Estimating space-time covariance functions: a composite likelihood approach*

M. BEVILACQUA[†], C. GAETAN[§], J. MATEU[‡], E. PORCU[‡] [†]

[†] Dipartimento di Scienze Statistiche, Università di Padova, Italy

[§] Dipartimento di Statistica, Università Ca' Foscari di Venezia, Italy

[‡] Department of Mathematics, Universitat Jaume I, Castellón, Spain

Abstract

In the last years there has been a growing interest in the construction space-time covariance functions. However, effective estimation methods for these models are somehow unexplored. In this paper we propose a composite likelihood approach and a weighted variant for the space-time estimation problem.

The proposed method can be a valid compromise between the computational burdens, induced by the use of a maximum likelihood approach, and the loss of efficiency induced by using a weighted least squares procedure. An identification criterion based on the composite likelihood is also introduced. The effectiveness of the proposed procedure is illustrated through an extensive simulation experiment, and by reanalising a data set on Irish wind speeds (Haslett and Raftery, 1989). We also address an important issue, which has been recently explored in the literature, on how to select an appropriate space-time model by accounting for the tradeoff between goodness-of-fit and model complexity.

Keywords: Asymmetry in time, Composite likelihood, Full symmetry, Irish wind speed data, Separability, Space-time geostatistics.

* Work partially funded by grant MTM2004-06231 from the Spanish Ministry of Science and Education.
[†]Corresponding author. *Email: porcu@mat.uji.es.* Fax: +34.964.728429

2000 Mathematics Subject Classification: 60G55, 60K35.

1. INTRODUCTION

In the last years there has been a great demand for models describing the evolution of environmental or geophysical processes in space and time. In particular, there is a considerable need for models able to catch the simultaneous behaviour of the spatial and temporal components. If we considered them separately, a large amount of information would be probably lost. The geostatistical approach models such dependence through data coming from a limited number of monitoring stations observed over time. These observations are considered as a partial realisation of a spatio-temporal random field (RF) defined on a continuum space per time. A typical assumption is that the field is Gaussian, so the modelling task entails to the specification and estimation of space-time covariance functions.

Literature on space-time covariance models has been continuously increasing in the last ten years. By starting from different settings and mathematical instruments, several authors have produced considerable efforts in order to build nonseparable covariance functions. Among many others, Jones and Zhang (1997), Cressie and Huang (1999), Christakos (2000), De Cesare et al. (2001), Gneiting (2002) and Ma (2002), should be cited. Most of these contributions deal with stationary fully symmetric space-time covariance functions, *e.g.* isotropic in the spatial component and symmetric in the temporal one. Recently, new covariance models have been proposed to face more realistic situations, such as anisotropy in space or asymmetry in time (Stein, 2005*a*; Porcu et al., 2006).

However, space-time estimation is not so thoroughly developed in the literature. Maximum likelihood (ML) and related techniques are generally considered the best method for estimating the parameters of space-time covariance models. For Gaussian RF with a given parametric covariance function, exact computation of the likelihood requires calculation of the inverse and determinant of the covariance matrix, and this evaluation is slow when the number of observations is large. This fact motivates the search for approximations to the likelihood function that require smaller computational burdens. For instance, if we assume that the RF is stationary in time and data are collected at regular time intervals on a fixed monitoring network, Stein (2005*b*) shows, by extending the conditional approach proposed by Vecchia (1988), that there are considerable computational gains in exploiting the natural order induced by time.

Due to its ease of implementation, Weighted Least Squares (WLS) (Cressie, 1985) is currently the most used estimation method (see, for instance, Gneiting et al. (2007)). In general, WLS method requires a preliminary step, *e.g.* obtaining a nonparametric estimate of the variogram (the method of moments is the most popular). Thus, a heteroschedastic non-linear regression model is estimated, using the nonparametric estimates as response variable. The last step requires computation of the correlation between the empirical variogram estimates, and this operation needs caution, as emphasised by several authors (Genton, 1998; Pardo-Iguzquiza and Dowd, 2001).

Another problem, often overcome by the literature, is the bias of the empirical variogram when binning (typically for irregular grids) and its consequence on the WLS estimates. Moreover, estimation of not fully symmetric models, as in Stein (2005*a*), or of anisotropic (or nonstationary) in space and symmetric in time covariance functions, as in Porcu et al. (2006), needs adapted versions of the method of moments variogram estimator.

Due to the aspects that have been stressed previously, the literature encourages the proposal for new methodologies devoted to the approximation of the likelihood function for a space-time data set.

In this paper, we suggest the use of composite likelihood approaches (Lindsay, 1988) to estimate space-time dependence. Composite likelihood (CL for short) is a likelihood-based approximation indicating a general class of pseudo-likelihoods based on the likelihood of marginal or conditional events. In a pure spatial setting, Curriero and Lele (1999) introduced CL for semivariogram estimation and showed that this method shares some of the best properties of the WLS and ML methods. In particular, it does not depend on the choice of lag bins and it is robust towards miss-specification of the distribution.

The use of CL methods in the spatio-temporal context has never been proposed by the literature and this fact motivates our research.

We show, through an extensive simulation study, that our method induces considerable gains in efficiency with respect to WLS methods. These gains can be justified using well-known results about the estimating equations. We also address an important issue, which has been recently explored in the literature, on how to select an appropriate space-time model by taking into account the tradeoff between goodness-of-fit and model complexity (Huang et al., 2007). In this paper the authors used the Akaike Information Criterion and the Bayesian Information Criterion which are based on the likelihood function. Here, we propose an information criteria for model selection based on composite likelihood, exploiting recent results in Varin and Vidoni (2005).

The plan of the paper is the following. Space-time covariance models and their estimation methods are briefly reviewed in Sections 2 and 3. In Section 4 we introduce the CL and WCL methods for space-time data and the adopted model selection criteria. In Section 5 we present a simulation study and in Section 6 we re-analyse the Irish-wind speed dataset (Haslett and Raftery, 1989). Finally Section 7 summarises our findings.

2. SPACE-TIME COVARIANCE MODELS

Let $\{Z(\boldsymbol{s},t), \boldsymbol{s} \in \mathbb{R}^d, t \in \mathbb{R}\}$ be a real-valued Gaussian space-time RF and assume that $\mathbb{V}\text{ar}[Z(\boldsymbol{s},t)] < \infty$, for all $\boldsymbol{s} \in \mathbb{R}^d$, $t \in \mathbb{R}$. The very high order of complexity of space-time interactions calls for simplifying assumptions, such as those of intrinsic or weak stationarity, that have implications on the existence of the moments of the RF.

Under the assumption of intrinsic stationarity, we have that $\mathbb{E}[Z(\boldsymbol{s},t)] = \mu$ and $\mathbb{V}\text{ar}(Z(\boldsymbol{s},t) - Z(\boldsymbol{s}',t')) = 2\gamma(\boldsymbol{h},u)$, where $\boldsymbol{h} = \boldsymbol{s} - \boldsymbol{s}' \in \mathbb{R}^d$ and $u = t - t' \in \mathbb{R}$ are respectively the spatial and temporal separation vectors. The mapping $2\gamma(\boldsymbol{h},u) : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$ is called the space-time variogram associated to the intrinsically stationary RF.

A RF is weakly stationary if $\mathbb{E}[Z(\boldsymbol{s},t)] = \mu$, for all $\boldsymbol{s}$, t and the covariance between $Z(\boldsymbol{s},t)$ and $Z(\boldsymbol{s}',t')$ depends only on the spatial and temporal separation vectors, namely $\mathbb{C}\text{ov}(Z(\boldsymbol{s},t), Z(\boldsymbol{s}',t')) = \mathrm{C}(\boldsymbol{h},u)$. It is easy to see that weak stationarity implies intrinsic stationarity, whilst the converse is in general not true. A covariance function is called isotropic if it depends on the separation vectors only through their lengths $\|\boldsymbol{h}\|$ and $|u|$, *i.e.* $\mathrm{C}(\boldsymbol{h},u) := \tilde{\mathrm{C}}(\|\boldsymbol{h}\|, |u|)$, for $\tilde{\mathrm{C}}$ positive definite.

Covariance functions which are not isotropic are called anisotropic. In the literature the assumption of isotropy is very popular because of its simplicity and interpretability.

For weakly or intrinsically stationary RF's, other simplifying assumptions are those of *separability* and *full symmetry*. A separable model is obtained through the tensorial product between a merely spatial covariance function and a temporal one. Following Mitchell et al. (2005), we have

$$\mathrm{C}(\boldsymbol{h}, u) = \frac{\mathrm{C}(\boldsymbol{h}, 0)\mathrm{C}(\mathbf{0}, u)}{\mathrm{C}(\mathbf{0}, 0)}, \quad \forall (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}.$$

A simple example of separable model is the doubly exponential model, namely

$$C(\boldsymbol{h}, u) = \sigma^2 \exp\left(-c\|\boldsymbol{h}\| - a|u|\right), \tag{1}$$

where c, a are positive scale parameters and σ^2 is the variance of the RF.

Separability gives important computational benefits and consequently separable models have been used even in situations in which they are not physically justifiable. Therefore many reseachers have recently proposed nonseparable covariance classes for spatial-temporal processes. For instance, Gneiting (2002) proposed the following model

$$C(\boldsymbol{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{\beta d/2}} \exp\left(-\frac{c\|\boldsymbol{h}\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta\gamma}}\right), \tag{2}$$

where a, c are positive scale parameters, $\alpha, \gamma \in [0, 1]$ are respectively temporal and spatial smoothing parameters, $\beta \in [0, 1]$ is a space-time interaction parameter, and σ^2 is the variance of the RF.

The product with the purely temporal covariance $(a|u|^{2\alpha}+1)^{-1}$, fixing $d = 2$ and through a convenient reparametrisation, gives an interesting variation of the previous model

$$C(\boldsymbol{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)} \exp\left(-\frac{c\|\boldsymbol{h}\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta\gamma}}\right). \tag{3}$$

Note that if $\beta = 0$ we have a separable model.

The models defined by equations (1-2) are *fully symmetric*, *e.g.* they are rotationally invariant (or isotropic) in the spatial component and symmetric in the temporal one, *i.e.* $\mathrm{C}(\boldsymbol{h}, u) = \tilde{\mathrm{C}}(\|\boldsymbol{h}\|, |u|)$. This implies the obvious identity,

$$\mathrm{C}(\boldsymbol{h}, u) = \mathrm{C}(-\boldsymbol{h}, u) = \mathrm{C}(\boldsymbol{h}, -u) = \mathrm{C}(-\boldsymbol{h}, -u),$$

for $(\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R}$, as noticed by Gneiting (2002).

The assumption of full symmetry is a useful restriction for constructing positive definite functions, but the real world offers a number of situations in which this assumption is frequently violated. A substantive improvement of the previous construction can be achieved by working in two different directions: on the one hand, allowing anisotropy in space, and on the other one, building models that are asymmetric in time. In this direction, new covariance functions can be obtained by working in the Lagrangian framework (Cox and Isham, 1988).

Roughly speaking, the covariance is obtained as the expected value of a stationary covariance function C_S on $\mathbb{R}^d$

$$C(\boldsymbol{h}, u) = \mathbb{E}_{\mathbf{V}}(C_S(\boldsymbol{h} - \mathbf{V}u)),$$

where the random vector $\boldsymbol{V}$ incorporates some physical knowledge through a probabilistic distribution and the expectation is taken with respect to its distribution. For instance, $\boldsymbol{V}$ could represent the velocity vector in $\mathbb{R}^3$ and could be uniformly distributed on the unit sphere. A Lagrangian version (Stein, 2005*a*) of the model (3) is

$$C(\boldsymbol{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha}+1)} \frac{\exp(-c\|\boldsymbol{h} - \varepsilon u \boldsymbol{v}\|)}{(a|u|^{2\alpha}+1)^{\beta/2}}, \tag{4}$$

where a, c are positive scale parameters, $\alpha \in [0, 1]$ is a smoothing parameter, $\beta \in [0, 1]$ is a space-time interaction parameter, $\varepsilon \in \mathbb{R}$ controls the degree of asymmetry, σ^2 is the variance, and $\boldsymbol{V}$ is a degenerated distribution with mass on $\boldsymbol{v}$. If $\boldsymbol{v}$ is a vector of the canonical basis for $\mathbb{R}^d$, the choice introduces an asymmetry only in one direction. Successful application for modelling prevailing winds can be found in the previously mentioned papers.

3. ESTIMATION METHODS

In this section we review those estimation methods that are used throughout this paper. This section is largely expository and for a more complete review, we refer the reader to the textbooks of Cressie (1993) and Stein (1999).

The popularity of Least Squares methods (LS for short) is due to their ease of computation and the fact that they are free of distributional assumptions. Let $\boldsymbol{Z} = (Z(\boldsymbol{s}_1, t_1), \dots Z(\boldsymbol{s}_N, t_N))'$ be N observations from a intrinsically stationary space-time RF.

A nonparametric estimate of the semi-variogram is typically obtained through the method of moments (Matheron, 1965). Assuming that points are disposed on a regular lattice, the estimate is

$$\hat{\gamma}(\boldsymbol{h}, u) = \frac{1}{2|\mathrm{N}(\boldsymbol{h}, u)|} \sum_{(i,j)\in \mathrm{N}(\boldsymbol{h},u)} (Z(\boldsymbol{s}_i, t_i) - Z(\boldsymbol{s}_j, t_j))^2,$$

where $\mathrm{N}(\boldsymbol{h}, u) = \{(i, j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h};\, t_i - t_j = u\}$.

Then, the following generalised sum of squares is minimised

$$\left(\widehat{\boldsymbol{\gamma}}(\boldsymbol{h}, \boldsymbol{u}) - \gamma(\boldsymbol{h}, \boldsymbol{u}; \boldsymbol{\theta})\right)' \boldsymbol{R}(\boldsymbol{h}, \boldsymbol{u}; \boldsymbol{\theta})^{-1} \left(\widehat{\boldsymbol{\gamma}}(\boldsymbol{h}, \boldsymbol{u}) - \gamma(\boldsymbol{h}, \boldsymbol{u}; \boldsymbol{\theta})\right),$$

where $\widehat{\boldsymbol{\gamma}}(\boldsymbol{h}, \boldsymbol{u}) = (\hat{\gamma}(\boldsymbol{h}_1, u_1), \ldots, \hat{\gamma}(\boldsymbol{h}_m, u_m))'$, $\boldsymbol{\gamma}(\boldsymbol{h}, \boldsymbol{u}; \boldsymbol{\theta}) = (\gamma(\boldsymbol{h}_1, u_1; \boldsymbol{\theta}), \ldots, \gamma(\boldsymbol{h}_m, u_m; \boldsymbol{\theta}))'$ and $\boldsymbol{R}$ is a definite positive matrix. A natural choice for $\boldsymbol{R}$ is $\mathrm{var}(\widehat{\boldsymbol{\gamma}}(\boldsymbol{h}, \boldsymbol{u}))$, obtained for instance through the Pardo-Iguzquiza and Dowd (2001) approximation. Cressie (1985) suggested an approximation of $\mathrm{var}(\widehat{\boldsymbol{\gamma}}(\boldsymbol{h}, \boldsymbol{u}))$ assuming a Gaussian RF and uncorrelation among $\hat{\gamma}(\boldsymbol{h}_j, u_j)$. The matrix $\boldsymbol{R}$ is replaced by a diagonal matrix where the diagonal entries are calculated as

$$\mathrm{var}(\hat{\gamma}(\boldsymbol{h}_k, u_k)) \approx 2\frac{\gamma(\boldsymbol{h}_k, u_k; \boldsymbol{\theta})^2}{|N(\boldsymbol{h}, u_k)|}, \qquad k = 1, \ldots, m.$$

Thus, instead of minimising the generalised sum of squares, Cressie (1985) proposed to minimise the weighted sum of squares

$$\sum_{k=1}^{m} \frac{|N(\boldsymbol{h}_k, u_k)|}{\gamma(\boldsymbol{h}_k, u_k; \boldsymbol{\theta})^2} \left(\hat{\gamma}(\boldsymbol{h}_k, u_k) - \gamma(\boldsymbol{h}_k, u_k; \boldsymbol{\theta})\right)^2.$$

Barry et al. (1997) show that minimising this function corresponds to solving an estimating equation whose bias is of order $N(\boldsymbol{h}_k, u_k)^{-1}$. The WLS estimate has evident computational gains but it depends on the $\hat{\boldsymbol{\gamma}}(\boldsymbol{h}_k, u_k)$, that are correlated.

Whenever the data are not regularly spaced, a tolerance region is needed in order to get reasonable estimates, otherwise the number of points lying on regular separations would be too small. If the data are regularly collected over time, but not over space, an adapted version of $\mathrm{N}(\boldsymbol{h}, u)$ could be $\mathrm{N}_\Delta(\boldsymbol{h}, u) = \{(i, j) : \boldsymbol{s}_i - \boldsymbol{s}_j \in Tol(\boldsymbol{h}_\Delta);\, t_i - t_j = u\}$, where $Tol(\boldsymbol{h}_\Delta)$ is some specified tolerance region Δ around $\boldsymbol{h}$. Anyway, any tolerance region introduces a subjective choice in the estimation, and a nonnegligible bias that strongly depends on the choice of the lag bins, $\mathrm{N}_\Delta(\boldsymbol{h}_k, u_k)$, that are chosen arbitrarily or on the base of heuristic considerations. This drawback is even bigger when we want to estimate anisotropic parametric models (Porcu et al., 2006). Moreover, in the case of space-time covariance functions built throughout the Lagrangian framework (Stein, 2005a), it is not clear how to proceed with WLS estimation. Computation of the bias and its impact on WLS estimates has been overcome by the literature. In a pure spatial setting, Muller (1999) noted that a subjective choice of lag bins could have critical impact on the parameter estimates. In this sense, he suggested that fitting based on the variogram cloud would avoid this drawback.

ML methods require the distribution of the underlying RF to be known and in literature it is only developed the case of Gaussian RF (Mardia and Marshall, 1984; Pardo-Iguzquiza, 1998). Let $\boldsymbol{Z} = (Z(\boldsymbol{s}_1, t_1), \dots Z(\boldsymbol{s}_N, t_N))'$ be N observations from a zero-mean Gaussian RF. The log-likelihood function can be written as

$$l(\boldsymbol{\theta}) = -\log|\Sigma(\boldsymbol{\theta})| - \mathbf{Z}'\Sigma(\boldsymbol{\theta})^{-1}\boldsymbol{Z},$$

where $\Sigma(\boldsymbol{\theta}) = \mathbb{V}\mathrm{ar}(\boldsymbol{Z})$ depends on $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$. The ML estimator in a spatial setting has desirable asymptotic properties like consistency and asymptotic normality under increasing domain asymptotics (Mardia and Marshall, 1984). However applications of ML methods to spatial data have been criticised for computational reasons. A potential drawbak is the multimodality of the likelihood surface (Warnes and Ripley, 1987). Moreover, the most critical part in likelihood calculation is to evaluate the determinant and inverse of Σ. This evaluation could be theoretically carried out in $O(N^{2.81})$ steps (Aho et al., 1974) but the most widely used algorithms such as Cholesky decomposition require $O(N^3)$ steps. This can be prohibitive if N is large. This motivated Vecchia (1988); Stein et al. (2004); Stein (2005b); Caragea and Smith (2005); Fuentes (2007) to look for approximations to the likelihood function that require less than $O(N^3)$ steps to evaluate, but that still have reasonable statistical properties.

4. COMPOSITE LIKELIHOOD METHODS

Let $\{f(\boldsymbol{Z}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}$ be a parametric family of joint densities for the observations $\boldsymbol{Z} \in D \subseteq \mathbb{R}^N$ and consider a set of events $\{A_i : A_i \subseteq \mathfrak{F}, i \in I \subseteq \mathbb{N}\}$, where $\mathfrak{F}$ is a σ-algebra on D. The logarithm of the composite likelihood (CL) (Lindsay, 1988) is defined as

$$CL(\boldsymbol{\theta}) = \sum_{i \in I} \log f(\boldsymbol{Z} \in A_i, \boldsymbol{\theta}).$$

The pioneering example in spatial statistics was the Besag pseudo-likelihood function for lattice data (Besag, 1974), and we can find applications of CL in geostatistics (Curriero and Lele, 1999), point processes (Guan, 2007), binary spatial data (Heagerty and Lele, 1998) and image models (Nott and Ryden, 1999) amongst many others.

Let $Z(\boldsymbol{s},t)$ be an intrinsically stationary RF with variogram $2\gamma(\cdot,\cdot;\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is an unknown $p-$dimensional parameter vector. We follow Curriero and Lele (1999) and consider all possible differences $U(\boldsymbol{s},\boldsymbol{s}',t,t')=Z(\boldsymbol{s},t)-Z(\boldsymbol{s}',t')$, $\boldsymbol{s}\neq\boldsymbol{s}'$, $t\neq t'$. We assume that the distribution of these differences is Gaussian, *i.e.*

$$U(\boldsymbol{s},\boldsymbol{s}',t,t')\sim N(0,2\gamma(\boldsymbol{s}-\boldsymbol{s}',t-t';\boldsymbol{\theta})), \tag{5}$$

that gives the negative log-likelihood

$$cl(\boldsymbol{s},\boldsymbol{s}',t,t';\boldsymbol{\theta})=\log\gamma(\boldsymbol{s}-\boldsymbol{s}',t-t';\boldsymbol{\theta})+\frac{1}{2}\frac{U^2(\boldsymbol{s},\boldsymbol{s}',t,t')}{\gamma(\boldsymbol{s}-\boldsymbol{s}',t-t';\boldsymbol{\theta})}.$$

Summing up all the possible differences, we get the composite likelihood function

$$\mathrm{CL}(\boldsymbol{\theta})=\sum_{i=1}^{N}\sum_{j>i}^{N}cl(\boldsymbol{s}_i,\boldsymbol{s}_j,t_i,t_j;\boldsymbol{\theta}). \tag{6}$$

Note that the method does not request any inversion matrix and the order of computations is $O(N^2)$. Computational gains can be further achieved if we consider a weigthed version of composite likelihood in equation (6), namely

$$\mathrm{WCL}(\boldsymbol{\theta})=\sum_{i=1}^{N}\sum_{j>i}^{N}cl(\boldsymbol{s}_i,\boldsymbol{s}_j,t_i,t_j;\boldsymbol{\theta})w_{i,j}, \tag{7}$$

where $w_{i,j}=1$ when $\|s_i-s_j\|\leq h$ and $|t_i-t_j|\leq u$, and 0 otherwise, for specified positive values h and u defining the bandwidth of a space-time neighborhood. Note that the weights should be adapted for dealing with geometric anisotropies. Estimates $\hat{\boldsymbol{\theta}}_{\mathrm{WCL}}$ are obtained by minimising $\mathrm{WCL}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. For this, we need to introduce some minimal notation and write $\nabla\gamma(\cdot,\cdot;\boldsymbol{\theta}):=\frac{\delta^p}{\delta\theta_1,\cdots,\delta\theta_p}\gamma(\cdot,\cdot;\theta_1,\cdots,\theta_p)$.

The associated estimating function

$$\mathbf{s}(\boldsymbol{\theta})=\sum_{i=1}^{N}\sum_{j>i}^{N}\frac{\nabla\gamma(\boldsymbol{s}_i-\boldsymbol{s}_j,t_i-t_j;\boldsymbol{\theta})}{\gamma(\boldsymbol{s}_i-\boldsymbol{s}_j,t_i-t_j;\boldsymbol{\theta})}\left(1-\frac{U^2(\boldsymbol{s}_i,\boldsymbol{s}_j,t_i,t_j)}{2\gamma(\boldsymbol{s}_i-\boldsymbol{s}_j,t_i-t_j;\boldsymbol{\theta})}\right)w_{i,j}, \tag{8}$$

can be readily verified to be zero-unbiased, irrespectively of the distributional assumptions imposed on $U(\boldsymbol{s}_i, \boldsymbol{s}_j, t_i, t_j)$. In Section 5 we show that (7) gives more efficient estimates than (6). This can be also illustrated by considering a simple example in a spatial framework, imposing that the spatial dependence can be catched by using an exponential variogram, having equation $\gamma(\boldsymbol{h};\theta) = 1 - \exp(-3\|\boldsymbol{h}\|/\theta)$, $\theta > 0$. Thus, in this case θ is a positive scalar representing the range of the spatial correlation, *e.g.* the minimum lag at which the spatial correlation of the underlying RF is negligible.

Assuming a spatial Gaussian RF, the inverse of the Godambe information (Godambe, 1991) associated to (8), using the relation

$$\mathbb{C}ov(U(\boldsymbol{s}_i, \boldsymbol{s}_j, t_i, t_j)^2, U(\boldsymbol{s}_l, \boldsymbol{s}_k, t_l, t_k)^2) = 2(\gamma_{il} - \gamma_{jl} - \gamma_{jk} + \gamma_{ik})^2,$$

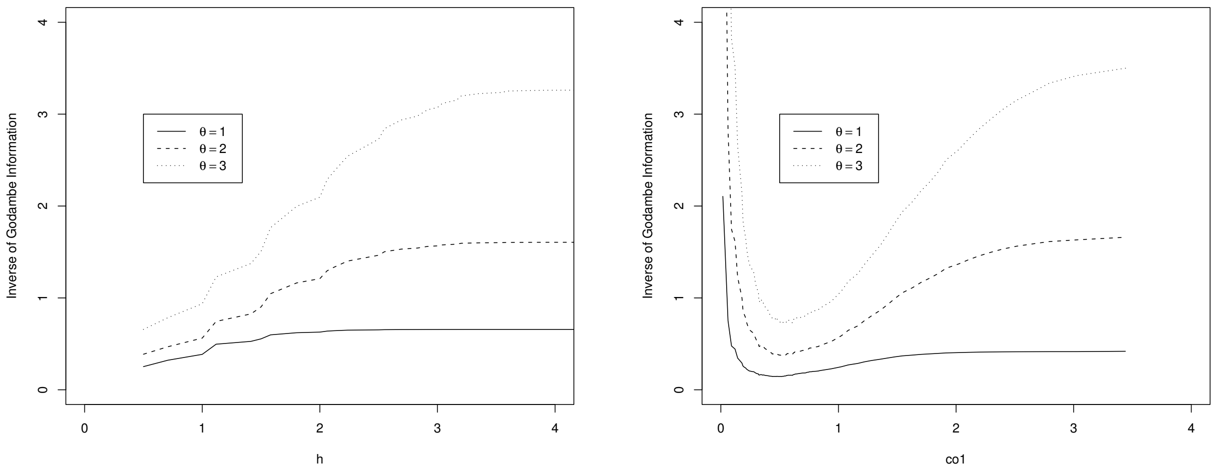
where $\gamma_{ij} = \gamma(||\boldsymbol{s}_i - \boldsymbol{s}_j||, \theta)$, reduces to

$$G(\theta) = \frac{\mathbb{E}[s(\theta)^2]}{[\mathbb{E}\nabla s(\theta)]^2} = \frac{\sum_i \sum_{j>i} \sum_l \sum_{k>l} \frac{\nabla\gamma_{ij}\nabla\gamma_{lk}}{\gamma_{ij}^2\gamma_{lk}^2}(\gamma_{il} - \gamma_{jl} - \gamma_{jk} + \gamma_{ik})^2 w_{i,j} w_{l,k}}{2\left[\sum_i \sum_{j>i} \left(\frac{\nabla\gamma_{ij}}{\gamma_{ij}}\right)^2 w_{i,j}\right]^2}. \tag{9}$$

It can be appreciated that, for this example, the inverse of the Godambe information only implies smoothness assumptions of the first order on the associated variogram, and a very broad class of variograms satisfies this assumption.

Figure 1-a shows $G(\theta)$ as a function of h on a 7×7 regular grid in a square $[1, 4]^2$ for different values of θ. Pairs of locations separated for distances larger than h are ignored. We can see that the efficiency increases with the spatial lag h, and attains the minimum at the first lag available ($h = 0.5$). The gain is more evident when the spatial dependence, θ, increases. Figure 1-b considers the same setting but under an irregular grid. Note that in this spatial scheme, it seems dangerous to take h too small, and that the optimal h is proportional to the spatial dependence, as could be expected.

An important issue in the geostatistical approach is the selection of an appropriate covariance model, taking into account the tradeoff between goodness-of-fit and model complexity. Model selection criteria as AIC and BIC have been proposed in a spatial (Hoeting et al., 2006) and space-time (Huang et al., 2007) setting. Nevertheless, they depend on the computation of the likelihood function. For the framework proposed in this paper, we can follow Varin



(a) (b)

Figure 1: Inverse of Godambe information matrix for an exponential variogram model, $\gamma(\boldsymbol{h};\theta) = 1 - \exp(-3\|\boldsymbol{h}\|/\theta)$. We considered 49 spatial locations on $[1,4]^2$ and $\theta = 1,2,3$; (a) a regular 7×7 grid, (b) 49 sites randomly located.

and Vidoni (2005), who introduce a model selection criterion based on the CL estimating function.

More precisely, the CL information criterion selects the model maximising

$$CLIC(\hat{\boldsymbol{\theta}}_{\mathrm{WCL}}) = WCL(\hat{\boldsymbol{\theta}}_{\mathrm{WCL}}) + \mathrm{tr}(\hat{J}\hat{H}^{-1}), \tag{10}$$

where $\hat{J}$ and $\hat{H}$ are consistent and first-order unbiased estimators for $J(\boldsymbol{\theta}) = \mathbb{E}(\mathbf{s}(\boldsymbol{\theta})\mathbf{s}(\boldsymbol{\theta})')$ and $H(\boldsymbol{\theta}) = \mathbb{E}(\nabla \mathbf{s}(\boldsymbol{\theta}))$, respectively.

These matrices could be estimated by

$$\hat{H} = \sum_i \sum_{j>i} \frac{\nabla\hat{\gamma}_{ij}\nabla\hat{\gamma}'_{ij}}{\hat{\gamma}^2_{ij}} w_{i,j}$$

and

$$\hat{J} = \frac{1}{2}\sum_i \sum_{j>i}\sum_l \sum_{k>l} \frac{\nabla\hat{\gamma}_{ij}\nabla\hat{\gamma}'_{lk}}{\hat{\gamma}^2_{ij}\hat{\gamma}^2_{lk}}(\hat{\gamma}_{il} - \hat{\gamma}_{jl} - \hat{\gamma}_{jk} + \hat{\gamma}_{ik})^2 w_{i,j}w_{l,k},$$

where $\hat{\gamma}_{ij} = \gamma(\boldsymbol{s}_i - \boldsymbol{s}_j, t_i - t_j; \hat{\boldsymbol{\theta}}_{WCL})$ but evaluation of $\hat{J}$ is computationally expensive for large datasets.

Assume that the RF is stationary in time and data are collected at regular time intervals, $t = 1, \ldots, T$, on a fixed monitoring network of n sites and define $\boldsymbol{Z}_t = (Z(\boldsymbol{s}_1, t), \ldots, Z(\boldsymbol{s}_n, t))'$. In this case we can use a sub-sampling method (Heagerty and Lele, 1998; Heagerty and Lumley, 2000). The idea of sub-sampling is to consider overlapping sub-sample $\boldsymbol{Z}_i^{(m)} = (\boldsymbol{Z}_i', \ldots, \boldsymbol{Z}_{i+m-1}')'$, $i = 1, \ldots, T - m + 1$. In our case we can estimate J by

$$\hat{J} = \frac{1}{(nT - mn + 1)} \sum_{i=1}^{nT-mn+1} \frac{m}{T} \boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{WCL}) \boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{WCL})',$$

where $\boldsymbol{s}_i(\hat{\boldsymbol{\theta}}_{WCL})$ is calculated using the sub-samples $\boldsymbol{Z}_i^{(m)}$. As far as the choice of the temporal window size m, we refer the reader to section 2.6 of Heagerty and Lumley (2000).

Finally, we aim to remark that the definition of (10) relies on asymptotic properties of the composite likelihood estimator, that are not fully developed for the space-time setting. In this case, as data are regularly spaced on a lattice, some results can be obtained following the approach in the appendix of Heagerty and Lele (1998), even if with tedious calculus.

5. SIMULATED DATA

In this section we describe a simulation study with the aim of examining and comparing the performances of CL and WCL estimates with respect to WLS and ML ones. Simulations were performed including the separable and nonseparable spatio-temporal covariance models, respectively defined in equations (1) and (2).

The methods are compared in terms of (relative) mean squared errors (MSE), calculated over 150 RF realisations, $Z(\boldsymbol{s}, t)$, $\boldsymbol{s} \in \{1, \ldots, N\}^2$, $t \in \{1, \ldots, T\}$, with $N = 4, 7, 10$ and $T = 15, 30, 45$. The number of the observations was chosen in order to keep feasible the computation of the full likelihood. We considered two different models:

1. The model in equation (1), that is a separable exponential model, with $c = 2$, $a = 0.8$, $\sigma^2 = 1.7$.

2. The model in equation (2), that is a nonseparable and fully symmetric model, with $c = 2$, $a = 0.8$, $\sigma^2 = 1.7$ (smoothing parameter equal to 0.5).

Tables 1 and 2 show the relative efficiency of the WLS and CL with respect to ML. CL and WLS present almost the same efficiency, with a slight advantage for CL especially for

		$N=4$		$N=7$		$N=10$	
		WLS	CL	WLS	CL	WLS	CL
$T=15$	c	3.08	2.91	4.74	4.70	7.43	7.44
	a	1.36	1.27	1.82	1.80	2.10	2.10
	σ^2	1.04	1.03	1.01	1.02	1.04	1.03
$T=30$	c	3.64	3.55	5.32	5.23	8.34	8.33
	a	1.36	1.34	2.10	2.08	2.91	2.94
	σ^2	1.03	1.02	1.03	1.01	1.02	1.02
$T=45$	c	5.19	5.02	5.44	5.39	9.01	9.02
	a	1.40	1.12	2.14	2.14	3.10	3.11
	σ^2	1.05	1.03	1.04	1.04	1.03	1.02

Table 1: Relative efficiency, based on MSE, for WLS and CL estimation methods with respect to ML, when model (1) is used with $c = 2, a = 0.8, \sigma^2 = 1.7$.

small sample sizes. We also found that this advantage is even stronger for irregular grids since the binning introduces a bias in the empirical variogram, as expected. As the choice of the grids is somewhat arbitrary, we do not report the results.

The differences among the estimates seem to be clearer for the scale parameter, whereas the three methods behave very similarly in terms of variance estimates.

Now we consider the weighted version of CL as in equation (7). In Table 3 we report the MSE relative efficiencies for WCL for model (2) observed on a regular grid with $N = 4$, $T = 45$ for several combinations of spatial and temporal lags, respectively h and u lags. Similar results (not reported here) have been obtained for different sample sizes. The remarkable fact is that a careful lag selection leads to a significative improvement. We obtained similar results for models (1) and (2) when fixing the scale parameters and estimating the smoothness ones.

Finally, we show a small example of model identification using the CL information criterion based on WCL. We considered 100 independent simulations from the Stein model in equation (4) with $N = 4$ and $T = 150$ and three different settings:

		$N=4$		$N=7$		$N=10$	
		WLS	CL	WLS	CL	WLS	CL
	c	7.88	7.21	11.84	11.80	15.23	15.26
$T=15$	a	6.99	6.80	11.73	11.65	16.11	16.10
	σ^2	2.23	2.22	2.24	2.21	2.31	2.30
	c	9.67	9.59	13.11	13.10	18.39	18.37
$T=30$	a	10.23	10.12	14.25	14.25	18.11	18.09
	σ^2	2.30	2.20	2.29	2.28	2.41	2.40
	c	11.21	11.02	15.94	15.92	20.12	20.12
$T=45$	a	12.09	11.91	14.98	14.99	19.61	19.60
	σ^2	2.39	2.39	2.21	2.22	2.65	2.65

Table 2: Relative efficiency, in terms of MSE, for WLS and CL estimation methods when model (2) is used with $c = 2, a = 0.8, \gamma = 0.5, \alpha = 0.5, \beta = 0.5, \sigma^2 = 1.7$.

		$\|u\| \leq 1$	$\|u\| \leq 2$	$\|u\| \leq 3$
	c	1.04	1.19	1.21
$\|\|\boldsymbol{h}\|\| \leq 1$	a	1.21	2.80	3.65
	σ^2	1.01	1.20	1.21
	a	3.15	3.45	3.71
$\|\|\boldsymbol{h}\|\| \leq 2$	c	6.64	6.48	6.41
	σ^2	3.32	3.04	2.91
	a	2.10	2.59	3.07
$\|\|\boldsymbol{h}\|\| \leq 3$	c	4.53	4.91	5.29
	σ^2	2.22	2.29	2.35

Table 3: Relative efficiency, based on MSE, for WCL at different spatial and time lags when model (2) is used with parameters: $c = 2, a = 0.8, \sigma^2 = 1.5$, with $N = 4$ and $T = 45$.

1. $c = 1$, $a = 1$, $\alpha = 0.5$, $\sigma^2 = 0.6$, $\beta = 0$, $\varepsilon = 0$;

2. $c = 1$, $a = 1$, $\alpha = 0.5$, $\sigma^2 = 0.6$, $\beta = 0.7$, $\varepsilon = 0$;

3. $c = 1$, $a = 1$, $\alpha = 0.5$, $\sigma^2 = 0.6$, $\beta = 0.7$, $\varepsilon = -0.5$.

Note that these models are nested. We chose a small number of locations and a sufficiently large number of times in order to estimate the matrix J. This is a common setting in environmental studies. Calculation of the empirical covariances yielded values $\hat{C}(h, 3) \leq 0.15$, for any spatial lags and identically equal to 3 for the temporal lag, so we set $u = 3$ and considered all spatial lags in the WCL computation. The J matrix was estimated through sub-sampling method with a temporal window $m = 4$. The outcomes of these simulations are summarised in Table 4. These results seem promising because they indicate that at least 80% of the models are correctly identified.

		Identified		
		A	B	C
	A	80	19	1
True	B	10	81	9
	C	0	5	95

Table 4: CLIC identification between different models on a 4×4 regular grid, and $T = 150$.

6. REAL DATA

We reanalysed the Irish wind speed data originally described in Haslett and Raftery (1989) and further analysed in Gneiting (2002); De Luna and Genton (2005); Stein (2005*a*,*b*); Gneiting et al. (2007). In this data set daily wind speeds are collected over 18 years (1961-1978) at 12 sites in Ireland. Following Haslett and Raftery (1989) we omitted the Rosslare station, and then considered a square root transformation to have deseasonalised data. The seasonal component was estimated by calculating the average of the square roots of the daily means over all years and stations for each day of the year, and regressing the result on a set of annual harmonics.

As in Gneiting et al. (2007) we considered only the first ten years as a training set and the next 8 years as a validation set. We considered a spatial nugget effect, as Gneiting (2002) did, thus a discontinuity in $C(\boldsymbol{h}, u)$ at $\mathbf{h} = \mathbf{0}$, for all u due to measurement errors and/or small scale variability. A simple way to model a spatial nugget effect is to write the covariance function in the following form

$$\bar{C}(\mathbf{h}, u) = \sigma^2(1-\nu)C(\mathbf{0}, u)\left(\frac{C(\mathbf{h}, u)}{C(\mathbf{0}, u)} + \frac{\nu}{1-\nu}\delta_{\mathbf{h}=\mathbf{0}}\right), \tag{11}$$

with $\nu \in [0, 1]$ and where $\sigma^2 > 0$ is the variance or sill of the model.

As outlined in Gneiting (2002), data showed asymmetry in time in the east-west direction.

Thus, using CL, we estimated the spatio-temporal structure of the Irish wind speed data using three covariance models for C$(\mathbf{h}, u)$:

- (a) A separable structure, obtained by considering the model in equation (2) with $\beta = 0$.
- (b) A nonseparable model, obtained by considering the same model in equation (2), but with a non-vanishing β.
- (c) a nonseparable, and not fully symmetric model as in equation (4).

We used the WCL with $u = 3$ since correlations in time decay in a fast way (see Figure 5 in Gneiting (2002)). The estimation results are shown in Table 6. Estimates of scale, smoothing, and variance parameters were quite similar for the three considered models and gave a good fit for the spatial margins (see Figure 2)

To choose the best model we considered two approaches. The former is based on prediction in time performance. Following Gneiting et al. (2007), we considered the velocity measures during the 8 years of the test period (1971-1978) and we calculated for each station $365 \times 8 = 2920$ one day ahead predictions, $\hat{z}(\boldsymbol{s}, t)$, using the simple kriging predictor, with the covariance function estimated in the training period. To assess and rank the point forecasts, we used the root-mean-square error (RMSE), defined as $RMSE(\boldsymbol{s}) = \sum_{t=3651}^{6570}(z(\boldsymbol{s}, t) - \hat{z}(\boldsymbol{s}, t))^2/2920$.

Station	(a)	(b)	(c)
Roche's Point	0.484	0.478	0.477
Valentia	0.501	0.501	0.501
Kilkenny	0.436	0.430	0.430
Shannon	0.464	0.464	0.464
Birr	0.477	0.476	0.476
Dublin	0.446	0.442	0.442
Claremorris	0.490	0.491	0.491
Mullingar	0.428	0.425	0.425
Clones	0.484	0.480	0.479
Belmullet	0.495	0.495	0.496
Malin Head	0.496	0.491	0.491

Table 5: RMSE for the stations of the Irish wind speed data

Looking at the RMSE results in Table 5, there is not a clear advantage from a particular model. In particular, (b) and (c) models perform similarly.

The latter approach is based on CLIC values as introduced in Section 4. Here, we chose a window size of four days for the sub-sampling. Results in Table 6 show that this criterion selects the asymmetric in time model.

7. CONCLUSIONS AND DISCUSSION

We have introduced a CL-based estimation of covariance models in a general space-time setting. The method is computationally cheaper with respect to ML, presents some advantages with respect to the classical WLS, and allows us to introduce a model selection criterion. We have also discussed a weighted version of CL in a space and space-time setting that allows to improve estimation from computational and efficiency points of view.

Our results are consistent with those of other authors in different settings. For instance Heagerty and Lele (1998) recommended to consider only *significant* lags but only to save computations. Varin and Vidoni (2007) noted that taking only fixed time lags helps to

	(a)	(b)	(c)
a	0.80333	0.90942	0.97137
c	0.00130	0.00125	0.00136
α	0.84481	0.90714	0.93208
υ	0.03145	0.03800	0.03994
β	-	0.72098	0.80170
ε	-	-	0.00419
σ^2	0.32416	0.31446	0.30466
l_{CL}	921116	2748165	2749202
$CLIC$	920155	2747212	2748367

Table 6: CL estimates for the Irish wind speed data: (a) separable model, (b) Gneiting nonseparable model and (c) nonseparable and not fully symmetric model.

highly improve efficiency of CL estimation.

There exist other ways to construct CL in a space-time setting. For instance the approximating likelihoods on selected blocks introduced by Caragea and Smith (2005) are special cases of the CL. We can borrow their idea for regular monitoring data with respect to time. Define $\boldsymbol{Z}_t = (Z(\boldsymbol{s}_1, t), \ldots, Z(\boldsymbol{s}_n, t))'$, $t = 1, \ldots, T$, and the log-likelihoods $cl_t(\boldsymbol{\theta}) = \log f(\boldsymbol{Z}_t, \cdots, \boldsymbol{Z}_{t-p}; \boldsymbol{\theta})$. A composite likelihood function is $CL(\boldsymbol{\theta}) = \sum_{t=p+1}^{T} cl_t(\boldsymbol{\theta})$. Assuming a Gaussian distribution and stationarity on time, we have to invert a $(p+1)n \times (p+1)n$ matrix for T times, and so the complexity is $O(T(p+1)^3 n^3)$. Thus, the method is useful for data sets with few locations observed many times.

References

Aho, A., Hopcroft, J., and Ullmann, J. (1974), *The Design and Analysis of Computer Algorithms*, : Addison Wesley.

Barry, J., Crowder, M., and Diggle, P. (1997), Parametric estimation of the variogram,, Technical report, Dept Maths & Stats, Lancaster University.

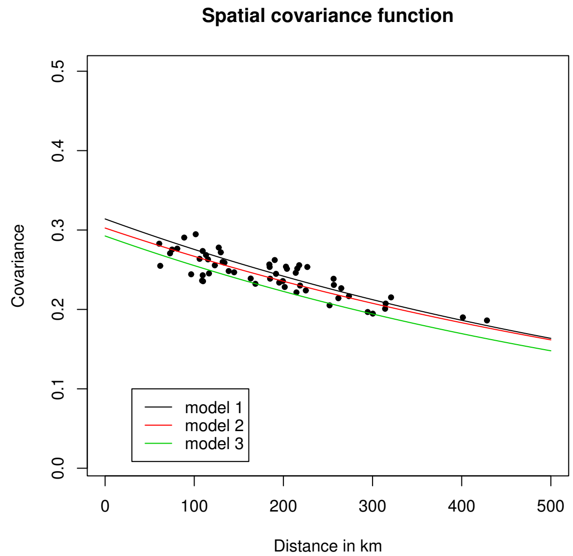


Figure 2: Empirical and fitted marginal (spatial) covariances for models

Besag, J. (1974), "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society B*, 36, 192–236.

Caragea, P., and Smith, R. (2005), Approximate likelihoods for spatial processes,, Technical report, Department of Statistics, Iowa State University.

Christakos, G. (2000), *Modern Spatiotemporal Geostatistics*, Oxford: Oxford University Press.

Cox, D., and Isham, V. (1988), "A simple spatial-temporal model of rainfall," *Proceedings of the Royal Society of London, Ser. A*, 415, 317–328.

Cressie, N. (1985), "Fitting variogram models by weighted least squares," *Mathematical Geology*, 17, 239–252.

Cressie, N. (1993), *Statistics for Spatial Data*, revised edn, New York: Wiley.

Cressie, N., and Huang, H. (1999), "Classes of nonseparable, spatiotemporal stationary covariance functions," *Journal of the American Statistical Association*, 94, 1330–1340.

Curriero, F., and Lele, S. (1999), "A composite likelihood approach to semivariogram estimation," *Journal of Agricultural, Biological and Environmental Statistics*, 4, 9–28.

De Cesare, L., Myers, D., and Posa, D. (2001), "Product-sum covariance for space-time modeling: an environmental application," *Environmetrics*, 12, 11–23.

De Luna, X., and Genton, M. (2005), "Predictive spatio-temporal models for spatially sparse environmental data," *Statistica Sinica*, 15(547-568).

Fuentes, M. (2007), "Approximate likelihood for large irregularly spaced spatial data," *Journal of the American Statistical Association*, to appear.

Genton, M. (1998), "Variogram fitting by generalized least squares using an explicit formula for the covariance structure," *Mathematical Geology*, 30, 323–345.

Gneiting, T. (2002), "Stationary covariance functions for space-time data," *Journal of the American Statistical Association*, 97, 590–600.

Gneiting, T., Genton, M. G., and Guttorp, P. (2007), "Geostatistical space-time models, stationarity, separability and full symmetry," in *Statistical Methods for Spatio-Temporal Systems*, eds. B. Finkenstadt, L. Held, and V. Isham, Boca Raton: Chapman & Hall/CRC, pp. 151–175.

Godambe, V. (1991), *Estimating Functions*, New York: Oxford University Press.

Guan, Y. (2007), "A composite likelihood approach in fitting spatial point process models," *Journal of the American Statistical Association*, to appear.

Haslett, J., and Raftery, A. (1989), "Space-time modelling with long-memory dependence: assessing Ireland's wind power resource," *Applied Statistics*, 38, 1–50.

Heagerty, P., and Lele, S. (1998), "A composite likelihood approach to binary spatial data," *Journal of the American Statistical Association*, 93, 1099–1111.

Heagerty, P., and Lumley, T. (2000), "Window subsampling of estimating functions with application to regression models," *Journal of the American Statistical Association*, 95, 197–211.

Hoeting, A., Davis, A., Merton, A., and Thompson, S. (2006), "Model selection for geostatistical models," *Ecological Applications*, 16, 87–98.

Huang, H., Martinez, F., Mateu, J., and Montes, F. (2007), "Model comparison and selection for stationary space-time models," *Computational Statistics and Data Analysis*, to appear.

Jones, R., and Zhang, Y. (1997), "Models for continuous stationary space-time processes," in *Modelling Longitudinal and Spatially Correlated Data*, eds. T. Gregoire, D. Brillinger, P. Diggle, E. Russek-Cohen, W. Warren, and R. Wolfinger, New York: Springer Verlag, pp. 289–298.

Lindsay, B. (1988), "Composite likelihood methods," *Contemporary Mathematics*, 80, 221–239.

Ma, C. (2002), "Spatio-temporal covariance functions generated by mixtures," *Mathematical Geology*, 34, 965–974.

Mardia, K., and Marshall, R. (1984), "Maximum likelihood estimation of models for residual covariance in spatial regression," *Biometrika*, 71, 135–146.

Matheron, G. (1965), *Les variables régionalisées et leur estimation*, Paris: Masson.

Mitchell, M., Genton, M., and Gumpertz, M. (2005), "Testing for separability of space-time covariances," *Environmetrics*, 16, 819–831.

Muller, W. (1999), "Least-squares fitting from the variogram cloud," *Statistics and Probability Letters*, 43, 93–98.

Nott, J., and Ryden, T. (1999), "Pairwise likelihood methods for inference in image models," *Biometrika*, 86, 661–676.

Pardo-Iguzquiza, E. (1998), "Maximum likelihood estimation of spatial covariance parameters," *Mathematical Geology*, 30, 95–108.

Pardo-Iguzquiza, E., and Dowd, P. (2001), "Variance covariance matrix of the experimental variogram: assessing variogram uncertainty," *Mathematical Geology*, 33, 397–419.

Porcu, E., Gregori, P., and Mateu, J. (2006), “Nonseparable stationary anisotropic space time covariance functions,” *Stochastic Environmental Research and Risk Assessment*, 21, 113–122.

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

Stein, M. (2005*a*), “Space-time covariance functions,” *Journal of the American Statistical Association*, 100, 310–321.

Stein, M. (2005*b*), “Statistical methods for regular monitoring data,” *Journal of the Royal Statistical Society B*, 67, 667–687.

Stein, M., Chi, Z., and Welty, L. (2004), “Approximating likelihoods for large spatial data sets,” *Journal of the Royal Statistical Society B*, 66, 275–296.

Varin, C., and Vidoni, P. (2005), “A note on composite likelihood inference and model selection,” *Biometrika*, 52, 519–528.

Varin, C., and Vidoni, P. (2007), “Pairwise likelihood inference for general state space Models,” *Econometric Reviews*, . to appear.

Vecchia, A. (1988), “Estimation and model identification for continuous spatial processes,” *Journal of the Royal Statistical Society B*, 50, 297–312.

Warnes, J., and Ripley, B. (1987), “Problems with likelihood estimation of covariance functions of spatial Gaussian processes,” *Biometrika*, 74, 640–642.