

## Estimating the spatial covariance structure using the geoadditive model

Peer-reviewed author version

VANDENDIJCK, Yannick; FAES, Christel & HENS, Niel (2017) Estimating the spatial covariance structure using the geoadditive model. In: ENVIRONMENTAL AND ECOLOGICAL STATISTICS, 24(2), p. 341-361.

DOI: 10.1007/s10651-017-0373-3

Handle: <http://hdl.handle.net/1942/23825>

# Estimating the Spatial Covariance Structure using the Geoadditive Model

Yannick Vandendijck

Interuniversity Institute for Biostatistics and statistical Bioinformatics,  
Hasselt University, B3590 Diepenbeek, Belgium

and

Christel Faes

Interuniversity Institute for Biostatistics and statistical Bioinformatics,  
Hasselt University, B3590 Diepenbeek, Belgium

and

Niel Hens

Interuniversity Institute for Biostatistics and statistical Bioinformatics,  
Hasselt University, B3590 Diepenbeek, Belgium

Centre for Health Economic Research and Modeling Infectious Diseases,  
Vaccine and Infectious Disease Institute, University of Antwerp,  
B2610 Wilrijk, Belgium

## Abstract

In geostatistics, both kriging and smoothing splines are commonly used to generate an interpolated map of a quantity of interest. The geoadditive model proposed by Kammann and Wand (2003) represents a fusion of kriging and penalized spline additive models. Complex data issues, including non-linear covariate trends, multiple measurements at a location and clustered observations are easily handled using the geoadditive model. We propose a likelihood based estimation procedure that enables the estimation of the range (spatial decay) parameter associated with the penalized splines of the spatial component in the geoadditive model. We present how the spatial covariance structure (covariogram) can be derived from the geoadditive model. In a simulation study, we show that the underlying spatial process and prediction of the spatial map are estimated well using the proposed likelihood based estimation procedure. We present several applications of the proposed methods on real-life data examples.

*Keywords:* Covariogram; Geoadditive Model; Kriging; Mixed Model; Penalized Splines

# 1 Introduction

The objective of geostatistics (Mathéron, 1963) is to produce a map of a variable of interest on a specified domain based on observations which are measured with or without noise. Geostatistics emerged in geology, mining and hydrology, but is now a widely spread statistical tool. Consider the geostatistical model

$$y(\mathbf{s}_i) = z(\mathbf{s}_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where the  $y(\mathbf{s}_i)$  are observed data values from the underlying true values  $z(\mathbf{s}_i)$ . These data values are possibly noise-corrupted by measurement error  $\varepsilon_i$ . It is assumed that the  $\varepsilon_i$  are white-noise error terms. The spatial locations  $\mathbf{s}_i$  belong to a specified continuous domain  $D \subset \mathbb{R}^d$ . The idea of geostatistics is to use the data  $y(\mathbf{s}_i)$  to make a prediction of  $z(\mathbf{s}_0)$  where  $\mathbf{s}_0 \in D$ . Both kriging and spline methods can be used to handle geostatistical problems.

In the case of kriging the values  $z(\mathbf{s}_i)$  are assumed to be the realisations of a spatially correlated random process (Mathéron, 1963; Cressie, 1993). Kriging yields the optimal - in terms of minimizing the mean squared error - unbiased linear predictor of the spatial process  $z(\cdot)$  at a specified location. Smoothing splines assume that the  $z(\mathbf{s}_i)$  are the values of smooth non-parametric function (see *e.g.*, Hutchinson and Gessler, 1994). The objective of both methods is different. In kriging, the predicted surface does not need to be aesthetic, only accurate (Dubrule, 1984). The covariation of the spatially correlated process is modelled explicitly. The goal of smoothing splines is to create a prediction surface that is as smooth as possible, meaning that the roughness of the predicted surface is taken into account in the fitting procedure. There is no account for possible spatial autocorrelation and therefore it is argued that smoothing splines are build for surfaces that are deterministic or deterministic with white noise (Cressie, 1993).

Several authors have compared the predictive performance of kriging and splines in geostatistics (Altman, 2000; Dubrule, 1984; Hutchinson and Gessler, 1994; Laslett, 1994). The conclusions of this research about the merits of both approaches vary. The formal equivalence between kriging and splines has been discussed often in literature (Cressie, 1993; Dubrule, 1983, 1984; Handcock et al., 1994; Mathéron, 1981; Nychka, 2000; Salkauskas,

1982; Watson, 1984). In summary, a kriging estimate will be identical to a spline if the appropriate (generalized) covariance is used. One well studied example is the equivalence of natural thin plate splines (Wahba and Wendelberger, 1980; Green and Silverman, 1994) and the use of specific generalized covariance functions (see *e.g.*, Dubrule, 1984).

More recently, several authors have investigated low rank approximations of the kriging approach. Low rank techniques to perform kriging have emerged from two main problems related with kriging: (1) Kriging large spatially indexed datasets is often computationally infeasible or impossible; (2) Standard kriging approaches do not allow for the incorporation of non-linear covariate effects.

Most of the approaches proposed to perform kriging on large datasets rely on a low rank approximation of the spatial covariance matrix. In fixed rank kriging a low rank is achieved by representing the spatial covariance through a small set of basis functions (Cressie and Johannesson, 2008). Covariance tapering imposes sparsity in the spatial covariance matrix (Kaufman et al., 2008). Kleiber and Nychka (2015) propose equivalent kriging which is based on approximating the kriging weight function using an equivalent kernel. Other approaches to low rank kriging can be found in, for example, Banerjee et al. (2008), Lindgren et al. (2011), Sang and Huang (2012), Stein (2014) and Nychka et al. (2015).

On the other hand, the geoaddivitive model by Kammann and Wand (2003) is such a low rank approach in geostatistics that allows for the incorporation of non-linear covariate effects. The geoaddivitive model consists of a penalized spline additive model with a geostatistical extension. Radial basis functions at a smaller set of data locations are used as spline functions to model the geostatistical component of the model. In addition, since an additive model is used, non-linear relationships between covariates and the response variable are easily incorporated in geoaddivitive models. This is an important extension since linearity of the covariate effects is usually assumed in kriging approaches. The geoaddivitive model can be expressed as a linear mixed model which allows for fitting and diagnosis using standard methodology. Unlike kriging approaches, the covariation of the underlying spatial process is not modelled explicitly by using the geoaddivitive model. This inability of the spline methods to take proper account of spatial autocorrelation of the underlying process has been a main argument against splines in geostatistics (Cressie, 1993).

In this paper the focus is on the geoadditive model and we add to the literature a likelihood based estimation procedure that enables the estimation of covariogram parameters from the geoadditive model. This is achieved by estimating the spatial decay parameter of the radial basis functions describing the spatial component in the geoadditive model. The proposed estimation procedure enables the estimation of the underlying spatial process in terms of a covariogram. We present how the (implicitly) estimated spatial covariance structure can be derived from the geoadditive model. The advantages the described method are: (1) In a geoadditive model extensions such as non-linear covariate effects, multiple measurements at the same location and clustered observations, which are difficult to account for in kriging methods, can easily be integrated due to the mixed model representation of the model; (2) Standard software can be used; (3) Predictions have similar performance as standard kriging methods; and (4) The implied spatial structure of the geoadditive model can easily be represented by a covariogram function. Previous work (*e.g.*, Stein, 1999; Gelfand et al., 2010) has described likelihood based estimation of covariogram parameters, however, linear effects in the mean function must be assumed and possible incorporation of random effects is not discussed.

Section 2 gives a review of two important concepts used throughout this paper: the covariogram and the geoadditive model introduced by Kammann and Wand (2003). In Section 3 we describe the proposed methodology in detail. A simulation study and real-life data applications are provided in Sections 4 and 5, respectively. We close the paper with a discussion in Section 6.

## 2 Important Concepts

The covariogram is an often used measure of association for spatially dependent data. The covariance for a spatially correlated random process  $z(\mathbf{s})$  is defined as

$$\text{Cov}[z(\mathbf{s} + \mathbf{h}), z(\mathbf{s})] = E[z(\mathbf{s} + \mathbf{h}), z(\mathbf{s})] - E[z(\mathbf{s} + \mathbf{h})] E[z(\mathbf{s})]. \quad (2)$$

The random process is said to be weakly stationary if  $\text{Cov}[z(\mathbf{s} + \mathbf{h}), z(\mathbf{s})] = C(\mathbf{h})$ . The function  $C(\cdot)$  is the so-called covariogram (Gelfand et al., 2010). In this paper, we only consider isotropic processes which imply that  $C(\cdot)$  depends only on the difference vector  $\mathbf{h}$

through its Euclidean length  $\|\mathbf{h}\|$ . The novelty of this paper is that we propose a likelihood based estimation procedure which enables estimation of covariogram parameters using the geoadditive model.

The geoadditive model proposed by Kammann and Wand (2003) is an additive model which incorporates a geographical component, and that is expressed as a linear mixed model. For simplicity, suppose that the data are  $(y_i, \mathbf{s}_i, a_i, b_i), 1 \leq i \leq n$ , where  $y_i$  is the value of the response at the  $i$ th location,  $a_i$  and  $b_i$  are the values of two predictor variables  $a$  and  $b$ , and  $\mathbf{s}_i$  represents the geographical location. Suppose the predictor  $a$  enters the model linearly and that the predictor  $b$  enters the model non-linearly. The geoadditive model is

$$y_i = \beta_0 + \beta_a a_i + f(b_i) + S(\mathbf{s}_i) + \varepsilon_i, \quad 1 \leq i \leq n, \quad (3)$$

where  $\varepsilon_i \sim \mathcal{N}(0, \sigma_\epsilon^2)$ ,  $f$  is a smooth function of  $b$  and  $S$  is the geographical component of the model. Both  $f$  and  $S$  are modelled using penalized spline functions. Kammann and Wand (2003) use the truncated spline basis functions to define  $f$ . Here, we use the approximating thin plate spline family. More specifically,

$$f(b) = \beta_b b + \sum_{k=1}^{K_b} u_k^b |b - \kappa_k^b|^3, \quad (4)$$

where  $\kappa_1^b, \dots, \kappa_{K_b}^b$  are  $K_b$  pre-specified knots of  $b$  and  $u_1^b, \dots, u_{K_b}^b$  are knot coefficients. Guidelines for knot selection in one dimension can be found in Ruppert et al. (2003). The coefficients  $u_1^b, \dots, u_{K_b}^b$  are penalized to overcome overfitting of the trend in the  $b$ -direction.

Although spline basis functions of the class of Matérn-splines (Handcock et al., 1994) were recommended, no firm prescriptions regarding the choice of basis functions for the spatial component  $S$  in the geoadditive model (3) were given by Kammann and Wand (2003). In this paper we investigate radial spline basis functions of the form

$$S(\mathbf{s}) = \sum_{k=1}^{K_s} u_k^s g_\tau(\mathbf{s} - \boldsymbol{\kappa}_k^s) \quad (5)$$

where  $g_\tau$  can be any of the proper covariance or generalized covariance functions used in Kriging (French et al., 2001). The subscript in  $g_\tau$  is used to denote a possible dependence on a (strictly positive) spatial decay parameter  $\tau$ . For example, Kammann and Wand (2003) presented the geoadditive model with the Matérn covariance function

$g_\tau(\mathbf{x}) = \exp(-\|\mathbf{x}\|/\tau) (1 + \|\mathbf{x}\|/\tau)$ . Another often used covariance function is the exponential function represented by  $g_\tau(\mathbf{x}) = \exp(-\|\mathbf{x}\|/\tau)$ . An overview of important (generalized) covariance functions  $g_\tau$  used in (5) are given in the Supplementary Materials. The vector  $(u_1^s, \dots, u_{K_s}^s)$  contains the  $K_s$  unknown knot coefficients that are penalized to overcome overfitting. The  $K_s$  knots  $\boldsymbol{\kappa}_1^s, \dots, \boldsymbol{\kappa}_{K_s}^s$  are a representative subset of  $(\mathbf{s}_1, \dots, \mathbf{s}_n)$  used for the construction of the basis functions. A representative subset of knots can be obtained via an efficient space filling algorithm (*e.g.*, Johnson et al., 1990).

To allow for fitting of the geoadditive model using standard mixed model software Kammann and Wand (2003) proposed to choose  $\tau$  via the simple rule

$$\hat{\tau} = \max_{1 \leq i, j \leq n} \|\mathbf{s}_i - \mathbf{s}_j\|. \quad (6)$$

It is important to notice that this choice does not have to be made for the geoadditive model. Other rules would be equally possible. In this paper, we propose to estimate  $\tau$  via a likelihood based approach. In the simulation study in Section 4 we observe that this latter choice leads to less variable predictions.

Choosing  $\tau$  upfront via a simple rule, *e.g.* using (6), has the nice feature that the geoadditive model given in (3) can be expressed as a linear mixed model. Using linear mixed models, penalization of the  $u_k^b$  and  $u_k^s$  coefficients is equivalent to treating them as random effects (Brumback et al., 1999; Ruppert et al., 2003). Specifically, let  $\boldsymbol{\beta} = (\beta_0, \beta_a, \beta_b)^T$ ,  $\mathbf{u} = (u_1^b, \dots, u_{K_b}^b, u_1^s, \dots, u_{K_s}^s)^T = (\mathbf{u}_b^T, \mathbf{u}_S^T)^T$ ,  $\mathbf{X} = (1 \ a_i \ b_i)_{1 \leq i \leq n}$  and  $\mathbf{Z} = (\mathbf{Z}_b | \mathbf{Z}_S)$ , where

$$\mathbf{Z}_b = (|b_i - \kappa_k^b|^3)_{1 \leq i \leq n, 1 \leq k \leq K_b} \text{ and } \mathbf{Z}_S = (g_\tau(\mathbf{s}_i - \boldsymbol{\kappa}_k^s))_{1 \leq i \leq n, 1 \leq k \leq K_s}, \quad (7)$$

then the geoadditive model in (3) can be expressed as  $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$ . For fitting purposes, one should reparametrize to  $\tilde{\mathbf{Z}} = (\tilde{\mathbf{Z}}_b | \tilde{\mathbf{Z}}_S)$  and  $\tilde{\mathbf{u}} = (\tilde{\mathbf{u}}_b^T, \tilde{\mathbf{u}}_S^T)^T$ , where  $\tilde{\mathbf{Z}}_b = \mathbf{Z}_b \boldsymbol{\Omega}_b^{-1/2}$ ,  $\tilde{\mathbf{Z}}_S = \mathbf{Z}_S \boldsymbol{\Omega}_S^{-1/2}$ ,  $\tilde{\mathbf{u}}_b = \boldsymbol{\Omega}_b^{1/2} \mathbf{u}_b$ ,  $\tilde{\mathbf{u}}_S = \boldsymbol{\Omega}_S^{1/2} \mathbf{u}_S$ ,  $\boldsymbol{\Omega}_b = (|\kappa_k^b - \kappa_{k'}^b|^3)_{1 \leq k' \leq K_b, 1 \leq k \leq K_b}$  and  $\boldsymbol{\Omega}_S = (g_\tau(\boldsymbol{\kappa}_k^s - \boldsymbol{\kappa}_{k'}^s))_{1 \leq k' \leq K_s, 1 \leq k \leq K_s}$ , which yields

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \tilde{\mathbf{Z}}\tilde{\mathbf{u}} + \boldsymbol{\varepsilon}, \quad \begin{pmatrix} \tilde{\mathbf{u}} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \sigma_b^2 \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_S^2 \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma_\varepsilon^2 \mathbf{I} \end{pmatrix} \right). \quad (8)$$



The matrices of  $\mathbf{\Omega}_b^{-1/2}$  and  $\mathbf{\Omega}_S^{-1/2}$  are calculated using singular value decomposition. For example,  $\mathbf{\Omega}_S^{-1/2} = \mathbf{U} \text{diag}(1/\sqrt{\mathbf{d}}) \mathbf{V}^T$ , where  $\mathbf{U}$ ,  $\mathbf{d}$  and  $\mathbf{V}$  correspond to the singular value decomposition of  $\mathbf{\Omega}_S$  and  $\sqrt{\mathbf{d}}$  is obtained by replacing the entries of  $\mathbf{d}$  by their non-negative square roots. The linear mixed model given in (8) can be fitted using standard linear mixed model software. The incorporation of more linear or non-linear covariates is straightforward. In addition, clustering of observations can be incorporated by the introduction of a random intercept.

### 3 Methodology

In this section, we describe the proposed likelihood based approach to estimate the spatial decay parameter  $\tau$ . In the simulation study in Section 4, we will observe that this enables us to estimate the covariogram parameters with minor bias. Next, we present how predictions can be made using the geoadditive model. Finally, we discuss how the covariogram can be derived from the geoadditive model.

#### Maximum Likelihood Estimation

Consider the geoadditive model in (3), where (4) and (5) are used to model  $f$  and  $S$ , respectively. A two-stage iterative likelihood based estimation method is proposed. At the first stage the linear mixed model in (8) is estimated fixing  $\tau$  at its current value, and in the second stage the  $\tau$  parameter is optimized. This process is iterated until convergence is attained. To be more specific, the estimation method consists of the following steps:

- [i] Set an initial value,  $\hat{\tau}^{(0)}$ , for the  $\tau$  parameter.
- [ii] Fixing  $\tau$  at  $\hat{\tau}^{(k)}$  in the  $g_\tau$  function used in (5), fit the linear mixed model (8) using maximum (ML) or restricted maximum likelihood (REML). This yields estimates of the variance parameters  $\boldsymbol{\theta}^{(k+1)} = \{(\hat{\sigma}_b^2)^{(k+1)}, (\hat{\sigma}_S^2)^{(k+1)}, (\hat{\sigma}_\varepsilon^2)^{(k+1)}\}$  and of the fixed and random effects parameters  $\hat{\boldsymbol{\beta}}^{(k+1)}$  and  $\hat{\mathbf{u}}^{(k+1)}$ , respectively.
- [iii] Using  $\boldsymbol{\theta}^{(k+1)}$  and  $\hat{\boldsymbol{\beta}}^{(k+1)}$  maximize, in case of ML in step (ii), the likelihood function

$$l_{ML}(\tau) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\tilde{\boldsymbol{\Sigma}}(\tau)| - \frac{1}{2} \left( \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(k+1)} \right)^T \tilde{\boldsymbol{\Sigma}}^{-1}(\tau) \left( \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(k+1)} \right), \quad (9)$$

or in case of REML in step (ii) the restricted likelihood function

$$l_{REML}(\tau) = -\frac{n-p}{2} \log(2\pi) - \frac{1}{2} \log |\tilde{\Sigma}(\tau)| - \frac{1}{2} \log |\mathbf{X}^T \tilde{\Sigma}^{-1}(\tau) \mathbf{X}| - \frac{1}{2} \left( \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(k+1)} \right)^T \tilde{\Sigma}^{-1}(\tau) \left( \mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^{(k+1)} \right) \quad (10)$$

with respect to the parameter  $\tau$ , where  $p$  denotes the number of fixed effects in the linear mixed model (8),  $\tilde{\Sigma}(\tau) = \Sigma(\tau; \hat{\boldsymbol{\theta}}^{(k+1)}) = \tilde{\mathbf{Z}} \mathbf{G} \tilde{\mathbf{Z}}^T + (\hat{\sigma}_\varepsilon^2)^{(k+1)} \mathbf{I}$  and  $\mathbf{G} = \begin{pmatrix} (\hat{\sigma}_b^2)^{(k+1)} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & (\hat{\sigma}_S^2)^{(k+1)} \mathbf{I} \end{pmatrix}$ . The value of  $\tau$  that maximizes these functions is denoted by  $\hat{\tau}^{(k+1)}$ .

- [iv] For  $k = 0, 1, \dots$ , iterate between steps [ii] and [iii] until the difference between two successive  $\hat{\tau}$  value is smaller than a prespecified tolerance level, namely  $|\hat{\tau}^{(k)} - \hat{\tau}^{(k+1)}| < c$ . We set  $c = 1e-6$ .

An R module to fit geoaddivitive models using this iterative approach has been developed (see Supplementary Materials Section 4).

## Predictions

An important aspect of geostatistics is predicting the value of the outcome at an arbitrary location  $\mathbf{s}_0 \in D$ . Suppose the covariate values  $a_0$  and  $b_0$  of  $a$  and  $b$  are available at location  $\mathbf{s}_0$ . The prediction using model (8) has the form  $\hat{y}_0 = \mathbf{x}_0 \hat{\boldsymbol{\beta}} + \tilde{\mathbf{Z}}_0 \hat{\mathbf{u}}$ , where  $\mathbf{x}_0 = [1 \ a_0 \ b_0]$ ,  $\tilde{\mathbf{Z}}_0 = [\tilde{\mathbf{Z}}_{0,b}, \tilde{\mathbf{Z}}_{0,S}]$ ,  $\tilde{\mathbf{Z}}_{0,b} = \mathbf{Z}_{0,b} \boldsymbol{\Omega}_b^{-1/2}$ ,  $\tilde{\mathbf{Z}}_{0,S} = \mathbf{Z}_{0,S} \boldsymbol{\Omega}_S^{-1/2}$ ,  $\mathbf{Z}_{0,b} = (|b_0 - \kappa_k^b|^3)_{1 \leq k \leq K_b}^T$  and  $\mathbf{Z}_{0,S} = (g_\tau(\mathbf{s}_0 - \boldsymbol{\kappa}_k^s))_{1 \leq k \leq K_s}^T$ . Note that the prediction  $\hat{y}_0$  can be written as  $\hat{y}_0 = \mathbf{c}_0 [\hat{\boldsymbol{\beta}}^T, \hat{\mathbf{u}}^T]^T$  where  $\mathbf{c}_0 = [\mathbf{x}_0, \tilde{\mathbf{Z}}_0]$ .

## Prediction Variance

The variance associated with the prediction  $\hat{y}_0$ ,  $\text{var}(\hat{y}_0)$ , can be calculated using the covariance matrix

$$\text{Cov} \left( \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} - \tilde{\mathbf{u}} \end{bmatrix} \right) = \left( \frac{1}{\sigma_\varepsilon^2} \mathbf{C}^T \mathbf{C} + \mathbf{B} \right)^{-1}, \quad \text{where } \mathbf{B} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix} \quad (11)$$

and  $\mathbf{C} = [\mathbf{X} | \tilde{\mathbf{Z}}]$  (Ruppert et al., 2003). Thus the prediction variance is given by  $\text{var}(\hat{y}_0) = \mathbf{c}_0 \left( \frac{1}{\sigma_\varepsilon^2} \mathbf{C}^T \mathbf{C} + \mathbf{B} \right)^{-1} \mathbf{c}_0^T$ . An estimate of the variance is obtained by plugging in the (restricted) maximum likelihood estimates of the variance parameters. A confidence interval (CI) is calculated by adding and subtracting the square root of the variance from the

prediction using the normal distribution as reference. We will observe in the simulation study in Section 4 that CIs based on this analytical expression of the variance do not reach acceptable nominal coverages. Gelfand et al. (2010) note that, indeed, uncertainty estimation is problematic with penalized spline methods in geostatistics, especially if the standard deviation of the measurement errors is not large compared to the changes in the underlying function.

We propose a bootstrap method to calculate the prediction variance that properly accounts for the uncertainty associated with the spatial component  $S$ . The procedure resembles bootstrap methods used in kriging to obtain bootstrapped prediction variances (Schelin and Sjöstedt-de Luna, 2010). The procedure consists of the following steps:

- [i] From the original data, use the estimation method described above to obtain the estimates  $\hat{\tau}$ ,  $(\hat{\sigma}_b^2, \hat{\sigma}_S^2, \hat{\sigma}_\varepsilon^2)$ ,  $\hat{\boldsymbol{\beta}}$  and  $\hat{\mathbf{u}} = (\hat{\mathbf{u}}_b^T, \hat{\mathbf{u}}_S^T)^T$ .
- [ii] Compute  $\boldsymbol{\mu}^* = \begin{pmatrix} \mathbf{X} \\ \mathbf{x}_0 \end{pmatrix} \hat{\boldsymbol{\beta}} + \begin{pmatrix} \tilde{\mathbf{Z}}_b \\ \tilde{\mathbf{Z}}_{0,b} \end{pmatrix} \hat{\mathbf{u}}_b$ . If the model in (8) also contains a random intercept, this is included in  $\boldsymbol{\mu}^*$  where the random intercepts are drawn from a normal distribution with mean zero and variance equal to the estimated random intercept variance.
- [iii] Let  $\mathbf{s}_{n+1} = \mathbf{s}_0$  and construct a  $(n+1) \times (n+1)$  matrix  $\boldsymbol{\Xi}$  with  $ij$ th entry  $\hat{\sigma}_S^2 g_\tau(\mathbf{s}_i - \mathbf{s}_j)$ . The Cholesky decomposition yields  $\boldsymbol{\Xi} = \mathbf{L}\mathbf{L}^T$ , where  $\mathbf{L}$  is a lower triangular  $(n+1) \times (n+1)$  matrix (Stewart, 1973). Values  $\mathbf{S}^*$  representing the spatial component can be simulated by  $\mathbf{S}^* = \mathbf{L}\boldsymbol{\epsilon}$ , where  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_{n+1})^T$  is a vector of uncorrelated normally distributed random variables with zero mean and unit variance.
- [iv] Let  $(\tilde{\mathbf{y}}^{*T}, \tilde{y}_0^{*T})^T = \boldsymbol{\mu}^* + \mathbf{S}^*$  and  $(\mathbf{y}^{*T}, y_0^{*T})^T = \boldsymbol{\mu}^* + \mathbf{S}^* + \boldsymbol{\varepsilon}$ , where  $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_{n+1})^T$  and  $\varepsilon_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, \hat{\sigma}_\varepsilon^2)$  for  $1 \leq i \leq n+1$ .
- [v] Using the  $\mathbf{y}^*$  as response values, fit model (8) using the estimation method described above to obtain the estimates  $\hat{\tau}^*$ ,  $(\hat{\sigma}_b^{*2}, \hat{\sigma}_S^{*2}, \hat{\sigma}_\varepsilon^{*2})$ ,  $\hat{\boldsymbol{\beta}}^*$  and  $\hat{\mathbf{u}}^* = (\hat{\mathbf{u}}_b^{*T}, \hat{\mathbf{u}}_S^{*T})^T$ . Define  $\hat{y}_0^* = \mathbf{x}_0 \hat{\boldsymbol{\beta}}^* + \tilde{\mathbf{Z}}_0 \hat{\mathbf{u}}^*$ .
- [vi] Using steps [ii]-[v], construct  $B$  bootstrap replicates  $(\tilde{y}_0^{*(b)}, \hat{y}_0^{*(b)})_{1 \leq b \leq B}$ . The bootstrap prediction variance of  $\hat{y}_0$  is  $\frac{1}{B} \sum_{b=1}^B \left( \hat{y}_0^{*(b)} - \tilde{y}_0^{*(b)} \right)^2$ .

The time consuming part of the bootstrap procedure is the model fitting in step [v]. We consider an approximation to the proposed bootstrap procedure by fixing the scale and

variance parameters in step [v] at their original estimates, namely  $\hat{\tau}$  and  $(\hat{\sigma}_b^2, \hat{\sigma}_S^2, \hat{\sigma}_\varepsilon^2)$ . Consequently, the computation of estimates  $\hat{\beta}^*$  and  $\hat{\mathbf{u}}^*$  involve only matrix operations, namely  $\hat{\beta}^* = (\mathbf{X}^T \tilde{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \tilde{\Sigma}^{-1} \mathbf{y}^*$  and  $\hat{\mathbf{u}}^* = \mathbf{G} \tilde{\mathbf{Z}}^T \tilde{\Sigma}^{-1} (\mathbf{y}^* - \mathbf{X} \hat{\beta}^*)$ . In this manner, a bootstrapped prediction variance (with  $B=500$ ) is obtained fast (*e.g.*, within three seconds on a laptop personal computer with Intel Core i5-4210M @ 2.60HGz processor). In Section 4, we investigate the performance of this approximated bootstrap procedure in a simulation study.

The estimated parameters  $\hat{\beta}^*$  and  $\hat{\mathbf{u}}^*$  can also be used to calculate 95% pointwise bootstrap confidence intervals of the non-linear effect and the spatial component  $S$ .

### Deriving the Covariogram from the Geoadditive Model

From model (3) and (8) it follows that the spatial covariance of  $\mathbf{y}$  implied by the spatial component  $S$  is given by  $\Sigma_S = \tilde{\mathbf{Z}}_S (\sigma_S^2 \mathbf{I}) \tilde{\mathbf{Z}}_S^T = \sigma_S^2 \tilde{\mathbf{Z}}_S \tilde{\mathbf{Z}}_S^T$ . Let  $\mathbf{y}_k = (y_1, \dots, y_{K_s})^T$  be the response values at the  $K_s$  knots  $(\kappa_1^s, \dots, \kappa_{K_s}^s)^T$ . It follows that the spatial covariance between  $\mathbf{y}_k$  and  $y_0$  at an arbitrary spatial location  $s_0$  is given by  $\text{cov}^{(S)}(\mathbf{y}_k, y_0) = \sigma_S^2 \tilde{\mathbf{Z}}_{k,S} \tilde{\mathbf{Z}}_{0,S}^T$ , where  $\tilde{\mathbf{Z}}_{k,S} = \mathbf{Z}_{k,S} \Omega_S^{-1/2}$  and  $\mathbf{Z}_{k,S}$  is a  $K_s \times K_s$  matrix with  $kk'$ th entry  $g_\tau(\kappa_k^s - \kappa_{k'}^s)$ . We thus have that  $\mathbf{Z}_{k,S} = \Omega_S$  (Section 2) which implies  $\tilde{\mathbf{Z}}_{k,S} = \Omega_S \Omega_S^{-1/2}$ . If  $\Omega_S$  is positive semi-definite it further follows that  $\tilde{\mathbf{Z}}_{k,S} = \Omega_S^{1/2}$  (Searle, 1982). Since  $\tilde{\mathbf{Z}}_{0,S} = \mathbf{Z}_{0,S} \Omega_S^{-1/2}$  it follows that

$$\text{cov}^{(S)}(\mathbf{y}_k, y_0) = \sigma_S^2 \Omega_S^{1/2} \Omega_S^{-1/2} \mathbf{Z}_{0,S}^T = \sigma_S^2 \mathbf{Z}_{0,S}^T \quad (12)$$

and thus

$$\text{cov}^{(S)}(\mathbf{y}_k, y_0) = \sigma_S^2 \{g_\tau(\mathbf{s}_0 - \kappa_k^s)\}_{1 \leq k \leq K_s}. \quad (13)$$

Thus  $\text{cov}^{(S)}(\mathbf{y}_k, y_0)$  is fully defined by  $\sigma_S^2$ ,  $\tau$  and the function  $g_\tau$  used as basis function for the spatial component  $S$  in (3). The covariance  $\text{cov}^{(S)}(\mathbf{y}_k, y_0)$  is similar in form as the covariogram functions used in kriging. Since the considered functions  $g_\tau$  only depend on the distance  $\|\mathbf{h}\|$  between two spatial points, the implied covariogram can be calculated by  $\sigma_S^2 g_\tau(\mathbf{h})$  for every distance  $\|\mathbf{h}\|$ . We further note that the value of the implied covariogram at a distance zero is  $\sigma_S^2 + \sigma_\varepsilon^2$ . This is equivalent as the covariogram in kriging when it is assumed that the nugget of the covariogram is entirely made out of the measurement error variance.

To be able to derive (13), we used the condition that  $\mathbf{\Omega}_S$  is positive semi-definite. Most of the typically used covariogram functions imply a positive semi-definite matrix  $\mathbf{\Omega}_S$  (see Supplementary Materials). Without this condition, the implied covariance in (13) does not hold. For example, thin plate splines do not imply a positive semi-definite matrix  $\mathbf{\Omega}_S$  when used to model the spatial component  $S$ . Therefore, a covariogram cannot be derived for thin plate splines, however they imply a generalized covariance function (Cressie, 1993).

## 4 Simulation Study

In order to evaluate the proposed methodology we carried out a simulation study. We focus on two scenarios: (1) Data is simulated from a spatial process with constant mean; and (2) Data is simulated from a spatial process with a mean function depending on two covariates.

### No Covariates

We consider as spatial domain the unit square with a resolution of  $0.01 \times 0.01$ . Data at a spatial location  $\mathbf{s} = (s_x, s_y)$  on this square is simulated using the model

$$y_{\mathbf{s}} = S(\mathbf{s}) + \varepsilon_{\mathbf{s}}, \quad (14)$$

where  $\varepsilon_{\mathbf{s}} \sim \mathcal{N}(0, \sigma_{\varepsilon}^2)$  and  $S(\mathbf{s})$  is a zero-mean Gaussian Random Field (GRF) (Gelfand et al., 2010) with one of the following four distinct covariogram models:

- Exponential covariogram:  $C(\mathbf{h}) = c_s \exp\left(-\frac{\|\mathbf{h}\|}{\tau}\right)$
- Matérn covariogram ( $\nu = 3/2$ ):  $C(\mathbf{h}) = c_s \exp\left(-\frac{\|\mathbf{h}\|}{\tau}\right) \left(1 + \frac{\|\mathbf{h}\|}{\tau}\right)$
- Spherical covariogram:  $C(\mathbf{h}) = c_s \left(1 - \frac{3}{2} \frac{\|\mathbf{h}\|}{\tau} + \frac{1}{2} \frac{\|\mathbf{h}\|^3}{\tau^3}\right) I_{\|\mathbf{h}\| \leq \tau}$
- Gaussian covariogram:  $C(\mathbf{h}) = c_s \exp\left(-\frac{\|\mathbf{h}\|^2}{\tau^2}\right)$

Note that  $\varepsilon_{\mathbf{s}}$  represents the so-called nugget effect in kriging which consist entirely out of measurement error in this case. As parameter values we consider  $\sigma_{\varepsilon}^2 = 0.10$  and in all four covariogram models  $c_s = 0.50$  and  $\tau = 0.15$ . For each covariogram model we obtained 250 simulated realizations using (14). In each realization a random sample is drawn. For the exponential and Matérn covariogram we consider a sample of size  $n = 200$ , and for the spherical and Gaussian covariogram  $n = 500$ .

For each simulated dataset, the covariogram parameters  $(c_s, \tau)$  and the measurement error parameter  $\sigma_\varepsilon^2$  were estimated using seven different methods. [1] Direct maximum likelihood ( $D_{ML}$ ) parameter estimation for GRFs, and [2] Direct restricted maximum likelihood ( $D_{REML}$ ) parameter estimation for GRFs. We use the term *direct* to indicate to all parameters are estimated jointly. [3] Weighted least squares (WLS) estimation using Cressie's (1985) weights of the empirical semivariogram. [4] Maximum likelihood estimation of the geoadditive model with  $\tau$  fixed at (6) ( $GM1_{ML}$ ), and [5] restricted maximum likelihood estimation of the geoadditive model with  $\tau$  fixed at (6) ( $GM1_{REML}$ ). [6] Maximum likelihood estimation of the geoadditive model using the estimation procedure proposed in Section 3 ( $GM2_{ML}$ ). [7] Restricted maximum likelihood estimation of the geoadditive model using the estimation procedure proposed in Section 3 ( $GM2_{REML}$ ). For each estimation method the mean function contains only the intercept.  $D_{ML}$ ,  $D_{REML}$  and WLS are often used methods in kriging to obtain covariogram parameter estimates (for more details on these estimation methods we recommend Gelfand et al., 2010).

In addition to estimation performance of the covariogram parameters, the prediction performance corresponding to each of the seven methods is also evaluated. The predictive performance is investigated at five locations on the considered spatial domain, namely  $(0.50, 0.50)$ ,  $(0.05, 0.05)$ ,  $(0.05, 0.95)$ ,  $(0.95, 0.05)$  and  $(0.95, 0.95)$ . For  $D_{ML}$ ,  $D_{REML}$  and WLS, the standard ordinary kriging equations are used to obtain predictions and kriging variances (Cressie, 1993). For  $GM1_{ML}$ ,  $GM1_{REML}$ ,  $GM2_{ML}$  and  $GM2_{REML}$ , predictions are made as described in Section 3. The prediction variances are calculated using the analytical formula using the covariance matrix (11). For  $GM2_{ML}$  and  $GM2_{REML}$ , we also use the bootstrap approach presented in Section 3 to obtain prediction variances. The prediction performance was measured by the mean squared error averaged over the five locations. Using the prediction variances, 95% pointwise confidence intervals (CIs) are calculated using the normal distribution as reference (i.e.,  $\pm 1.96 \times$  the square root of the prediction variance). The reported nominal coverages of the CIs are calculated by averaging over the five prediction locations.

For each covariogram model used to simulate the GRF, we use the same covariogram model to perform parameter estimation. For example, when the exponential covariogram

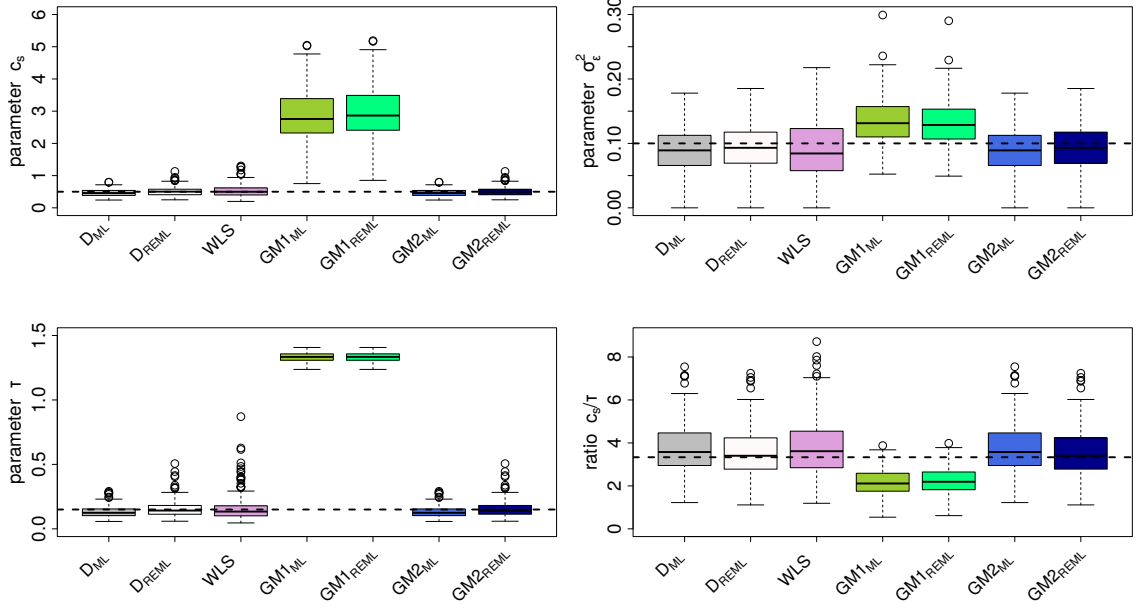


Figure 1: Boxplots of the estimated covariogram parameters  $(c_s, \tau)$  and the measurement error parameter  $\sigma_\epsilon^2$  over 250 simulations using seven methods (see text). Data is simulated according to model (14) with the exponential covariogram model.

is used to simulate data for the GRF  $S$ , the exponential function  $g_\tau$  is also used in the geoadditive model. For the simulations using the exponential, Matérn and spherical covariogram models all sampled locations are used as knots. For the simulations using the Gaussian covariogram model only 150 knots are used since surfaces obtained from a Gaussian covariogram are more smooth.

Results are displayed in Figure 1 and Tables 1-2. Results for the ratio  $c_s/\tau$  are also presented since this ratio is known to affect predictions more than either the parameters  $c_s$  and  $\tau$  do individually in kriging (Stein and Handcock, 1989). As expected, estimated covariogram parameters using  $GM1_{ML}$  and  $GM1_{REML}$  are seriously biased, which also affects the estimation of the measurement error parameter  $\sigma_\epsilon^2$ .  $GM2_{ML}$  and  $GM2_{REML}$  perform well and yield very similar results as the direct likelihood approaches  $D_{ML}$  and  $D_{REML}$ . Results for other covariogram models are qualitatively similar (see Supplementary Materials Figures 1-3)

With respect to MSE (Table 1), it is observed that the geoadditive model in which  $\tau$  is estimated using the proposed procedure ( $GM2_{ML}$  and  $GM2_{REML}$ ) performs well and similar

Table 1: MSE ( $\times 100$ ) of the covariogram parameters ( $c_s$ ,  $\tau$ ), the measurement error parameter  $\sigma_\varepsilon^2$  and the ratio  $c_s/\tau$  over 250 simulations using seven methods (see text). Data is simulated according to model (14) with four different covariogram models.

	$D_{ML}$	$D_{REML}$	WLS	$GM1_{ML}$	$GM1_{REML}$	$GM2_{ML}$	$GM2_{REML}$
Exponential covariogram							
$c_s$	1.27	1.83	115.88	626.45	678.70	1.27	1.83
$\tau$	0.21	0.40	41.72	139.58	139.58	0.21	0.40
$\sigma_\varepsilon^2$	0.13	0.13	0.21	0.25	0.22	0.13	0.13
$c_s/\tau$	140.29	122.72	222.59	177.82	160.24	140.35	122.76
Matérn covariogram							
$c_s$	4.81	9.49	138.76	$> 10^3$	$> 10^3$	4.81	9.51
$\tau$	0.16	0.23	1.31	138.80	138.80	0.16	0.23
$\sigma_\varepsilon^2$	0.02	0.02	0.03	0.03	0.03	0.02	0.02
$c_s/\tau$	83.92	110.44	212.08	$> 10^3$	$> 10^3$	83.93	110.54
Spherical covariogram							
$c_s$	0.43	0.43	0.62	$> 10^3$	$> 10^3$	0.43	0.43
$\tau$	0.02	0.02	0.03	147.77	147.77	0.02	0.02
$\sigma_\varepsilon^2$	0.04	0.04	0.08	0.05	0.05	0.04	0.04
$c_s/\tau$	19.32	19.32	35.15	28.16	29.33	19.33	19.32
Gaussian covariogram							
$c_s$	1.32	1.45	2.91	$> 10^3$	$> 10^3$	1.33	1.47
$\tau$	0.01	0.01	0.05	147.52	147.52	0.01	0.01
$\sigma_\varepsilon^2$	0.00	0.00	0.01	4.46	3.87	0.00	0.00
$c_s/\tau$	44.23	47.36	78.35	$> 10^3$	$> 10^3$	44.66	48.02

as the direct likelihood methods. This implies that the proposed doubly iterative estimation procedure performs as good as direct likelihood methods for this simple simulation setting (no covariates or clustering). Thus although both methods are different, in  $D_{ML}$  and  $D_{REML}$  the spatial variation is modelled via a Gaussian process and in  $GM2_{ML}$  and  $GM2_{REML}$  the spatial variation is modelled through splines in the geosadditive model, they perform similar for this simulation setting. ML approaches yield slightly better results than REML approaches for the parameters  $c_s$  and  $\tau$ . However, for the ratio  $c_s/\tau$  there is no clear answer



Table 2: MSE ( $\times 100$ ) of the predictive performance and the nominal coverage of the 95% confidence intervals over 250 simulation using seven methods (see text). Data is simulated according to model (14) with four different covariogram models.

	$D_{ML}$	$D_{REML}$	WLS	$GM1_{ML}$	$GM1_{REML}$	$GM2_{ML}$	$GM2_{REML}$
Exponential covariogram							
MSE ( $\times 100$ )	18.24	18.25	18.25	18.89	18.84	18.24	18.25
95% coverage	94.0	93.8	94.1	67.1	67.0	56.8	57.9
95% coverage <sup>a</sup>						94.1	94.0
Matérn covariogram							
MSE ( $\times 100$ )	4.82	4.78	4.94	4.99	4.97	4.82	4.78
95% coverage	94.1	94.2	93.1	89.1	89.7	90.9	91.2
95% coverage <sup>a</sup>						94.2	94.0
Spherical covariogram							
MSE ( $\times 100$ )	16.08	16.08	16.27	16.42	16.41	16.08	16.08
95% coverage	94.7	94.8	94.6	63.0	62.8	65.4	65.4
95% coverage <sup>a</sup>						94.6	94.5
Gaussian covariogram							
MSE ( $\times 100$ )	2.35	2.35	2.47	15.46	14.32	2.35	2.35
95% coverage	95.7	95.7	94.9	56.7	58.2	94.6	94.6
95% coverage <sup>a</sup>						95.9	96.0

a: Based on the bootstrap procedure described in Section 3.

whether ML or REML is better. The WLS approach does not perform as well as the direct likelihood estimators, which is in line with other simulations studies comparing  $D_{ML}$  and WLS (*e.g.*, Zimmerman and Zimmerman, 1991; Zhang and Zimmerman, 2007).  $GM1_{ML}$  and  $GM1_{REML}$  yield marked increases in the MSE of  $c_s$ ,  $\tau$  and  $c_s/\tau$ . This is expected since methods  $GM1_{ML}$  and  $GM1_{REML}$  do not pursue estimation of the covariogram parameters. However, this also leads to worse MSE results for the parameter  $\sigma_\epsilon^2$ . This is clearly observed for the Gaussian covariogram case. Results are qualitatively similar over the different covariogram models.

Differences amongst the different approaches in terms of predictive performance (Table 2) are less pronounced. In particular,  $GM2_{ML}$  and  $GM2_{REML}$  perform similar as

the kriging methods using direct likelihood estimation ( $D_{ML}$  and  $D_{REML}$ ) and perform (slightly) better than  $GM1_{ML}$  and  $GM1_{REML}$ . However, for the Gaussian covariogram model there is a clear difference in MSE. Looking at the coverage of the predictions (Table 2), it is observed that the bootstrap prediction variance is necessary for  $GM2_{ML}$  and  $GM2_{REML}$  to obtain CIs with close to 95% nominal coverage.

Additional results with respect to prediction bias are given in the Supplementary Materials Figure 4. In Figure 5 in the Supplementary Materials, we also present the estimated covariogram functions.

### Covariates

The simulation scenarios in the previous section all assumed a constant mean. In this section, we introduce covariate effects, namely one linear effect and one non-linear effect. Data on the unit square are now simulated using the model

$$y_{\mathbf{s}} = S(\mathbf{s}) - 0.5x_{1\mathbf{s}} + \sin(2\pi x_{2\mathbf{s}}) + \varepsilon_{\mathbf{s}}, \quad (15)$$

where again  $\varepsilon_{\mathbf{s}} \sim \mathcal{N}(0, \sigma_{\varepsilon}^2 = 0.10)$ ,  $x_{1\mathbf{s}} \sim U(0, 1)$ ,  $x_{2\mathbf{s}} \sim U(0, 1)$ , and  $S(\mathbf{s})$  is a GRF with a Gaussian covariogram model or a circular covariogram model, namely  $C(\mathbf{h}) = c_s \left\{ 1 - \frac{2}{\pi} \left( \vartheta \sqrt{1 - \vartheta^2} + \arcsin \vartheta \right) \right\} I_{\|\mathbf{h}\| \leq \tau}$ , where  $\vartheta = \min \left( \frac{\|\mathbf{h}\|}{\tau}, 1 \right)$ . Again, we consider  $c_s = 0.50$ ,  $\tau = 0.15$  and obtain 250 simulated realizations from (15) from which a random sample of size  $n = 500$  is obtained. The estimation methods  $D_{ML}$ ,  $D_{REML}$ , WLS,  $GM1_{ML}$ ,  $GM1_{REML}$ ,  $GM2_{ML}$  and  $GM2_{REML}$  are used to obtain estimates of the covariogram parameters, the measurement error parameter  $\sigma_{\varepsilon}^2$  and predictions at five spatial locations. For  $D_{ML}$ ,  $D_{REML}$  and WLS, both covariates  $x_1$  and  $x_2$  enter the mean function linearly (only considering linear effects is a common choice for kriging methods). For the geoaddivitive models ( $GM1_{ML}$ ,  $GM1_{REML}$ ,  $GM2_{ML}$  and  $GM2_{REML}$ ), the covariate  $x_1$  enters the model linearly and  $x_2$  non-linearly. The simulations using the circular covariogram model use all sampled locations as knots, and the simulations using the Gaussian covariogram model use 150 knots.

The results are displayed in Tables 3 and 4. It is observed that  $GM2_{ML}$  and  $GM2_{REML}$  perform the best for the estimation  $c_s$ ,  $\tau$  and  $\sigma_{\varepsilon}^2$ . As expected, entering the non-linear covariate  $x_2$  linearly in the mean function for  $D_{ML}$ ,  $D_{REML}$  and WLS has an effect on the estimation of the covariogram parameters. The higher MSE values for  $D_{ML}$ ,  $D_{REML}$  and

Table 3: MSE ( $\times 100$ ) of the covariogram parameters ( $c_s$ ,  $\tau$ ), the measurement error parameter  $\sigma_\varepsilon^2$  and the ratio  $c_s/\tau$  over 250 simulations using seven methods (see text). Data is simulated according to model (15) with two different covariogram models.

	$D_{ML}$	$D_{REML}$	WLS	$GM1_{ML}$	$GM1_{REML}$	$GM2_{ML}$	$GM2_{REML}$
Gaussian covariogram							
$c_s$	1.67	1.82	2.03	$> 10^3$	$> 10^3$	1.51	1.55
$\tau$	0.02	0.02	0.06	148.52	148.52	0.01	0.01
$\sigma_\varepsilon^2$	3.76	3.83	3.88	3.71	3.73	0.01	0.01
$c_s/\tau$	58.97	62.24	75.33	$> 10^3$	$> 10^3$	49.11	50.06
Circular covariogram							
$c_s$	4.75	5.28	1.02	$> 10^3$	$> 10^3$	1.70	1.71
$\tau$	1.07	1.14	0.06	147.73	147.73	0.12	0.12
$\sigma_\varepsilon^2$	3.91	3.96	4.10	0.17	0.13	0.03	0.03
$c_s/\tau$	38.21	38.83	55.38	201.10	155.19	22.96	22.96

Table 4: MSE ( $\times 100$ ) of the predictive performance and the nominal coverage of the 95% confidence intervals over 250 simulation using seven methods (see text). Data is simulated according to model (15) with two different covariogram models.

	$D_{ML}$	$D_{REML}$	WLS	$GM1_{ML}$	$GM1_{REML}$	$GM2_{ML}$	$GM2_{REML}$
Gaussian covariogram							
MSE ( $\times 100$ )	23.65	23.64	23.88	14.61	14.61	2.39	2.40
95% coverage	57.2	57.2	57.6	63.4	64.4	94.8	94.8
95% coverage <sup>a</sup>						95.0	95.3
Circular covariogram							
MSE ( $\times 100$ )	37.68	37.67	37.67	15.50	15.10	14.40	14.42
95% coverage	80.6	80.7	80.2	62.3	64.6	68.5	68.5
95% coverage <sup>a</sup>						93.7	93.8

a: Based on the bootstrap procedure described in Section 3.

WLS are mainly caused by an increase in the variability of the obtained estimates (see Supplementary Material Figures 9 and 10). The parameter estimates of  $\sigma_\varepsilon^2$  are biased for  $D_{ML}$ ,  $D_{REML}$  and WLS (see Supplementary Material Figures 9 and 10).

In terms of prediction, the geoadditive models in which  $\tau$  is estimated using the proposed procedure (GM2<sub>ML</sub> and GM2<sub>REML</sub>) outperform the other considered approaches (Table 4). The higher MSE values for D<sub>ML</sub>, D<sub>REML</sub>, WLS, GM1<sub>ML</sub> and GM1<sub>REML</sub> are associated with higher variability of the obtained predictions (see Supplementary Material Figure 8). The prediction bias is negligible for all considered methods. For the circular covariogram model, we observe that the bootstrap approach for variance estimation using GM2<sub>ML</sub> and GM2<sub>REML</sub> is necessary to obtain acceptable nominal coverage values for the 95% CIs.

The covariate effects of  $x_1$  and  $x_2$  are both well estimated for the geoadditive model approaches (see Supplementary Material Figures 11-13). Similar as the predictive performance, GM2<sub>ML</sub> and GM2<sub>REML</sub> yield less variable estimates of the covariate effects.

## 5 Data Application

We apply our proposed methodology on two data examples. In the Supplementary Materials Section 3, additional tables and graphical results are provided, as well as an additional example.

### Paraná State, Brazil, Rainfall Data

In this first application, average precipitation levels throughout Paraná State, Brazil, are considered. Data is available on 143 locations and the response represents average winter (dry season) rainfall. Figure 2 (a) presents the data. No covariate information is available. This dataset is available from the R-package `geoR`. This data was also used by Diggle and Ribeiro (2002) to illustrate methodology on Bayesian inference of Gaussian model-based geostatistics. We will analyse this data using the geoadditive model which is fitted using the proposed procedure in Section 3 and compare results with results in Diggle and Ribeiro (2002).

The data will be analysed using the following geoadditive model

$$y_i = \beta_0 + \beta_1 s_{1i} + \beta_2 s_{2i} + S(\mathbf{s}_i) + \varepsilon_i, \quad 1 \leq i \leq 143, \quad (16)$$

where  $s_1$  and  $s_2$  are the latitude and longitude of the spatial location  $\mathbf{s}$ . We consider different choices for  $g_\tau$  to model the spatial component  $S$  in (16) and compare different model fits using AIC (Akaike, 1973). All data locations are used as knots in the analysis.

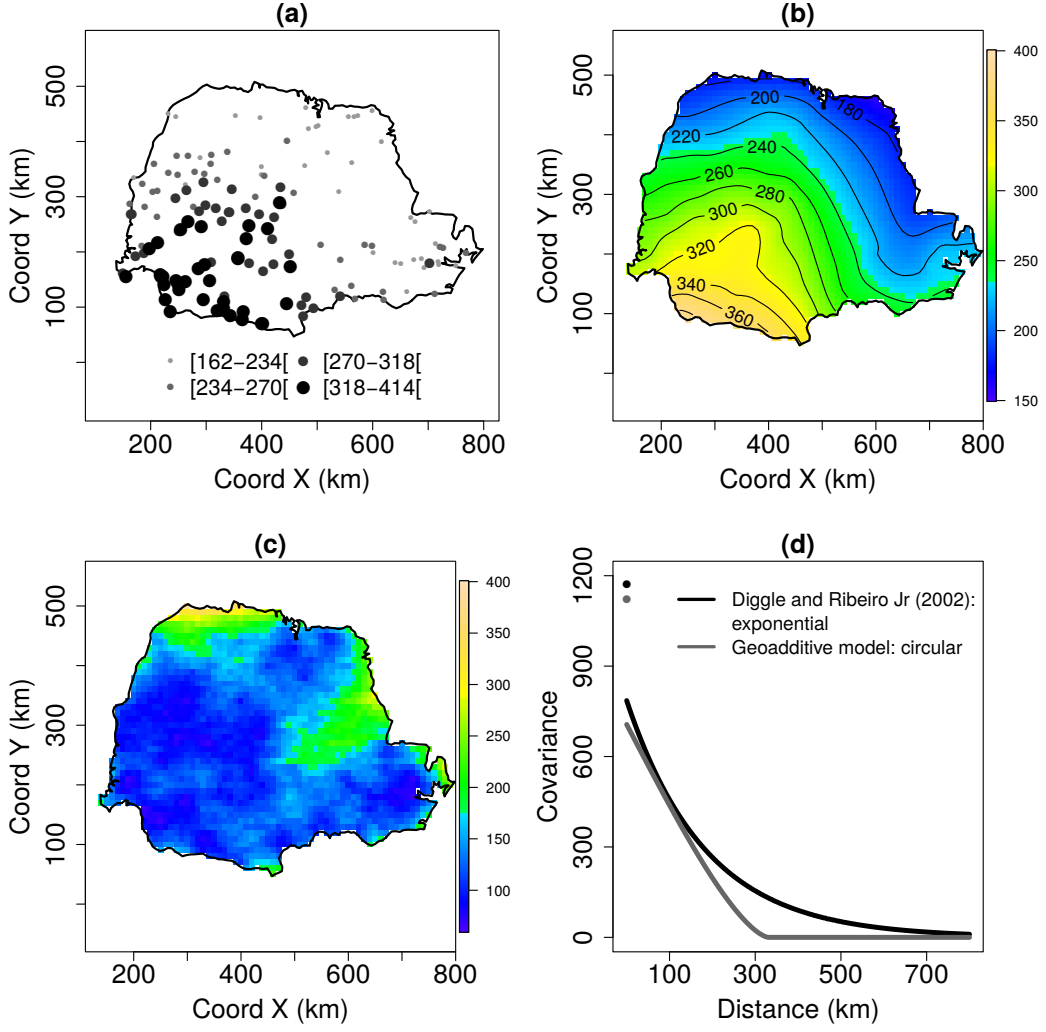


Figure 2: (a) Map of Paraná State, Brazil, showing the locations and corresponding recorded average precipitation values; (b) Predicted surface based on geoadditive model (16) fitted using the proposed estimation procedure in Section 3; (c) Prediction variance using the bootstrap approach described in Section 3; (d) Estimated Covariogram obtained by Diggle and Ribeiro (2002) and based on geoadditive model (16).

The lowest AIC value for the geoadditive model fitted using the proposed methodology is observed for a circular  $g_\tau$  function (see Supplementary Materials). The predicted surface and associated prediction variances based on the bootstrap approach for the geoadditive model are given in Figures 2 (b)-(c). The model chooses 29.9 degrees-of-freedom for the spatial component  $S$  in (16). This indicates that there is a strong non-planar spatial trend in the data. The prediction map is very similar to the one obtained by Diggle and

Ribeiro (2002) which is based on a Bayesian geostatistical analysis using an exponential covariogram. Figure 2 (d) presents the implied covariogram function of the best fitting geoadditive model, together with the estimated covariogram (based on direct maximum likelihood estimation using an exponential covariogram) obtained by Diggle and Ribeiro (2002).

We note that the geoadditive model using an exponential  $g_\tau$  function fitted using the proposed methodology yields similar results as obtained using direct maximum likelihood estimation in Diggle and Ribeiro (2002). Furthermore, for each choice of the function  $g_\tau$ , all estimated covariogram parameters using the geoadditive model fitted using the proposed methodology were similar as results obtained by direct likelihood estimation (see Supplementary Materials). This can be expected, since model (16) only includes linear covariate effects. It was also observed that geoadditive models fixing  $\tau$  at (6) have higher AIC values than the geoadditive models fitted using the proposed estimation procedure in Section 3.

### Meuse Dataset

The Meuse data is a classical geostatistical dataset used frequently to demonstrate various geostatistical analyses. The dataset comprises of four heavy metals measured in the top soil in a flood plain along the river Meuse. Polluted sediment is carried by the river, and mostly deposited close to the river bank. In total, 155 heavy metal concentration measurements (ppm) are available along with a number of soil and landscape variables. This dataset is available in the R-package `gstat`.

The goal is to create a prediction map of the lead concentrations using the distance to the river as covariate. Additionally, the landscape class variable *landuse* is used in the model. The *landuse* variable takes 15 different values and is considered as a clustering variable which is accounted for by a random intercept. One measurement has no information on the *landuse* value and is, consequently, deleted from the analysis. Thus, in total, 154 lead concentration measurements are available. In Figures 3 (a) and (b) the locations of the recorded log-lead values and the associated *landuse* variable are shown over the area of interest. The lead concentrations are log-transformed to remove skewness.

Two geoadditive models are fit to the data, one without taking into account the clus-

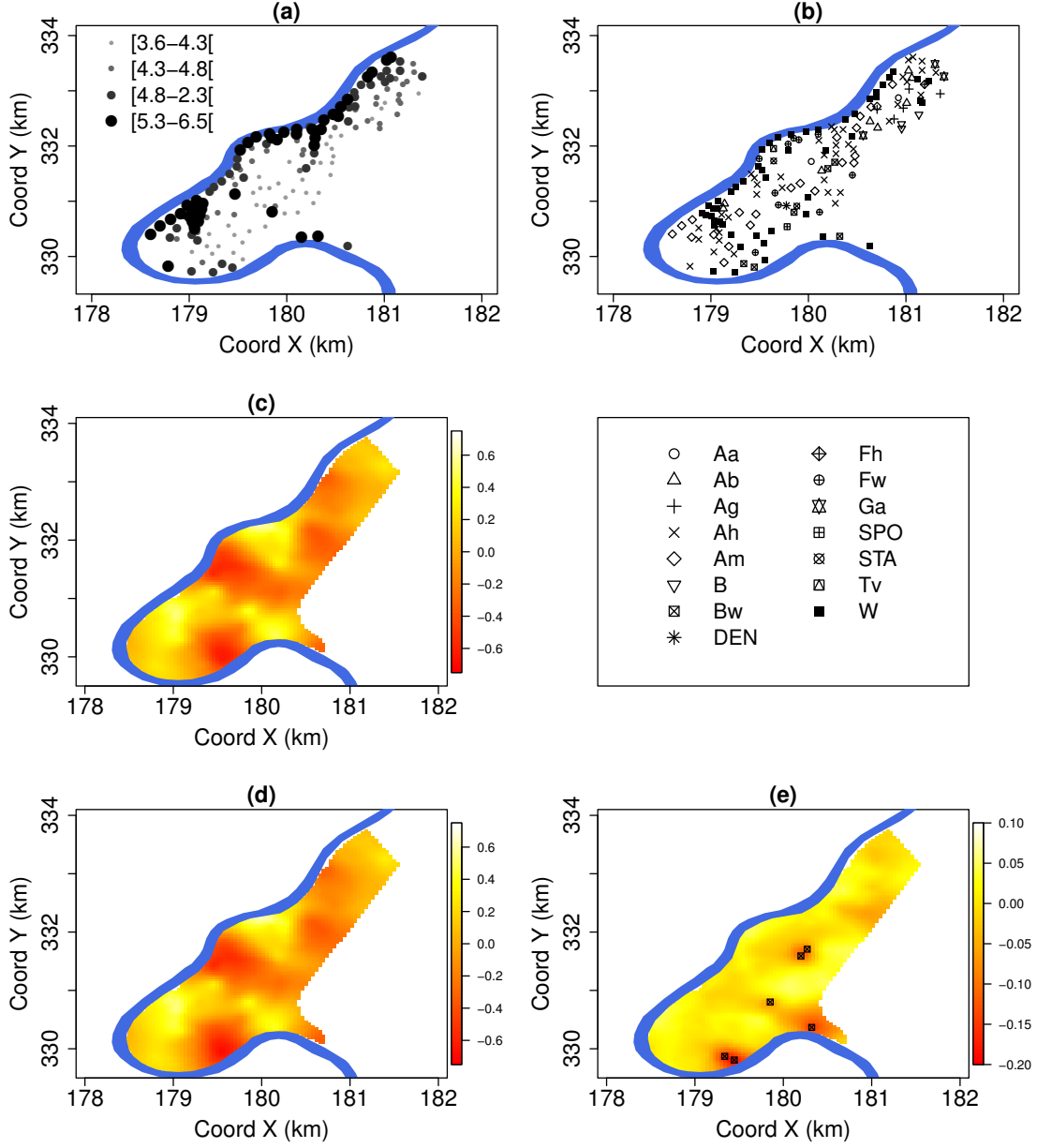


Figure 3: (a) Map showing the Meuse river and the locations and corresponding recorded log-lead values; (b) Map showing the locations of the different landuse categories; (c) Estimated spatial component using geoadditive model (17) with a circular function  $g_\tau$  fitted using the proposed estimation procedure in Section 3; (d) Estimated spatial component using geoadditive model (18) with a circular function  $g_\tau$  fitted using the proposed estimation procedure in Section 3; (e) Difference between the surface in (d) and (c) with marks at locations with landuse level 'Bw'.

tering

$$\log(\text{lead}_i) = \beta_0 + f(\text{dist}_i) + S(\mathbf{s}_i) + \varepsilon_i, \quad 1 \leq i \leq 154, \quad (17)$$

and one accounting for the clustering

$$\log(\text{lead}_i) = \beta_0 + f(\text{dist}_i) + S(\mathbf{s}_i) + \alpha_{k(i)} + \varepsilon_i, \quad 1 \leq i \leq 154, \quad (18)$$

where  $\alpha_{k(i)}$  is a random intercept for *landuse* with  $\alpha_k \sim \mathcal{N}(0, \sigma_\alpha^2)$ , for  $k = 1, \dots, 15$ . For the spatial component  $S$  we used the exponential, Matérn, circular, spherical and Gaussian as  $g_\tau$  function. The circular function yielded the best fit in terms of AIC, namely 174.6 for model (17) and 174.0 for model (18). Thus the model accounting for the landuse by a random intercept has a slightly better fit.

Using the circular function, the estimated covariogram parameters are ( $\hat{c}_s = 0.14, \hat{\tau} = 0.76$ ) and ( $\hat{c}_s = 0.13, \hat{\tau} = 0.76$ ) for model (17) and (18), respectively. The cluster variance parameter  $\sigma_\alpha^2$  in model (18) is estimated to be 0.03. Figures 3 (c) and (d) present the estimated spatial component  $S$  for models (17) and (18). Figure 3 (e) shows the difference between the estimated surfaces in Figures 3 (c) and (d). The estimated random intercept was highest, namely 0.371, at the landuse level 'Bw'. It is observed that the spatial component  $S$  estimated by model (18) has lower values in the surroundings of landuse 'Bw' (Figure 3 (e)) to compensate for the estimated value of the random intercept. Thus, from this application we observe that higher (smaller) estimated values of the spatial component  $S$  by a model that does not account for clustering could be induced by a clustering variable. A geoadditive model fitted using the proposed estimation procedure in Section 3 can account for this clustering and also estimate the spatial covariation in terms of a covariogram.

## 6 Discussion

The geoadditive model Kammann and Wand (2003) is useful to analyse geostatistical data where the data could be accompanied with (non-linear) covariates, the data could exhibit clustering and multiple measurements at a spatial location could be available. We described a likelihood based estimation procedure that estimates the spatial decay parameter of the radial basis functions describing the spatial component in the geoadditive model. The



estimation of the spatial decay parameter enables one to derive the covariogram function implied by the geoadditive model. We emphasize once more that a covariogram can only be derived in geoadditive models where the function  $g_\tau$  implies a positive semi-definite matrix  $\Omega_S$ .

Only Gaussian data were considered. In the case where the response is binary, a count variable or heavily skewed, the geoadditive should be defined within the framework of generalized linear mixed models. Future research will investigate the use and practicalities of the proposed estimation procedure in this context.

Hypothesis testing of parameters in the geoadditive model can be performed within the likelihood ratio paradigm due to the linear mixed model representation. Testing for a linear effect of  $f$  or the spatial effect of  $S$  can be done by hypotheses where the variance parameter  $\sigma_b^2$  and  $\sigma_S^2$  are set to zero in the null hypothesis, respectively. These kind of hypotheses are complicated because the variance parameter in the null hypothesis is on the boundary of the parameter space and thus asymptotic theory from Self and Liang (1987) is required. However, Crainiceanu and Ruppert (2004) show that this theory does not hold for penalized spline models and propose a simulation algorithm to derive the null distribution. This should be further investigated in the context of geoadditive models.

Towards model selection, the Akaike's information criterion (AIC) (Akaike, 1973) does not take proper account of the degrees-of-freedom in penalized spline models, therefore a modified AIC called corrected AIC ( $AIC_c$ ) could be used instead (Ruppert et al., 2003). Nevertheless, we used the standard AIC to perform model selection across the different fitted models in Section 5.

The considered functions  $g_\tau$  for the radial basis functions are only able to describe isotropic spatial processes. Many geostatistical applications exhibit, however, anisotropy. It remains a topic of future research to investigate how geoadditive models can be used to describe anisotropic processes. The use of tensor product splines could be one option within this context (Wood, 2006; Lee and Durbán, 2009, 2011). Another interesting and related future research topic is the investigation of the spatial dependency assumption. In this paper, we only considered weak stationary spatial processes. It is worthwhile to investigate how the proposed methods can be extended to other stationary processes.

No theoretical justification is given for the performance of the described likelihood based estimation procedure. We only investigated the performance of the proposed procedure using elaborate simulation studies. A theoretical justification, especially with regard to basis selection, knot selection and the number of knots, would be interesting. We also want to indicate that the convergence rate of the proposed procedure was very high ( $> 99\%$ ).

Some recently developed geostatistical approaches focus on the analysis of very large spatial datasets and attempt to overcome the computational burdens of kriging in the presence of large datasets (Cressie and Johannesson, 2008; Kaufman et al., 2008; Kleiber and Nychka, 2015). In this paper, we did not investigate the performance of our proposed estimation procedure on very large spatial datasets. However, since geoadditive models make use of penalized splines that are of low rank, this is an interesting topic to be studied.

The simulation study revealed that, in terms of prediction, the geoadditive model where the spatial decay parameter  $\tau$  is estimated using the proposed procedure in Section 3 performs better than when  $\tau$  is estimated using (6). Although both approaches yield unbiased predictions at the spatial locations, the obtained predictions using the proposed procedure exhibit less variability which, consequently, results in lower MSE values. When compared with estimation methods used for kriging purposes, the proposed methods perform similar or better for the estimation of covariogram parameters. In the Supplementary Materials we also present simulation results for data that is simulated from a spatial process with constant mean in which clustering is introduced. Overall, results are comparable with the results without clustering. In addition, we also give simulation results on the predictive performance when thin plate splines are used for the spatial component. These spline basis functions are often used to compare the performance of kriging and splines (Dubrule, 1984; Hutchinson and Gessler, 1994; Laslett, 1994; Altman, 2000).

The bootstrap approach to calculate the prediction variance described in Section 3 is needed to obtain confidence intervals with close to 95% nominal coverage. Within the bootstrap approach, we do not re-estimate the variance parameters and spatial decay parameter in step [v] but fix them at the original estimates. Although we recognize that this approximation is not without danger, it results in huge gains in computational time and does not lead to worrisome losses in the nominal coverages of the confidence intervals.

**SUPPLEMENTARY MATERIAL:** (xxxxxxx.pdf) This file contains additional information concerning the amount of smoothing in geoadditive models and plotting issues. Additional results of the simulation study in Section 4 and the applications in Section 5 are provided. Some information on the R-module to fit geoadditive models using the proposed estimation procedure is also provided.

**ACKNOWLEDGEMENTS:** Support from a doctoral grant of Hasselt University is acknowledged (BOF11D04FAEC to YV). Support from the National Institutes of Health is acknowledged [award number R01CA172805 to CF]. Support from the University of Antwerp scientific chair in Evidence-Based Vaccinology, financed in 2009-2014 by a gift from Pfizer, is acknowledged [to NH]. Support from the IAP Research Network P7/06 of the Belgian State (Belgian Science Policy) is gratefully acknowledged. For the simulation study we used the infrastructure of the VSC - Flemish Supercomputer Center, funded by the Hercules Foundation and the Flemish Government - department EWI.

## References

- Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* 60, 255–265.
- Altman, N. (2000). Krige, smooth, both or neither. *Australian and New Zealand Journal of Statistics* 42(4), 441–461.
- Banerjee, S., A. E. Gelfand, A. O. Finley, and H. Sang (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B* 70.
- Brumback, B. A., D. Ruppert, and M. P. Wand (1999). Comment on Shively, Kohn and Wood. *Journal of the American Statistical Association* 94, 794–797.
- Crainiceanu, C. M. and D. Ruppert (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society: Series B* 66, 165–185.

- Cressie, N. A. C. (1985). Fitting variogram models by weighted least squares. *Journal of the International Association for Mathematical Geology* 5, 563–586.
- Cressie, N. A. C. (1993). *Statistics for spatial data*. New York: Wiley.
- Cressie, N. A. C. and G. Johannesson (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B* 70(1), 209–226.
- Diggle, P. J. and P. Ribeiro (2002). Bayesian inference in Gaussian model-based geostatistics. *Geographical & Environmental Modelling* 6, 129–146.
- Dubrule, O. (1983). Two methods with different objectives: splines and kriging. *Journal of the International Association for Mathematical Geology* 15, 245–257.
- Dubrule, O. (1984). Comparing splines and kriging. *Computers & Geosciences* 10, 327–338.
- French, J. L., E. E. Kammann, and M. P. Wand (2001). Comment on paper by Ke and Wang. *Journal of the American Statistical Association* 96, 1285–1288.
- Gelfand, A. E., P. J. Diggle, M. Fuentes, and P. Guttorp (Eds.) (2010). *Handbook of Spatial Statistics*. Boca Raton: Chapman & Hall/CRC.
- Green, P. J. and B. W. Silverman (1994). *Nonparametric regression and generalized linear models: A roughness approach*. London: Chapman & Hall.
- Handcock, M. S., K. Meier, and D. Nychka (1994). Comment on paper by (Laslett). *Journal of the American Statistical Association* 89, 401–403.
- Hutchinson, M. F. and P. E. Gessler (1994). Splines – more than just a smooth interpolator. *Geoderma* 62, 45–67.
- Johnson, M. E., L. M. Moore, and D. Ylvisaker (1990). Minimax and maximin distance designs. *Journal of Statistical Planning and Inference* 26, 131–148.
- Kammann, E. E. and M. P. Wand (2003). Geoadditive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52(1), 1–18.

- Kaufman, C. G., M. J. Schervish, and D. W. Nychka (2008). Covariance tapering for likelihoodbased estimation in large spatial data sets. *Journal of the American Statistical Association* 103, 1545–1555.
- Kleiber, W. and D. W. Nychka (2015). Equivalent kriging. *Spatial Statistics* 12, 31–49.
- Laslett, G. M. (1994). Kriging and splines: An empirical comparison of their predictive performance in some applications (with discussion). *Journal of the American Statistical Association* 89, 392–409.
- Lee, D.-J. and M. Durbán (2009). Smooth-CAR mixed models for spatial count data. *Computational Statistics and Data Analysis* 53, 2968–2979.
- Lee, D.-J. and M. Durbán (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling* 11, 46–69.
- Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(4), 423–498.
- Mathéron, G. (1963). Principles of geostatistics. *Economic Geology* 58, 1246–1266.
- Mathéron, G. (1981). Splines and griging: their formal equivalence. Down to Earth Statistics: Solutions looking for geological problems. *Syracuse University, geology contribution* 8, 77–95.
- Nychka, D. (2000). *Spatial process estimates as smoothers*. In *Smoothing and Regression* (ed. M. Schimek). Heidelberg: Springer.
- Nychka, D., S. Bandyopadhyay, D. Hammerling, F. Lindgren, and S. Sain (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics* 24(2), 579–599.
- Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computers & Geosciences* 30, 683–691.

- R Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ribeiro Jr, P. J. and P. J. Diggle (2001). geoR: A package for geostatistical analysis. *R-NEWS* 1(2), 14–18. ISSN 1609-3631.
- Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric Regression*. Cambridge: University Press.
- Salkauskas (1982). Some relationships between surface splines and kriging. In W. Schempp and K. Zeller (Eds.), *Multivariate Approximation Theory II*, pp. 313–325. Basel: Birkhauser.
- Sang, H. and J. Z. Huang (2012). A full scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 74(1), 111–132.
- Schelin, L. and S. Sjöstedt-de Luna (2010). Kriging prediction intervals based on semiparametric bootstrap. *Mathematical Geosciences* 42(8), 985–1000.
- Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. New York: Wiley.
- Self, S. G. and K. Y. Liang (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* 82, 605–610.
- Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.
- Stein, M. L. (2014). Limitations on low rank approximations for covariance matrices of spatial data. *Spatial Statistics* 8, 1–19.
- Stein, M. L. and M. S. Handcock (1989). Some asymptotic properties of kriging when the covariance function is misspecified. *Mathematical Geology* 21(2), 171–190.
- Stewart, G. W. (1973). *Introduction to matrix computations*. London: Academic Press.

- Wahba, G. and J. Wendelberger (1980). Some new mathematical methods for variational objective analysis using splines and cross-validation. *Monthly Weather Review* 108, 1122–1145.
- Watson, G. S. (1984). Smoothing and interpolation by kriging and with splines. *Mathematical Geology* 16, 601–615.
- Wood, S. N. (2006). Low-rank scale-invariant tensor product smooth for generalized additive mixed models. *Biometrics* 62, 1025–1036.
- Zhang, H. and D. L. Zimmerman (2007). Hybrid estimation of semivariogram parameters. *Journal of the International Association for Mathematical Geology* 39, 247–260.
- Zimmerman, D. L. and M. B. Zimmerman (1991). A comparison of spatial semivariogram estimators and corresponding ordinary kriging predictions. *Technometrics* 33, 77–91.