular Hypothesis Test," submitted.

Kenward, M. G., and Roger, J. H. (1997), "Small Sample Inference for Fixed Effects From Restricted Maximum Likelihood," *Biometrics*, 53, 983–997.

Kitanidis, P. K. (1983), "Statistical Estimation of Polynomial Generalized Covariance Functions and Hydrological Applications," *Water Resources Research*, 19, 909–921.

Lark, R. M., and Cullis, B. R. (2004), "Model-Based Analysis Using REML for Inference from Systematically Sampled Data on Soil," *European Journal of Soil Science*, 55, 799–813.

Matheron, G. (1973), "The Intrinsic Random Functions and their Applications," *Journal of Applied Probability*, 5, 439–468.

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

Stein, M. L., Chi, Z., and Welty, L. J. (2004), "Approximating Likelihoods for Large Spatial Datasts," *Journal of the Royal Statistical Society*, Series B, 66, 275–296.

Stram, D. O., and Lee, J. W. (1994), "Variance Components Testing in the Longitudinal Mixed Effects Model," *Biometrics*, 50, 1171–1177.

Vecchia, A. V. (1998), "Estimation and Model Identification for Continuous Spatial Processes," *Journal of the Royal Statistical Society*, Series B, 50, 297–312.

Verbyla, A. P. (1990), "A Conditional Derivation of Residual Maximum Likelihood," *Australian Journal of Statistics*, 32, 227–230.

Webster, R., and Oliver, M. A. (2001), *Geostatistics for Environmental Scientists*, Chichester: Wiley.