Composite likelihood inference for space-time covariance models

Moreno Bevilacqua

January 29, 2008

Abstract

Modelisation and prediction of environmental phenomena, which typically show dependence in space and time, has been one of the most important challenges over the last years.

The classical steps of the spatial modeling approach can be resumed as follows: (1) a model-oriented step, in which a random fields assumption is considered; (2) estimation of the objects defining homogeneity and variability of the random field (i.e the trend and the covariance structure); (3) prediction, which is classically implemented through universal or ordinary kriging procedures.

This thesis focuses on the first and second steps. Specifically, it consists of three major parts.

First part consists in spatio-temporal covariance modeling. In the geostatistical approach spatial and temporal structure are entangled in the covariance structures and it is not easy to model these two parts simultaneously. The classical kriging predictor, depends crucially on the chosen parametric covariance function. This fact motivates the request for more candidate models of covariance functions that can be used for space-time data. In particular, the wide variety of practical situations that one may face in the space-time domain motivates the request for flexible models of space-time covariance functions, in order to cover several settings such as non-separability or asymmetry in time. We introduce a new class of stationary space-time covariance model which allows for zonal spatial anisotropies.

Second part consists in space and space time covariance models estimation. Maximum likelihood and related techniques are generally considered the best method for estimating the parameters of space-time covariance models. For a spatial Gaussian random field 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. The problem increases dramatically in a space-time setting. This fact motivates the search for approximations to the likelihood function that require smaller computational burden and that perform better than classical least square estimation. We introduce a weighted composite likelihood estimation (WCL) for space and space time covariance model estimation. The method induces gains in statistical efficiency with respect to the least squares estimation and from the computational point of view with respect to maximum likelihood estimation.

Third part consists in a set of application of WCL. Specifically we apply the method in the estimation of particular spatial covariance functions which allow for negative values and in the estimation of covariance function describing residuals dependence in dynamic life tables. A simulation based test to verify separability of some parametric covariance models in a space time setting is proposed.

Abstract

La modellizzazione e la previsione di fenomeni ambientali che esibiscono dipendenza nello spazio e nel tempo hanno rappresentano una sfida molto importante negli ultimi anni.

I passi della modellizzazione spazio temporale possono essere riassunti come segue: (1) un primo passo in cui vengono fatte assunzioni sul campo aleatorio; (2) stima degli oggetti che definiscono le caratteristiche principali del campo aleatorio (media e covarianza); (3) previsione che di solito viene effettuata attraverso il kriging ordinario o universale.

Questa tesi si focalizza in particolare nel primo e nel secondo punto. In particolare è composta da tre parti principali.

La prima parte consiste nella modellizzazione della covarianza spazio-temporale. Nell'approcio geostatistico la dipendenza spaziale e temporale viene traferita nella struttura di covarianza che di solito viene modellizzata attraverso modelli parametrici validi. Il kriging dipende crucialmente dal modello di covarianza scelto e questo fatto motiva la necessità di modelli di covarianza sempre più flessibili. In particolare nello spazio-tempo si richiedono modelli che possano descrivere situazioni particolari tipiche del contesto spazio-temporale come la non separabilità o l'asimmetria nel tempo. Qui presentiamo una nuova classe parametrica di covarianza spazio-temporale che permette di modellare anisotropia nello spazio.

La seconda parte consiste nella stima della funzione di covarianza. La massima verosimiglianza in generale è considerato il miglior metodo per stimare questo tipo di oggetto. Tuttavia il calcolo della verosimiglianza richiede l'inversione e il calcolo del determimante della matrice di covarianza. Per numerosità campionarie elevate questo diventa un problema da un punto di vista computazionale e spesso si ripiega su stimatori meno efficienti come ad esempio quelli basati sui minimi quadrati pesati. Qui presentiamo un nuovo metodo di stima basato sulla verosimiglianza composita pesata (WCL), per la stima di funzioni di covarianza spaziale e spazio temporale che induce vantaggi dal punto di vista dell'efficienza statistica (rispetto ai classici stimatori dei minimi quadrati) e dal punto di vista del carico computazionale (rispetto alla massima verosimiglianza).

Il metodo di stima proposto viene poi applicato per la stima di particolari funzioni di covarianza che ammettono valori negativi e per la stima di modelli di covarianza che descrivono i residui in modelli dinamici tipo life tables. Infine un test basato sulla verosimiglianza composita pesata

viene proposto per verificare l'ipotesi di separabilità della funzione di covarianza in un contesto parametrico.

Contents

List of Figures

List of Tables

Acknowledgements

This dissertation would not have been possible without the support of many people. First of all, I would like to express my gratitude to my advisor Prof. Carlo Gaetan, for his encouragement, insightful guidance and endless support throughout my graduate school experience. It has been a privilege, both from an academic and personal point of view, to work with him.

I am extremely grateful for having Prof. Jorge Mateu and Prof. Emilio Porcu as my advisors. During my staying in Castellon they provided much help to me on spatial statistics and oriented me in the right research direction.

Their enthusiasm, knowledgeable guidance, and constant support have been very important for me. They introduced to me in the space-time covariance function subject and pass to me the passion for the research. I thank them also for giving me the opportunity to collaborate with Prof. Hao Zhang, Prof. George Christakos, Prof. Paco Montes and Prof. Alessandro Zini.

I would like also to thank the director of PHD School of Statistics of Padova Prof. Alessandra Salvan for her helpfulness and my colleagues Padoan Simone, Daniela Zugna, Sofia Massa, Federico Tedeschi, Giuliana Cortese, Francesca Santello, Antonio Parisi to share with me this PHD experience.

Thanks go as well to Cristiano Varin and Nicola Sartori with whom I had interesting discussions on composite likelihood topic.

Last but not least, I wish to dedicate this work to my family for their love, understanding and support.

Chapter 1

Overview

Data in areas such as environmental health, ecology, epidemiology biology and geology, often have a geographical and temporal label associated with them. Usually, data that are close together in space (and time) are more alike than those that are far apart.

There are three types of spatial data: point-referenced data, areal data, and point-process data. Point-referenced data are also known as geostatistical data where data are observed at a set of locations in a set D in $\mathbb{R}^d$ and d is the number of dimensions. Typically, the locations are represented in two or three spatial coordinates, e.g. longitude, latitude, and altitude.

Conceptually, observations can be taken at every location in D so we should consider uncountably many data points. The observations are considered as a realization from a stochastic spatial process. We denote this process by $Z(\mathbf{s})$, where $\mathbf{s} \in D$ indexes location. Often, the goals are to make statistical inference about $Z(\mathbf{s})$ and predict the process at new locations based on the available data through kriging techniques.

For areal data, the study region D is again a fixed subset in $\mathbb{R}^d$, but now partitioned into a finite number of areal units with well-defined boundaries. For example, in an environmental health investigation, for purpose of confidentiality, counts of some adverse health outcome (e.g. lung cancer) are aggregated by county in a particular state and environmental risk factors are supplied for these areal units to explain the counts.

Point-process data describe the locations of some events of interests. Examples of this kind of data include incidence of disease, sightings or births of a species, or the occurrences of fires, earthquakes, lightning strikes, tsunamis, or volcanic eruptions. A spatial point process is a collection

of random points, where each point indicates the location of an event. A point process $N(.)$ is defined as a random measure on $D \subset \mathbb{R}^d$, taking non-negative integer values. So $N(A)$ means the number of points falling in the set $A \subset D$.

In the last years the temporal component has been added in the analysis of these three kinds of data. For point-referenced data the process is typically indexed by $\mathbf{s}$ and t where $t \in T \subset \mathbb{R}$ represent the time domain.

The focus of this thesis is on spatial and spatio-temporal point-referenced data. In general, it is very challenging to study dependent data in space and time in the geostatistical setting. One of the problem in dealing with this kind of data is to model the covariance structure. Specially for space-time data, there is a growing literature on covariance function models. The wide variety of practical situations that one may face in the space-time domain motivates the request for flexible models, in order to cover several settings, such as non-separability, asymmetry in time or anisotropy in space. Thus, several nonseparable space-time covariance models have been proposed in the last years.

Chapter two introduces the basic tools for spatial modeling, while chapter three describes recent developments in space-time covariance models. A new class of stationary nonseparable space-time covariance functions that can be used for both geometrically and zonally anisotropic data, is introduced. In addition, we show some desirable mathematical features of this class.

Since kriging techniques depends crucially on covariance function, the estimation of this object is one of the most important step in the geostatistical analysis.

Maximum likelihood and related techniques are generally considered the best method for estimating the parameters of covariance models. However, for a Gaussian random field 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. Obviously for space time data the problem gets worse dramatically. The Irish wind speed data-set (Haslett and Raftery, 1989) is an example. This data set consists of time series of daily average wind speed at eleven synoptic meteorological stations in Ireland during the period 1961-1978. In this case, we dispose of $11 \times 365 \times 18 = 72270$ observations, thus ML is infeasible. The classical remedy is to use less efficient but computationally feasible estimation methods such as weighted least squares.

Chapter four reviews classical covariance function estimation methods and the new recent pro-

posal to solve the computational problem. A weighted composite likelihood approach estimation in introduced. The goal of this estimation methods is basically one: to offer a "good" quality inference method for covariance parametric models when maximum likelihood for computational reasons is infeasible. The method can be considered as 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 classical weighted least squares procedure. The method introduces the concept of "optimal distance" which we have to identify before performing parameter estimation. It is optimal because it allows to maximize the Godambe Information associate to weighted composite likelihood function. It ensures to obtain estimates more efficient than classical methods such as weighted least square and feasible for huge dataset.

Chapter 5 contains a set of applications of WCL to different contexts: the former consists of estimation of particular spatial covariance functions which allow for negative values, the latter consists of estimating covariance functions describing residuals dependence in dynamic life tables. Finally, we propose a simulation based test to verify separability of some parametric covariance models in a space time setting.

Chapter 2

Spatial modelling

2.1. SPATIAL PROCESSES

We now review a few important basics of spatial processes. We denote as $Z(\mathbf{s})$, where $\mathbf{s}$ indexes location and $\mathbf{s} \in D \subset \mathbb{R}^d$, the quantity we are studying. For each $\mathbf{s}$, $Z(\mathbf{s})$ is a random variable. The collection, $Z(\mathbf{s})$, when $\mathbf{s}$ varies over all its possible values, is called a spatial process or random field. In practice, $Z(\mathbf{s})$ is a random function indexed by the symbol $\mathbf{s}$ which belongs to some index set D. When $d = 1$ it is usually called random stochastic or random process, while when $d \geq 2$ it is defined as random field.

$Z(\mathbf{s})$ is simply a random variable for each $\mathbf{s}$ and its properties (e.g. mean and variance) can be described by its distribution function. More generally we are interested in studying the whole collection of random variables $\{Z(\mathbf{s})\}$ and its joint distribution function.

The Kolmogorov consistency theorem states that, under fairly general conditions, the probability structure of $Z(\mathbf{s})$ is fully specified if the joint distribution of $\{Z(\mathbf{s}_1), Z(\mathbf{s}_2), ..., Z(\mathbf{s}_n)\}$, *i.e.* the finite-dimensional distribution, is given for arbitrary choice of n and $\mathbf{s}_1, .., \mathbf{s}_n$. Usually simplifying assumptions are considered on the probability structure.

2.2. STATIONARITY

When making inference on the probability structure of the spatial process based on what we observe (often just a single realization of the process) a simplifying assumption often made is the stationarity assumption. Stationarity in simple terms means that the random field looks similar in different parts of the domain. There are different kinds of stationarity. Now suppose $D = \mathbb{R}^d$

Definition 1. *A process $Z(\mathbf{s})$ is said strictly stationary if for all $\mathbf{s}_1, ..., \mathbf{s}_n$ and any $\mathbf{h} \in \mathbb{R}^d$, the joint distribution of $Z(\mathbf{s}_1), ..., Z(\mathbf{s}_n)$ is identical with the joint distribution of $Z(\mathbf{s}_1 + \mathbf{h}), ..., Z(\mathbf{s}_n + \mathbf{h})$, i.e.,*

$$Pr(Z(\mathbf{s}_1) \leq z_1, ..., Z(\mathbf{s}_n) \leq z_n) = Pr(Z(\mathbf{s}_1 + \mathbf{h}) \leq z_1, ..., Z(\mathbf{s}_n + \mathbf{h}) \leq z_n), \tag{2.1}$$

where $z_1, ..., z_n \in \mathbb{R}$.

That is the probability law of a strictly stationary process is invariant under a shift in space. A lighter type of stationarity is the weak stationarity:

Definition 2. *A process $Z(\mathbf{s})$ is weakly stationary (WS) if:*

$$E(Z(\mathbf{s})) = \mu$$

and

$$Cov(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = \mathrm{C}(\mathbf{s}_2 - \mathbf{s}_1) = \mathrm{C}(\mathbf{h}).$$

Thus a spatial process which mean does not depend on the spatial location and which covariance is a function of the separation lag $\mathbf{h}$, is a WS process. $\mathrm{C}(\mathbf{h})$ is called the covariance function of $Z(\mathbf{s})$.

For a WS process $Z(\mathbf{s})$, the correlation between $Z(\mathbf{s}_1)$ and $Z(\mathbf{s}_2)$ is defined as:

$$Corr(Z(\mathbf{s}_1), Z(\mathbf{s}_2)) = \frac{\mathrm{C}(\mathbf{h})}{\mathrm{C}(\mathbf{0})} = \rho(\mathbf{h}).$$

Strict stationarity (if exist moments of second order) implies WS while the reverse is not true.

A common hypothesis regarding the finite-dimensional distribution of the random fields $Z(\mathbf{s})$, is Gaussianity.

Definition 3. *$Z(\mathbf{s})$ is called Gaussian process if for all n and admissible $\mathbf{s}_1, ..., \mathbf{s}_n$, the joint distribution of $Z(\mathbf{s}_1), ..., Z(\mathbf{s}_n)$ is multivariate normal.*

A multivariate normal distribution is characterized by its mean and covariance matrix, so the first two moments of a Gaussian process completely specify its probability structure. Thus for Gaussian processes, WS implies strict stationarity.

Last kind of stationarity regards the increments of the process.

Definition 4. *A process $Z(\mathbf{s})$ is intrinsic stationary (IS) if:*

$$E(Z(\mathbf{s})) = \mu$$

and

$$Var(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)) = 2\gamma(\mathbf{s}_2 - \mathbf{s}_1) = 2\gamma(\mathbf{h}).$$

The function $2\gamma(\mathbf{h})$ is called variogram. IS is a weaker property than WS. If the process is WS, it is easy to verify that:

$$Var(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)) = 2\mathrm{C}(\mathbf{0}) - 2\mathrm{C}(\mathbf{h})$$

and so $\gamma(\mathbf{h}) = \mathrm{C}(\mathbf{0}) - \mathrm{C}(\mathbf{h})$. Conversely, in general IS does not imply weak stationarity. For instance if $Z(\mathbf{s})$ is the standard Brownian motion in one dimension, the variogram function is $Var(Z(\mathbf{s}_1) - Z(\mathbf{s}_2)) = |\mathbf{s}_1 - \mathbf{s}_2|$. However, we can not recover the covariance function since the variogram is unbounded for $\mathbf{h}$ tends to infinity.

Gaussian processes play a central role in modeling spatial data. The advantages of the Gaussian process assumption are obvious: it allows convenient distribution theory (for instance, conditional distributions are easily obtained from the joint distributions). Gaussian processes have a rich, detailed and very well understood general theory. Furthermore, in most applications, we observe a single realization of the process at a finite set of locations. It is not easy to criticize a Gaussian assumption since we only have a sample size of one from a finite dimensional distribution. Nevertheless, there are situations in which it is more appropriate to use other processes to model spatial data.

2.3. COVARIANCE FUNCTIONS AND VARIOGRAMS PROPERTIES

For a real-valued mapping $\mathrm{C} : \mathbb{R}^d \to \mathbb{R}$, necessary conditions for C to be a covariance function are:

(i) $\mathrm{C}(\mathbf{0}) \geq 0$;

(ii) $\mathrm{C}(\mathbf{h}) = \mathrm{C}(-\mathbf{h})$, i.e., C is an even function;

(iii) $\mathrm{C}(\mathbf{0}) \geq |\mathrm{C}(\mathbf{h})|$;

(iv) If $\mathrm{C}_j(\mathbf{h})$ are valid covariance function $j = 1 \ldots k$ then $\sum_{j=1}^{k} b_j \mathrm{C}_j(\mathbf{h})$ is a valid covariance function, if $b_j \geq 0, \quad \forall j$;

(v) If $\mathrm{C}_j(\mathbf{h})$ are valid covariance function $j = 1 \ldots k$ then $\prod_{j=1}^{k} \mathrm{C}_j(\mathbf{h})$ is a valid covariance function;

(vi) If $\mathrm{C}(\mathbf{h})$ is a valid covariance function in $\mathbb{R}^d$, then it is also a valid covariance function in $\mathbb{R}^p$, $p \leq d$.

Analogous properties of the variogram are:

(i) $\gamma(\mathbf{0}) = 0$;

(ii) $\gamma(\mathbf{h}) = \gamma(-\mathbf{h})$, i.e., γ is an even function;

(iii) $\gamma(\mathbf{h}) \geq 0$;

(iv) If $\gamma_j(\mathbf{h})$ are valid variograms $j = 1 \dots k$ then $\sum_{j=1}^{k} b_j \gamma_j(\mathbf{h})$ is a valid variogram,, if $b_j \geq 0, \quad \forall j$;

Valid (or permissible) covariance function or variogram means that they must respect some mathematical constraints.

Indeed, one cannot define a spatial covariance or variogram function in a totally arbitrary way. The key property which has to satisfy is semi-positive definiteness. For a spatial process $Z(\mathbf{s})$ with covariance matrix $\mathrm{C} = \{\mathrm{C}(\mathbf{s}_i, \mathbf{s}_j)\}_{i,j}^{n}$, it means that:

$$\sum_{i}^{n} \sum_{j}^{n} a_i a_j \mathrm{C}(\mathbf{s}_i, \mathbf{s}_j) \geq 0 \tag{2.2}$$

for any set of $\mathbf{s}_1, ..., \mathbf{s}_n$ and all real $a_1, ..., a_n$.

The positive semi-definite condition is necessary for the existence of a random field with finite second moments. This condition guarantees that the variance of spatial predictions is non-negative. This simply follows noting that (2.2) is $Var(\sum_i^n a_i Z(\mathbf{s}_i))$.

On the other hand, if C is positive semi-definite, there exists a Gaussian random field with covariance matrix K and mean $E(Z(\mathbf{s})) = m < \infty$. Thus, positive definiteness is a necessary and sufficient condition for a covariance function.

Bochner's theorem (1933) provides necessary and sufficient conditions for a covariance function $\mathrm{C}(\mathbf{h})$ of a WS process to be positive semi-definite.

Theorem 1. *(Bochner's Theorem). For a real-valued WS process on* $\mathbb{R}^d$, $\mathrm{C}(\mathbf{h})$ *is positive semi-definite if and only if it can be represented as:*

$$\mathrm{C}(\mathbf{h}) = \int e^{i\boldsymbol{\omega}^T \mathbf{h}} dF(\boldsymbol{\omega}) \tag{2.3}$$

where F is a positive, symmetric, and finite measure and is called the spectral measure of $\mathrm{C}(\mathbf{h})$. *If F is absolutely continuous with respect to Lebesgue measure, i.e., $dF(\boldsymbol{\omega}) = f(\boldsymbol{\omega})d\boldsymbol{\omega}$, $f(\boldsymbol{\omega})$ is called the spectral density.*

Analogously to the covariance function, the variogram must respect some conditions to be permissible. Specifically, for any set of $\mathbf{s}_1, ..., \mathbf{s}_n$ and any set of real $a_1, ..., a_n$ such that $\sum_i^n a_i = 0$,

$$\sum_i^n \sum_j^n a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) \leq 0. \tag{2.4}$$

This follows by noting:

$$\sum_i^n \sum_j^n a_i a_j \gamma(\mathbf{s}_i - \mathbf{s}_j) = -E(\sum_i^n (a_i Z(\mathbf{s}_i))^2).$$

The variogram and covariance functions are parameters of the spatial process and play a critical role in the geostatistical method of spatial data analysis. Both are important ingredients of the kriging methods for spatial prediction. Statisticians are more familiar with covariance functions, while geostatisticians prefer the variogram. Under WS, use $\mathrm{C}(\mathbf{h})$ or $\gamma(\mathbf{h})$ for statistical or prediction purpose is equivalent. However $\gamma(\mathbf{h})$ presents some advantages from estimation point of view (see chapter 4).

Note that $\gamma(\mathbf{h})$ only describes the first two moments (not the probability law) of the spatial process so it is not possible to make likelihood-based inference on its basis. However, it can be used for approximations of likelihood as in WCL proposed in Chapter 4.

2.4. ISOTROPY AND SOME PARAMETRIC MODELS

A WS random field is said isotropic if its covariance function $\mathrm{C}(\mathbf{h})$ only depends on $||\mathbf{h}||$, where $||.||$ indicates the Euclidean distance. The class of all valid continuous covariance functions on $\mathbb{R}^d$ can be characterized by the Fourier transforms of all finite positive measures on $\mathbb{R}^d$ (see Theorem 1). There is an analogous characterization for isotropic covariance functions (see Yaglom (1987), Section 22). Specifically,

Theorem 1. . *For $d \geq 2$ a function $\mathrm{C}(\mathbf{h})$ is a continuous isotropic covariance function of a WS random field on $\mathbb{R}^d$ if and only if it can be represented as:*

$$\mathrm{C}(\mathbf{h}) = 2^{\frac{d-2}{2}} \Gamma(\frac{d}{2}) \int_0^\infty (\boldsymbol{\omega}||\mathbf{h}||)^{-\frac{d-2}{2}} J_{\frac{d-2}{2}}(\boldsymbol{\omega}||\mathbf{h}||) dG(\boldsymbol{\omega}) \tag{2.5}$$

Here J_k is the Bessel function of the first kind of order k (Abromowitz and Stegun, 1967) and the measure $G(.)$ is nondecreasing bounded in $\mathbb{R}^+$ and $G(\mathbf{0}) = 0$.

A general form of an isotropic variogram function is:

$$\gamma(\mathbf{h}, \boldsymbol{\theta}) = \begin{cases} \theta_0 + \sigma^2(1 - \rho(||\mathbf{h}||, \boldsymbol{\theta})), & ||\mathbf{h}|| > 0 \\ 0, & ||\mathbf{h}|| = 0 \end{cases} \tag{2.6}$$

where θ_0 is the nugget parameter. This parameter describes the behavior of the variogram near the origin. A phenomenon quite common in applications is that the variogram at the origin does not attain the zero. This is due to microscale variability (variability of a spatial process operating at lag distances shorter than the smallest lag observed in the data) or/and measurement error. In geostatistical literature θ_0 is the nugget, $\sigma^2 + \theta_0$ is the sill and σ^2 is the partial sill or variance.

In (2.6) $\rho(||\mathbf{h}||, \boldsymbol{\theta})$ is a parametric correlation function which depends on $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$. Typically parametric correlation models depend on few parameters . Following is a list of popular parametric isotropic correlation functions:

1. Exponential covariance function:

$$\rho(||\mathbf{h}||, \theta) = e^{-\theta||\mathbf{h}||}, \qquad \theta > 0. \tag{2.7}$$

   Here θ is the decay parameter which tells us how quickly the correlation decays as the distance $||\mathbf{h}||$ increases. The decay parameter is related with a notion of effective range which is often used geostatistics. Effective range is the distance at which there is essentially no lingering spatial correlation. In practice, it is commonly defined as the distance at which the correlation drops to only 0.05. In the exponential correlation function case, the effective range is attained approximatively at $3/\theta$.

2. Gaussian covariance function:

$$\rho(||\mathbf{h}||, \theta) = e^{-\theta||\mathbf{h}||^2}, \qquad \theta > 0. \tag{2.8}$$

3. Powered exponential covariance function:

$$\rho(||\mathbf{h}||, \theta, \alpha) = e^{-\theta||\mathbf{h}||^\alpha}, \qquad \theta > 0, 0 < \alpha \leq 2. \tag{2.9}$$

4. Matérn covariance function:

$$\rho(||\mathbf{h}||, \theta, \nu) = \frac{1}{2^{v-1}\Gamma(\nu)}(\theta||\mathbf{h}||)^\nu \kappa_\nu(\theta||\mathbf{h}||), \qquad \theta > 0, \nu > 0, \tag{2.10}$$

where θ is the range parameter and ν is the smoothing parameter (the larger is ν the smoother the corresponding process is). Different parametrization has been proposed for this model (see for instance (Handcock and Wallis, 1994) . The parametrization in (2.10) has the advantage to not depend on d. The flexibility in parameterizing the smoothness of the process by changing ν is the main reason why this family has been advocated as a default covariance model for most spatial applications (Stein (1999)).

Note that the exponential covariance function and the Gaussian covariance function are two special cases of the Matern class with $\nu = 1/2$ and $\nu = \infty$, respectively.

If the covariance function of a WS stationary process is anisotropic, the spatial structure is directional dependent. Anisotropy is generally difficult to deal with but there are special cases that are tractable yet still interesting:

- Geometrical Anisotropy: variograms in two or more directions have different ranges, but the same sill value. That is, the variability of an observation is the same, but they are correlated over longer or shorter ranges, depending on the direction.
- Zonal Anisotropy: here the variogram has different sills and different ranges in two or more directions.

To correct for geometric anisotropy, a linear transformation of the coordinates is performed in order to reduce to an isotropic space. A linear transformation may correspond to rotation or stretching of the coordinate axes. Thus, in general if $\rho_0(.,\theta)$ is an isotropic parametric covariance function, then

$$\rho(\mathbf{h},\theta) = \rho_0(||Ah||,\theta),$$

is a geometrically anisotropic covariance function. Here A is a $d \times d$ matrix describing the linear transformation. There are different approach to face zonal anisotropy. Basically they are based on weighted linear combination of valid variograms (Rouhani and Hall, 1989). A special kind of anisotropy is attained considering variogram that are function of the sub-vectors $\mathbf{h} = (|h_1|, |h_2|)$, that is considering isotropy in the two main directions (Shapiro and Botha, 1991). This kind of anisotropy will be extended to space-time context in chapter 3.

2.5. SPATIAL CONTINUITY AND DIFFERENTIABILITY

Continuity and differentiability of a random fields are important since they are informative about the structure and the smoothness properties of the random field.

Assume the process have 0 mean and finite second-order moments.

Definition 5. *A process $Z(\mathbf{s})$ is L_2 continuous at $\mathbf{s}_0$ if and only if* $\lim_{\mathbf{s}\longrightarrow\mathbf{s}_0} E[Z(\mathbf{s}) - Z(\mathbf{s}_0)]^2 = 0$

Continuity in L_2 sense is also referred to as mean square continuity and will be denoted by $Z(\mathbf{s}) \stackrel{L_2}{\rightarrow} Z(\mathbf{s}_0)$

Definition 6. *A process $Z(\mathbf{s})$ is almost surely continuous at $\mathbf{s}_0$ if $Z(\mathbf{s}) \longrightarrow Z(\mathbf{s}_0)$ a.s. as $\mathbf{s} \longrightarrow \mathbf{s}_0$. If the process is almost surely continuous for every $\mathbf{s}_0 \in \mathbb{R}^d$ then the process is said to have continuous realizations.*

In general, one form of continuity does not imply the other since one form of convergence does not imply the other. However, if $Z(\mathbf{s})$ is a bounded process then *a.s.* continuity implies L_2 continuity. Of course, each implies $Z(\mathbf{s}) \stackrel{P}{\rightarrow} Z(\mathbf{s}_0)$. It is easy to show that for a WS random field, mean square continuity at $\mathbf{s}$ implies that:

$$\lim_{\mathbf{h}\longrightarrow 0} E[Z(\mathbf{s}) - Z(\mathbf{s}+\mathbf{h})]^2 = 0$$

Thus, it is easily shown that for a WS random field mean square continuity is equivalent to the covariance function C($\mathbf{h}$) being continuous at 0. That is mean square continuity can be verified through the behavior of the covariance function near 0. As explained in section 2.4 some process appear to have a variogram for which $\gamma(\mathbf{h}) \longrightarrow c > 0$ as $\mathbf{h}. \longrightarrow 0$, *i.e* the nugget effect.

Means square continuity by itself does not convey much about the smoothness of the process and how it is related to the covariance function. The smoothness concept is brought into focus by studying the partial derivatives of the random field and the introducing the mean square differentiability.

In $\mathbb{R}^1$ we can define the process:

$$Z_\delta(t) = \frac{Z(t+\delta) - Z(t)}{\delta}$$

with $t, \delta \in \mathbb{R}^1$ and we say that the process $Z(t)$ is mean square differentiable if the process $Z_\delta(t)$ converges in L_2. More formally, we have the following definition.

Definition 7. *A process $Z(t)$ on R^1 is (mean square) differentiable if there exists a process $Z'(t)$ such that the following holds:* $\lim_{h \longrightarrow 0} E[Z_h(t) - Z'(t)]^2 = 0$

It can be shown that the stationary process $Z_\delta(h)$ has the covariance function $C_\delta(h)$ such that $\lim_{h \longrightarrow 0} C_\delta(h) = -C''(h)$ provided $C(h)$ is twice differentiable. This also shows that $-C''(h)$ is positive definite. Stein (1999) proves that $Z(t)$ is m-times mean square differentiable if and only if $[\frac{d^{2m}C(h)}{dh^{2m}}]_0$ exists and is finite. That is there is a strong relation between the mean square differentiability of a process and the derivative of its covariance function.

From this point of view it can be possible to make the selection of a particular correlation function based upon theoretical considerations. This possibility arises from the powerful fact that the choice of correlation function determines the smoothness of realizations from the spatial process. In this sense Stein (1999) recommends the Matèrn class as a general tool for building spatial models since the parameter ν can control the degree of smoothness.

2.6. SPECTRAL AND CONVOLUTION REPRESENTATIONS

In this thesis we do not propose approaches based on spectral methods, however it is a powerful tool for studying random processes. One explanation is through Bochner's Theorem which builds a correspondence between a covariance function and a spectral density. Some theory properties of random fields are better studied using spectral analysis (Stein (1999)) and spectral measures. Inference, even if with certain problems, is possible in the frequency domain. Moreover to every WS process $Z(\mathbf{s})$ there can be assigned a process $Y(\boldsymbol{\omega})$ with orthogonal increments, such that we have for each fixed $\mathbf{s}$ the spectral representation:

$$Z(\mathbf{s}) = \int_{\mathbb{R}^d} e^{is^T\boldsymbol{\omega}} dY(\boldsymbol{\omega}) \tag{2.11}$$

$Y(\boldsymbol{\omega})$ is called the random measure corresponding to $Z(\mathbf{s})$. It has the following properties:

- $E(Y(A)) = 0$ for all measurable set A.
- $E(Y(A, \bar{B})) = 0$ for disjoint measurable sets A and B.
- $Z(A \cup B) = Z(A) + Z(B)$ for disjoint measurable sets A and B.

If $E(|Z(A)|^2) = F(A)$ for some positive finite measure F, then the covariance function associated with Z(s) can be expressed as:

$$\mathrm{C}(\mathbf{h}) = \int e^{i\boldsymbol{\omega}^T\mathbf{h}} dF(\boldsymbol{\omega}) \tag{2.12}$$

The function F is called the spectral measure of Z. If F has a density with respect to Lebesgue measure, this density is the spectral density, $f = F'$, defined as the Fourier transform of the autocovariance function:

$$f(\boldsymbol{\omega}) = \frac{1}{(2\pi)^d} \int e^{i\boldsymbol{\omega}^T \mathbf{h}} C(\mathbf{h}) d\mathbf{h} \tag{2.13}$$

Inference on the frequency domain pass through estimation of the spectral density.

Another useful representation is the convolution representation. To generate or represent a WS process $\{Z(\mathbf{s}) : \mathbf{s} \in D \subset \mathbb{R}^d\}$, we can consider a white noise process $\epsilon(\mathbf{s})$ such that $E(\epsilon(\mathbf{s})) = \mu$, $Var(\epsilon(\mathbf{s})) = \sigma^2$ and $Cov(\epsilon(\mathbf{s}), \epsilon(\mathbf{s} + \mathbf{h})) = 0$. Then it is possible to describe a Gaussian stationary process by convolving $\varepsilon(\mathbf{s})$, with a square-integrable smoothing kernel function $k(\mathbf{s})$.

$$Z(\mathbf{s}) = \int_{\mathbb{R}^d} k(\boldsymbol{u} - \mathbf{s}) \varepsilon(\boldsymbol{u}) d\boldsymbol{u}$$

This representation is useful since the covariance associated depend only on the kernel. Higdon (1998), Fuentes (2002) and Paciorek and Schervish (2006) start from this setting to build closed form of stationary and non stationary covariance function.

2.7. MODELING SPATIAL DATA

A common spatial process model is constructed as follows

$$Y(\mathbf{s}_i) = \mu(\mathbf{s}_i) + Z(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \tag{2.14}$$

where $\mu(\mathbf{s}_i)$ is the mean of the response $Y(\mathbf{s}_i)$, typically of the form $X^T(\mathbf{s}_i)\beta$. $X(\mathbf{s}_i)$ is a p-dimensional vector of explanatory variables at location $\mathbf{s}_i$ and β is a p-dimensional vector of parameters. $Z(\mathbf{s})$ is a zero mean spatial process ($Z(\mathbf{s})$ is often assumed to be a WS Gaussian process with a parametric covariance function. and $\epsilon(\mathbf{s})$ is a pure error process with mean 0 and variance nugget). Thus, the spatial signal is decomposed into a determinist trend, the pure spatial variability explained by the covariance function and a pure error process explained by the nugget.

The model (2.14) can be viewed as a hierarchical model with a conditionally independent first stage given $Z(s)$ and $\mu(s)$. In the second stage, usually we assume $Z(s)$ to be a Gaussian random field with mean zero and certain parametric covariance structure.

In some situations, the response variable $Y(s)$ (even after transformation) is not appropriate to be treated as a normal random variable. For instance, $Y(s)$ might be a binary variable or a count

variable. It is natural to consider an extension of the model (2.14) analogous to the generalized linear model and considering at first stage the distribution of $Y(s_i)$ conditionally β and $W(s_i)$ belonging to the exponential family and consider $W(s_i)$ as a WS process at second stage (Diggle et al., 1998).

2.8. KRIGING

The main goal in spatial statistics is often interpolation. There exists different kinds of interpolators but kriging presents relevant advantages. For instance it provides some measure of the accuracy of the prediction with respect to deterministic interpolator such as splines or inverse distance method.

Kriging is a geostatistical interpolation technique that considers both the distance and the degree of variation between known data points when estimating values in unknown location points. A kriged estimate is a weighted linear combination of the known sample values around the point to be estimated. Applied properly, kriging allows the user to derive weights that result in optimal and unbiased estimates. Let us consider a random field $Z(\mathbf{s})$, $\mathbf{s} \in D \subset \mathbb{R}^d$ and the linear model

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + \epsilon(\mathbf{s}) \quad \mathbf{s} \in D, \tag{2.15}$$

where $\mu(\mathbf{s})$ is a deterministic function and $\epsilon(\mathbf{s})$ is a WS random process. We observe the process at n different locations, Z= $(Z(\mathbf{s}_1) \dots Z(\mathbf{s}_n))$, and wish to predict the process Z at an unobserved location $\mathbf{s}_0$. Let us denote with $p(\mathbf{s}_0, Z)$ the kriging interpolator at $\mathbf{s}_0$. The main properties of this object are:

- $p(\mathbf{s}_0, Z) = \sum_i^n \lambda_i Z(\mathbf{s}_i)$.
- $E(p(\mathbf{s}_0, Z)) = \mu(\mathbf{s}_0)$.
- $E((Z(\mathbf{s}_0) - p(\mathbf{s}_0, Z)^2)$ is minimum.

That is, kriging interpolator is optimal in the class of linear interpolator, *i.e* it is unbiased and with minimum variance error. Relaxing first condition, it is easy to show that the best interpolator is the conditional expectation of $Z(\mathbf{s}_0)$ given the observed data:

$$p(\mathbf{s}_0, Z) = E(Z(\mathbf{s}_0)|Z) \tag{2.16}$$

However, in general $p(\mathbf{s}_0, Z)$ is not a linear function of the data and establishing the statistical properties of the best predictor under squared-error loss can be difficult. Fortunately, if the random

field is Gaussian the best linear interpolator is also the best interpolator. Kriging appears in many forms and flavors, distinguished by whether the mean is known or not, what the distribution of is, whether predictions are made for points or areas and so forth. Here we describe classical ordinary kriging.

Let be $\mu(\mathbf{s}) = \mu$ and $\epsilon(\mathbf{s}) \sim (0, \mathrm{C})$ in (2.15), where μ is a unknown constant and C is known. We consider linear predictor of the form $p(\mathbf{s}_0, Z) = \lambda_0 + \lambda^T Z$, where $\lambda = (\lambda_1 \dots \lambda_N)$. Since we are looking for unbiased interpolator it is easy to show that it is equivalent to set $\lambda_0 = 0$ and $\sum_i^N \lambda_i = 1$.

Thus the problem now is to choose the λ weights that minimize:

$$E((\lambda^T Z - Z(\mathbf{s}_0))^2) \qquad \text{subject to} \quad \sum_i^N \lambda_i = 1.$$

This can be accomplished as an unconstrained minimization problem introducing the Lagrange multiplier m:

$$\underset{\lambda}{\operatorname{argmin}} E((\lambda^T Z - Z(\mathbf{s}_0))^2) - 2m(\sum_i^N \lambda_i - 1). \tag{2.17}$$

It can be shown that (Cressie, 1993) the solution to this problem is:

$$\lambda^T = c + 1(\frac{1 - 1^T C^{-1} c}{1^T C^{-1} 1})^T \mathrm{C}^{-1}$$

$$m = \frac{1 - 1^T C^{-1} c}{1^T C^{-1} 1}.$$

Thus the optimal linear predictor is,

$$p(\mathbf{s}_0, Z)_{OK} = \hat{\mu} + c^T C^{-1}(Z - 1\mu),$$

and the minimized mean-square prediction error, *i.e.* the ordinary kriging variance is:

$$\sigma^2(\mathbf{s}_0) = \mathrm{C}(0) - c^T C^{-1} c + \frac{1 - 1^T C^{-1} c}{1^T C^{-1} 1}.$$

When μ is known ordinary kriging is called simple kriging. Cressie (1993) discusses more complicated versions, such as lognormal and trans-Gaussian kriging, and universal kriging, used in a presence of non-stationary mean field model.

Chapter 3

Space-time covariance models

3.1. INTRODUCTION

Spatiotemporal models arise whenever data are collected across time as well as space. Therefore they have to be analyzed by considering both their spatial and temporal structures. Until recently, there has not been a theory of spatial-temporal processes separated from the established theories of spatial statistics and time series analysis. Unfortunately, statistical tools for the analysis of spatio-temporal processes are not as fully developed as methods for time series or spatial data alone. In Kyriakidis and Journel (1999) the geostatistical approach for space-time data is well described. In particular, two conceptual viewpoints are identified:

- approaches involving vectors of space random functions or vectors of time series.
- approaches involving a single spatiotemporal random function model.

In the first approach space-time data are analyzed through models initially developed for spatial (temporal) distributions and the joint space-time dependence is not fully modeled. It does not include the temporal (spatial) dependence and prediction at unknown locations (unknown time instants) is not possible. See for instance Rouhani and Wackernagel (1990) and Bogaert and Christakos (1997).

The second approach allow to consider the interactions between the spatial and the temporal components and thus interpolation of observations in a continuous space-time is possible. The natural domain for a geostatistical space-time model is $\mathbb{R}^d \times \mathbb{R}$, where $\mathbb{R}^d$ stands for space and $\mathbb{R}$ for time. As outlined in Gneiting et al. (2007) physically, there is clear-cut separation between

the spatial and the time dimensions, and a realistic statistical model will take account thereof. This contrasts with a purely mathematical perspective in which $\mathbb{R}^d \times \mathbb{R} = \mathbb{R}^{d+1}$ with no differences between the coordinates. The latter equality has important implications. In particular, all technical results on spatial covariance functions or on least-square prediction, or kriging, in Euclidean spaces apply directly to space-time problems, simply by separating a vector into its spatial and temporal components.

While literature on covariance models in the spatial context is consolidated, in the last ten years by starting from different settings and mathematical frameworks, several authors have produced considerable efforts in order to build valid and flexible space-time covariance models. Among many others, (Jones and Zhang (1997), Cressie and Huang (1999), Christakos (2000), De Cesare et al. (2001*a*), Gneiting (2002) and Stein (2005*a*)). The temporal component adds complexity to the covariance structure and new kinds of concepts have to be introduced such as separability or asymmetry in time.

In this chapter we review the main features and some parametric covariance models. We then introduce a new class of covariance function useful when dealing with anisotropy in space.

3.2. SPACE-TIME COVARIANCE FUNCTION FEATURES: STATIONARITY, SEPARABILITY AND FULL SYMMETRY

Let $Z(\mathbf{s}, t)$, $\mathbf{s} \in \mathbb{R}^d$ and $t \in \mathbb{R}$, be a real-valued spatio-temporal RF with mean $\mu(\mathbf{s}, t)$ and with constant and finite variance. Then, the function $(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2) \mapsto \mathrm{C}(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2)$ defined on the product space $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}$ is called the spatio-temporal covariance function of the process and, if no further assumptions are made, depends on the space-time coordinates $(\mathbf{s}_1, \mathbf{s}_2, t_1, t_2)$. As well known, a real valued function C defined on the product space $\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}$ is the covariance function associated to a spatio-temporal RF if and only if

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \mathrm{C}(\mathbf{s}_i, \mathbf{s}_j, t_i, t_j) \geq 0,$$

for all finite sets of real coefficients a_i and points $(\mathbf{s}_i, t_i) \in \mathbb{R}^d \times \mathbb{R}$, $i = 1, \ldots, n$. In practice, however, estimation and modeling of covariance models call for simplifying assumptions. Under the assumption of weak stationarity, we have that $\mu(\mathbf{s}, t) = \mu$ independently of the spatio-temporal coordinates and that the covariance function $\mathrm{cov}(Z(\mathbf{s}_i, t_i), Z(\mathbf{s}_j, t_j)) = \mathrm{C}(\mathbf{h}, u)$ is defined for $(\mathbf{h}, u) = (\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) \in \mathbb{R}^d \times \mathbb{R}$, thus it exclusively depends on the spatial and temporal

separation vectors, respectively $\mathbf{h}$ and u.

A stationary covariance function is called isotropic if it is rotation and translation-invariant, *i.e.*

$$\mathrm{C}(\mathbf{h}, u) = \widetilde{\mathrm{C}}(\|\mathbf{h}\|, |u|),$$

where with $\|.\|$ we denote the usual Euclidean norm and $\widetilde{\mathrm{C}}$ is a positive definite function.

Other very popular assumptions are that of separability and full symmetry.

A space-time covariance function is called *separable* if we can factor (Mitchell *et al.*, 2004)

$$\mathrm{C}(\mathbf{h}, u) = \frac{\mathrm{C}(\mathbf{h}, 0)\mathrm{C}(\mathbf{0}, u)}{\mathrm{C}(\mathbf{0}, 0)} \tag{3.1}$$

In other words, separability means that the spatio-temporal covariance structure factors into a purely spatial and a purely temporal component and thus it is very easy build these kind of models. Consequently, separable covariance models have been used even in situations in which they are not physically justifiable.

An interesting definition coined in Gneiting (2002) is that of full symmetry, which happens if

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

for every $(\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}$. Atmospheric, environmental and geophysical processes are often under the influence of prevailing air or water flows, resulting in a lack of full symmetry. For instance Gneiting (2002), Stein (2005*a*) and De Luna and Genton (2005), who considered the Irish wind data of Haslett and Raftery (1989), described the lack of full symmetry for these kind of data.

3.3. SOME SPACE-TIME COVARIANCE MODELS

Since it is very easy build valid space-time covariance models using (3.1), the first models used in the application were separable. Moreover they have computational gains since it can be shown that the matrix associated to this kind of covariance can be factorizes in the Kronecker product of the two matrix associated to the spatial and temporal covariances.

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

$$\mathrm{C}(\boldsymbol{h}, u) = \sigma^2 \exp\left(-c\|\mathbf{h}\| - a|u|\right), \tag{3.2}$$

where c, a are positive scale parameters and σ^2 turns out to be the variance of the RF. However the separable models are very limited because they do not model space-time interaction.

One of the first contributions to non-separable space-time covariance models is in Cressie and Huang (1999). It is based on the following theorem:

Theorem 2. *(Cressie and Huang) Let* C *a continuous, bounded, integrable and symmetric function on* $\mathbb{R}^d \times \mathbb{R}$. *Then* C *is a stationary covariance if and only if*

$$\rho(\boldsymbol{\omega}, u) = \int \exp(-i\mathbf{h}^T\boldsymbol{\omega})\mathrm{C}(\mathbf{h}, u)d\mathbf{h} \qquad u \in \mathbb{R} \tag{3.3}$$

is positive definite for almost all $\boldsymbol{\omega} \in \mathbb{R}^d$.

Using this result they builded valid models through closed form Fourier inversion in $\mathbb{R}^d$ and choosing a continuous positive definite function $\rho(\boldsymbol{\omega}, u)$ with $u \in \mathbb{R}$ for all $\boldsymbol{\omega} \in \mathbb{R}^d$. Gneiting (2002) gave a criterion that is based on this construction but does not depend on closed form Fourier inversion and does not require integrability. Specifically, using tools as completely monotonic functions and Bernstein functions, he proposes the following class:

$$\mathrm{C}(\mathbf{h}, u) = \frac{\sigma^2}{(\psi(|u|^2))^{d/2}}\varphi\left(\frac{\|\mathbf{h}\|^2}{\psi(|u|^2)}\right), \tag{3.4}$$

where φ is completely monotonic function, ψ a Bernstein function and σ^2 is the variance, while d denotes the dimension of the spatial domain. This class is particularly important as constitutes one of the main contributions for stationary nonseparable covariances for space-time data. Recall that a function $t \mapsto \varphi(t), t > 0$ is said to be completely monotonic if it is positive and

$$(-1)^n\varphi^n(t) \geq 0$$

for all n, a natural number, while Bernstein functions are positive real functions defined on the positive real line whose derivative is completely monotonic.

The specific choices $\varphi(t) = \sigma^2 exp(-ct^\gamma)$ and $\psi(t) = (1 + at^\alpha)^\beta$ yield the parametric family:

$$\mathrm{C}(\mathbf{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)^{\beta d/2}} \exp\left(-\frac{c\|\boldsymbol{h}\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta\gamma}}\right), \tag{3.5}$$

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

$$\mathrm{C}(\mathbf{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)} \exp\left(-\frac{c\|\boldsymbol{h}\|^{2\gamma}}{(a|u|^{2\alpha} + 1)^{\beta\gamma}}\right). \tag{3.6}$$

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

Instead of integration in the frequency domain, nonseparable covariance function can be builded by summation or integration in the spatiotemporal domain. Specifically we can obtain valid covariance functions by mixing spatial and temporal covariance function. For instance Ma (2002) considers a probability mass function π_{ij}, with $(i, j) \in Z_+$, of $[U, V]$ a bivariate discrete random vector with support on the non-negative integers. Then starting from a purely spatial correlation function $R_s(\mathbf{h})$ and a purely time correlation function $R_t(u)$ it can be shown that:

$$R(\mathbf{h}, u) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} R_s^i(\mathbf{h}) R_t^j(u) \pi_{ij} \tag{3.7}$$

is a valid nonseparable correlation function. This is possible since if $R(u)$ is a valid correlation in $\mathbb{R}^d$ then $R(u)^i$ is also valid for every positive integer i.

A related approach consists in consider product of covariance function making the spatial and temporal coordinates depend on one another. Let $[U, V]$ be a bivariate random vector with distribution function $F(u, v)$ (not necessarily discrete). If $[U, V]$ is independent of the spatial and temporal processes $Z_s(\mathbf{s})$ and $Z_t(t)$ which are independent of each other, then the scale mixture process $Z(\mathbf{s}, t) = Z_s(\mathbf{s}U) Z_t(tV)$ has covariance function:

$$\mathrm{C}(\mathbf{h}, u) = \int \mathrm{C}_s(\mathbf{s}v_1) \mathrm{C}_t(uv_2) dF(v_1, v_2) \tag{3.8}$$

and reducing to the univariate case we obtain:

$$\mathrm{C}(\mathbf{h}, u) = \int \mathrm{C}_s(\mathbf{s}v) \mathrm{C}_t(uv) dF(v) \tag{3.9}$$

Covariance models of this kind can be found in De Cesare et al. (2001*b*), Ma (2002) and De Cesare et al. (2002), Porcu et al. (2005) for instance. We will follow this approach in section 3.4 in building covariance function that are anisotropic in space.

The models defined by equations (3.2-3.5) are *fully symmetric*, *e.g.* they are isotropic in the spatial component and symmetric in the temporal one.

However this assumption is frequently violated, particularly when dealing with environmental data-sets. A substantive improvement of the previous construction can be achieved by 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), where the covariance is obtained

as the expected value of a stationary covariance function C_S on $\mathbb{R}^d$

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

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 (3.6) is:

$$\mathrm{C}(\boldsymbol{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha} + 1)} \exp\left(-\frac{c\|\boldsymbol{h} - \varepsilon u \boldsymbol{v}\|^{2\gamma})}{(a|u|^{2\alpha} + 1)^{\beta\gamma}} \right). \tag{3.10}$$

Here $\varepsilon \in \mathbb{R}$ controls the degree of asymmetry 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.

3.4. SPACE TIME COVARIANCE MODELS ANISOTROPIC IN SPACE: THE BERNSTEIN CLASS

Here we sketch our proposal, for which more details can be found in Porcu, Mateu and Bevilacqua (2007).

Consider $Z_i(s)$, $i = 1, \ldots, (d+1)$, $s \in \mathbb{R}$ and $d \in \mathbb{Z}_+$, univariate mutually independent continuous weakly stationary Gaussian random processes defined on the real line. In particular, let the process Z_{d+1} be continuously indexed by time. Consider also a $(d+1)$-dimensional nonnegative random vector $\boldsymbol{W} = (W_1, \ldots, W_{d+1})'$ with W_i independent of Z_i. Let the univariate covariances C_{s_i} and the temporal covariance C_t be respectively associated with Z_i, $i = 1, \ldots, d$ and Z_{d+1}. In the following we shall impose these covariances to be stationary, symmetric, and of the type $\mathrm{C}_{s_i}(h_i) = \exp(-\psi_i(|h_i|))$, $i = 1, \ldots, d$, and $\mathrm{C}_t(u) = \exp(-\psi_t(|u|))$, with ψ_i, ψ_t Bernstein functions. Positive definiteness of this construction is guaranteed by direct application of Schoenberg (Schoenberg, 1938), Theorem 1.

Then, we are interested in inspecting the properties of the following stationary spatio-temporal scale mixture-based random field, defined on $\mathbb{R}^d \times \mathbb{R}$,

$$Z(\mathbf{s}, t) = Z_{d+1}(tW_{d+1}) \prod_{i=1}^{d} Z_i(s_i W_i), \tag{3.11}$$

with $\mathbf{s} = (s_1, \ldots, s_d)' \in \mathbb{R}^d$ and $t \in \mathbb{R}$. It can be easily seen that the covariance structure associated

to this random field is nonseparable, as

$$\mathrm{C}(\mathbf{h},u)=\int_{\mathbb{R}_+^{d+1}}\exp\left(-\sum_{i=1}^{d}\psi_i(|h_i|)w_i-\psi_t(|u|)w_{d+1}\right)\mathrm{d}F(\boldsymbol{\omega}), \tag{3.12}$$

with $\mathbf{h}=(h_1,\ldots,h_d)^{'}\in\mathbb{R}^d$, $u\in\mathbb{R}$ and $\boldsymbol{\omega}=(w_1,\ldots,w_{d+1})^{'}\in\mathbb{R}^{d+1}$, and where F is the $(d+1)$-variate distribution function associated to the random vector $\boldsymbol{W}$. If F is absolutely continuous with respect to the Lebesgue measure, then previous representation can be reformulated with respect to the $(d+1)$-variate density, say f, that is

$$\mathrm{C}(\mathbf{h},u)=\int_{\mathbb{R}_+^{d+1}}\exp\left(-\sum_{i=1}^{d}\psi_i(|h_i|)w_i-\psi_t(|u|)w_{d+1}\right)f(\boldsymbol{\omega})\mathrm{d}\boldsymbol{\omega}.$$

Observe that this construction allows for the case of separability if and only if the integrating $(d+1)$-dimensional measure F (or equivalently its associated density f) factorises into the product of $(d+1)$ marginal ones, *i.e.* if the nonnegative random vector $\boldsymbol{W}$ has mutually independent components. This representation seems to be general enough and justifies its treatment for the construction of new space-time covariances.

We shall use this setting, or special cases of it, in order to build new classes of space-time covariance functions satisfying some desirable requirements. In particular, it would be nice to find criteria to obtain, through simple procedures, some closed forms that are analytically tractable and physically interpretable. In the following we shall argue that the construction we propose allows to introduce anisotropic components in the spatial lag vector. This aspect needs to be stressed properly, as only few contributions in the recent literature are devoted to the analysis of spatial anisotropy for nonseparable space-time covariance models.

The following result (Porcu, Mateu and Bevilacqua, 2007) gives a first insight in what we are looking for.

Proposition 1. *Let $\mathcal{L}$ be the Laplace transform of a nonnegative random vector (W_1,W_2). Let ψ_i, $i=1,\ldots,d$ and ψ_t be either positive Bernstein functions defined on the positive real line, or continuous, increasing and concave functions on the positive real line. Then,*

$$\mathrm{C}(\mathbf{h},u)=\mathcal{L}\left(\sum_{i=1}^{d}\psi_i(|h_i|),\psi_t(|u|)\right) \tag{3.13}$$

is a stationary nonseparable spatio-temporal covariance function in $\mathbb{R}^d\times\mathbb{R}$, for d any positive natural number.

The proof of Proposition 1 can be found in the Appendix.

It can be shown that (3.13) is a special case of a wider class of covariance functions introduced by Ma (2005) Theorem 3. Let us call this subclass Bernstein class. In the following we shall justify the reasons of a detailed analysis about this class, that possesses some features that need to be stressed properly:

1 The construction in Proposition 1 represents a special case of the general representation in equation (3.11). In particular, Proposition 1 shows the existence of a stationary random field $Z(\mathbf{s},t) = Z_t(tW_2)\prod_i Z_i(s_iW_1)$, with covariance function (3.13), and where F is the distribution function associated to $(W_1, W_2)'$.

2 Bernstein class can be used for both geometrically and zonally anisotropic models. This aspect has been somehow overcome by recent literature. Zonal anisotropy is traditionally modelled through the *sum model* (Rouhani and Hall, 1989), which has a very simple and easy-to-implement construction, but considers separately the dependence in different directions. Besides, this construction is variogram-based and not covariance-based. Dimitrakopoulos and Lou (1994) propose models for geometric anisotropy, and Fernández-Casal (2003) propose anisotropic models that are not easily interpretable as they are not obtainable in closed form, but can only be calculated numerically. Thus, Bernstein class can give some insights in this direction. It will be shown in the sequel that very simple closed forms can be obtained with straightforward procedures.

3 Covariance functions belonging to Bernstein class are not necessarily based on L_p metrics, which are desirable because of their easier interpretability, availability in software packages and well-established theoretical properties. At the same time, Euclidean distance is not always the best solution and its impact upon the statistical estimation and prediction can be significant, as emphasised by Banerjee (2003). Thus, Bernstein class can be of help even in this direction.

4 Bernstein class can be easily extended to the case of temporally asymmetric covariance functions, as their members admit an easy extension with respect to the Lagrangian framework as in Cox and Isham (1988). Indeed, suppose that the integrating measure F of the bivariate Laplace transform $\mathcal{L}$ is concentrated on the line $w_1 = w_2$ (thus $\mathcal{L} = \varphi$, where φ is a completely monotone function). Let the spatial covariance be of the form

$\mathrm{C}_s(h_1,\ldots,h_d) = \varphi\left(\sum \psi_i(|h_i|)\right)$. Now, by using the Lagrangian framework, *i.e.* by introducing a perturbation random variable V, that can be physically interpreted as a velocity scalar indicator, one easily obtains that $\mathrm{C}(h_1,\ldots,h_d,u) = \mathbb{E}_V \mathrm{C}_s(|h_1|,\ldots,|h_i - Vu|,\ldots,h_d) = \mathbb{E}_V \varphi(\sum_{j\neq i} \psi_j(|h_j|) + \psi_i(|h_i - Vu|))$ is a space-time covariance function, still belonging to the Bernstein class, but with asymmetry in time.

5 The permissibility condition in Proposition 1 is much less restrictive than that proposed in Theorem 3 of Ma (2005). It can be easily shown that it is much easier to check a function to be continuous, increasing and concave on the positive real line, than to check conditional negative definiteness in $\mathbb{R}^d$. Thus, implementation of this class is much easier for the practitioner who needs to implement new classes of anisotropic covariance functions.

6 Bernstein class has some mathematical properties that will be shown in the sequel, and that are linked to the use of linear operators, such as partial derivatives.

Some interesting special cases of Bernstein class can be obtained by imposing the integrating measure F of the bivariate Laplace transform $\mathcal{L}$ to be concentrated on the line $w_1 = w_2$. We present them as corollaries.

Corollary 1. *Let φ be a completely monotone function defined on the positive real line and such that $\varphi(0) = 1$. Let ψ_1, ψ_2, ψ_3 Bernstein functions, and γ_s and γ_t respectively, spatial and temporal intrinsically stationary variograms. Then, the functions*

$$(\mathbf{h}, u) \mapsto \varphi(\psi_1(\gamma_s(\mathbf{h})) + \psi_2(\gamma_t(u))) \tag{3.14}$$

and

$$(\mathbf{h}, u) \mapsto \varphi\left(\psi_3\left(\psi_1(\gamma_s(\mathbf{h})) + \psi_2(\gamma_t(u))\right)\right) \tag{3.15}$$

are stationary covariance functions on $\mathbb{R}^d \times \mathbb{R}$.

Corollary 2. *Let φ completely monotonic such that $\varphi(0) = 1$ and $\varphi^{(2k)}(0)$ exists and is finite for some positive integer k. For γ_s, γ_t respectively merely spatial and temporal intrinsically stationary variograms and ψ_1, ψ_2 Bernstein functions, the function*

$$(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, u) \mapsto \varphi^{(2k)}(\gamma_s(\mathbf{h}_1) + \gamma_t(u))\varphi(\psi_1(\|\mathbf{h}_2\|^2) + \psi_2(\|\mathbf{h}_3\|^2)) \tag{3.16}$$

is a stationary covariance function on $\mathbb{R}^d \times \mathbb{R}$, where $(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)' \in \mathbb{R}^d$.

The importance of Corollary 1 and 2 is that they allow, through a simple univariate scale mixture, for the construction of space-time covariances which are zonally anisotropic in the sense of Shapiro and Botha (1991) , who find zonally anisotropic models through componentwise isotropy. The advantage with respect to these authors approach is that, with our procedure, closed forms can be easily obtained. For instance, Corollary 2 can be used by those interested in the construction of anisotropic covariance models defined in $\mathbb{R}^3 \times \mathbb{R}$, as will be shown in the sequel.

Proceeding this way, various results can be obtained. For instance, an interesting property is that if χ is a negative definite function, indeed a function which admits the Lévy-Khinchin representation, and ψ a Bernstein function, then $\psi \circ \chi$ is negative definite. This fact allows to apply once again the above result (3.14), as for Schoenberg's theorem, $\exp(-\psi \circ \chi)$ is positive definite.

3.5. DIFFERENTIAL OPERATORS PRESERVING POSITIVE DEFINITENESS OF BERNSTEIN CLASS

As previously said, some nice mathematical features of the proposed structure can be obtained by using differential operators. This aspect is not new in the literature regarding permissible covariance functions that are isotropic, and that represent the characteristic function of a spherical symmetrically distributed random vector. Roughly speaking, the main problem was to find whether it is possible to preserve positive definiteness, even if not necessarily in the same d-dimensional Euclidean space, applying derivative or integral operators to positive definite radial functions with support in $\mathbb{R}^d$. For this reason Matheron (1965) coined the terms *Descente* and *Montée* to denote respectively differentiation and integration of a positive definite radial function. This topic is also important for the implications that these approaches have for the simulation of isotropic random fields by using the turning bands method.

The use of linear operators for the spatially anisotropic and spatio-temporal domain has been discussed in a recent paper by Porcu, Gregori and Mateu (2007). In this section, we shall show that partial derivative operators, applied to bivariate Laplace transforms, and in particular to the Bernstein class, induce new functions that preserve the permissibility in the spatio-temporal domain. The result is desirable as it gives further flexibility to the proposed class.

In order to understand the results, we need to introduce some standard notation for derivative

operators. Define

$$D_k\mathcal{L}(\theta_1,\theta_2)=\frac{d^{2k}}{d\theta_1^k d\theta_2^k}\mathcal{L}(\theta_1,\theta_2), \tag{3.17}$$

for the mixed derivative of $2k$-*th* order of the bivariate Laplace transform, with k a positive integer. Also, recall that any bivariate Laplace transform admit an integral representation as specified in formula (6.1) in the Appendix. We are then able to show the first result.

Proposition 2. *Let γ_s, γ_t be respectively purely spatial and temporal intrinsically starionary variograms. Also, let $\mathcal{L}$ as in (6.1) with F such that*

$$\int_0^\infty\int_0^\infty w_1^k w_2^k \mathrm{d}F(w_1,w_2)<\infty, \tag{3.18}$$

for some positive integer k. Then,

$$\nu(\mathbf{h},u)=\frac{D_k\mathcal{L}(\theta_1,\theta_2)|_{(\gamma_s(\mathbf{h}),\gamma_t(u))}}{D_k\mathcal{L}(\theta_1,\theta_2)|_{(0,0)}} \tag{3.19}$$

is a valid stationary nonseparable spatio-temporal covariance function on $\mathbb{R}^d\times\mathbb{R}$.

Next result shows another property of the *Bernstein* class in equation (3.13). Define

$$\mathcal{L}^*(\psi_1(s_1),\ldots,\psi_d(s_d),\psi_t(s_t))=\mathcal{L}\left(\sum_{i=1}^d\psi_i(s_i),\psi_t(s_t)\right),$$

and the operator $D_{d+1}\mathcal{L}^*$ as follows

$$D_{d+1}\mathcal{L}^*(s_1,\ldots,s_d,s_t)=\frac{d^{(d+1)}}{dx_1\ldots dx_d dx_t}\mathcal{L}^*\left(\psi_1(s_1),\ldots,\psi_d(s_d),\psi_t(s_t)\right), \tag{3.20}$$

where $s_1,\ldots,s_d,s_t$ are positive arguments and ψ_i, ψ_t Bernstein functions, $i=1,\ldots,d$.

Proposition 3. *Let $\psi_1,\ldots,\psi_d,\psi_t$ be Bernstein functions such that $\psi_i'(0)$, $\psi_t'(0)$ exist and are finite, for $i=1,\ldots,d$. Let $\mathcal{L}$ as in (6.1) with F such that*

$$\int_0^\infty\int_0^\infty w_1^d w_2 \mathrm{d}F(w_1,w_2)<\infty. \tag{3.21}$$

Also, let γ_i, γ_t be univariate variograms defined on the real line, such that $\gamma_i(0)=\gamma_t(0)=0$, $i=1,\ldots,d$. Then

$$\nu^*(h_1,\ldots,h_d,u)=\frac{D\mathcal{L}^*(\gamma_1(h_1),\ldots,\gamma_d(h_d),\gamma_t(u))}{D\mathcal{L}^*(0,\ldots,0,0)} \tag{3.22}$$

is a stationary nonseparable spatio-temporal covariance function on $\mathbb{R}^d\times\mathbb{R}$.

The proofs of both Propositions 2 and 3 can be found in the Appendix.

The following analytic examples show particular features of these new results.

Example 1. *In order to apply the result shown in equation (3.13), let us consider the following example. Following Shapiro and Botha (1991) and the extension to the spatio-temporal case (Fernández-Casal, 2003), a possible way to consider anisotropy is to split the lag vector* $\mathbf{h} \in \mathbb{R}^d$*, into two sub-vectors, say* $\mathbf{h}_1$*,* $\mathbf{h}_2$*, having respectively dimension* d_1 *and* $d_2 = d - d_1$*. With this setup it is possible to obtain, applying result in equation (3.13), an anisotropic spatio-temporal covariance function.*

Let $\mathcal{L}(\theta_1, \theta_2)$ *be the Laplace transform for the Frechet-Hoeffding lower bound of bivariate copulas, whose expression is*

$$\mathcal{L}(\theta_1, \theta_2) = \frac{\exp(-\theta_1) - \exp(-\theta_2)}{\theta_2 - \theta_1}, \tag{3.23}$$

with $\theta_1 \neq \theta_2$ *and* $\mathcal{L}(\theta, \theta) = \exp(-\theta)$*. Now, consider the following Bernstein functions*

$$\begin{aligned} \psi_1(t) &= (a_1 t^{\alpha_1} + 1)^{\beta} \\ \psi_2(t) &= \frac{(a_2 t^{\alpha_2} + b)}{b\,(a_2 t^{\alpha_2} + 1)} \\ \psi_t(t) &= t^{\rho}, \end{aligned}$$

where a_1, a_2 *are positive scale parameters,* $\alpha_1, \alpha_2 \in (0, 2]$*,* $\beta \in [0, 1]$*,* $b > 1$ *and* $\rho \in [0, 1)$*. With the proposed setup, and plugging the result from equation (3.13) into (3.23), we obtain that*

$$\mathrm{C}(\mathbf{h}_1, \mathbf{h}_2, u) = \sigma^2 k \frac{\exp\left(-(a_1\|\mathbf{h}_1\|^{\alpha_1} + 1)^{\beta} - \frac{a_2\|\mathbf{h}_2\|^{\alpha_2} + b}{b(a_2\|\mathbf{h}_2\|^{\alpha_2} + 1)}\right) - \exp(-|u|^{\rho})}{|u|^{\rho} - (a_1\|\mathbf{h}_1\|^{\alpha_1} + 1)^{\beta} - \frac{a_2\|\mathbf{h}_2\|^{\alpha_2} + b}{b(a_2\|\mathbf{h}_2\|^{\alpha_2} + 1)}} \tag{3.24}$$

is a permissible nonseparable space-time covariance which is spatially two-component anisotropic. Observe that with σ^2 *we denote the variance of the S/TRF, whilst* $k = (\frac{1}{2}(1 - exp(-2)))^{-1}$ *is a normalisation constant, so that if* $\sigma^2 = 1$*, then* $\mathrm{C}(\mathbf{0}, \mathbf{0}, 0) = 1$ *and thus (3.23) would be a space-time correlation function. The properties of this covariance function in terms of smoothness can be evaluated by looking at the margins* $\mathrm{C}(\mathbf{h}_1, \mathbf{0}, 0)$*,* $\mathrm{C}(\mathbf{0}, \mathbf{h}_2, 0)$ *and* $\mathrm{C}(\mathbf{0}, \mathbf{0}, u)$*. For instance, it is easy to see that none of these margins is differentiable at the origin.*

Example 2. *Following Corollary 2, completely monotonic functions can be used in order to implement easily a space-time dependence structure. For instance, the function* $t \to exp(-t)$ *is completely monotonic, but its characteristics in terms of non differentiability away from the origin*

makes it not very appealing, for the reasons explained by Stein (1999). A candidate completely monotone function having very nice mathematical features is the celebrated Cauchy class (Gneiting and Schlather, 2004), given by

$$\varphi(t) = (1 + ct^{\delta})^{-\nu}, \tag{3.25}$$

with $c > 0$, $\delta \in (0, 2]$, and ν a strictly positive parameter. This class is particularly important as it allows for decoupling the fractal dimension and the Hurst effect associated to the underlying RF. Unfortunately, for this class the result in Corollary 2 cannot be applied, as it is easy to show that the second derivative, which is still a completely monotone function, and of the form

$$\begin{aligned} \nu c\delta t^{-2}\{\left(1+ct^{\delta}\right)^{-\nu-2}\nu c\left(t^{2}\right)^{\delta}\delta-\left(1+ct^{\delta}\right)^{-\nu-1}t^{\delta}\delta+ \\ \left(1+ct^{\delta}\right)^{-\nu-1}t^{\delta}+\left(1+ct^{\delta}\right)^{-\nu-2}c\left(t^{2}\right)^{\delta}\delta\}, \end{aligned} \tag{3.26}$$

admits a finite limit at the origin only for some values of the parameters. By the way, one could use this class to obtain a covariance function that is anisotropic in space and asymmetric in time. For instance, a nice class of covariance models defined on $\mathbb{R}^2 \times \mathbb{R}$ can be obtained as follows: take $\psi_1(t) = (a_1 t^{\alpha_1} + 1)^{\beta}$ for the lag component h_1, and $\psi_3(t) = t^{\rho}$ for h_2, where the parameter setting is the same as in previous example. Thus, one easily gets that

$$\mathrm{C}(h_1, h_2, u) = \mathbb{E}_V(1 + c((a_1 h_1^{\alpha_1} + 1)^{\beta} + |h_2 - Vu|^{\rho})^{\delta})^{-\nu}$$

is a permissible class of covariance functions on $\mathbb{R}^2 \times \mathbb{R}$, that are spatially anisotropic and temporally asymmetric. The choice of the distribution function for the velocity variable V is crucial to obtain a closed form for the expectation above. Numerical solutions can be easily obtained, for instance by assuming that V is uniformly distributed on the unit interval. Finally, V could be fixed a priori, on the base of, for instance, physical knowledge of the process.

Example 3. *Consider the completely monotone function*

$$\varphi(t) = \left[e^{ct^{1/2}} + e^{-ct^{1/2}}\right]^{-\nu}, \tag{3.27}$$

where c, ν are positive parameters and $t > 0$. Writing $c = 1$, without loss of generality, we obtain

that the second derivative admits the expression

$$\varphi^{(2)}(t) = \frac{\frac{1}{4}\nu\left\{(\nu+1)\,\frac{1}{t}\left(e^{2\sqrt{t}}+e^{-2\sqrt{t}}-2\right)+t^{-3/2}\left(e^{2\sqrt{t}}-e^{-2\sqrt{t}}\right)-t^{-1/2}\left(e^{\sqrt{t}}+e^{-\sqrt{t}}\right)^2\right\}}{\left[\left(e^{\sqrt{t}}+e^{-\sqrt{t}}\right)\right]^{\nu+2}}, \tag{3.28}$$

which also admits a finite limit at the origin, whatever be the value of ν. Thus, applying Corollary 2, we obtain that for any choice of intrinsically stationary variograms γ_s, γ_t and Bernstein functions ψ_1, ψ_2,

$$\mathrm{C}(\mathbf{h},u) = \frac{\frac{1}{4}\nu\left\{\begin{array}{c}(\nu+1)\left(\gamma_s(\mathbf{h}_1)+\gamma_t(u)\right)^{-1}\left(e^{2\sqrt{\gamma_s(\mathbf{h}_1)+\gamma_t(u)}}+e^{-2\sqrt{\gamma_s(\mathbf{h}_1)+\gamma_t(u)}}-2\right)+\\ \left(\gamma_s(\mathbf{h}_1)+\gamma_t(u)\right)^{-3/2}\left(e^{2\sqrt{\gamma_s(\mathbf{h}_1)+\gamma_t(u)}}-e^{-2\sqrt{\gamma_s(\mathbf{h}_1)+\gamma_t(u)}}\right)\\ -\left(\gamma_s(\mathbf{h}_1)+\gamma_t(u)\right)^{-1/2}\left(e^{\sqrt{\gamma_s(\mathbf{h}_1)+\gamma_t(u)}}+e^{-\sqrt{\gamma_s(\mathbf{h}_1)+\gamma_t(u)}}\right)^2\end{array}\right\}}{\left[\left(e^{\sqrt{\gamma_s(\mathbf{h}_1)^2+\gamma_t(u)}}+e^{-\sqrt{\gamma_s(\mathbf{h}_1)+\gamma_t(u)}}\right)\right]^{\nu+2}\left[e^{(\psi_1(\mathbf{h}_2))^{1/2}}+e^{-(\psi_2(\mathbf{h}_3))^{1/2}}\right]^{\nu}}, \tag{3.29}$$

with $\mathbf{h} = (\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3)' \in \mathbb{R}^d$, is a nonseparable space-time covariance function, which is spatially three-component anisotropic.

Example 4. *In this example we propose some new covariance functions obtained through the following mixture*

$$\mathrm{C}(\mathbf{h},u) = \int_0^1\int_0^1 \exp(-w_1\gamma_s(\mathbf{h}) - w_2\gamma_t(u))d\mathcal{K}(w_1,w_2), \tag{3.30}$$

where $\mathcal{K}$ is a bivariate copula, i.e. *a bivariate distribution function defined on the unit square, with the peculiarity that its margins are uniform on the unit interval. As copulas are defined on a compact support, integration in (3.30) can be easy, as shown in the following lines. Consider, for instance, the following classes of copulas:*

1. *The product copula $\mathcal{K}(w_1,w_2) = w_1w_2$.*
2. *The copula B11 in Joe (1997) p. 148*

$$\mathcal{K}(w_1,w_2) = \delta min\left\{w_1,w_2\right\} + (1-\delta)w_1w_2, \tag{3.31}$$

where $\delta \in [0,1]$, so that (3.31) can be seen as a convex sum of two elementary copulas.

Applying (3.30) to the families of copulas just mentioned, the resulting covariances are of the form:

1. $$C_{\boldsymbol{s},t}(\boldsymbol{h},u) = \begin{cases} 1 & if \quad \boldsymbol{h}=\mathbf{0};\ \ u=0 \\ \frac{(1-e^{-\|\boldsymbol{h}\|^{\alpha}})(1-e^{-|u|^{\beta}})}{\|\boldsymbol{h}\|^{\alpha}|u|^{\beta}} & elsewhere, \end{cases} \tag{3.32}$$

   which is a separable spatio-temporal covariance function in $\mathbb{R}^d \times \mathbb{R}$, and where $\gamma_s(\mathbf{h}) = \|\mathbf{h}\|^{\alpha}$ and $\gamma_t(u) = |u|^{\beta}$, which are valid variograms for $0 \leq \alpha, \beta \leq 2$. This covariance function is differentiable at the origin for $\alpha, \beta > 1$ and it is integrable in $\mathbb{R}^d \times \mathbb{R}$ if and only if $\alpha > d$ and $\beta > 1$, with d the dimension of the spatial domain.

2. $$C_{\boldsymbol{s},t}(\boldsymbol{h},u) = \begin{cases} 1 & if \quad \boldsymbol{h}=\mathbf{0};\ \ u=0 \\ \delta\frac{\left(1-e^{-\gamma_s(\boldsymbol{h})-\gamma_t(u)}\right)}{\gamma_s(\boldsymbol{h})+\gamma_t(u)} + (1-\delta)\frac{\left(1-e^{-\gamma_s(\boldsymbol{h})}\right)\left(1-e^{-\gamma_t(u)}\right)}{\gamma_s(\boldsymbol{h})\gamma_t(u)} & elsewhere \end{cases} \tag{3.33}$$

   A similar structure for the covariance function can be obtained through the Kimeldorf and Sampson family (Joe, 1997).

Other families of copulas admit a more complicated form, which can be integrated only through computer intensive calculation. One such case is the Frank family (Genest, 1987; Joe, 1997).

Chapter 4

Covariance function estimation

4.1. INTRODUCTION

In the geostatistical framework, the estimation of the covariance is an important problem for space and space-time processes. For example, the best linear unbiased predictor (or, equivalently, the kriging predictor) of the process at a space-time location depends on knowledge of the covariance function of the process.

Since a covariance function must be conditionally positive definite, practical estimation generally requires the selection of some parametric class of covariance and the corresponding estimation of these parameters. Parametric covariance functions has been reviewed in chapter 2 and 3 in the space and space-time context.

Maximum likelihood (ML) and related techniques are generally considered the best method for estimating the parameters of covariance models. However, for a 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.

When the computation of the likelihood is infeasible, the most used estimation method is weighted least squares (WLS) (Cressie, 1985). This is often the case for space-time data (see, for instance, Cressie and Huang (1999) and Gneiting et al. (2007)).

Section 2 reviews classical methods for covariance function estimation in spatial context, such as maximum likelihood and least square estimation. For the sake of simplicity the methods are

described in the spatial context. The extension to space-time context is straightforward. Section 3 reviews recent proposal for covariance function estimation.

Section 4 introduces weighted composite likelihood (WCL) method which is builded starting from a proposed method by Curriero and Lele (1999). It is based on composite likelihood (Lindsay, 1988), *i.e.* a likelihood-based approximation indicating a general class of pseudo-likelihood based on the likelihood of marginal or conditional events.

The method introduce the concept of "optimal distance" to be consider before performing estimation. This is attained by considering cut-off weights and then maximizing the Godambe Information associated to WCL respect to these weights.

We argue WCL estimates are consistent and asymptotically Gaussian with asymptotic variance equal to the inverse of Godambe Information, under increasing domain regularity assumptions. Standard error estimation based on subsampling technique is suggested.

A model selection criteria based on WCL, exploiting recent results in Varin and Vidoni (2005), in introduced.

We show, through examples and simulation experiments, that WCL method represents a satisfactory approximation of ML and always induces gains in efficiency with respect to WLS method. At the same time, WCL allows for benefits from the computational point of view with respect to ML. The computational gains depend on the optimal distance.

A real data application on the Irish-wind speed dataset (Haslett and Raftery, 1989) is presented.

4.2. CLASSICAL ESTIMATION METHODS

4.2.1 Nonparametric estimation

Historically empirical estimation of the second order properties of a WS process has been performed on the variogram rather than the covariance function. This for practical reasons: for instance the empirical variogram does not require to estimate the mean and it is an unbiased estimator of the variogram.

The variogram and/or covariance function of the spatial process plays a key role in the geostatistical method. Non parametric technique as the empirical variogram estimator does not guarantee conditional negative definiteness. Finding an empirical estimate of the semivariogram is a science in itself, however. It is a basic tools in explorative data analysis in geostatistics. Among the non parametric estimators the empirical variogram obtained through the method of moment (Math-

eron, 1965), is the more famous. Moreover it is a necessary step in the least square estimation.

Assuming that points $(Z(\mathbf{s}_1), \dots Z(\mathbf{s}_N))$ are N observations from a WS stationary spatial RF and that they are disposed on a regular lattice, the estimate is

$$\hat{\gamma}(\boldsymbol{h}) = \frac{1}{2|\mathrm{N}(\boldsymbol{h})|} \sum_{(i,j)\in \mathrm{N}(\mathbf{h})} (Z(\boldsymbol{s}_i) - Z(\boldsymbol{s}_j))^2, \tag{4.1}$$

where $\mathrm{N}(\boldsymbol{h}) = \{(i,j) : \boldsymbol{s}_i - \boldsymbol{s}_j = \boldsymbol{h}\}$. 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. In this case an adapted version of $\mathrm{N}(\boldsymbol{h})$ could be $\mathrm{N}_\Delta(\boldsymbol{h}) = \{(i,j) : \boldsymbol{s}_i - \boldsymbol{s}_j \in Tol(\boldsymbol{h}_\Delta)\}$, 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. A practical rule sometimes met in practice is to consider only half the maximum distance in the data and at least 30 points for lag bins (Journel and Huijbregts, 1978).

As far as the statistical properties of the empirical variogram, we can note that it is unbiased since

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

However, when data are irregularly disposed and tolerance regions are considered a nonnegligible bias that strongly depends on the choice of the lag bins, is introduced.

It is difficult in general to determine distributional properties and moments of the variogram without some assumptions. However, if $Z(\mathbf{s})$ is a Gaussian random field, it can be shown (see for instance Smith (2001)) that:

$$Cov(\gamma(\hat{h}_1), \gamma(\hat{h}_2)) = \frac{2}{|N(h_1)||N(h_2)|} \sum_{(i,j)\in \mathrm{N}(\mathbf{h}_1)} \sum_{(l,k)\in \mathrm{N}(\mathbf{h}_2)} (\gamma(\mathbf{s}_i - \mathbf{s}_l) - \gamma(\mathbf{s}_j - \mathbf{s}_l) + \gamma(\mathbf{s}_j - \mathbf{s}_k) - \gamma(\mathbf{s}_i - \mathbf{s}_k))^2 \tag{4.2}$$

One objection to the classical empirical variogram is that, like many methods based on sample averages, it is not robust against outlying values of Z. In this direction go some methods for robust semivariogram estimation as for instance (Cressie and Hawkins, 1980).

4.2.2 Least square methods

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.

LS method requires a preliminary step, *e.g.* obtaining a nonparametric estimate of the variogram.

Then, the following generalised sum of squares is minimised

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

where $\widehat{\boldsymbol{\gamma}}(\mathbf{h}) = (\hat{\gamma}(\mathbf{h}_1), \ldots, \hat{\gamma}(\mathbf{h}_m))'$, $\boldsymbol{\gamma}(\mathbf{h};\boldsymbol{\theta}) = (\gamma(\mathbf{h}_1;\boldsymbol{\theta}), \ldots, \gamma(\mathbf{h}_m;\boldsymbol{\theta}))'$ and $\boldsymbol{R}$ is a definite positive matrix. Here $\boldsymbol{R}$ is $\mathbb{C}\mathrm{ov}(\widehat{\boldsymbol{\gamma}}(\mathbf{h}))$, obtained for instance through (4.2). Generalized least square (GLS) estimation is performed, as outlined by Muller (1999), with an iterative re-weighting scheme which requires the inversion of a $K \times K$ matrix, where K is the number of lags, at each step.

Cressie (1985), for computational reasons, proposed to replace the matrix $\boldsymbol{R}$ by a diagonal matrix where the diagonal entries are calculated as

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

Thus, instead of minimising the generalised sum of squares, minimise the weighted sum of squares

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

The WLS estimate has evident computational gains but it depends on the $\hat{\boldsymbol{\gamma}}(\mathbf{h}_k)$, that are correlated. Moreover it can be easily shown that the estimating function associated to WLS is biased.

When data are irregularly spaced, the nonnegligible bias deriving from using of subjective tolerance regions impact the estimation of the covariance parameter. This drawback is even bigger when we want to estimate anisotropic parametric models as in section 3.4. Moreover, in the case of space-time covariance functions built throughout the Lagrangian framework (Stein, 2005*a*), it is not clear how to proceed with WLS estimation. In this sense, when data are not regularly disposed, a fitting procedure based on variogram cloud could be preferred.

4.2.3 Maximum likelihood

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)). Let $Z = (Z(\mathbf{s}_1), \ldots Z(\mathbf{s}_N))'$ be N observations from a zero-mean Gaussian RF. The log-likelihood function can be written as

$$l(\boldsymbol{\theta}) = -\log|\Sigma(\boldsymbol{\theta})| - Z'\Sigma(\boldsymbol{\theta})^{-1}Z, \tag{4.3}$$

where $\Sigma(\boldsymbol{\theta}) = \mathbb{V}\mathrm{ar}(Z)$ 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 (Mardia and Marshall (1984)). However applications of ML methods to spatial data have been criticised for

computational reasons. 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. ML estimation have appealing statistical properties but in finite sample they are often biased. To correct for this drawback restricted maximum likelihood (REML) has been proposed. The idea of REML estimation is to estimate covariance parameters by maximising the likelihood of KZ where the matrix of contrasts K is chosen so that $E(KZ) = 0$.

4.3. ASYMPTOTIC CONSIDERATIONS

There are two quite different asymptotic frameworks in spatial statistics to which one can appeal: increasing domain asymptotics, in which the minimum distance between sampling points is bounded away from zero and thus the spatial domain of observation is unbounded, and infill asymptotics, in which observations are taken ever more densely in a fixed and bounded domain.

Asymptotic behavior of estimation methods of covariance function has been considered in both the framework or in a mixture of them.

ML estimation in classical statistical theory is considered the best estimation essentially for asymptotics reasons. In spatial settings classical asymptotics does not apply since the observations are tipycally correlated. However results under increasing domain are similar to the classical statistical theory framework.

Consider a spatial process $Z(s)$ that is observed on a set of n points D_n , with $D_1 \subset D_2 \subset \ldots \subset \mathbb{R}^d$, and whose distribution depends on the parameter $\theta \in \mathbb{R}^p$, where p and d are fixed positive integers. Let $l(\theta)$ be the likelihood function defined in (4.3) and let $\hat{\theta}_{ML}$ be the maximum likelihood estimator. If $Z(s)$ is a Gaussian random field, Mardia and Marshall (1984) showed that, under an increasing domain asymptotic framework and subject to some regularity conditions, $\hat{\theta}_{ML}$ is approximately normally distributed with mean θ and covariance the Fisher information matrix. Cressie and Lahiri (1996) showed that this result holds for the REML estimator under certain regularity conditions.

A similar general result do not exist on infill asymptotics. Consider a stationary, zero-mean, Gaussian process observed in $\mathbb{R}$ with covariance function as in (2.6) with $\theta_0 = 0$ and $\rho(h, \theta)$ as in (2.7), *i.e.* an exponential covariance model. When this process is observed in the unit interval,

Ying (1993) showed that, as $n \longrightarrow \infty$,

$$\sqrt{n}(\hat{\sigma^2}_{ML}\hat{\theta}_{ML} - \sigma^2\theta) \longrightarrow N(0, 2(\sigma^2\theta)^2) \tag{4.4}$$

However, the individual parameters (range and variance) are not consistently estimable under the infill asymptotic framework. Zhang (2004) generalized this result and showed that not all parameters in a Matern covariance function are consistently estimable under infill asymptotics.

Cressie and Lahiri (2002) describes asymptotic results of least square estimators under increasing domain and a mixed form of increasing and infill.

Typically in practice, spatial data are observed at a finite number of points with no intention or possibility of taking more observations, and it is not clear which asymptotic framework to appeal to. Stein (1999) argues that if interpolation of the spatial process is the ultimate goal, infill asymptotics must be preferred. Zhang and Zimmerman (2005) compared the two asymptotic frameworks and showed that for some parameters, the infill asymptotics seems preferable.

4.4. RECENT ESTIMATION METHODS

Gains in computation of (4.3) can be obtained exploiting the particular structure of the covariance matrix, thus without resorting to approximations. Zimmerman (1989) outlines that under certain sampling schemes it is possible to compute the likelihood more efficiently. For example, when the sampling locations form a regular lattice of R rows and C columns, the covariance matrix is block Toeplitz, and matrix inversion algorithms exist which reduce the required number of computations to $O(R^2C^3)$, from $O(R^3C^3)$ for an arbitrary $RC \times RC$ matrix (Akaike, 2003). If the sampling locations form a regular rectangular lattice and the covariance function is separable, then the covariance matrix is a Kronecker product of two symmetric Toeplitz matrices, and the inverse can be found using $O(R^2C^2)$ computations.

Another way in which calculations for lattice data can be simplified is via the spectral representation (2.11) of the process Z. Whittle (1954) proposed an approximation of the likelihood based on the frequency domain rather than the sampling domain.

Fuentes (2007) extended this idea to construct an approximation for irregularly space data, dividing the spatial domain into a lattice of blocks, then working with the process obtained by integrating Z over each of the blocks. However some concerns regarding the accuracy associated with approximation in the spectral domain (e.g., the likelihood of Whittle (1954), and with the

ad hoc creation of the periodogram (e.g., how many low frequencies are ignored) leads to a careful use of these tools for inference.

Kaufmann et al. (2007) focus directly on the covariance matrix and propose to compute (4.3) with a "tapered" version of the covariance matrix instead of the original covariance. Tapering is performed by multiplying (in the Schur sense) the covariance matrix by a particular compact correlation matrix. The goal is to obtain matrix forms which can be manipulated using efficient sparse matrix algorithms. Furrer et al. (2007) use this technique for data interpolation rather than estimation.

The computational problem has been attacked with likelihood approximations such as composite likelihood. Three kinds of composite likelihood have been proposed in the literature to approximate the full likelihood for large spatial data sets. (Curriero and Lele (1999), Caragea and Smith (2006), Vecchia (1988) and its modified version of Stein et al. (2004)).

Caragea and Smith (2006) propose to approximate the likelihood through three different methods called "Big block", "Small Block" and "Hybrid" methods. They consist in splitting the full region in sub-regions and then consider 1) the likelihood for the means over each of the subregions, 2) the likelihood for the observations, assuming subregions are independent, or 3) the likelihood for the observations, assuming they are conditionally independent given the means of the subregions.

Vecchia (1988) proposed to factor the full likelihood into a product of conditional densities and then reduce the size of conditioning sets as an approximation. Stein et al. (2004) improved and extended this idea to REML estimation and compared this kind of estimator with ML by using the associated Godambe Information, since the derivative of the approximate log-likelihood forms an unbiased estimating equation.

Curriero and Lele (1999) approach is the starting point for the WCL method and it is introduced in the next section.

4.5. WEIGHTED COMPOSITE LIKELIHOOD METHOD

4.5.1 Our proposal

Estimation of the variogram historically has been made with the empirical variogram. As outlined by Genton (1998) a practical rule sometimes met in practice is to consider only half the maximum distance in the data and at least 30 points, when building the empirical variogram. This rule of thumb seems to be based in tradition and it is at least questionable. However, it suggests that not

considering large distances in the estimation could improve the "quality" of the estimates.

Following this idea, it is of interest, for computational and efficiency reasons, looking for estimators based on differences considering an optimal distance (in some sense) in the estimation. For this purpose we need to define an objective criteria when looking for this optimal distance. It is probably better building estimators based on variogram cloud since the empirical variogram is not unbiased in general.

Curriero and Lele (1999) introduced a particular form of composite likelihood for covariance estimation based on variogram cloud, 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, it is robust towards misspecification of the distribution and its order of computation lies between the ML and WLS ones.

Let $Z(\mathbf{s}, t)$ be a weakly stationary RF with a parametric variogram function $2\gamma(\mathbf{h}, u; \boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \Theta \subset R^p$ is an unknown $p-$dimensional parameter vector. Following Curriero and Lele (1999), we consider all possible differences $U_{ij} = Z(\mathbf{s}_i) - Z(\mathbf{s}_j)$, $i \neq j$. We assume that the distribution of these differences is Gaussian, $U_{ij} \sim \mathcal{N}(0, 2\gamma_{ij}(\boldsymbol{\theta}))$, where $\gamma_{ij}(\boldsymbol{\theta}) = \gamma(\mathbf{s}_i - \mathbf{s}_j; \boldsymbol{\theta})$. The negative log-likelihood of a difference is:

$$l(U_{ij}; \boldsymbol{\theta}) = \log \gamma_{ij}(\boldsymbol{\theta}) + \frac{U_{ij}^2}{2\gamma_{ij}(\boldsymbol{\theta})}.$$

Summing up all the possible differences, we get the composite likelihood (CL for short) function as proposed in Curriero and Lele (1999),

$$\mathrm{CL}(\boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j>i}^{N} l(U_{ij}; \boldsymbol{\theta}). \tag{4.5}$$

This method does not require any inversion matrix and the order of computations is $O(N^2)$. To obtain estimates of $\boldsymbol{\theta}$ we can maximise the function $CL(\boldsymbol{\theta})$ or equivalently solve the estimating equation

$$\sum_{i=1}^{N} \sum_{j>i}^{N} \nabla l(U_{ij}; \boldsymbol{\theta}) = \sum_{i=1}^{N} \sum_{j>i}^{N} \frac{\nabla \gamma_{ij}(\boldsymbol{\theta})}{\gamma_{ij}(\boldsymbol{\theta})} \left(1 - \frac{U_{ij}^2}{2\gamma_{ij}(\boldsymbol{\theta})}\right) = 0, \tag{4.6}$$

where $\nabla \gamma_{ij}(\boldsymbol{\theta}) = \nabla \gamma(\mathbf{s}_i - \mathbf{s}_j; \boldsymbol{\theta})$ is the vector of partial derivative with respect to the vector $\boldsymbol{\theta}$. Note that this is an unbiased estimating equation, irrespectively of the distributional assumptions imposed on U_{ij}. If the fourth-order joint distributions of U_{ij} is known, it would be possible to come up with an optimal way of combining the individual score $\nabla l(U_{ij}; \boldsymbol{\theta})$. One can stack all the

individual score into a vector $v(\boldsymbol{\theta})$ and form the optimal estimating equation

$$(\mathbb{E}\nabla v(\boldsymbol{\theta}))^{T}[\mathbb{C}\mathrm{ov}(v(\boldsymbol{\theta}))]^{-1}v(\boldsymbol{\theta}) = 0.$$

However, the computation of the covariance matrix $\mathbb{C}\mathrm{ov}(v(\boldsymbol{\theta}))$ is of order $O(N^2) \times O(N^2)$, and its inversion is computationally prohibitive for large N.

Now we look at a modification to the composite likelihood (4.5), which aims to improve the efficiency of the estimates and reduce the computational burden. Instead of searching optimal weights we consider a modified version of (4.5) namely

$$\mathrm{WCL}(\boldsymbol{\theta}, d_s) = \frac{1}{W_{N,d_s}} \sum_{i}^{N} \sum_{j>i}^{N} l(U_{ij};\boldsymbol{\theta}) w_{ij}(d_s), \tag{4.7}$$

and its associated estimating equation

$$e_{\mathrm{WCL}}(\boldsymbol{\theta}, d_s) = \frac{1}{W_{N,d_s}} \sum_{i}^{N} \sum_{j>i}^{N} \nabla l(U_{ij};\boldsymbol{\theta}) w_{ij}(d_s) = 0, \tag{4.8}$$

where $W_{N,d_s} = \sum_{i}^{N} \sum_{j>i}^{N} w_{ij}(d_s)$ and $w_{ij}(d_s) = 1$ if $\|\mathbf{s}_i - \mathbf{s}_j\| \leq d_s$, , and 0 otherwise.

We define the Godambe information (Godambe, 1991) or sandwich information matrix of (4.8) as

$$G(\boldsymbol{\theta}, \boldsymbol{d}) = H(\boldsymbol{\theta}, d_s) J(\boldsymbol{\theta}, d_s)^{-1} H(\boldsymbol{\theta}, d_s)^{T},$$

with $H(\theta, d_s) = \mathbb{E}[\nabla e_{\mathrm{WCL}}(\theta, d_s)]$, $J(\boldsymbol{\theta}, d_s) = \mathbb{E}[e_{\mathrm{WCL}}(\boldsymbol{\theta}, d_s) e_{\mathrm{WCL}}(\boldsymbol{\theta}, d_s)^{T}]$ and

$$H(\boldsymbol{\theta}, d_s) = \mathbb{E}[\nabla e_{\mathrm{WCL}}(\theta, d_s)] = \frac{1}{W_{N,d_s}} \sum_{i}^{N} \sum_{j>i}^{N} \frac{\nabla \gamma_{ij}(\boldsymbol{\theta}) \nabla \gamma_{ij}(\boldsymbol{\theta})^{T}}{\gamma_{ij}^2(\boldsymbol{\theta})} w_{ij}(d_s), \tag{4.9}$$

$$\begin{aligned} J(\boldsymbol{\theta}, d_s) &= \mathbb{E}[e_{\mathrm{WCL}}(\boldsymbol{\theta}, d_s) e_{\mathrm{WCL}}(\boldsymbol{\theta}, d_s)^{T}] \\ &= \frac{1}{W_{N,d_s}^2} \sum_{i}^{N} \sum_{j>i}^{N} \sum_{l}^{N} \sum_{k>l}^{N} \frac{\nabla \gamma_{ij}(\boldsymbol{\theta}) \nabla \gamma_{lk}(\boldsymbol{\theta})^{T}}{4\,\gamma_{ij}^2(\boldsymbol{\theta}) \gamma_{lk}^2(\boldsymbol{\theta})} \mathbb{C}ov(U_{ij}^2, U_{lk}^2) w_{ij}(d_s) w_{lk}(d_s). \end{aligned} \tag{4.10}$$

Note that if the RF is Gaussian, it turns out to be $\mathbb{C}ov(U_{ij}^2, U_{l,k}^2) = 2(\gamma_{il}(\boldsymbol{\theta}) - \gamma_{jl}(\boldsymbol{\theta}) + \gamma_{jk}(\boldsymbol{\theta}) - \gamma_{ik}(\boldsymbol{\theta}))^2$ so we can evaluate readily expression (4.10).

In order to improve the efficiency we choose the 'lag' d_s minimising the inverse of $G(\boldsymbol{\theta}, d_s)$ in the partial order of nonnegative definite matrices. Using a scalar equivalence result (Heyde, 1997) we seek the optimal d_s^* such that

$$d_s^* = \underset{d_s \in \mathcal{D}}{\operatorname{argmin}}\, \mathrm{tr}(G^{-1}(\boldsymbol{\theta}, d_s)), \tag{4.11}$$

where $\mathcal{D} = \{(d_s) : \min_{i\neq j} \|s_i - s_j\| \leq d_s \leq \max_{i\neq j} \|s_i - s_j\| \}$ and $\text{tr}(A)$ denotes the trace of the matrix A. To save computation time, instead of computing the Godambe information for every observed lag, we could restrict the search space and choose a fixed number of lags.

Then the WCL estimator is defined as $\hat{\boldsymbol{\theta}}_{d_s^*} = \underset{\boldsymbol{\theta}\in\Theta}{\text{argmin}}\ WCL(\boldsymbol{\theta}, d_s^*)$

Optimal cut-off weights allow us to establish an objective criteria on leading up the "optimal distance" mentioned before.

Note that we need to know $\boldsymbol{\theta}$ to perform this task. In Section 4.5.3 we suggest a method to solve this problem.

Asympotics results for the WCL can be derived following the same arguments as in Heagerty and Lele (1998). In particular note that in their proof these authors have considered a weighted version of the pairwise likelihood that is exactly our weighted scheme. Thus WCL estimator $\hat{\boldsymbol{\theta}}_{N,d_s^*}$ turns out to be consistent and asymptotically normal under increasing domain asymptotics and the conditions stated in Heagerty and Lele (1998), namely for $N \to \infty$

$$J(\boldsymbol{\theta}, d_s^*)^{-1/2} H(\boldsymbol{\theta}, d_s^*)^T (\hat{\boldsymbol{\theta}}_{N,d_s^*} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(0, I_p). \tag{4.12}$$

4.5.2 The spatio-temporal case

A natural extension of the WCL method consists in considering all the spatio-temporal differences: $U_{ij} = Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j)$, $i \neq j$. Then, assuming Gaussianity of the differences $U_{ij} \sim \mathcal{N}(0, 2\gamma_{ij}(\boldsymbol{\theta}))$, where $\gamma_{ij}(\boldsymbol{\theta}) = \gamma(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j; \boldsymbol{\theta})$ is a valid space-time variogram model, we obtain the WCL for space-time data:

$$\text{WCL}(\boldsymbol{\theta}, \boldsymbol{d}) = \frac{1}{W_{N,\boldsymbol{d}}} \sum_{i}^{N} \sum_{j>i}^{N} l(U_{ij}; \boldsymbol{\theta}) w_{ij}(\boldsymbol{d}), \tag{4.13}$$

and its associated estimating equation

$$e_{\text{WCL}}(\boldsymbol{\theta}, \boldsymbol{d}) = \frac{1}{W_{N,\boldsymbol{d}}} \sum_{i}^{N} \sum_{j>i}^{N} \nabla l(U_{ij}; \boldsymbol{\theta}) w_{ij}(\boldsymbol{d}) = 0, \tag{4.14}$$

where $W_{N,\boldsymbol{d}} = \sum_i^N \sum_{j>i}^N w_{ij}(\boldsymbol{d})$ and $w_{ij}(\boldsymbol{d}) = 1$ if $\|\mathbf{s}_i - \mathbf{s}_j\| \leq d_s$, $|t_i - t_j| \leq d_t$, and 0 otherwise, $\boldsymbol{d} = (d_s, d_t)^T$.

Then, we propose to minimise $WCL(\boldsymbol{\theta}, \boldsymbol{d}^*)$ respect to $\boldsymbol{\theta}$, where

$$\boldsymbol{d}^* = \underset{\boldsymbol{d}\in\mathcal{D}}{\text{argmin}}\ \text{tr}(G^{-1}(\boldsymbol{\theta}, \boldsymbol{d})), \tag{4.15}$$

where $\mathcal{D} = \{(d_s, d_t) : \min_{i\neq j} \|s_i - s_j\| \leq d_{sk} \leq \max_{i\neq j} \|s_i - s_j\|, \min_{i\neq j} |t_i - t_j| \leq d_{tk} \leq \max_{i\neq j} |t_i - t_j|]\}$ and where $G(\boldsymbol{\theta}, \boldsymbol{d})$ is the Godambe Information associated to the weighted estimating function.

4.5.3 On $\boldsymbol{d}$ identification, standard error estimation and information criteria based on WCL

Before performing optimisation of $WCL(\boldsymbol{\theta}, \boldsymbol{d})$, we need to identify the optimal spatial lag $\boldsymbol{d}^*$. In our experience we have found useful choosing the WLS estimate $\hat{\boldsymbol{\theta}}$ as a preliminary and consistent estimate.

However the evaluations of the plug-in estimates $\hat{H} = H(\hat{\boldsymbol{\theta}}, \boldsymbol{d})$ and $\hat{J} = J(\hat{\boldsymbol{\theta}}, \boldsymbol{d})$ are of order $O(W^2_{N,\boldsymbol{d}})$ and $O(W^4_{N,\boldsymbol{d}})$ respectively, and these become computationally infeasible for large datasets.

In this case we can resort to subsampling techniques as in Heagerty and Lumley (2000). More precisely, let $D_1, \ldots, D_k$ be K overlapping subsets of $D = \{(s_1, t_1), \ldots, (s_N, t_N)\}$. If we assume that for $N \to \infty$,

$$W_{N,\boldsymbol{d}}\, J(\boldsymbol{\theta}, \boldsymbol{d}) \;\to\; J_\infty,$$

we can estimate J from an estimate of J_∞, namely

$$\widehat{J}_\infty = \frac{1}{K}\sum_{k=1}^{K} \frac{1}{W^{(k)}_{N,\boldsymbol{d}}} \sum_{(i,j)\in D_k} \sum_{(i',j')\in S_k} \nabla l(U_{ij}, \hat{\boldsymbol{\theta}})\, \nabla l(U_{i'j'}, \hat{\boldsymbol{\theta}})^T w_{ij}(\boldsymbol{d}) w_{i',j'}(\boldsymbol{d}) \tag{4.16}$$

where $W^{(k)}_{N,\boldsymbol{d}} = \sum_{(i,j)\in S_k} w_{ij}(\boldsymbol{d})$. To select an appropriate subset size in the spatial case, we refer to Heagerty and Lumley (2000) and the references therein.

In the space-time context much research had addressed the choice of optimal space-time subset. Since a typical situation in environmental studies is when we dispose of few spatial locations observed in many temporal instants, considering temporal windows of fixed dimensions seems the natural choice in this case. Li et al. (2007) extend Carlstein (1986) to determine the optimal length when using overlapping temporal windows:

$$l(n) = \left(\frac{2\beta}{1-\beta^2}\right)^{\frac{2}{3}} \left(\frac{3n}{2}\right)^{\frac{1}{3}}$$

where we estimate β by $\hat{\beta} = C(\hat{0}, 1)/C(\hat{0}, 0)$

This approach assumes that the statistic of interest is the simple mean and that the temporal correlation follows an $AR(1)$ with parameter β. This procedure often works well in practice.

Finally, the estimates of J and H allow us to propose a model selection criterion. An important issue in the geostatistical approach is the selection of an appropriate covariance model, taking into account the trade off between goodness-of-fit and model complexity. Model selection criteria as AIC and BIC have been proposed in both spatial (Hoeting et al., 2006) and space-time (Huang et al., 2007) settings. Nevertheless, they depend on the computation of the likelihood function. In

our framework we can follow Varin and Vidoni (2005) and we propose a model selection criterion based on the weighted composite likelihood. The WCL information criterion selects the model maximising

$$WCLIC(\widehat{\boldsymbol{\theta}}_{\boldsymbol{d}^*}, \boldsymbol{d}^*) = \text{WCL}(\widehat{\boldsymbol{\theta}}_{\boldsymbol{d}^*}, \boldsymbol{d}^*) + \text{tr}(\hat{J}\hat{H}^{-1}). \tag{4.17}$$

The effectiveness of this criterion will be illustrated in section 4.6.

4.5.4 Examples

Let us first consider an example for a process with a Cauchy variogram model of type :

$$\gamma(d_u; b) = 1 - \frac{1}{1 + d_u/b}, \quad b > 0, d_u \in \mathbb{R}. \tag{4.18}$$

where the parameter b controls the spatial range. Figure 4.1 shows $G^{-1}(b, d_u)$ as a function of d_u for different values of b on a transect $[0, 60]$ equally spaced by 0.5. We note that the best lag is always the smallest ($d_u^* = 0.5$) and that the gain of WCL versus CL is more evident when the spatial dependence is stronger ($b = 3$).

To analyse the asymptotic distribution of b, we simulated 1000 spatial random fields with a Cauchy model with $b = 1$ and then perform estimation with $WCL(0.5)$ and CL. Figure 4.2 shows the results by comparing the distribution of the estimates versus the theoretical ones.

Note that the estimated distribution of b under $WCL(0.5)$ match the asymptotic approximation (4.12).

Second, let us consider an exponential variogram with equation

$$\gamma(d_s; \theta) = 1 - \exp\left(-3\frac{d_s}{\theta}\right), \quad \theta > 0. \tag{4.19}$$

In this case θ is a positive scalar representing the practical range of the spatial correlation, *e.g.* the minimum lag at which the spatial correlation of the underlying RF is negligible.

Figure 4.3-a shows the inverse of the Godambe Information as a function of h for 49 location points on a 7×7 regular grid in a square $[1, 4]^2$ for different values of θ. We can see that the variance increases with the spatial lag h, and that attains the minimum at the first lag available ($d_s^* = 0.5$). Note that the gain is more evident when θ increases. Figure 4.3-b considers the same setting but under an irregular grid. Note that under this spatial scheme, it seems dangerous to take d_s too small, and that the optimal d_s^* is proportional to the spatial dependence, as could be expected. To check asymptotic results, we simulated 1000 spatial random fields, under the former regular

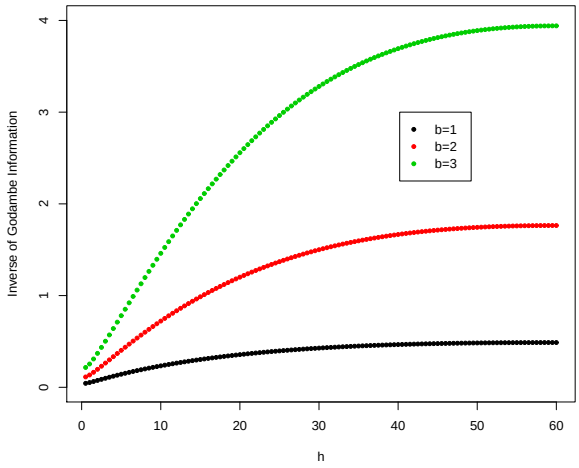


Figure 4.1: Inverse of the Godambe Information matrix for a Cauchy model on a transect $[0, 60]$ with sampling points equally spaced by 0.5, with $b = 1, 2, 3$.

scheme, with an exponential covariance model with $\theta = 2$ and then estimated with $WCL(0.5)$ and CL. Figure 4.4 describes the results of simulations by comparing the distribution of the estimates versus the theoretical one. Again the estimated distribution match the asymptotic approximation. The gain of WCL versus CL is evident in terms of variability.

Both examples shows a clear gain in efficiency when estimating with WCL and that asymptotic result in 4.2 seems to be a reasonable approximation for the distribution of the parameters.

4.6. SIMULATED DATA

In this section we describe a simulation study with the aim of examining and comparing the performances of CL, WLS and WCL with respect to ML ones.

The methods are compared in terms of mean squared errors (MSEs), calculated over 300 realisations of stationary Gaussian RF, $\{Z(\mathbf{s}, t)\}$, with zero mean and unitary variance, where $\mathbf{s} \in \{1, 1.5, 2, \ldots, n\}^2$ with $n = 3, 4, 5$ (*i.e.* 25, 49 and 81 points observed on a regular grid equally spaced), and $t \in \{1, \ldots, T\}$, with $T = 15, 30, 45$. The number of the observations was chosen in order to make rather simple evaluating the full likelihood. We have considered three different models for the covariance:

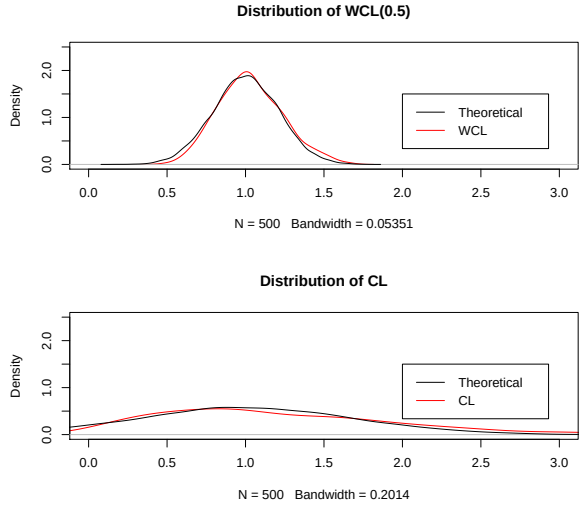
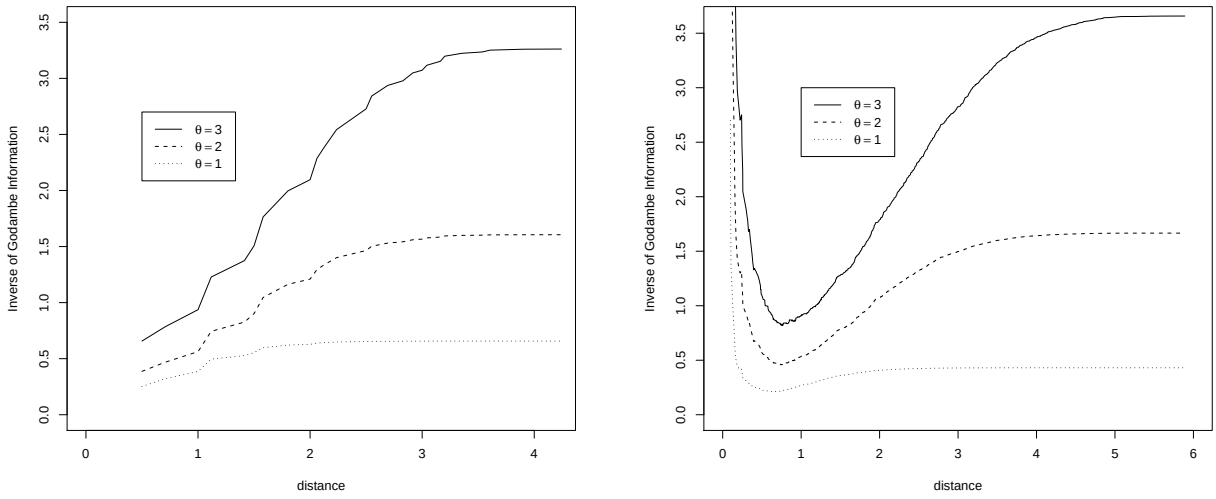


Figure 4.2: Distribution of 1000 WCL(0.5) estimates from the model model Cauchy with $b = 1$ ($mean = 1.031, var = 0.0424$) versus the theoretical distribution (mean=1 var=0.0439), and distribution of 1000 CL estimates ($mean = 1.215, var = 0.635$) versus the theoretical distribution ($mean = 1, var = 0.491$)



(a) (b)

Figure 4.3: Inverse of the Godambe Information matrix for an exponential variogram model. We considered 49 spatial locations on $[1, 4]^2$ and $\theta = 1, 2, 3$ with sampling points: (a) a regular 7×7 grid, (b) 49 sites randomly located.

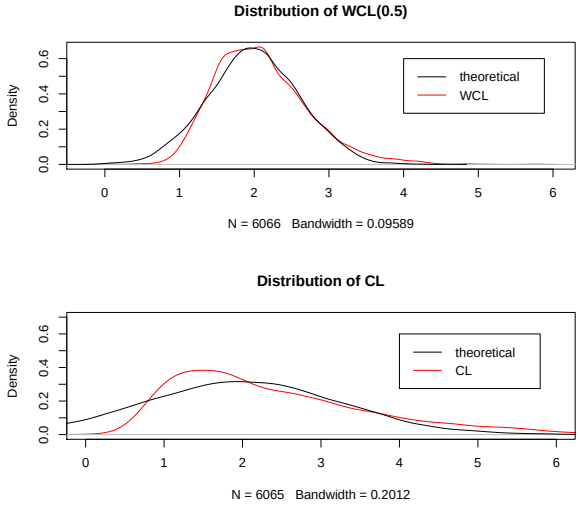


Figure 4.4: Distribution of 1000 WCL(0.5) estimates from an exponential model with $\theta = 2$ ($mean = 2.07, var = 0.391$) versus the theoretical distribution (mean=2 var=0.385), and distribution of 1000 CL estimates ($mean = 2.41, var = 1.87$) versus the theoretical distribution ($mean = 2, var = 1.60$)

1. a spatial model like (4.19) with $\theta = 1, 2, 3$;

2. a space-time separable model like (3.2), with range parameters $a = c = 2$;

3. a nonseparable and fully symmetric space-time model like (3.6), with $a = c = 2$, smoothing parameters $\alpha = \gamma = 0.5$ and nonseparating parameter $\beta = 0.5$.

WCL estimation is performed after the identification of $\boldsymbol{d}^*$ using (4.11) and WLS estimates as preliminary estimates

Table 4.1 shows the relative efficiency (i.e. the ration between the MSEs) of the WLS, CL, and WCL estimates with respect to ML ones for the exponential model with different values of parameters.

	$n = 25$			$n = 49$			$n = 81$		
	WLS	CL	WCL	WLS	CL	WCL	WLS	CL	WCL
$\theta = 1$	2.570	3.139	1.463	3.250	4.279	1.538	5.271	6.616	1.611
$\theta = 2$	2.743	3.202	1.252	3.916	4.510	1.271	6.133	7.504	1.303
$\theta = 3$	2.821	3.359	1.170	4.206	4.698	1.194	7.355	8.506	1.221

Table 4.1: Relative efficiency, based on MSE, for WLS, CL and WLC estimation methods with respect to ML, when model (4.19) is used with $\theta = 1, 2, 3$.

In this example we have plugged-in the WLS estimates into the Godambe information without resorting to subsampling for estimating J. The identified distance d_s^* always coincides with the optimal true one which is equal to 0.5 for this spatial design.

The table suggests that the gain of WCL estimates respect to CL ones increases when the spatial range is higher ($\theta = 3$). Moreover we note that if the number of points increases the relative efficiencies of CL and WLS estimates get worse while the WCL one is nearly constant.

Tables 4.2 and 4.3 show the relative efficiency of the WLS, CL and WCL estimates with respect to ML estimate for the separable and non separable covariance modes, respectively. In this cases identification of $\boldsymbol{d}^*$ is performed using subsampling tecniques with temporal windows as explained in section 4.5.3.. WCL* indicates the relative efficiency for the weighted likelihood when we choose $\boldsymbol{d}$ assuming known the true value $\boldsymbol{\theta}$.

Again WCL estimates have the best performances and its relative efficiency does not get worse

respect to ML increasing the observations in space and/or in time. Moreover the discrepancy between WCL* and WCL increases for the Gneiting model (3.6) mirroring the incertitude of the preliminary estimates.

In Table 4.4 (resp. 4.5) we give more details of the behaviour of the relative efficiency of WCL estimates with respect to some combination of spatial and temporal lags for model (3.2) (resp. model (3.6)) when $n = 49$ $T = 30$. The tables confirm that optimal space-time lag is $\boldsymbol{d} = (0.5, 1)^T$ and that when increasing the space-time lag relative efficiency get worse.

Finally, we show a small example of model identification using the weighted composite likelihood information criterion defined in (4.17). We have considered 100 independent simulations from a stationary Gaussian RF with zero mean, unitary variance and covariance model (3.10) with $n = 16$ points on a regular spaced grid $S = \{1, 2, 3, 4\}^2$ and $t = 1, \ldots, 150$. In our idea this choice (a small number of locations and a relatively large number of measures in times) reflects a common setting in environmental studies.

We set three different models. In (3.10) we fix:

Model A $\gamma = 0.5$, $\beta = 0$, $\varepsilon = 0$ fixed ; $c = 1$, $a = 1$, $\alpha = 0.5$ to be estimated

Model B $\gamma = 0.5$, $\varepsilon = 0$ fixed ; $c = 1$, $a = 1$, $\alpha = 0.5$, $\beta = 0.7$ to be estimated

Model C $\gamma = 0.5$ fixed; $c = 1$, $a = 1$, $\alpha = 0.5$, $\beta = 0.7$, $\varepsilon = -0.5$ to be estimated

We estimate with $WCL(1, 2)$. Note that these models are nested. The outcomes of these simulations are summarised in Table 4.6 that reports the times that a particular model is identified with respect to the model choosen for the simulations, i.e. the 'true' model. These results point out that the WCLIC seems a promising identification criterion because in this case at least 80% of the models have been correctly identified.

4.7. A REAL DATA EXAMPLE

We have considered the daily wind speeds collected over 18 years (1961-1978) at 12 sites in Ireland and originally analysed in Haslett and Raftery (1989). Due to his size and wholeness this dataset has been examined by several authors, for instance Gneiting (2002); De Luna and Genton (2005); Stein (2005*a*,*b*) and recently Gneiting et al. (2007) have found evidence that the data violate the assumptions of full symmetry and separability.

		$n = 25$				$n = 49$				$n = 81$			
		WLS	CL	WCL	WCL*	WLS	CL	WCL	WCL*	WLS	CL	WCL	WCL*
	c	6.25	7.72	2.29	1.63	8.75	12.81	1.84	1.73	10.70	13.28	1.84	1.68
$T = 15$	a	24.01	42.12	9.22	7.52	33.21	45.38	8.49	8.38	34.23	44.63	10.21	9.85
	c	7.03	7.96	2.11	2.01	9.20	10.68	1.98	1.96	11.28	14.95	2.15	2.03
$T = 30$	a	22.15	30.83	8.36	8.15	38.79	49.78	10.29	10.23	41.41	50.74	10.30	9.23
	c	6.83	7.62	2.01	1.96	7.14	7.76	1.70	1.69	11.25	12.77	2.27	2.02
$T = 45$	a	26.36	35.01	9.56	8.62	31.35	33.66	9.32	9.29	43.71	53.19	9.85	9.29

Table 4.2: Relative efficiency, based on MSE, for WLS, CL , WCL and WCL* estimation methods with respect to ML, when model (3.2) is used with $c = 2, a = 2$ and $\sigma^2 = 1$.

		$n = 25$				$n = 49$				$n = 81$			
		WLS	CL	WCL	WCL*	WLS	CL	WCL	WCL*	WLS	CL	WCL	WCL*
	c	12.96	16.33	4.17	2.74	21.83	28.69	4.39	2.72	28.51	37.61	7.10	2.72
$T = 15$	a	12.85	21.59	4.38	3.18	20.35	30.53	4.92	3.53	25.19	33.71	6.95	4.04
	c	19.23	23.95	4.32	3.17	26.01	32.59	6.79	2.73	33.53	41.81	7.30	2.95
$T = 30$	a	25.34	40.91	5.23	3.27	33.39	46.59	5.79	3.82	39.91	49.95	7.23	3.95
	c	20.25	25.05	4.85	2.95	33.04	41.54	6.56	2.64	39.10	47.97	8.01	2.84
$T = 45$	a	27.34	41.83	4.35	2.83	40.96	51.12	5.18	2.94	46.86	58.53	6.42	3.01

Table 4.3: Relative efficiency, in terms of MSE, for WLS, CL , WCL and WCL* estimation methods when model (3.6) is used with $c = 2, a = 2, \gamma = 0.5, \alpha = 0.5, \beta = 0.5, \sigma^2 = 1$.

		$d_t \leq 1$	$d_t \leq 2$	$d_t \leq 3$	$d_t \leq 4$
	c	1.96	2.02	2.03	2.05
$d_s \leq 0.5$	a	11.93	13.59	14.93	15.56
	c	2.88	3.05	3.09	3.11
$d_s \leq 1$	a	16.19	20.33	22.96	24.38
	c	4.70	5.12	5.29	5.36
$d_s \leq 1.5$	a	20.08	25.95	30.20	32.95
	c	6.20	6.91	7.23	7.49
$d_s \leq 2$	a	21.96	28.52	33.80	37.45

Table 4.4: Relative efficiency, based on MSE, for WCL estimates at different spatial and time lags when model (3.2) is used ($c = 2$, $a = 2$, ans $\sigma^2 = 1$), with $n = 49$ and $T = 30$.

		$d_t \leq 1$	$d_t \leq 2$	$d_t \leq 3$	$d_t \leq 4$
	c	2.73	3.16	3.44	3.66
$d_s \leq 0.5$	a	3.82	4.91	6.11	7.12
	c	4.23	5.02	5.53	5.93
$d_s \leq 1$	a	5.39	7.24	9.19	10.89
	c	7.00	8.53	9.53	10.31
$d_s \leq 1.5$	a	6.42	8.94	11.56	13.95
	c	9.04	11.22	12.69	13.87
$d_s \leq 2$	a	6.70	9.49	12.40	15.11

Table 4.5: Relative efficiency, based on MSE, for WCL estimates at different spatial and time lags when model (3.6) is used ($c = 2$, $a = 2$, and $\sigma^2 = 1$) with $n = 49$ and $T = 30$.

		Identified		
		A	B	C
	A	81	14	5
True	B	6	80	14
	C	3	11	86

Table 4.6: Times that a particular model is identified with respect to true model using the WCLIC.

Here we do not want propose a new analysis, but we would like illustrate the effectiveness of the weighted likelihood approach. Following Haslett and Raftery (1989) we have omitted the Rosslane station and as in Gneiting et al. (2007) we have considered only the first ten years of observations as a training set and the next 8 years as a validation set. Then we have considered a square root transformation of the data and removed a seasonal component. The seasonal component has been 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. After these tranformations, Haslett and Raftery (1989) argue convincingly that a stationary model seems sensible.

The inspection of the empirical marginal covariances suggests a spatial nugget effect, thus a discontinuity in $C(\mathbf{h}, u)$ 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 (Gneiting et al., 2007)

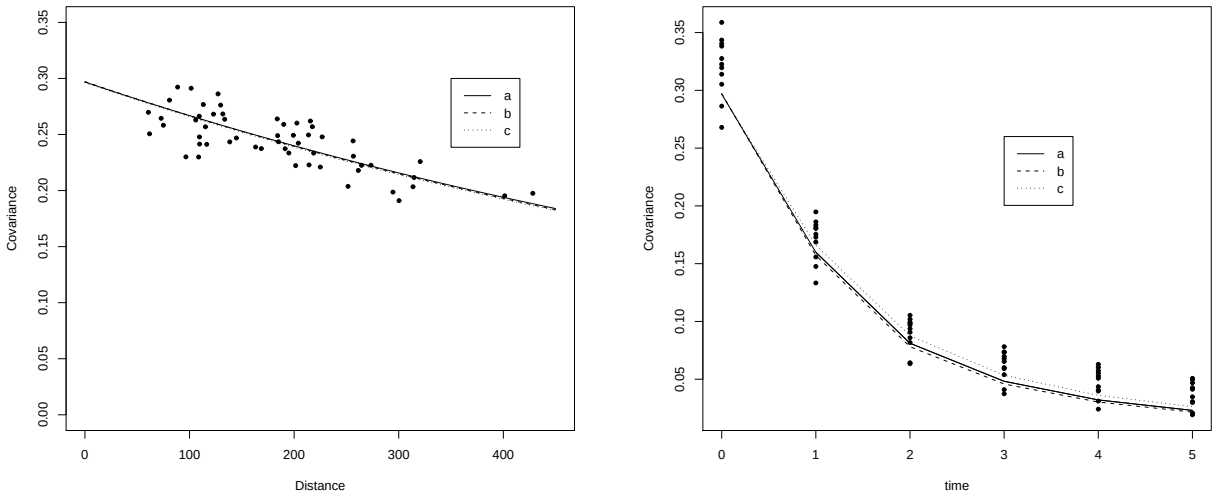
$$\bar{C}(\mathbf{h}, u) = \sigma^2(1-\nu)R(\mathbf{0}, u)\left(\frac{R(\mathbf{h}, u)}{R(\mathbf{0}, u)} + \frac{\nu}{1-\nu}\delta_{\mathbf{h}=\mathbf{0}}\right), \tag{4.20}$$

where $R(\mathbf{h}, u)$ is a space-time correlation function, $\nu \in [0, 1]$ and $\sigma^2 > 0$ 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 WCL, we estimated the spatio-temporal structure of the Irish wind speed data using three covariance models for R$(\mathbf{h}, u)$:

(a) a separable structure, obtained by considering the model in equation (3.6) with $\beta = 0$ and $\gamma = 0.5$;

(b) a nonseparable model, obtained by considering the same model in equation (3.6), with $\gamma = 0.5$



(a) (b)

Figure 4.5: Empirical and fitted marginal (spatial and time) covariance for model (a),(b),(c)

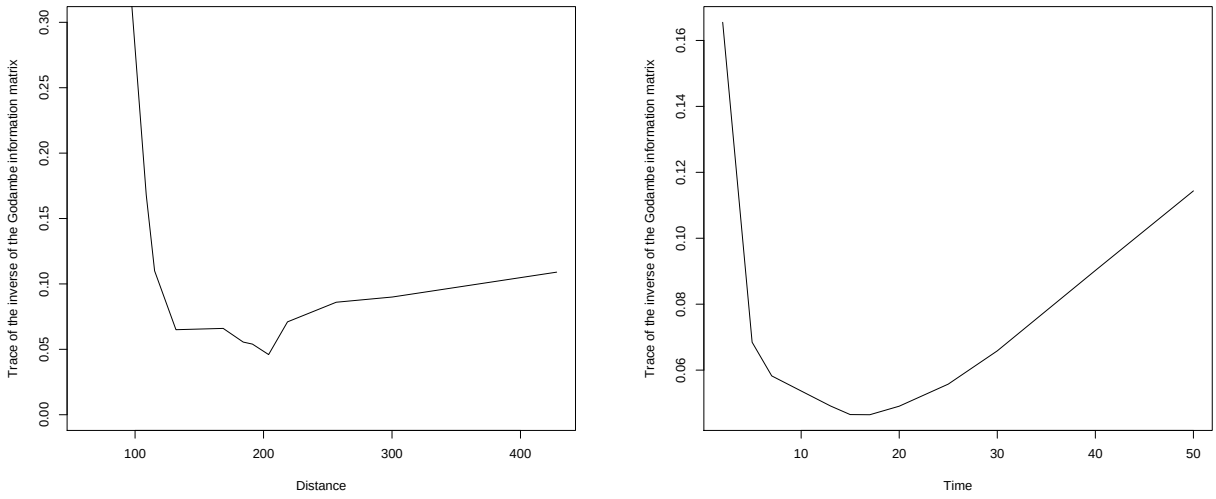
but with a non-vanishing β;

(c) a nonseparable, and not fully symmetric model as in equation (3.10) with $\gamma = 0.5$.

Because we have 40150 observations, ML estimate is infeasible and the aforemetioned authors resort on WLS estimates bounding the numbers of lags. In this case computation of CL is very slow. Using a preliminary WLS estimate we have identified $\boldsymbol{d} = (204, 15)^T$, where the spatial distance is expressed in miles. In Figures 4.6(a)-(b) we illustrate the behaviour of the Godambe information matrix for the nonseparable model with respect to the spatial distance h (at fixed time lag $u = 15$) and with respect to the time lag u (at fixed patial distance $h = 204$).

We summarize the estimation results for model (a),(b),(c) in Table 4.7. Standard errors are estimated using subsampling with a temporal window of 30 days. The fitted marginal covariances show a good agreement with respect to the empirical ones (see Figure 4.5). Note that estimation of spatial and time ranges, spatial nugget and variance are similar in the three models making the fitted covariances virtually indistinguishable, but the results indicate that the fully simmetric hypotesis ($\varepsilon = 0$) should be rejected.

These findings are confirmed considering the predictive performance of the three models. We



(a) (b)

Figure 4.6: Trace of the inverse of the Godambe information matrix for the nonseparable model: (a) with respect to the spatial distance h (at fixed time lag $u = 15$); (b) with respect to the time lag u (at fixed patial distance $h = 204$).

	(a)	(b)	(c)
a	0.8564	0.8912	0.7851
	$(5.00e-02)$	$(5.19e-02)$	$(2.56e-02)$
c	0.00106	0.00107	0.00108
	$(5.26e-05)$	$(5.35e-05)$	$(5.55e-05)$
α	0.81707	0.82666	0.79910
	$(8.71e-02)$	$(8.50e-02)$	$(4.57e-02)$
υ	0.0592	0.05805	0.0593
	$(6.15e-03)$	$(6.13e-03)$	$(1.30e-03)$
β	-	0.6428	0.7933
		$(2.14e-01)$	$(2.43e-01)$
ε	-	-	142.2
			$(5.42e-01)$
σ^2	0.3154	0.3155	0.3149
	$(7.08e-03)$	$(7.04e-03)$	$(5.67e-03)$
$WCLIC$	2656608	2656618	2656995

Table 4.7: $WCL(204, 15)$ estimates and relative standard errors for the Irish wind speed data: (a) separable model, (b) Gneiting nonseparable model and (c) nonseparable and not fully symmetric model.

have considered as in (Gneiting et al., 2007) 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, $\hat{z}(\mathbf{s}, t)$, using the simple kriging predictor, with the covariance function estimated with $WCL(\hat{d})$ in the training period. The predictor variables are the 33 velocity measures observed during the past three days at the eleven stations.

To assess and rank the point forecasts, we have used the prediction mean square error (PME) defined as $PME(\mathbf{s}) = (\sum_{t=3651}^{6570}(z(\mathbf{s}, t) - \hat{z}(\mathbf{s}, t))^2/2920)^{1/2}$.

Looking at the PME results in Table 4.8 the asymmetric in time model scores the best results.

Evidence in the rejection of the full symmetric hypothesis is given also by the WCLIC values

Station	(a)	(b)	(c)
Roche's Point	0.483	0.479	0.465
Valentia	0.500	0.500	0.500
Kilkenny	0.435	0.431	0.417
Shannon	0.463	0.463	0.460
Birr	0.476	0.475	0.470
Dublin	0.447	0.443	0.441
Claremorris	0.490	0.491	0.490
Mullingar	0.427	0.425	0.417
Clones	0.483	0.480	0.470
Belmullet	0.494	0.494	0.494
Malin Head	0.495	0.491	0.486

Table 4.8: PMEs for the stations of the Irish wind speed data for WCL

for the three different models as we can see considering the last row in Table 4.7.

4.8. DISCUSSION

Following Varin and Vidoni (2005) taxonomy, there are two classes of composite likelihood. The first one includes "subsetting methods" and it is based on marginal or subset "pieces" of full likelihood, while the second is based on "omission methods", since composite likelihood is obtained by omitting components to the full likelihood. "Omission methods" intuitively allow for computational benefits since we do not consider pieces of likelihood when computing the composite likelihood. Three kinds of composite likelihood have been proposed in the literature to approximate the full likelihood for large spatial data sets. Curriero and Lele (1999) and Caragea and Smith (2006) approaches belong to the class of subsetting methods, while the approximation of Vecchia (1988) and its modified version of Stein et al. (2004) belong to the second class.

The order of computation of the three "block" methods of Caragea and Smith (2006) is approximately $O(N^2)$, while the method proposed in Stein et al. (2004) can be made arbitrarily close to $O(N)$ depending on the size of the conditioning sets.

Both methods present some drawbacks: the method in Caragea and Smith (2006) can be infeasible for huge data sets, while in the method of Stein et al. (2004) there is a certain subjectiveness

in choosing conditional and unconditional sets. Both present a general trade-off between statistical efficiency and computational burdens.

WCL belongs to the class of "omission methods". Its statistical efficiency and computational burden depends on the optimal spatial lag. The order of computation for $WCL(\boldsymbol{\theta}, \boldsymbol{d}^*)$ is $O(W^2_{N,\boldsymbol{d}^*})$. That is, the computational gain depends on the position of $\boldsymbol{d}^*$ in the sort vector of empirical space-time lags . If $\boldsymbol{d}^*$ is the smallest lag, we have the best computational efficiency. If $\boldsymbol{d}^*$ is the largest empirical lags (maximum space-time distance), we have $W^2_{N,\boldsymbol{d}^*} = N(N-1)/2$, that is the classical CL.

For example, in the simulation study in Section 5 we consider 81 regularly spaced locations points observed in 45 times. For this particular setting, the optimal space-time lag is $\boldsymbol{d} = (0.5, 1)^T$. If $k = 81 \times 45$, ML requests in this case $\frac{1}{3}(k)^3$ floating point operations while classical CL requests $\frac{k(k-1)}{2}$ operations. WCL(0.5,1) involved 3645 operations. Thus WCL requires less than 3×10^{-7} of the floating point operations with respect to ML and 4×10^{-4} with respect to CL. Differently from Caragea and Smith (2006) and Stein et al. (2004) approaches , WCL seems do not suffer from the trade-off between computational burden and efficiency. For instance, in the examples 1 and 2 in Section 4.4, $\boldsymbol{d}^*$ is the smallest. At this lag we have both the best statistical efficiency and the best computational efficiency.

Moreover WCL does not suffer from any kind of subjectiveness in choosing "pieces" of likelihood. The price to pay is the estimation of optimal spatial lag before the estimation of covariance parameter. It could be interesting to compare these different likelihood approximations.

An clear evidence of this method is that tanking into account all possible differences between observed sites in the estimation of classical CL, does not necessarily increase the efficiency. These findings suggest a way for improving the efficiency of other kinds of estimators based on differences such as WLS. In this regard we could improve the results given by the usual practice (Journel and Huijbregts, 1978) of using lags which are half the maximum distance and have a certain number of points. This extension to WLS could be nontrivial, since the associated estimating function is biased and the empirical variogram depends on the empirical choice of lag bins. Note that even in the space-time context WLS has been implemented considering a fixed distance in space and time without a clear justification (see, for instance, Gneiting et al. (2007)).

Finally we remark that practical rules for choosing optimal lag are needed dealing with huge data-sets such as the case of space-time data and this will be a topic for future research.

Chapter 5

WCL applications

5.1. INTRODUCTION

In this section we present three application of WCL method: the first one consists in estimation of particular spatial covariance functions which allows for negative values, the second one consists in estimation of covariance function describing residuals dependence in dynamic life tables, the third one is a proposal for a simulation based test to verify separability of some parametric covariance models in a space time setting. In the first two applications WCL is used without resorting to optimal distance, that is we use the Curriero and Lele (1999) method. The test is based on WCL as described in section 4.

5.2. MODELLING RESIDUALS DEPENDENCE IN DYNAMIC LIFE TABLES: A GEOSTATISTICAL APPROACH

5.2.1 Introduction

In many countries all over the world, mortality forecasts are used to create and modify retirement pension schemes, disability insurance systems and other social security programmes. During the 20th century huge increases in life expectancy have followed medical and scientific breakthroughs. Forecasting mortality usually serves practical purposes because improvements therein could have enormous social and financial implications.

The graduation of mortality data by means of parametric methodology has been widely addressed in papers such as Forfar et al. (2002), Renshaw (1991) and Debón et al. (2005) and by means of non-parametric methodology in Gavin et al. (1993, 1994, 1995) and Debón, Montes and

Puig (2006). All these papers deal with static graduation, that is to say the influence of age on data graduation. However, mortality progresses over the years, but earlier methods do not take this fact into account, as they were designed to analyse data corresponding to one year in particular or, in the case of several years, they worked with aggregate figures. The concept of a dynamic table seeks to solve this problem by jointly analysing mortality data corresponding to a series of consecutive years. This approach allows the calendar effect's influence on mortality to be studied. A good sign that the representation of the evolution of mortality via dynamic models is an extremely important current issue, as much for actuaries as for statisticians and demographers, are Benjamin and Pollard (1991), Tabeau (2001), Pitacco (2004), Wong-Fupuy and Haberman (2004), Debón, Montes and Sala (2006*a*) and Booth (2006), where models that have arisen in the recent years are described. Most of these methods adapt traditional laws to the new situation and none of them take the dependence structure existing among the data into account, that some authors consider must be modeled (Booth et al., 2002; Renshaw and Haberman, 2003*b*). One of the most outstanding models, and also those more frequently used by actuaries, is the Lee-Carter model (Lee and Carter, 1992).

In Booth (2006) the author reviews the development of new and sophisticated methods for mortality forecasting. Debón et al. (2004) introduce, as an alternative to classical methods, methods of adjustment and prediction based on geostatistical techniques which exploit the dependence structure existing among the residuals. In that paper, the dynamic life table is considered as a two-way table on a grid equally spaced in either the vertical (age) or horizontal (year) direction, and the data are decomposed into a deterministic large-scale variation (trend) plus stochastic small-scale variation (error),

$$Z(s,t) = \mu(s,t) + \delta(s,t), \tag{5.1}$$

where $s \in \mathbb{R}_+$ denotes the age and t is time expressed in years.

The novelties of our approach can be resumed as follows:

1. We consider a nonstationary random field of the type (5.1), where the nonstationary component is exclusively explained by the trend, which is a function of the coordinates $(s,t) \in \mathbb{R}^2$. The residuals are shown to be zero mean and weakly stationary.

2. We model the trend function, for both males and females, by using three different methodologies, that are well known under the names of Lee-Carter model (LC for short), extended

Lee-Carter model($LC2$) and median polish (MP).

3. This allows to obtain six sets of residuals (as males and females are separated in order to highlight differences between them), that are modelled by means of geostatistical techniques.

4. We argue that these residuals exhibit an anisotropic component and model it using anisotropic models as described in chapter 3.

5. We forecast the mortality rates on a period for which we have no data.

To do this, we analyse Spain mortality data corresponding to the period 1980-2001. The first twenty years, from 1980 to 1999, are used to fit the models and the last two, 2000-2001, as validation set for measuring the accuracy of the prediction obtained with these models.

5.2.2 Models for trend estimation

The Lee-Carter Model, developed in (Lee and Carter, 1992), consists in adjusting the following function to the central mortality rates,

$$m_{xt} = \exp(a_s + b_s k_t + \delta(s,t))$$

or, its equivalent

$$\ln(m_{xt}) = a_s + b_s k_t + \delta(s,t), \tag{5.2}$$

and where the notation $\delta(s,t)$ is consistent with the treatment of the residuals by means of the geostatistical approach as described in Section 5.2.3.

In the previous two expressions, the double subscript refers to the age, s, and to the year or unit of time, t. a_s and b_s are age-dependent parameters and k_t is an specific mortality index for each year or unit of time. The errors $\delta(s,t)$, with 0 mean and variance σ_δ^2, reflect the historical influences of each specific age that are not captured by the model.

We are going to apply the extended version of this model to logit death probability q_{xt} as Debón, Montes and Sala (2006*b*)

$$\ln\left(\frac{q_{xt}}{1-q_{xt}}\right) = a_s + \sum_{i=1}^{r} b_s^i k_t^i + \delta(s,t). \tag{5.3}$$

The reasons of this change are two-fold: the former is the remark in (Lee, 2000), where the author points out that nothing ensures that the m_{xt} estimations obtained from (5.2) will not exceded 1, although this problem can be avoided by modelling the logit death rates. The latter is that Booth

et al. (2002) and Renshaw and Haberman (2003c) indicate that the interaction between age and time can be captured better by adding terms to (5.2).

Parameters need to be normalised in order to get a unique solution. Thus, one can show that, $\sum_s b_s^i = 1$ and $\sum_t k_t^i = 0$. The structure is then invariant under either of the parameter transformations, $(a_s, b_s/c, ck_t)$ or $(a_s + cb_s, b_s, k_t - c)$, for any constant c. In our application to the Spanish data of mortality we have used (5.3) with $r = 1$ and $r = 2$, consequently the corresponding models will be named LC and $LC2$, respectively. The estimation of the parameters in (5.3) is carried out by means of maximum likelihood methods (ML for short) as Brouhns and Vermunt (2002) propose.

considered as a two-way table on a grid equally spaced in either the vertical (age) or horizontal (year) direction. In this context, the deterministic trend of mortality rate can be decomposed as the sum of three effects

$$q_{xt} = \mu + r_s + c_t, \tag{5.4}$$

an overall effect, μ, a row effect due to age, r_s, and a column effect, c_t, due to year. A median-polish algorithm (Cressie, 1993) is used to estimate the overall effect, $\tilde{\mu}$, row effects, $\tilde{r}_s$, and column effects, $\tilde{c}_t$. For theoretical properties of median polish, we refer the interested reader to (Cressie, 1993). Here, it is worth stressing that it is a nonparametric method and that it is, with respect to other detrending methods, less sensitive to the presence of outliers and less biased than other methods based on the use of mean operators.

5.2.3 Residual analysis: geostatistical approach

Throughout the paper we shall make reference to real-valued weakly stationary zero mean Gaussian random fields $\{\delta(\mathbf{x}) : \mathbf{x} \in \mathcal{D} \subseteq \mathbb{R}^2\}$, where the coordinate $\mathbf{x} = (s, t)'$, *e.g.* (age,time).

Specifically, we shall be dealing with spatial processes that may exhibit a zonal and/or geometric anisotropic component. The approach we propose to face this situation follows that in chapter 3, where anisotropy is attained through isotropy between components. Thus, the covariance function C is said to be represented by a function $\mathrm{C}_0 : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that

$$\mathrm{C}(\mathbf{h}) := \mathrm{C}_0(|h_1|, |h_2|),$$

for $\mathbf{h} \in \mathbb{R}^2$.

5.2.4 Trend estimation

The models described in Section 2 have been used to adjust mortality data from Spain corresponding to the period 1980-1999, for a range of ages from 0 to 99. The adjustment has been performed separately for men and women. The crude estimates of q_{xt} have been obtained by means of the procedure used by the Spanish National Institute of Statistics (INE, Instituto Nacional de Estadística),

$$\dot{q}_{xt} = \frac{1/2(d_{xt} + d_{s(t+1)})}{P_{xt} + 1/2d_{xt}},$$

where d_{xt} are the deaths in year t at age s, $d_{s(t+1)}$ are the deaths in year $t+1$ at age s, and P_{xt} is the population that on December 31st of year t was aged s. The formula can be applied to all ages, except for age 0, due to the concentration of deaths in the first few months of life. The expression used for age 0 is,

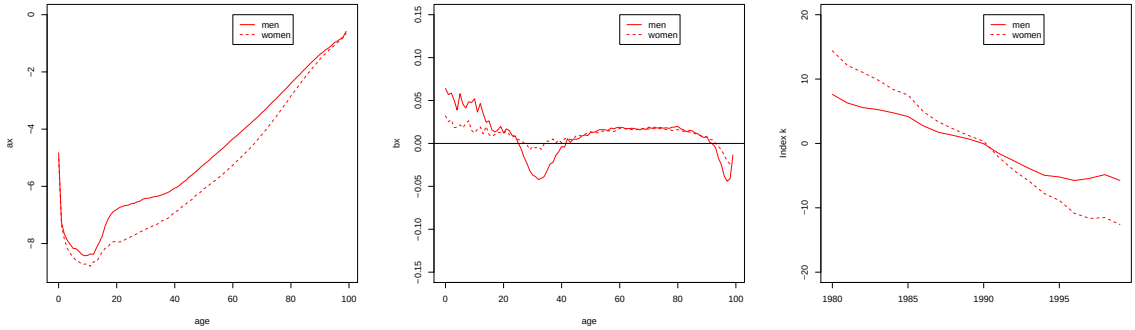
$$\dot{q}_{0t} = \frac{0.85d_{0t} + 0.15d_{0(t+1)}}{P_{0t} + 0.85d_{0t}}.$$

Trend estimation can be performed through the Lee-Carter models, LC and LC2, above described. Owing to the very high number of estimated parameters, a graphical representation seems to be more meaningful in order to represent the results in terms of estimates. These are resumed in Figure 5.1 and 5.2. As a general comment, the comparison of parameter a_s for both sexes shows that mortality for women is lower than for men. The hump in Figures 5.1(a) and 5.2(a) for men, that some authors call "the accident hump", reveals an increase of mortality in the range of ages from 11 to 40.

The negative values of parameter b_s and b_s^1 (Figures 5.1(b) and 5.2(b)) for intermediate ages and for advanced ages indicate that mortality in these age groups does increase over time. The second term, b_s^2, in $LC2$ model (Figure 5.2(d)), shows larger values for intermediate ages for both sexes, which implies that the effect of adding a second term acts more specifically on this age group.

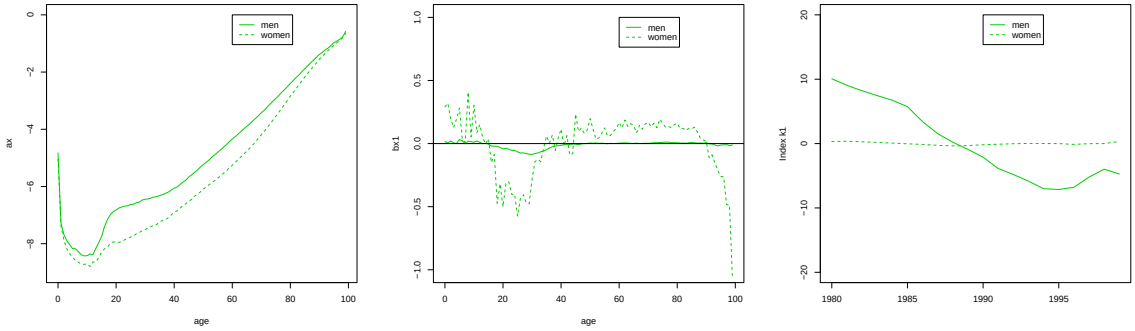
As far as the mortality index is concerned, k_t (Figure 5.1(c)) and k_t^2 (Figure 5.2(e)) show a clearly decreasing trend. With respect to k_t^1 (Figure 5.2(c)), it does not display any trend for women, whereas for men it decreases until year 1995 and increases for the later years.

The high number of effects estimated with the median-polish algorithm, 121, is also too large to be presented in a table. As in Lee-Carter, they are represented graphically in Figure 5.3. The

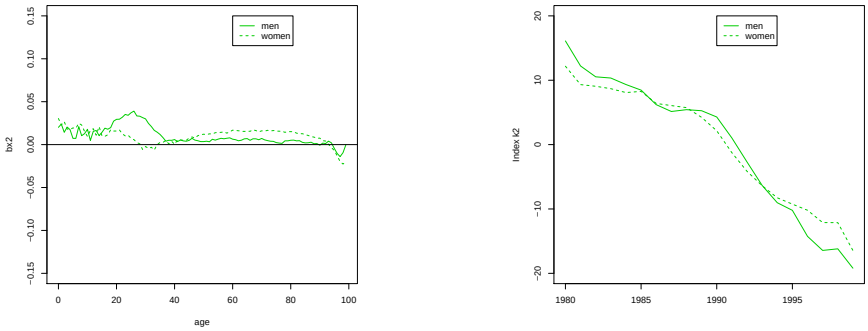


(a) a_s (b) b_s (c) k_t

Figure 5.1: Estimated values for LC model.



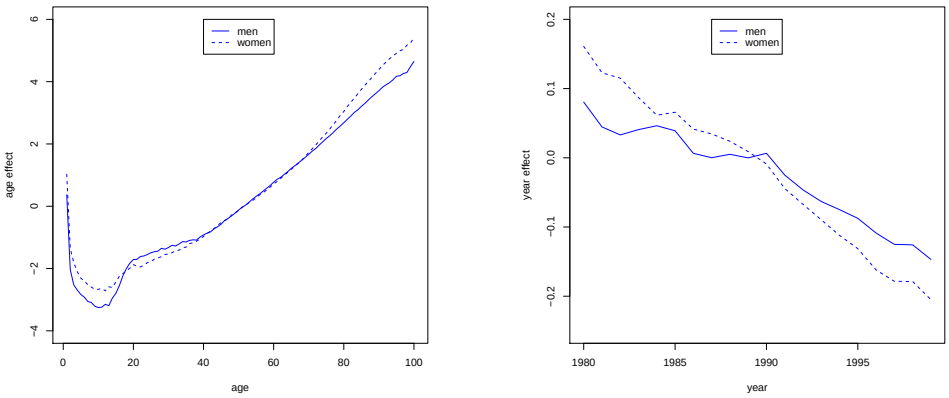
(a) a_s (b) b_s^1 (c) k_t^1



(d) b_s^2 (e) k_t^2

Figure 5.2: Estimated values for $LC2$ model.

Figure 5.3(a) reveals an increase of mortality for men in the group of intermediate age (accident hump) but not for women, this is a similar behaviour of parameter a_s in Lee-Carter model. With regards to the year effect, Figure 5.3(c)) shows a clearly decreasing trend for both sexes, with a similar profile to that of mortality indexes k_t and k_t^2.



(a) age effect (b) year effect

Figure 5.3: Estimated effects for median-polish trend.

5.2.5 Goodness-of-fit for trend

To our knowledge, only Felipe et al. (2002) and Guillen and Vidiella (2005) have studied the evolution of Spanish mortality with dynamic models. Felipe et al. (2002) use the Heligman-Pollards laws to evaluate the way in which the calendar time (1975-1993) affects mortality patterns in the Spanish population for a range of ages from 0 to 90. (Guillen and Vidiella, 2005) use a Poisson log-bilinear version of the Lee-Carter model, proposed by Wilmoth (1993) and Brouhns and Vermunt (2002), for Spanish mortality data during the 1975-1998 period and a range of ages from 0 to 105. Our results about the evolution of mortality rates are similar to those obtained by these authors.

In order to compare the goodness-of-fit for all models along the years, we have obtained $MAPE$ and MSE. These values are shown in Tables 5.1 and 5.2 in the columns headed by *trend*.

$$MAPE(\hat{q}_{xt}) = \frac{\sum_s \frac{|\dot{q}_{xt} - \hat{q}_{xt}|}{\dot{q}_{xt}}}{n}, \; t = 1980, \ldots, 1999$$

$$MSE(\hat{q}_{xt}) = \sqrt{\sum_s \frac{(\dot{q}_{xt} - \hat{q}_{xt})^2}{n}},\ t = 1980, \ldots, 1999$$

The values in the tables confirm the best behaviour of the $LC2$ model, the explanation of this fact is that the introduction of the second term better adapts the model for the ages involved in the accident hump. The MP model fits worse than the others. The adjustment for women is better than for men in LC and $LC2$, but similar for both sexes in MP.

5.2.6 Modelling residuals

Once the trend have been adjusted, six sets of logit residuals, $\delta(s,t) = logit(\dot{q}_{xt}) - \widehat{logit(q_{xt})}$, are obtained that are plotted in Figure 5.4. This figure provide information about the goodness-of-fit, similar to that obtained with Tables 5.1 and 5.2, but specially it confirms the existence of a dependence structure among residuals. The diagonal pattern that plots show could be interpreted as an effect cohort is present. In any case it is a proof of the anisotropy of the underlying random field. Booth et al. (2002) found similar residuals pattern with Australian mortality data.

We need to modify notation by denoting the distinct residuals as $\delta_{LCm}(s,t)$, $\delta_{LCw}(s,t)$, $\delta_{LC2m}(s,t)$, $\delta_{LC2w}(s,t)$, $\delta_{MPm}(s,t)$ and $\delta_{MPw}(s,t)$, where m and w at the end of every acronym stand for, respectively, men and women.

The perspective we are treating the residuals represents the novelty of this paper, as we consider them as realizations of a homogeneous and anisotropic Gaussian random field. There are several advantages induced by considering this setting. First, we can consider the dependence between the residuals as a function of the distance or separation between points in the considered domain Age $\times$ Time $\subseteq \mathbb{R}^2$. This is achieved by estimating the covariance structure from the estimated residuals. Then, the estimated covariance will be crucial for interpolation purpose, as we shall map the whole domain by using simple kriging techniques. Finally, the procedure allows for inference through cross-validation.

It is important to remark that the assumption of isotropy is somehow unrealistic for this kind of phenomena. In fact, it can be easily seen in Figure 5.5 that low and high ages, evaluated with respect to time, have a clear different behaviour respect the middle ones. Our explorative data analysis highlights the presence of a strong anisotropic component which is both zonal and directional. Thus, we believe that a significative improvement in terms of estimation could be achieved by considering a covariance function that allows for taking into account anisotropic components.

MEN

	LC				LC2				MP			
	trend		trend + residuals		trend		trend + residuals		trend		trend + residuals	
Year	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE
1980	5.24	0.011953	3.35	0.009835	4.89	0.011682	3.62	0.009603	14.38	0.024698	7.04	0.028679
1981	5.60	0.007366	2.89	0.008643	3.83	0.006823	3.07	0.006444	13.47	0.018803	3.85	0.015584
1982	5.91	0.009625	3.24	0.011045	3.53	0.009140	3.72	0.011542	12.47	0.017492	3.29	0.014537
1983	4.45	0.002862	2.31	0.004301	2.80	0.003339	2.86	0.003401	10.56	0.013439	2.95	0.009485
1984	4.34	0.004693	2.60	0.002935	2.70	0.005168	3.00	0.004837	9.10	0.012970	3.14	0.004701
1985	3.74	0.006440	2.61	0.003564	2.87	0.006556	3.16	0.005017	8.08	0.009696	2.80	0.003774
1986	4.55	0.008571	2.82	0.003212	4.41	0.008553	3.60	0.004400	6.65	0.007609	2.55	0.003667
1987	4.91	0.012863	3.07	0.003081	4.92	0.012740	3.75	0.006340	5.95	0.010562	3.19	0.004434
1988	5.69	0.018559	3.17	0.008077	4.54	0.018216	3.77	0.012191	6.16	0.016287	3.17	0.007305
1989	6.60	0.014533	2.81	0.009494	4.18	0.013939	3.54	0.012233	6.26	0.012821	2.88	0.010528
1990	6.94	0.004934	2.70	0.005852	3.75	0.004516	3.22	0.004787	6.91	0.002575	3.14	0.009666
1991	6.66	0.005111	2.61	0.002755	2.90	0.004893	2.86	0.003674	7.64	0.005328	3.02	0.004912
1992	5.49	0.010265	2.45	0.004559	3.39	0.010307	3.06	0.005601	7.62	0.009703	2.73	0.003975
1993	5.30	0.005232	2.80	0.002741	4.21	0.005433	3.66	0.004628	8.18	0.007708	2.78	0.005805
1994	5.30	0.005896	2.76	0.003209	4.67	0.005922	3.52	0.004694	9.32	0.008608	3.07	0.006434
1995	5.54	0.005867	3.08	0.004666	5.38	0.005706	3.81	0.005140	9.41	0.010096	3.18	0.006233
1996	5.30	0.007191	2.80	0.004411	4.31	0.007148	3.29	0.002903	9.57	0.020280	3.17	0.005994
1997	8.07	0.005293	3.11	0.003723	2.92	0.004700	3.42	0.002923	8.55	0.019077	3.29	0.005859
1998	10.11	0.004916	2.68	0.004204	3.79	0.004819	2.78	0.004597	8.47	0.016941	3.30	0.005949
1999	11.43	0.004672	3.51	0.002252	5.49	0.004978	3.94	0.002051	9.01	0.012217	5.36	0.009305
mean	6.06	0.007842	2.87	0.005128	3.97	0.007729	3.38	0.005850	8.89	0.012846	3.40	0.008341
std. Dev.	1.90	0.004001	0.31	0.002722	0.87	0.003860	0.36	0.003081	2.34	0.005560	1.03	0.005841

Table 5.1: MAPE and MSE for each year for men.

WOMEN

	LC				LC2				MP			
	trend		trend + residuals		trend		trend + residuals		trend		trend + residuals	
Year	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE	MAPE	MSE
1980	5.32	0.006799	4.11	0.008872	4.38	0.003779	3.97	0.005153	12.00	0.026236	4.54	0.010471
1981	4.16	0.006836	3.37	0.005926	2.76	0.003634	3.14	0.005343	10.63	0.024369	3.64	0.007058
1982	4.86	0.005690	3.97	0.004338	3.55	0.002175	3.94	0.002808	9.61	0.021095	4.44	0.004158
1983	4.57	0.003611	2.92	0.003035	3.45	0.002280	3.57	0.002945	8.94	0.016777	3.38	0.003578
1984	4.99	0.001813	3.14	0.002522	4.71	0.002302	3.75	0.002094	8.52	0.011889	3.69	0.002926
1985	4.25	0.004532	2.90	0.002388	4.66	0.003669	3.08	0.002573	6.73	0.008039	3.39	0.002458
1986	4.58	0.005180	3.37	0.002230	4.40	0.002901	3.89	0.002239	5.50	0.004961	3.98	0.002704
1987	5.02	0.008708	2.86	0.002672	3.79	0.003093	2.92	0.002316	4.88	0.005641	3.32	0.002868
1988	5.01	0.011732	3.05	0.003429	3.39	0.004146	3.54	0.004079	4.48	0.009010	3.62	0.003788
1989	5.11	0.009113	2.69	0.005223	2.91	0.002738	3.01	0.003601	4.85	0.007020	3.05	0.005405
1990	4.73	0.003660	2.71	0.005916	3.57	0.004371	2.90	0.005524	4.82	0.003375	3.43	0.007421
1991	4.72	0.003243	3.00	0.003810	4.34	0.004944	3.46	0.003025	5.37	0.003199	3.46	0.004106
1992	5.04	0.005071	2.79	0.003148	5.10	0.005527	3.89	0.003433	5.81	0.003869	3.24	0.003531
1993	5.29	0.005142	2.86	0.003506	5.32	0.004586	3.46	0.003501	6.68	0.006200	3.15	0.003716
1994	5.50	0.003542	2.95	0.003500	5.37	0.003304	3.35	0.002617	8.12	0.009184	3.44	0.003328
1995	5.84	0.004416	3.40	0.005671	5.78	0.004408	4.14	0.004195	8.55	0.011270	3.98	0.005677
1996	4.81	0.006650	3.01	0.007696	4.92	0.004390	4.13	0.005453	8.36	0.021706	3.48	0.008423
1997	6.05	0.003399	3.37	0.005067	5.66	0.003304	4.38	0.002920	7.95	0.020975	3.95	0.004732
1998	8.11	0.004741	3.68	0.003702	7.39	0.004819	4.18	0.004001	9.18	0.021929	3.97	0.003233
1999	9.02	0.006116	5.11	0.003991	6.86	0.003454	4.85	0.004273	9.24	0.014734	5.98	0.004411
mean	5.35	0.005500	3.26	0.004332	4.62	0.003691	3.68	0.003605	7.51	0.012574	3.76	0.004700
std. Dev.	1.20	0.002344	0.59	0.001783	1.23	0.000950	0.53	0.001111	2.16	0.007729	0.66	0.002131

Table 5.2: MAPE and MSE for each year for women.

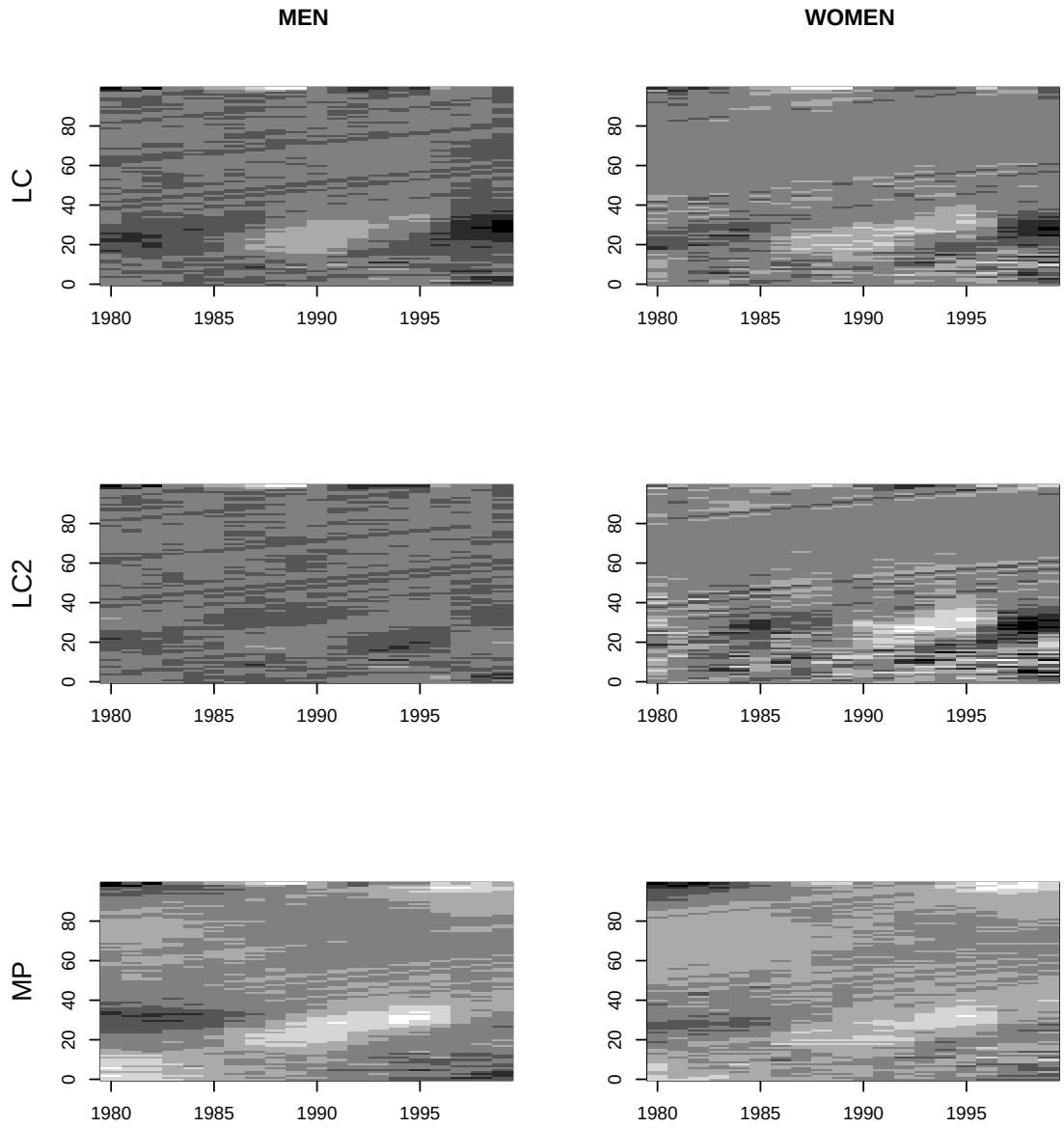


Figure 5.4: Residuals for the three models

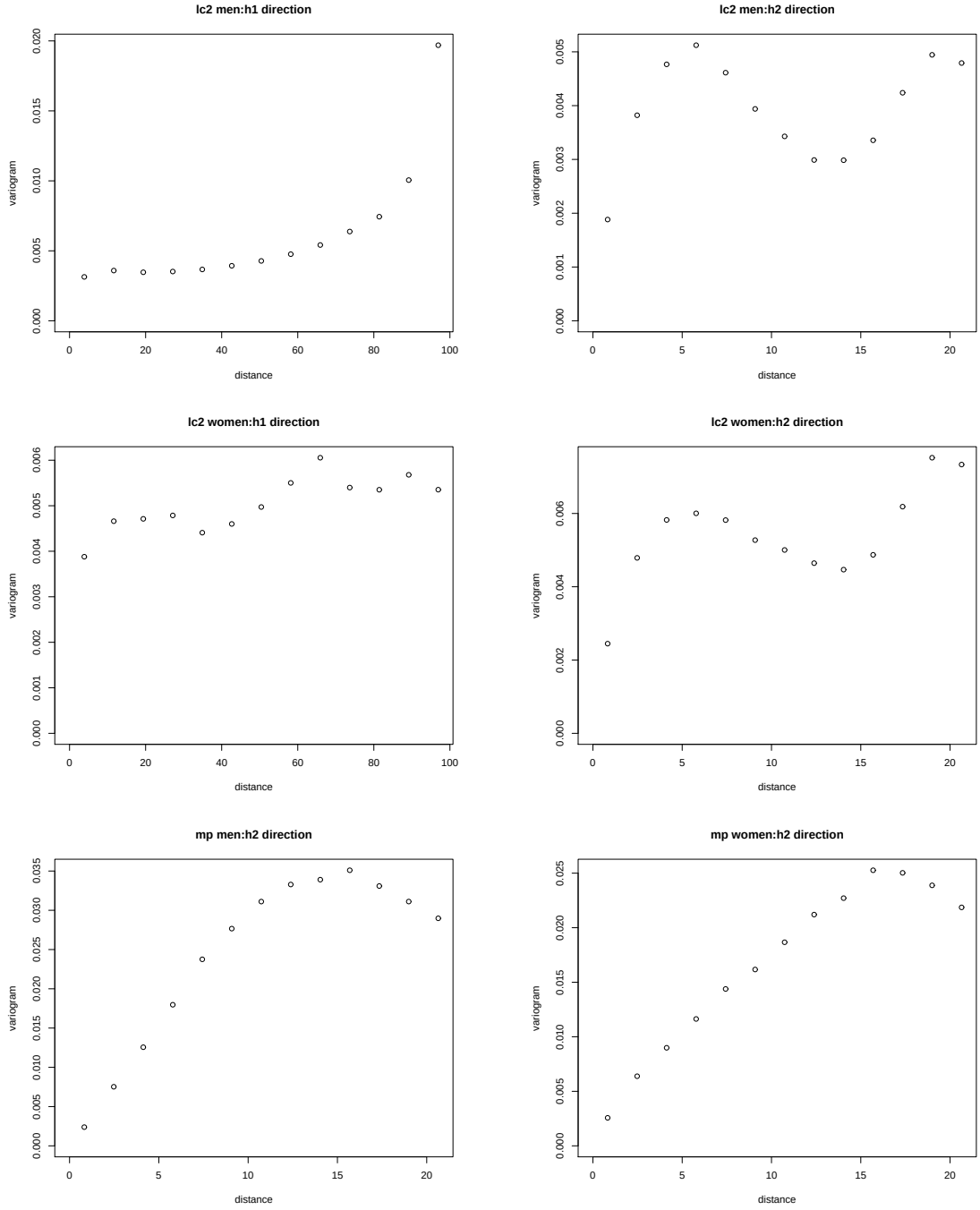


Figure 5.5: Comparison of some of the empirical variograms calculated from the six set of residuals. From up-left to down-right: LC2 for men, direction h_1; LC2 for men, direction h_2; LC2 for women, direction h_1; LC2 for women, direction h_2; MP for men, direction h_2; MP for women, direction h_2

	a_1	a_2	α	β	σ^2
δ_{LCm}	0.7792	0.5393	0.4608	1.2242	0.0110
δ_{LCw}	0.8130	1.0438	0.6223	1.7418	0.0076
δ_{LC2m}	1.7181	0.5862	0.6020	0.1450	0.0060
δ_{LC2w}	0.4272	0.3102	0.4131	0.7547	0.0055
δ_{MPm}	4.5892	6.0768	1.1289	1.9271	0.0255
δ_{MPw}	2.6139	5.3326	0.6536	1.3436	0.0219

Table 5.3: Estimates for the parameters of the covariance function in equation (5.5), for the six residuals

Our proposal is to attain anisotropy through isotropy between components. In particular, in this paper we shall focus on models as described in chapter 3. Specifically we use the model:

$$\mathrm{C}(h_1,h_2)=k\sigma^2\left(\frac{\exp(-\left(1+\frac{|h_1|^\alpha}{a_1}\right)-\left(1+\frac{|h_2|^\beta}{a_2}\right))}{-2-\frac{|h_1|^\alpha}{a_1}-\frac{|h_2|^\beta}{a_2}}\right), \tag{5.5}$$

where k is a normalization constant so that $\mathrm{C}(h_1,h_2)=\sigma^2$, which is the variance of the underlying random field. The parameters a_i, $i=1,2$ measure the scale over directions h_i and $0<\alpha,\beta\leq 1$ are smoothing parameters. This covariance function allows for geometric and zonal anisotropy and it is easily interpretable as every parameter has a specific meaning in terms of range and smoothness.

Estimation of the five parameters characterizing dependence in (5.5) may be performed through least squares as well as likelihood methods. In this work, estimation will be performed through composite likelihood methods, Curriero and Lele (1999) as described in chapter 4. The so obtained estimates are resumed in Table 5.3.

Once estimate are obtained, they can be used to implement the empirical covariance function, that is crucial for prediction purposes. In particular, we assume the mean known, thus the best linear unbiased predictor (BLUP for short) takes the name of simple kriging predictor in geostatistics, even if in this case it is more appropriate to call it kriging of the residuals. For the remainder of the paper, we shall use equivalently this nomenclature. Thus, for every age s fixed the simple kriging predictor for time t, say μ_t, can be written as $\mu_t\boldsymbol{c}_0'\boldsymbol{C}^{-1}\boldsymbol{\delta}_t$, with $\boldsymbol{C}$ the var-cov matrix of the predictor variables, $\boldsymbol{c}_0$ is the vector containing the covariance between the predictand and the predictor, and $\boldsymbol{\delta}_t$ is the vector of the observed residuals. For well known results, the associated kriging variance is

$$\sigma_t^2=\sigma_t^2-\boldsymbol{c}_0'\boldsymbol{C}^{-1}\boldsymbol{c}_0, \tag{5.6}$$

with σ_0^2 unconditional variance for the residual measure at the fixed age where we do prediction

for time t.

Goodness-of-fit for trend + residuals In order to evaluate the goodness-of-fit of our models when the residuals are modelled by choosing the anisotropic covariance function in equation (5.5), we proceed by calculating $MAPE$ and MSE. It must be point out that $\hat{q}_{xt}$ is now the antilogit of the sum of the estimated trend with the estimated residuals, taking into account the correction suggested by (**?**) for avoiding bias when a transformation of original data is needed. Columns headed by *trend + residuals* in Tables 5.1 and 5.2 show that modelling residuals has substantially improved the adjustment for all models, specially for MP which has the worst behaviour when only trend was used. It must be point out that all three models have, after modelling residuals, a similar performance, summarized (mean and standard deviation for $MAPE$ and MSE) in the last two rows in both tables.

5.2.7 Prediction

The prediction of mortality rates with the Lee-Carter model needs preliminary operations on a time series to mortality indexes, k_t, which is carried out by means of the Box-Jenkins methodology using an R code developed by Shumway and Stoffer (2006). The most appropriate model seems to be an $ARIMA$ model, as we consider reasonable the hypotheses of stationarity for the first differences, for both sexes. The $ARIMA$ fittings for each model are shown in Table 5.4. The time series $\{\hat{k}_t\}$, $\{\hat{k}_t^1\}$ and $\{\hat{k}_t^2\}$, $t = 1980, \ldots, 1999$, are used to forecast k_t, k_t^1 and k_t^2, respectively, for $t > 1999$. The underlying assumption is that, for $LC2$ model, the two time series corresponding to k_t^1 and k_t^2 are independent. This assumption, also made by Booth et al. (2002) and by Renshaw and Haberman ($2003c$), represents a potencial drawback as Renshaw and Haberman ($2003c$) recognize. An example of how the model can be expanded to include dependence and co-integration effects is given in Renshaw and Haberman ($2003a$).

	Men	**Women**
LC	$k_t - k_{t-1} = -0.7031 + \varepsilon_t$	$k_t - k_{t-1} = -1.4203 + \varepsilon_t$
LC2	$k_t^1 - k_{t-1}^1 = -0.7844 + \varepsilon_t + \varepsilon_{t-1}$ $k_t^2 - k_{t-1}^1 = -1.8584 + \varepsilon_t$	$k_t^1 - k_{t-1}^1 = \varepsilon_t$ $k_t^2 - k_{t-1}^2 = -1.5065 + \varepsilon_t$

Table 5.4: ARIMA models for mortality index.

When using a median polish algorithm for the adjusted trend, the associated predictor $\hat{q}_{xt}$ of

q_{xt} admits analytic expression:

$$\hat{q}_{xt} = \tilde{\mu} + \tilde{r}_s + \tilde{c}_{t_{1999}} + (t - t_{1999})(\tilde{c}_{t_{1999}} - \tilde{c}_{t_{1998}}), \quad \forall s \in D,\ t > 1999. \tag{5.7}$$

Analogously to that we have done in above Section, we obtain also now trend + residuals predictions. These predictions are obtained by considering the sum of the predicted trend with the predicted residuals. This implies that the underlying process is nonstationary, as the predicted trend is a function of the coordinates age and time, whilst the residuals are considered as homogeneous and modelled through geostatistical procedures. The adopted procedure is consistent with Perrin and Senoussi (1999), who argue that nonstationarity can be specified with respect to the trend function, the covariance function or both.

In order to assess the performance of the models in terms of prediction, we calculate the predictions for every age s at years 2000 and 2001 and we compare them with the crude estimates of $q_{s,2000}$ and $q_{s,2001}$. Table 5.5 shows the MAPE and MSE resulting of this comparison.

Men

		MSE		MAPE	
		trend	*trend + residuals*	*trend*	*trend + residuals*
2000	LC	0.01154	0.00859	15.51	7.83
	LC2	0.01212	0.01195	7.91	7.41
	MP	0.00937	0.00471	10.07	5.23
2001	LC	0.01608	0.01388	18.21	12.22
	LC2	0.01711	0.01696	9.06	8.79
	MP	0.00545	0.00866	10.78	8.13

Women

		MSE		MAPE	
		trend	*trend + residuals*	*trend*	*trend + residuals*
2000	LC	0.01465	0.00781	11.32	5.26
	LC2	0.00847	0.00865	8.85	6.83
	MP	0.00913	0.00448	10.12	4.95
2001	LC	0.01959	0.01481	13.08	8.88
	LC2	0.01330	0.01342	10.09	8.84
	MP	0.00766	0.00523	9.98	6.96

Table 5.5: MSE and MAPE of $\hat{q}_{s,2000}$ and $\hat{q}_{s,2001}$ for all models.

It should be stressed that Tables 5.1, 5.2 and 5.5 highlight a clear trade-off between goodness-of-fit and accuracy of prediction, as the model that performs better in the former, are then the worst in the latter. It seems that MP method is the one which performs better among the proposed ones.

Figure 5.6 shows the prediction of mortality ratios for years 2000 to 2010 obtained with the different models. In order to facilitate the interpretation of these results we have represented the ages from 10 to 50 years, for every 10 years. In Figure 5.6 it can be appreciated that, although LC performs quite well in terms of fitting, it is quite poor in terms of prediction, particularly in the age class 20-40 years. It seems in fact inconsistent to have higher mortality ratios for lower ages respect to higher ones. On the other hand, predictions performed through MP are a clear instance of trade-off between smoothness (plausibility of the projected predictions) and goodness-of-fit and it avoid any crossover in the projected rates. Similar results and conclusions can be argued by considering performance predictions on trend + residuals through the three methods.

Pedroza (2006) argue that measures of uncertainty of mortality forecasts are needed and points out these must be carried out by means of confidence intervals expressing probabilistic uncertainty of the point forecast. We have obtained confidence intervals for predicted death rates using bootstrapping techniques for predictions made only with trend (Koissi et al., 2006) or using variance expression (5.6) for predicted residuals in the case of predictions made with trend + residuals.

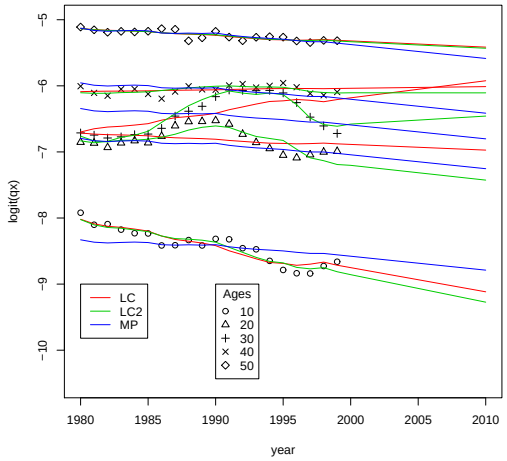
For obtaining bootstrap confidence intervals, N bootstrap samples, $\hat{\delta}^n(s,t), i = 1, \ldots, N$, are simulated from the residuals obtained with the original data. Each sample furnishes an estimated $\widehat{logit(q_{xt})}^n$ by means of

$$\widehat{logit(q_{xt})}^n = logit(\dot{q}_{xt}) - \hat{\delta}^n(s,t),$$

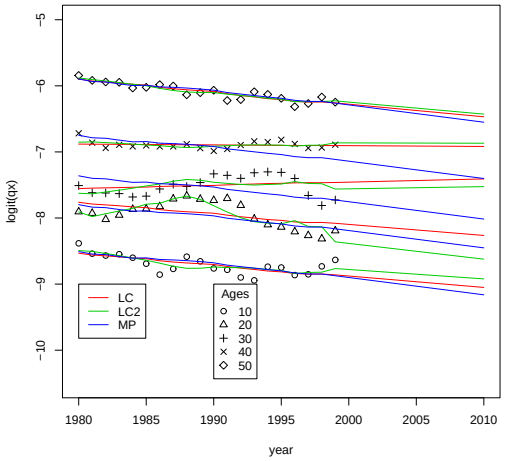
where $logit(\dot{q}_{xt})$ are computed from the original data. For each bootstrap sample a model is adjusted, in our case the $LC2$ or the MP model, that provides predicted values for q_{xt}. This procedure yields N predictions of q_{xt}. The confidence intervals are the percentile intervals, $IC_{95} = [p_{0.025}, p_{0.975}]$. We did not attain confidence interval for LC estimation as previous results confirm a better behavior of the $LC2$ model.

Figures 5.7 and 5.8 show the confidence interval for q_{xt}, for the following ages that are considered to be enough representative of the whole dataset: $s = 10, 20, 30, 97, 98, 99$. It is worth remarking the major magnitude of the confidence intervals obtained through MP. As far as differences between sexes, it can be appreciated that confidence intervals are narrower for women in all ages when using MP. This is confirmed only for some ages when using $LC2$.

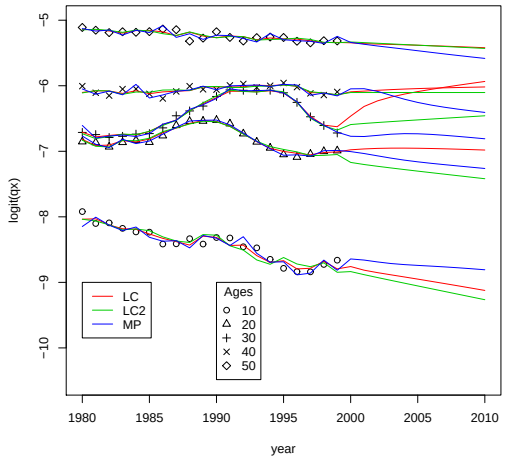
Is should be stressed that MP highlights a decreasing trend for all the ages, whilst $LC2$ emphasises exactly the opposite, above all for middle and old ages. The reason of this may be found in the fact that predictions performed through MP are based on contiguous years, whilst



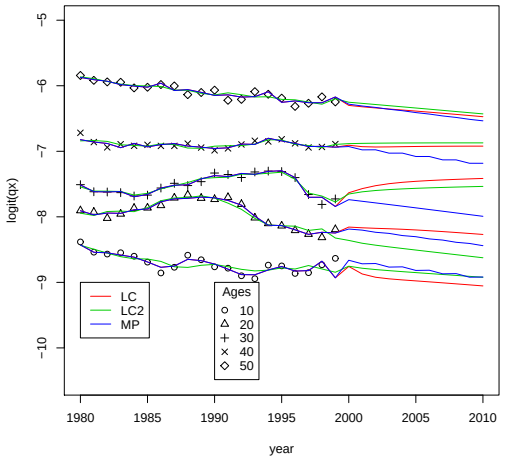
(a) Men (trend)



(b) Women (trend)

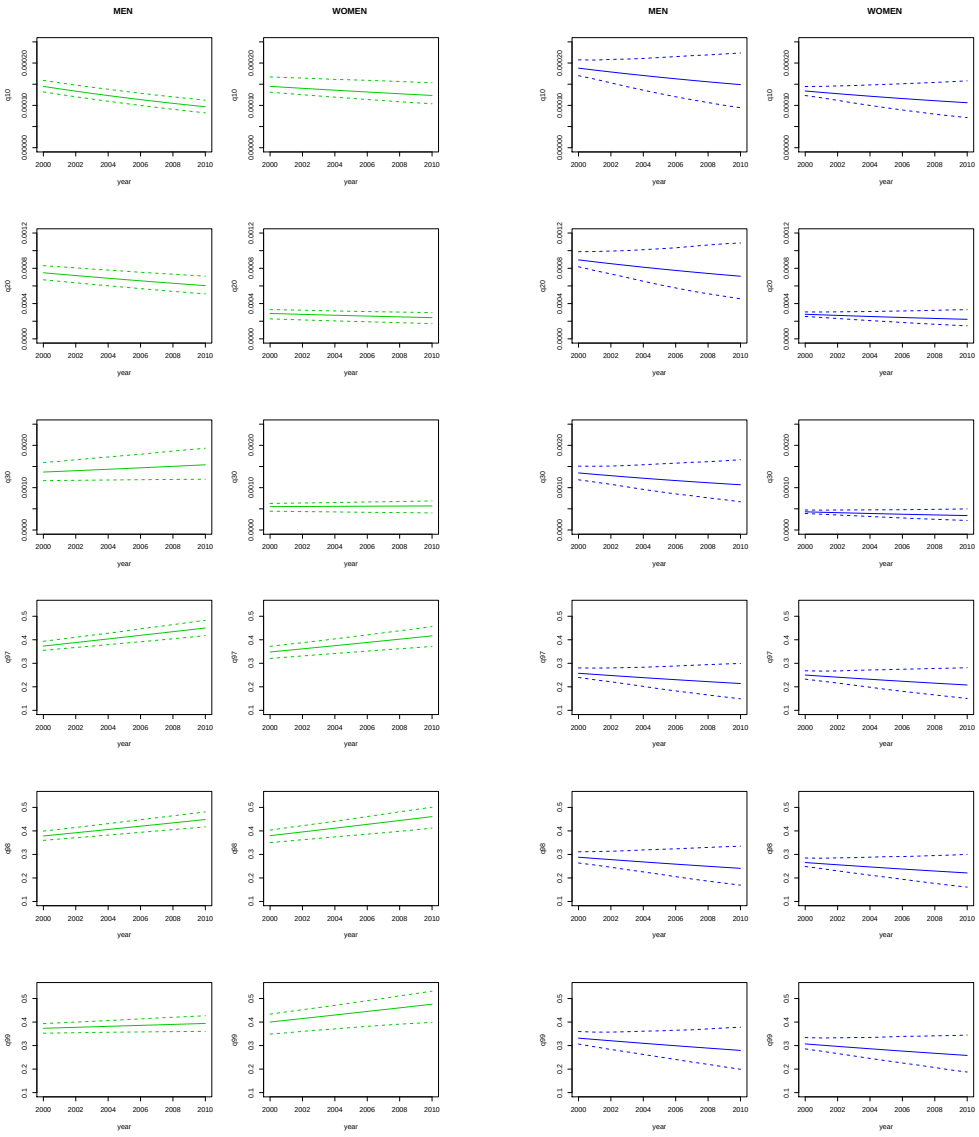


(c) Men (trend +residuals)



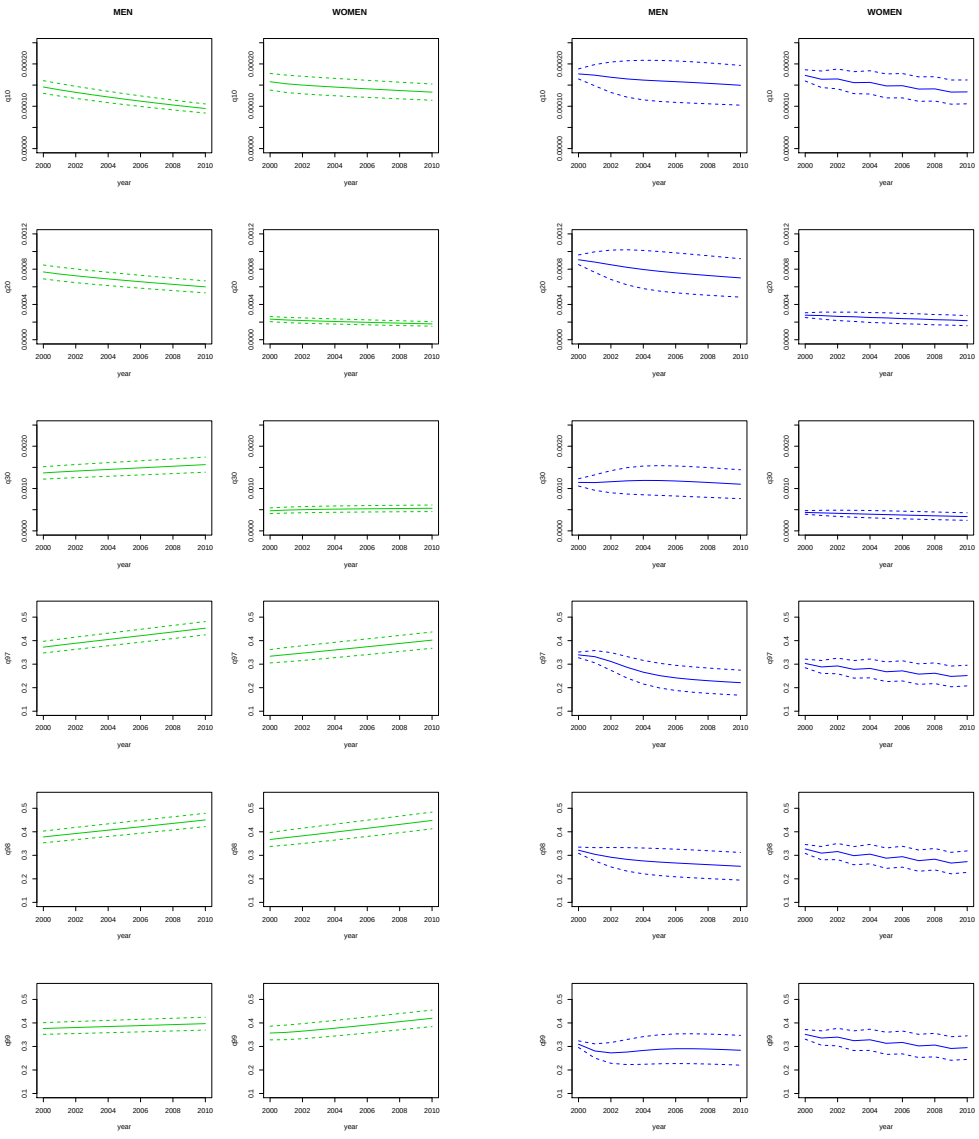
(d) Women (trend+residuals)

Figure 5.6: Predicted rates versus year (1980-2010) for some ages.



(a) LC2 (b) MP

Figure 5.7: Bootstrap confidence intervals for death probabilities q_{xt} for ages 10, 20, 30, 97, 98 and 99.



(a) LC2 (b) MP

Figure 5.8: Confidence intervals for death probabilities q_{xt} for ages 10, 20, 30, 97, 98 and 99 for models trend + residuals.

$LC2$ considers all past observations. Similar conclusions can be drawn by considering trend + residuals simultaneously.

5.2.8 Conclusions

We propose a new model for modelling and predict mortality rates. This model is based on a median-polish algorithm, which is easily interpretable and allows for observing the evolution along years and ages of the observed phenomenon, with a relatively low computational cost. Also, the results in terms of prediction performance are very satisfactory.

The residuals have been modelled through a geostatistical approach, where the novelty is represented by the fact that we took into account an anisotropic component by considering componentwise isotropy. The results were considerably improved both in terms of fitting and prediction performances, as confirmed by the diagnostics obtained through $MAPE$ and MSE in Tables 5.1, 5.2 and 5.5. The fact that these measures are significantly lower clearly implies a better behaviour for any considered ages. This result is particularly important if taking into account the intrinsic problems associated to prediction over very old ages. As pointed out by Wong-Fupuy and Haberman (2004), *Inaccuracies in recording ages in official statistics and high variability in the estimates due to small exposures to risk are common problems when estimating mortality rates for the oldest age groups.*

Modelling the residuals allows us to interpolate and predict missing data through kriging. In addition, as we know the exact expression for the variance of the prediction, the calculation of the corresponding confidence intervals for future predictions is clearly simplified. The so obtained confidence intervals are almost equivalent (in terms of widths) to those obtained through bootstrap techniques, as highlighted in Figures 5.8 and 5.7, even though the bootstrap-based intervals take into account any uncertainty inherent to prediction, as stated by Koissi et al. (2006).

We believe that our approach can be very useful for modelling mortality rates as it has been shown to be very flexible, allowing to incorporate, through several approaches, a deterministic component with covariates, and a stochastic one, exhibiting complex forms of interaction.

5.3. FITTING NEGATIVE COVARIANCES TO GEOTHERMAL FIELD TEMPERATURES IN NEA KESSANI (GREECE)

5.3.1 The geothermal field of Nea Kessani and statistical-modeling interest

The geothermal field of Nea Kessani (NE Thrace, Greece) is located within the Ksanthi-Komotini basin, next to the Aegean Sea and Vistonis Lake, as shown in the geological map in Figure 5.9 (Papantonopoulos and Modis, 2005) . The geothermal area is located near the southwestern margin of the Ksanthi-Komotini basin which is a post-orogenic Tertiary sedimentary basin, and covers an area of about 1600 km^2 between the Rhodope Mountains and the Aegean coast (Figure 5.10, *left*). It is mainly constituted of clastic sediments, and reaches maximum depth at the foot of the Rhodope chain, and minimum depth in the vicinity of the coast, where the Nea Kessani geothermal field is located (Thanassoulas and Tsokas, 1990). During the Paleogene, predominantly molassic sediments were deposited over the highly fractured substratum, which are represented by the metamorphic Paleozoic basement, mainly outcrops in the Rhodope Mountains. Local outcroppings of the basement, which is composed of gneiss, amphibolites with interbedded marbles, micaschists, migmatites and intruded granites, are located to the southwest of the study area. The Eocene-Oligocene sequence is mainly consisting of basal breccias and conglomerates, nummulitic limestones and arkosic sandstones, which make way to flysch formations in the upper part of the sequence. These formations are overlain by Pliocene lacustrine sediments, such as marls, clays and sands, as well as alluvial Quaternary deposits. During the Tertiary, as a result of the sinkage of the African plate below the European plate, an andesitic magmatism developed in the basin, with emplacement of sub-volcanic rocks. Small intrusions of such rocks are inter-bedded with the clastic horizons in the study area. Two major tensile fault systems, striking N 160° and N 70°, developed from the Miocene on. The most active is the N 160° system, which is probably related to the movements of the North Anatolic Fault.

The geothermal field is created by distribution of heat in the geological formations of the ground at some depth due to thermal fluid circulation. This causes an increase of the earth natural geothermal gradient, which is 30$^\circ C$/km at average, by several degrees reaching 35$^\circ C$/km. The hot reservoir created by the thermal fluids has an average temperature of $75-80^\circ C$ and covers an area of 5 km^2. The roof of the reservoir is found in a depth of 100 to 120 meters at its south part near Aegean Sea, while its basement in this area is located at the depth of 450 meters. At the north part of the reservoir its roof is located at the depth of $300-350$ meters and its basement

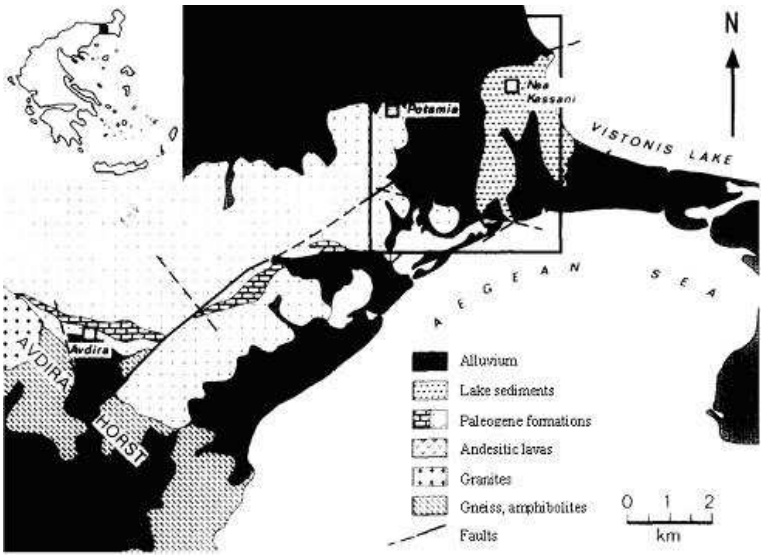


Figure 5.9: Geological map of Nea Kessani geothermal area (rectangle) and vicinity

at greater depth than 1 km. A schematic representation of the reservoir and the thermal fluid circulation can be seen in Figure 5.10 (*right*) where a N-S vertical section of the area is presented.

5.3.2 Drill-hole data

In order to explore and study the geothermal anomaly in the area, 25 exploratory boreholes up to a depth of 500 meters were drilled during 1980-1991 by IGME (Institute of Geological and Mineral Investigations, Athens, Greece). The first phase of the drilling campaign began in 1980 and produced 11 drill holes in depths varying from 65 to 475 meters. The second phase, which started in 1990, produced 14 drill holes to depths varying from 200 to 500 meters. The drill-hole data consist of temperature measurements taken in the drill wells by an electrical resistance thermometer with $\pm 0,1^{\circ}C$ error. Thus, we have a three-dimensional data set formed by the longitude and latitude (measured in meters) of the drill hole together with the depth at which the temperature is measured. In Figure 5.11 the locations of the drill-hole collars are shown, along with two vertical sections showing the temperature distribution in the area.

To resume, we dispose of 162 data distributed in 25 exploratory drills in this geothermal field case and the domain of the study is $2250m \times 2250m \times 400m$. This is shown in Figure 4.

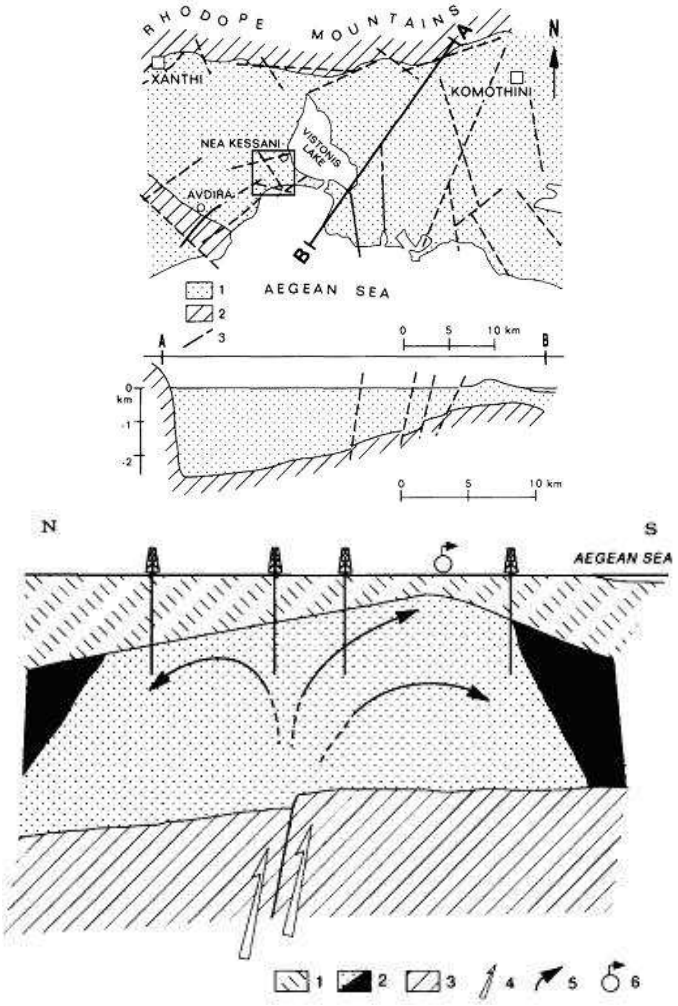


Figure 5.10: *Left*: Geological features of Ksanthi-Komotini basin and the geothermal area (rectangle) along with a vertical section: (1) Quaternary and Tertiary formations; (2) Paleozoic basement; (3) Fault. *Right*: A N-S section of the area, showing a thermal fluid circulation model and the morphology of the thermal field: (1) Cover formations; (2) Arkosic reservoir with hydrogeological boundaries (shaded area); (3) Basement; (4) Deep geothermal fluids; (5) Fluid circulation within the arkosic reservoir; (6) Thermal springs

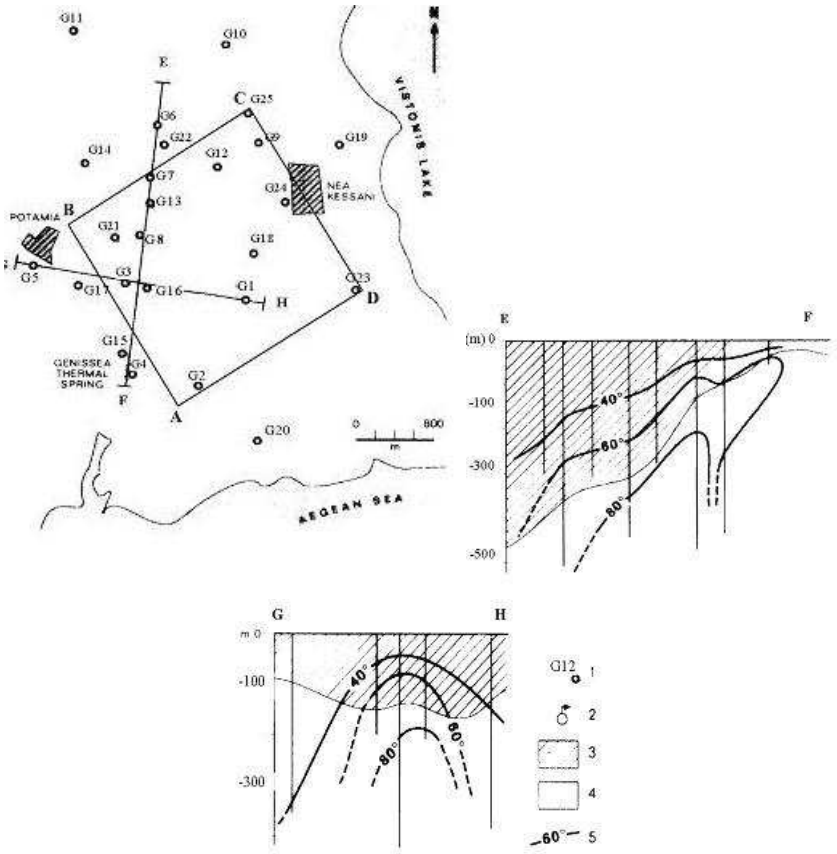


Figure 5.11: Drill hole (G's) locations and vertical cross-sections in area of interest (dashed parallelogram). In sections AB and CD the temperature distribution derived from drill data is presented. (1) Drill collar, (2) Thermal spring, (3) Roof formations, (4) Reservoir, (5) Isotherms in °C

5.3.3 Scientific motivation

The scientific interest on the data lies upon answering the question why should a spatial model of the geothermal field be created. The importance of geothermal energy as an energy source is easily understandable, especially during the current trends of oil prices. In addition, geothermal energy is a clean and environmental friendly form of energy. Thus it is of worth to explore deposits of thermal energy, since it can be effectively used in many applications as heating and or cooling of buildings, power generation etc. The parameter of interest in exploring a geothermal field is the temperature and its distribution in the underground area, since the subsequent exploitation of the deposit is tightly connected to it. For example, a power generator application can be implemented only if the temperatures of the field exceed 100°C, which can lead to the sufficient generation of steam. In the Nea Kessani area the temperature is known in certain positions due to the drilling campaign, so the aim of our work is to create a three-dimensional model of the geothermal field to describe the temperature in a domain as large as possible. As in the case of a metal deposit where the ore grade must be known almost everywhere in order to efficiently plan the exploitation, the temperature distribution in a geothermal field must be also known in underground space, in order to obtain a complete image of the phenomenon and identify the possible exploitation sites. In the Nea Kessani case, the drilling campaign produced only a small number of data. These data are insufficient to model the field using deterministic tools like inverse distance squares (IDS). Thus, the application of more sophisticated methods, like Geostatistics, was clearly necessary.

5.3.4 Exploratory analysis

Stationarity-isotropy-ergodicity are most often modeling assumptions that are justified on the basis of the outcome. On the basis of our study of the physical information underlying the data, we found that they obey an isotropic heat transfer equation, i.e. a Laplace equation. In such an equation the heat conductivity is isotropic and the transfer ability for the heat is also the same at all directions. The assumption of a stationary and isotropic covariance is thus consistent with this physical knowledge. As far as trend analysis is concerned, and before the calculation of the covariance, a nonparametric meantrend was calculated and removed from the data. In particular, we used a nonparametric moving average method similar to the moving window method that uses a spatial moving window to obtain the spatial moving average based on the data within its specified neighborhood. In this case, the spatial window was a square cubic with 100 meters long sides.

Also, an exponential smoothing technique was applied to the calculated moving average to have a smoother mean trend. The detendred data, or residuals, constituted the base data to be spatially modeled. The empirical estimation of the spatial covariance of the residuals was based on a 200 meters lag, except for few very short lags at the origin. This resulted in a quite smooth empirical covariance easy to be fitted by a theoretical model. Once the data was detrended, a QQ plot highlighted a slight left tail, which did not prevent the residuals from being Gaussian distributed. Indeed, we performed Kolmogorov-Smirnov and Shapiro-Wilks tests, and the results, in terms of p-values, confirmed our initial mood, so that we could assume to work with residuals that were stationary and approximately Gaussian.

Calculation of the residuals allowed us to obtain the empirical covariance, which highlighted the presence of negative values. This could be a problem for fitting procedures, as the great majority of celebrated models of covariance functions for spatial data only attain positive values. This strong motivation prompted our research. We are looking for spatial covariance models satisfying two main features:

1. They may be negative or oscillate between negative and positive values.

2. They would preferably allow for an easy interpretation.

We believe it is reasonable to select a class of models satisfying property 2, and then answer, if possible, the natural question: can we obtain negative covariances starting from a class which is easy-to-implement and interpretable? The answer is yes and a satisfactory solution for this data set will here be exposed.

5.3.5 Fitting negative covariances: methodology

In a recent paper, Gregori et al. (2007) proposed a novel and very general approach to building space-time covariance functions that attain negative values or oscillate between positive and negative ones. The approach, called Generalized Sum of the Products, is based on a linear combination of pairwise products of continuous spatial and temporal covariance functions. Some permissibility criteria are given to allow some of the weights in the linear combination to be negative. Here we are only involved in spatial analysis, but the results contained therein can be very useful in order to implement the following idea. Consider an arbitrary natural number n of continuous covariances defined on $\mathbb{R}^d$, and additionally integrable on their domain. Thus, we propose the following

covariance function

$$\mathrm{C}(\mathbf{h}) = \sum_{i=1}^{n} k_i \mathrm{C}_i(\mathbf{h}), \qquad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}, \tag{5.8}$$

whose permissibility is in principle guaranteed whenever the weights k_i, $i = 1, \ldots, n$ are nonnegative. By using Proposition 1 in Gregori et al. (2007) one can show the result proposed subsequently, for which we restrict to the case $n = 2$. We skip the proof as it follows exactly the same arguments of the previously cited authors.

Proposition 1. *Let* C_i *be spatial continuous and integrable covariance models,* $i = 1, 2$*. For* $\vartheta \in \mathbb{R}$*, let us define the function*

$$\mathrm{C}(\mathbf{h}) = \vartheta \mathrm{C}_1(\mathbf{h}) + (1 - \vartheta)\mathrm{C}_2(\mathbf{h}), \qquad \mathbf{h} \in \mathbb{R}^d.$$

Let us denote by f_i *the Fourier transforms of covariances* C_i*, assume* f_2 *does not vanish and write*

$$m := \inf_{\boldsymbol{\omega} \in \mathbb{R}^d} \frac{f_1(\boldsymbol{\omega})}{f_2(\boldsymbol{\omega})}, \quad M := \sup_{\boldsymbol{\omega} \in \mathbb{R}^d} \frac{f_1(\boldsymbol{\omega})}{f_2(\boldsymbol{\omega})}.$$

Then, C *is a valid spatial covariance if and only if*

$$[1 - \max(1, M)]^{-1} \leq \vartheta \leq [1 - \min(1, m)]^{-1}$$

(where $0^{-1} = -\infty$ *and* $(-\infty)^{-1} = 0$ *in the left hand side, and* $0^{-1} = +\infty$ *in the right hand side).*

The advantage of the result above is that it allows for building covariance functions that, thanks to the negative weights, attain negative values or may oscillate between positive and negative ones. The procedure is very simple: one must chose two parametric covariance functions that admit, possibly, an explicit closed form for the associated spectral density. Then, the range of ϑ can be easily computed.

Here we shall be involved in linear combinations of Whittle-Matérn type covariance functions (see Matérn, 1986), that is

$$\mathrm{C}(h) = \sigma^2\{\vartheta \mathrm{C}_{\mathcal{M}}(h; \alpha_1, \nu_1) + (1 - \vartheta)\mathrm{C}_{\mathcal{M}}(h; \alpha_2, \nu_2)\}, \tag{5.9}$$

with $h = \|\mathbf{h}\|$, $\mathbf{h} \in \mathbb{R}^d$, and where we use the Whittle-Matérn correlation functions $t \mapsto \mathrm{C}_{\mathcal{M}}(t; \alpha_i, \nu_i) = (\alpha_i t)^{\nu_i} \mathcal{K}_{\nu_i}(\alpha_i t)$, $i = 1, 2$, with $\mathcal{K}_\nu$ the modified Bessel function of the second kind and order ν (Abromowitz and Stegun, 1967). The nonnegative parameters α_i represent the scale of the spatial dependence, while $\nu_i > 0$ governs the level of smoothness of the associated Gaussian random field.

Table 5.6: Results of inf's and sup's needed in Proposition 1 for the particular case *Matérn vs. Matérn* correlation functions, respectively, in terms of their parameters values. Here d denotes the spatial dimension, and we assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$.

Matérn$_1$/Matérn$_2$		
Parameters	m_{M_1,M_2}	M_{M_1,M_2}
$\nu_1 < \nu_2$ $\frac{\alpha_2^2}{\alpha_1^2} \leq \frac{\nu_2+\frac{d}{2}}{\nu_1+\frac{d}{2}}$	$\frac{\Gamma\left(\nu_1+\frac{d}{2}\right)}{\Gamma\left(\nu_2+\frac{d}{2}\right)} 2^{\nu_1-\nu_2} \left(\frac{\alpha_2}{\alpha_1}\right)^d$	$+\infty$
$\nu_1 < \nu_2$ $\frac{\alpha_2^2}{\alpha_1^2} > \frac{\nu_2+\frac{d}{2}}{\nu_1+\frac{d}{2}}$	$\frac{\Gamma\left(\nu_1+\frac{d}{2}\right)}{\Gamma\left(\nu_2+\frac{d}{2}\right)} \frac{\left(\nu_1+\frac{d}{2}\right)^{\nu_2+\frac{d}{2}}}{\left(\nu_2+\frac{d}{2}\right)^{\nu_1+\frac{d}{2}}} \times \left(\frac{\alpha_2^2-\alpha_1^2}{2(\nu_2-\nu_1)}\right)^{\nu_2-\nu_1} \frac{\alpha_1^{2\nu_1}}{\alpha_2^{2\nu_2}}$	$+\infty$
$\nu_1 = \nu_2$ and $\alpha_2 \geq \alpha_1$	$\left(\frac{\alpha_1}{\alpha_2}\right)^{2\nu_1}$	$\left(\frac{\alpha_2}{\alpha_1}\right)^d$
$\nu_1 = \nu_2$ and $\alpha_2 < \alpha_1$	$\left(\frac{\alpha_2}{\alpha_1}\right)^d$	$\left(\frac{\alpha_1}{\alpha_2}\right)^{2\nu_1}$
$\nu_1 > \nu_2$ $\frac{\alpha_2^2}{\alpha_1^2} < \frac{\nu_2+\frac{d}{2}}{\nu_1+\frac{d}{2}}$	0	$\frac{\Gamma\left(\nu_1+\frac{d}{2}\right)}{\Gamma\left(\nu_2+\frac{d}{2}\right)} \frac{\left(\nu_1+\frac{d}{2}\right)^{\nu_2+\frac{d}{2}}}{\left(\nu_2+\frac{d}{2}\right)^{\nu_1+\frac{d}{2}}} \times \left(\frac{\alpha_2^2-\alpha_1^2}{2(\nu_2-\nu_1)}\right)^{\nu_2-\nu_1} \frac{\alpha_1^{2\nu_1}}{\alpha_2^{2\nu_2}}$
$\nu_1 > \nu_2$ $\frac{\alpha_2^2}{\alpha_1^2} \geq \frac{\nu_2+\frac{d}{2}}{\nu_1+\frac{d}{2}}$	0	$\frac{\Gamma\left(\nu_1+\frac{d}{2}\right)}{\Gamma\left(\nu_2+\frac{d}{2}\right)} 2^{\nu_1-\nu_2} \left(\frac{\alpha_2}{\alpha_1}\right)^d$

The parameter σ^2 is the variance of the Gaussian random field represented by the function (5.9). Thus, we are assuming that the random field under study is a linear combination of two random fields with a Matérn-type correlation structure, and that additionally have the same variance. Ergo, they can be distinguished only through the scale (α_i) and smoothing (ν_i) parameters characterizing their spatial dependence. This is also desirable by the fact that the range of permissibility for ϑ negative is a function of these parameters. In particular, one can appreciate this fact in Table 1, where the permissibility conditions for model (5.9) are resumed. We shall denote the model in equation (5.9) with the acronym *GSM*, by meaning *Generalized Sum of Matérn* models.

5.3.6 Fitting negative covariances to Nea Kessani data

In this section we illustrate the basic procedure for fitting negative covariances by using composite likelihood method as described in chapter 4. The aim of this section is two-fold. On the one hand,

we want to illustrate that the methodology we propose is useful for fitting data that exhibit negative empirical covariances. On the other hand, we need to show that it is necessary, for this dataset, to use this methodology. For the latter purpose, the modus operandi is necessarily comparative, so that the reader must be convinced that a covariance model that attains negative values has, for these data, a somehow better performance with respect to other (strictly positive) covariance models that one can easily find in the classic literature. As for the comparative performance of other covariance functions fitted to the same data, we shall evaluate some measures of error for the associated best linear unbiased predictor (BLUP, for short). This will be explained in detail subsequently.

As our procedure is necessarily comparative, we have selected two alternative models that attain only positive values:

1. The Generalized Cauchy model (*GC*, for short) introduced by Gneiting and Schlather (2004), who study its properties in terms of decoupling of the local and global behavior for the associated Gaussian random field

$$\mathrm{C}(h) = \sigma^2 \left(1 + ah^{\alpha}\right)^{-\beta}, \tag{5.10}$$

where $a > 0$ is the scaling parameter, $h = \|\mathbf{h}\|$ and $\mathbf{h} \in \mathbb{R}^d$. Necessary and sufficient conditions for the permissibility of this model are $\alpha \in (0, 2]$ and β positive.

2. The Gaussian model (*Gau*, for short), with equation

$$\mathrm{C}(h) = \sigma^2 \mathrm{e}^{-ah^2}, \tag{5.11}$$

with positive scaling parameter a.

Our selection was oriented towards these models because of their wide use in the geostatistical literature. In particular, a wide number of applications involving the use of Gaussian models can be found in the literature. The Generalized Cauchy model allows for identifying separately the Haussdorff dimension and the Hurst effect, which is desirable for those interested in modeling physical and geological phenomena with a geostatistical approach.

As far as model (5.9) is concerned, we had a practical problem with the estimation of the parameter ϑ. In particular, the problem was finding the bounds for this parameter in order to preserve the permissibility of (5.9), as $\vartheta := \vartheta(\alpha_1, \alpha_2, \nu_1, \nu_2)$. An empirical useful procedure that can overcomes this problem is the following:

Table 5.7: Summary of the estimates obtained by fitting the models *GC*, *Gau* and *GSM*.

Model	**Parameters**
GC	$\mathrm{C}(h) = 52.48\left(1 + 207079h^{1.54}\right)^{-337.2791}$
Gau	$\mathrm{C}(h) = 52.40734\mathrm{e}^{-3178h^2}$
GSM	$\mathrm{C}(h) = 50.67\{2.22\mathrm{C}_{\mathcal{M}}(h; 162, 1.12) - 1.22\mathrm{C}_{\mathcal{M}}(h; 211.44, 1.12)\}$

(i) We fix, for computational convenience, $\nu_1 = \nu_2 = \nu$. The number of parameters makes computation difficult, above all because of the small number of available data. Thus, we assume the same level of smoothness for the Matérn covariances in (5.9);

(ii) Estimate through CL the parameter vector $\boldsymbol{\Psi} := (\alpha_1, \alpha_2, \nu, \sigma^2)' \in \mathbb{R}^4_+$, *i.e.* find

$$\underset{\boldsymbol{\Psi}}{\mathrm{Argmax}} CL_1(\boldsymbol{\Psi}; \vartheta) := CL_2(\vartheta; \hat{\boldsymbol{\Psi}});$$

(iii) Find the corresponding bounds for the possible estimate of ϑ by using equation (5.9) and Table 1;

(iv) Maximize, with respect to ϑ, the function $CL_2(\vartheta; \hat{\boldsymbol{\Psi}})$.

This procedure is consistent to that of profile likelihood. Estimation results are reported in Table 2 together with those related to *GC* and *Gau* models, whilst Figure 5.12 illustrates the fittings for the three proposed models.

One can notice that estimates for the variance are very similar for the three models, but very different in terms of range. This is probably due to the fact that we assume the random field to be obtained through a weighted sum of independent random fields with different ranges. This procedure is sometimes used in order to model zonal anisotropy, but in this case we assume the covariances in the linear combination in equation (5.9) to be functions of the same argument. Subsequently we show that this procedure is useful, easy to implement and very effective in terms of goodness of fit performance.

Figure 5.12 illustrates the empirical covariance together with the three fitted models. It can be appreciated that there is substantially no graphical difference between *Gau* and *GC* models, as both

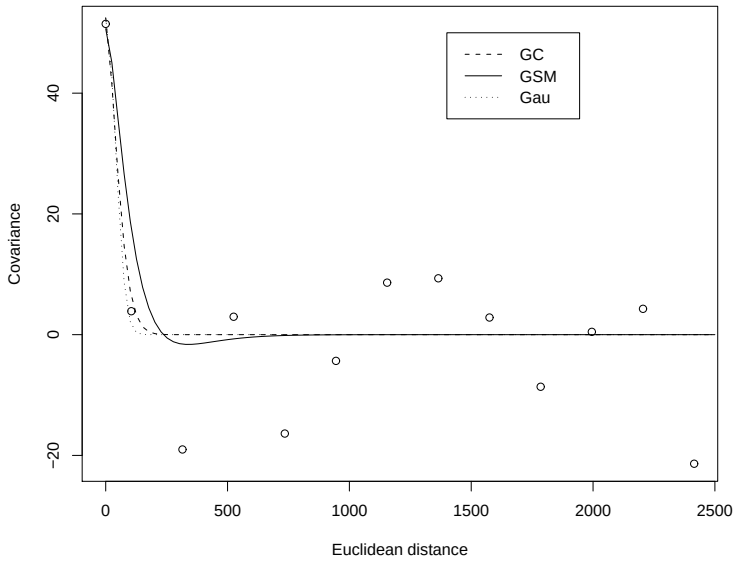


Figure 5.12: Empirical and fitted spatial covariances, with continuous line for *GSM*, dashed for *GC* and dotted for *Gau*

decay very fast to zero at low spatial lags. The *GSM* model attains negative values and then goes to zero. Actually, it would have been desirable, for having better diagnostic results, to have a model with lower negative values, but it was not possible, owing to the already mentioned constraints on the bound. It is also worth remarking that implementation and estimation of *GSM* model are very handy and that the model allows for an easy physical interpretation of the phenomena under study.

5.3.7 Diagnostics

In order to compare the performances of the proposed models in terms of fitting, we followed the classical cross-validation procedure. In particular, we use the cross-validation statistics $CRV1$, $CRV2$ and $CRV3$ proposed by Huang and Cressie (1996) and Carroll and Cressie (2002), whose expressions can be found therein. Roughly speaking, to every of the n points and corresponding realizations, we associate the simple kriging predictions (we assume the mean is known and constant, as the data was detrended) obtained by using, for every point $\mathbf{x}_i$ in the sampling design, the remaining $(n-1)$ points. Then, for every point one can calculate the prediction error and the

Table 5.8: Cross-validation results for the models GC, Gau and GSM

Model	*GC*	*Gau*	*GSM*
CRV1	-0.02981434	-0.08080574	-0.06923005
CRV2	0.9030204	2.547413	1.23477
CRV3	4.005985	5.598049	3.525365

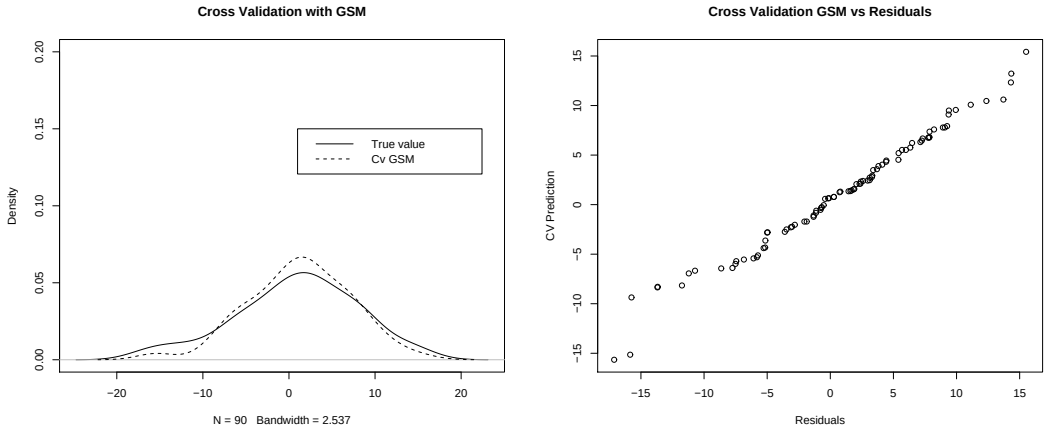


Figure 5.13: *Left*: Kernel density estimate for original residuals (continuous line) and GSM (dotted). *Right*: Observed versus predicted residuals, using *GSM*

quadratic prediction, that is, respectively, the difference and the squared difference between realization and prediction. Thus, $CRV1$ checks the relative unbiasedness of the prediction error and should be close to zero, while $CRV2$ checks the accuracy of standard deviation of the prediction error and should be close to one. We particularly focus on $CRV3$, as recommended by Huang and Cressie (1996), as it is a measure of goodness of prediction.

Cross-validation results are reported in Table 3. It is quite evident that, for the index $CRV3$, which is the most important, *GSM* model outperforms *GC* and *Gau*, whilst the *GC* is slightly better in CRV1, which measures the relative unbiasedness. In Figure 5.13 (*right*) we show the plot of observed residuals versus predicted ones, obtained with the GSM model, as well as the plot of the nonparametric kernel estimate of the density of the original residuals together with that of the predicted residuals obtained using the GSM model (Figure 5.13, *left*).

We also mapped the kriging values at some fixed depth levels. In Figure 5.14 one can appreciate

that there is a general trend resulting in lower temperatures when sampling at low depth. This is consistent with the information we have concerning the temperature of the geothermal field.

Figure 5.14.1

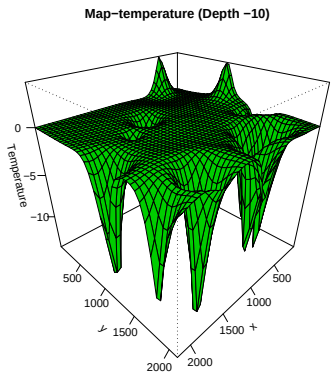


Figure 5.14.2

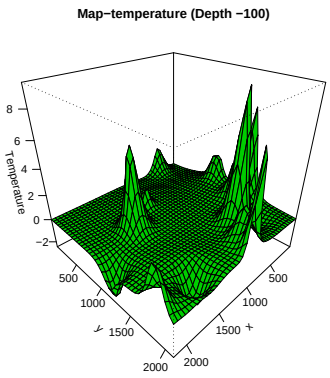


Figure 5.14.3

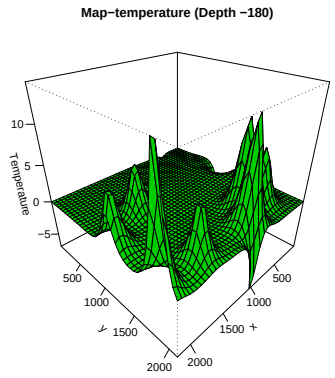


Figure 5.14.4

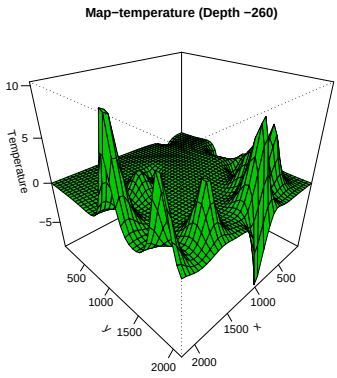


Figure 5.14: Simple Kriging predictions of Nea Kessani temperatures at different depths

5.3.8 Conclusions and discussion

We have shown a valid approach to fit negative empirical covariances to a gaussian random fields. The proposed method is easy to implement and allows to check directly the permissibility conditions in terms of bounds for the parameter ϑ that induces negative values on the corresponding covariance model. We showed that, for the presented case study, the proposed model outperforms in terms of goodness-of-fit other covariance models that are classically used in the geostatistical framework.

A future theme of research could be the following: find some permissibility criteria in order to

ensure negative weights on a linear combination of covariance functions that are additionally compactly supported. This is very challenging, as the arguments that can be used to show Proposition 1 in our approach are in general difficult to verify for compactly supported covariance functions, as they rarely admit a closed form for the associated Fourier transform. Note the Fourier transform associated to a covariance function with support on the unit ball of $\mathbb{R}^d$ is analytic and computation of m and M as in Proposition 1 is often infeasible.

5.4. TESTING SEPARABILITY FOR SPACE-TIME COVARIANCE FUNCTION

In recent years, it has become more prevalent to study some phenomena (such as air quality and climate change) across both space and time. Consequently, space-time data are being collected in many studies in an increasing number of disciplines. Modeling the space-time correlation is indispensable in the analysis of space-time data. A considerable number of space-time models have been proposed by recent literature by several authors (Cressie and Huang (1999) and Gneiting (2002) among others). However, the sample size of space-time data is usually quite large. Therefore cumbersome operations on the large covariance matrix are inevitable if one has to employ the likelihood-based inferences or Bayesian inferences. There is one case when the computation can be reduced, this is when the space-time covariance function is separable. Therefore it is always of interest to test the separability of space-time covariance function.

Let $\{Z(\mathbf{s},t),(\mathbf{s},t)\in\mathbb{D}=\mathbb{S}\times\mathbb{T}\subseteq\mathbb{R}^d\times\mathbb{R}\}$ be a weakly stationary random field with stationary covariance function $\mathrm{C}:\mathbb{D}\to\mathbb{R}$ that depends on the space-time lag $(\mathbf{h},u):=(\mathbf{s}_1-\mathbf{s}_2,t_1-t_2)\in\mathbb{D}$. Very popular simplifying assumptions are those of full symmetry and separability. The former means that $\mathrm{C}(\cdot,\cdot)$ is radially symmetric (or isotropic) in the spatial component and symmetric in the temporal one. $C(\mathbf{h},u)$ is separable if it can be represented as a product of a spatial covariance function $C_s(\mathbf{h})$ and a temporal covariance function $C_t(u)$, i.e.,$C(\mathbf{h},u)=C_s(\mathbf{h})C_t(u)$ for any $\mathbf{h}$ and u. The separability of $C(\mathbf{h},u)$ allows for ease of computation as the space-time covariance matrix factorizes into the Kronecker product of merely spatial and temporal covariance matrices. However, in most applications the lack of separability and full symmetry have been argued by several authors (see, *e.g.*, Gneiting et al. (2007) and a test of these assumptions is of great interest.

A number of tests for separability has been proposed recently. For instance Mitchell et al. (2005) proposed a test based on likelihood in a context of multivariate repeated measures, while Li et al. (2007) proposed a test based on asymptotic properties of empirical estimator of space-time

covariance function. Other proposed tests based on spectral densities can be found in Fuentes (2002), Fuentes (2005) and Scaccia and Martin (2005).

All of these tests are based on asymptotic results which are established under the assumption of separability. Instead of relying on asymptotic results, we propose a computational approach to test the separability, that is we approximate the distribution of the test statistic through simulations. Our approach can be reviewed as a parametric montecarlo test, that is we work under parametric assumptions, *i.e.* we suppose that the space-time dependence is described by a parametric family of covariance functions. The choice of the simulation based approach comes out basically from two reasons.

The former concerns the applicability of asymptotic to spatial or spatio-temporal data. There are two quite different asymptotic frameworks to which one can appeal in the space: increasing domain asymptotics, in which the minimum distance between sampling points is bounded away from zero and thus the spatial domain of observation is unbounded, and infill asymptotics, in which observations are taken ever more densely in a fixed and bounded domain.

The asymptotic results can be quite different under the two frameworks. For instance it is known that in the pure spatial setting (i.e., no temporal sampling takes place), not all parameters in the covariance functions are consistently estimable if the spatial domain is bounded (Zhang, 2004), but all parameters are consistently estimable under the increasing domain asymptotic framework (Mardia and Marshall, 1984). Zhang and Zimmerman (2005) compared the two asymptotic frameworks and showed that for some parameters, the fixed-domain asymptotic framework is preferred. In the space-time case, we expect all parameters are consistently estimable as long as the temporal domain is increasing. However, different asymptotic distributions are expected depending whether the spatial domain is bounded or increasing. A finite sample never tells us if the spatial domain is increasing or bounded. We therefore face a possibility of applying different asymptotic frameworks and hence different asymptotic results for the test. This results in different rejection regions for the test and potential contradictory conclusions.

The latter lies in the fact that for some space-time models, the separability corresponds to the parameter lying on the boundary of the parameter space (see for instance Cressie and Huang (1999), Gneiting (2002) and Porcu, Mateu and Christakos (2007)). Even using the full likelihood function, no simple solution to this inferential issue is available. For instance in the classical case, the asymptotic distribution of standard tests is a mixture of chi-squared rather than the usual

chi-squared distribution.

Clearly this is a problem for approximations of the likelihood as well such as WCL method. Simulation approach can be useful in this case. For instance Bellio and Varin (2005) propose a montecarlo test based on composite likelihood, to verify variance components of a generalized linear models with crossed random effects when lying on the bound. We know that WCL estimation method yields consistent and asymptotically gaussian estimates under increasing domain in space and time but this result are builded under regular conditions included the true parameter must lies on the interior of parameter space. Thus there are difficulties to establish the asymptotic distribution of the test statistic.

Since the montecarlo test approach is hard from computational point of view, ML method is ineffective. We use WCL as described in chapter 4.

5.4.1 Examples of space-time covariance functions

Here we consider some families of space-time covariance functions for which the separability corresponds to a parameter lying on the boundary of parameter space. Gneiting (2002) showed the following model is a valid space-time covariance function

$$\mathrm{C}(\mathbf{h}, u) = \frac{\sigma^2}{(a|u|^{2\alpha}+1)} \exp\left(-\frac{c\|\boldsymbol{h}\|^{2\gamma}}{(a|u|^{2\alpha}+1)^{\beta\gamma}}\right), \tag{5.12}$$

where σ^2 is the variance, $a > 0, \alpha > 0, \gamma > 0$ and $\beta \in [0, 1]$. It is obvious that the covariance function is separable if and only if $\beta = 0$, which is on the boundary of the interval.

Another family of space-time covariance functions to be considered is new. We consider a class of covariance functions that is obtained by composing two *margins*, that is, a merely spatial covariance function $\mathrm{C}_S(\cdot;\theta_S)$ and a temporal covariance function $\mathrm{C}_T(\cdot;\theta_T)$. The space-time covariance function is of the form

$$C(\mathbf{h}, u) \mapsto \psi_\delta\left(\mathrm{C}_S(\mathbf{h};\theta_S), \mathrm{C}_T(u;\theta_T)\right), \qquad (\mathbf{h}, u) \in \mathbb{D} \tag{5.13}$$

where

$$\psi_\delta(v_1, v_2) := \left(\frac{v_1^{-\delta} + v_2^{-\delta}}{2}\right)^{-1/\delta}, \qquad v_1, v_2 > 0, \quad \delta > 0. \tag{5.14}$$

We show in the Appendix that if both C_S and C_T are the generalized Cauchy, i.e.,

$$\mathrm{C}_S(\mathbf{h}) = (1 + a_1\|\mathbf{h}\|^{\gamma_1})^{2\beta_1}, \ \mathrm{C}_T(\mathbf{h}) = (1 + a_2\|\mathbf{h}\|^{\gamma_2})^{2\beta_2} \tag{5.15}$$

then

$$\begin{aligned} \mathrm{C}_\delta(\mathbf{h}, u; \theta_S, \theta_T) &= \sigma^2 \psi_\delta \left(\mathrm{C}_S(\mathbf{h}; \theta_S), \mathrm{C}_T(u; \theta_T) \right) \\ &= \sigma^2 \left(1/2 \left(1 + a_1 \|\mathbf{h}\|^{\gamma_1} \right)^{-2\beta_1 \delta} + 1/2 \left(1 + a_2 |u|^{\gamma_2} \right)^{-2\beta_2 \delta} \right)^{-1/\delta}, \end{aligned} \quad (5.16)$$

is a valid space-time covariance function, where σ^2 is the variance associated to a real-valued Gaussian weakly stationary RF defined on some domain of $\mathbb{R}^d \times \mathbb{R}$. The space-time parameter vector is $\theta = (\theta_S; \theta_T; \delta)' = (a_1, \gamma_1, \beta_1; a_2, \gamma_2, \beta_2; \delta)$, where the permissibility conditions are $\gamma_i \in]0, 2]$, $\beta_i < 1/2\delta$ and δ positive. Details for the proof of this permissibility conditions are contained in the Appendix.

The space-time parameter δ governs the type of interaction between space and time. Also, it can be readily verified that this class is increasing in δ.

Through a simple properties of the power means we obtain:

$$\lim_{\delta \to 0+} \mathrm{C}_\delta(\mathbf{h}, u) = \mathrm{C}_S(\mathbf{h}) \times \mathrm{C}_T(u), \quad (5.17)$$

for C_S and C_T belonging to the class (5.15) , and where the equality is intended in the limit, for $\delta \to 0^+$. We will write $C_0(\mathbf{h}, t) = \lim_{\delta \to 0+} \mathrm{C}_\delta(\mathbf{h}, u)$. Therefore, the separability of the space-time covariance again corresponds to the parameter δ lying on the boundary of the interval $(0, \infty)$.

5.4.2 Tests for separability

We consider a family of space-time covariance functions $C(\mathbf{h}, u, \theta), \theta \in \Theta$ which includes separable covariance functions as special cases. We assume that separable covariance functions correspond to $\theta \in \Theta_0 = \{\theta \in \Theta, \nu = 0\}$, where ν is an element of θ and describes the separability of the covariance model, while non separability correspond to $\theta \in \Theta_1 = \{\theta \in \Theta\}$. The hypothesis of separability of $C(\mathbf{h}, u, \theta)$ is therefore

$$H_0 : \ \nu = 0.$$

In model (5.16), $\Theta = \{(a_1, \gamma_1, \beta_1; a_2, \gamma_2, \beta_2; \delta), a_i > 0, \gamma_i > 0, 0 < \beta_i < 1/(2\delta), i = 1, 2 \text{ and } \delta \geq 0\}$ and $\Theta_0 = \{\theta \in \Theta, \delta = 0\}$. The null hypothesis can therefore be written as $H_0 : \delta \downarrow 0$. Similarly, for model (5.12), the hypothesis of separability is $H_0 : \ \beta = 0$. In both cases, the hypothesized value is on the boundary of the parameter space.

Here we adopt a parametric montecarlo test approach to develop tests of separability. On the base of the WCL estimation method proposed in chapter 4, two candidate tests can be easily built

by assuming, without loss of generality, that the covariance associated to the underlying space-time Gaussian RF belongs to the parametric family defined in equation (5.16) or in equation (5.12) for instance.

Recall that basically WCL is a two step estimation. First we have to identify the optimal space-time distance $\boldsymbol{d}^*$, maximising the Godambe Information associated to WCL estimating equation (it is usually done using a preliminaries WLS estimate see chapter 4). Then we maximize the $WCL(.,\boldsymbol{d}^*)$ function using the distance identified at previous step.

Note that in general $WCL(.,\boldsymbol{d}_1^*) \neq WCL(.,\boldsymbol{d}_2^*)$ if $\boldsymbol{d}_1^* \neq \boldsymbol{d}_2^*$ that is if we want to build through simulation the distribution of a statistics test based on WCL, we need to choose a fixed distance.

We propose to identify this fixed distance through simulation, that is we simulate under H_0 and find the minimum of the inverse of Godambe Information respect to the space-time distance, P times. Thus we consider the P identified space-time distances:

$$\boldsymbol{d}_k^* = \underset{\boldsymbol{d}_k \in \mathcal{D}}{\operatorname{argmin}} \operatorname{tr}(G^{-1}({}_{WLS}\hat{\boldsymbol{\theta}}_k, \boldsymbol{d}_k)), \qquad k = 1 \dots P \tag{5.18}$$

where $\mathcal{D} = \{(d_s, d_t) : \min_{i\neq j} \|s_i - s_j\| \leq d_s \leq \max_{i\neq j} \|s_i - s_j\|, \min_{i\neq j} |t_i - t_j| \leq d_t \leq \max_{i\neq j} |t_i - t_j|]\}$ and $G(.,.)$ is the Godambe Information associated to WCL as defined in chapter 4. Then we consider the mode of the P space-time distances and we indicate it as ${}_M\boldsymbol{d}^*$. Optimal distance allows for efficient estimation ad shown in chapter 4.

Then we simulate, R Gaussian random field with separable covariance function under the null hypothesis and we consider the WCL estimates under H_0 and H_1:

$${}_r\hat{\theta}_{H_0}^{WCL} = \underset{\theta\in\Theta_0}{\arg\max}\ {}_r\mathrm{WCL}(\theta;{}_M\,\boldsymbol{d}^*) \qquad r = 1 \dots R.$$

and

$${}_r\hat{\theta}_{H_1}^{WCL} = \underset{\theta\in\Theta_1}{\arg\max}\ {}_r\mathrm{WCL}(\theta;{}_M\,\boldsymbol{d}^*) \qquad r = 1 \dots R.$$

Then we consider two tests. The former is based on the scalar ${}_r\hat{\nu}_{H_1}^{WCL}$ an element of the vector ${}_r\hat{\theta}_{H_1}^{WCL}$. It is the parameter of separability such as the β parameter in (5.12) or the δ parameter in (5.16). The latter is a likelihood ratio type but builded with the WCL function:

$${}_rRWCL_{H_0H_1}^{WCL} = 2(\mathrm{WCL}({}_r\hat{\theta}_{H_0}^{WCL};{}_M\,\boldsymbol{d}^*) - \mathrm{WCL}({}_r\hat{\theta}_{H_1}^{WCL};{}_M\,\boldsymbol{d}^*)). \tag{5.19}$$

Let us indicate with ${}_R\nu^{WCL} = ({}_1\hat{\nu}_{H_1}^{WCL}, \dots, {}_R\,\hat{\nu}_{H_1}^{WCL})$ and ${}_RRWCL^{WCL} = ({}_1RWCL_{H_0H_1}^{WCL}, \dots, {}_R\,RWCL_{H_0H_1}^{WCL})$ two vectors containing the R estimates of the two tests proposed.

We approximate the quantiles of interest of the distribution of the two tests with the empirical quantile $\hat{q}({}_R\hat{\nu}^{WCL})_{1-\alpha}$ and $\hat{q}({}_R\hat{RWCL}^{WCL})_{1-\alpha}$

As a general remark, we need some stability properties of the sample quantile defined above, that is, for a sufficiently large number of observations, the tests are supposed to converge to a random variable whose probability distribution is unknown in the analytic form.

If the parameter setting identifying the null hypothesis belonged to the interior of the parametric space, then the distribution of the first test can be obtained by using asymptotic Gaussianity of the WCL estimates, while Guyon (1995) describes asymptotic distribution regarding the second test (a mixture of chi squares). Both asymptotic distribution works under increasing domain in space and time.

5.4.3 Simulation study

The goals of this simulation study is to explore the properties of the test, that is evaluate the probability of error of the first type $\hat{\alpha}$ with respect to the nominal value α fixed a priori, and evaluate the power of the proposed tests.

We perform this simulation using the covariance model (5.16) and we adopt the following space-time simulation setting: we consider a spatial network of 6×6 monitoring sites on the square $[1, 6]^2$, so that the horizontal and vertical separations of contiguous points are identically equal to one; each spatial point is observed over respectively $40, 90, 100$ and 110 temporal instants, which covers small and large temporal scales.

For the sake of simplicity, the scale parameters and the variance (which are beyond the interest of this study) are kept fixed and identically equal to one. As for the other parameters, we shall proceed either by

a Fixing γ_i and estimating β_i, or

b Fixing β_i and estimating γ_i, $i = 1, 2$

This setting finds a physical interpretation in the fact that in the case (**a**) the fractal dimension of the associated margins is kept fixed and the Hurst parameter is estimated, and viceversa for the case **b**. Here it is worth noticing that throughout this simulation study we use, for computational convenience, the Cauchy reparametrisation $t \mapsto \mathrm{C}(t) = (1 + t^{\gamma})^{\xi}$, for $\xi = -\beta/2\gamma$, so that the equivalent permissibility condition becomes $\beta_i < 2\gamma_i/\delta$, for $i = 1, 2$. This choice allows for more stable

Table 5.9: For the class in equation (5.16), identification of the optimal space-time distance for P=100 simulated Gaussian Random Fields. The simulation parameter setting is the scenario (**a**).

$d_s \vert d_t$	1	2	3	4	5	6	> 6
1	50	23	4	4	1	3	1
1.41	6	3	0	0	0	0	0
2	0	0	1	0	0	0	0
2.23	2	0	0	0	0	0	0
> 2.23	2	0	0	0	0	0	0

Table 5.10: For the class in equation (5.16), identification of the optimal space-time distance for P=100 simulated Gaussian Random Fields. The simulation parameter setting is the scenario (**b**)

$d_s \vert d_t$	1	2	3	4	5	6	> 6
1	47	6	7	2	1	0	1
1.41	5	3	3	1	1	2	1
2	5	1	1	0	0	0	1
2.23	2	0	0	0	1	0	0
> 2.23	6	1	1	1	0	0	0

.

estimates and for a larger interval over which β_i can be simulated. Using this reparametrisation we set for the case (**a**) $(\gamma_1, \gamma_2, \beta_1, \beta_2)$=$(0.5, 0.5, 1, 1)$ with β_i fixed and γ_i to be estimated and for the case (**b**) $(\gamma_1, \gamma_2, \beta_1, \beta_2)$=$(1, 1, 0.5, 0.5)$ with γ_i fixed and β_i estimated.

Before performing the test we need to identify the optimal space-time distance to use in the WCL estimates. Thus we simulate $P = 100$ Gaussian Random Fields under separability and at each step we minimize the Godambe Information respect to the space-time distance exploiting a preliminaries WLS estimates. Recall that Godambe Information is computed using subsampling technique with temporal windows as described in chapter 4. Table 5.9 and 5.10 reports the results for the case (**a**) and the case (**b**). We can see that the identified space-time distance is ${}_M d^* = (1, 1)^T$ for both cases. Throughout the simulation study we use this distance for the WCL estimation.

Then we compute via simulation the distribution of the tests under the null hypothesis. Thus we simulate R=600 Gaussian Random Fields and at each step we compute the statistics ${}_r\hat{\delta}^{WCL}_{H_1}$ and ${}_r RWCL^{WCL}_{H_0 H_1}$, $(r = 1 \dots 600)$ as defined in (5.19). Tables 5.11 and 5.12 resume the results and

Table 5.11: For the class in equation (5.16), estimates of quantiles of interest for test 1 (based on $\hat{\delta}_{H_1}^{WCL}$) and test 2 (based on $RWCL_{H_0H_1}^{WCL}$) for increasing domain in time scenario with $R = 600$. The simulation parameter setting is the scenario (**a**).

T	${}_1Mean$	${}_1\hat{q}_{0.90}$	${}_1\hat{q}_{0.95}$	${}_1\hat{q}_{0.99}$	${}_1Var$	${}_2Mean$	${}_2\hat{q}_{0.90}$	${}_2\hat{q}_{0.95}$	${}_2\hat{q}_{0.99}$
40	0.136	0.392	0.513	0.732	0.0365	2.734	8.423	13.499	23.223
90	0.091	0.287	0.354	0.483	0.0162	3.168	10.843	15.844	27.782
100	0.086	0.275	0.349	0.475	0.0146	3.173	10.573	16.363	29.674
110	0.083	0.271	0.342	0.454	0.0144	3.453	12.892	17.984	28.677

Table 5.12: For the class in equation (5.16), estimates of empirical quantiles of interest for test 1 (based on $\hat{\delta}_{H_1}^{WCL}$) and test 2 (based on $RWCL_{H_0H_1}^{WCL}$) for increasing domain in time scenario with $R = 600$. The simulation parameter setting is the scenario (**a**).

T	${}_1Mean$	${}_1\hat{q}_{0.90}$	${}_1\hat{q}_{0.95}$	${}_1\hat{q}_{0.99}$	${}_1Var$	${}_2Mean$	${}_2\hat{q}_{0.90}$	${}_2\hat{q}_{0.95}$	${}_2\hat{q}_{0.99}$
40	0.133	0.405	0.532	0.719	0.0375	1.27	4.118	6.672	10.442
90	0.085	0.264	0.336	0.505	0.0147	1.34	4.453	6.723	13.124
100	0.072	0.227	0.328	0.454	0.0116	1.16	3.463	6.925	12.354
110	0.067	0.224	0.327	0.443	0.0108	1.15	3.903	6.429	13.028

show the statistics of interests of the estimated distribution of the two test in a increasing time setting. In both tables we can appreciate that increasing the time domain, mean and variability of the separability parameter tend to decrease as we could expect. Quantiles of both tests are similar among 90 100 and 110.

5.4.4 Evaluation of Type I error

To evaluate the first type error, we simulate, under H_0, $K = 300$ weakly stationary zero-mean Gaussian random fields with covariance function belonging to the class (5.16) for the scenario (**a**) and scenario (**b**). To compute the empirical first type error we compare the quantile of α order of the estimates and compare with the quantile computed in the previous section for both tests. Tables 5.13 and 5.14 present the results. As it can be seen, the empirical probability of error of the first type are concordant with the respective nominal value α when approaching to 90 times. Moreover the estimates of first type error are very similar for the two tests.

Table 5.13: For the class in equation (5.16), estimated empirical error of the type I ($\hat{\alpha}$), and nominal value (α) for test 1 and test 2, under $T = 40, 90, 100, 110$. The simulation parameter setting is the scenario (**a**).

Test 1			
T	$\hat{\alpha}$ $(\alpha = 0.1)$	$\hat{\alpha}$ $(\alpha = 0.05)$	$\hat{\alpha}$ $(\alpha = 0.01)$
40	0.132	0.072	0.019
90	0.105	0.053	0.015
100	0.106	0.053	0.013
110	0.093	0.048	0.012
Test 2			
T	$\hat{\alpha}$ $(\alpha = 0.1)$	$\hat{\alpha}$ $(\alpha = 0.05)$	$\hat{\alpha}$ $(\alpha = 0.01)$
40	0.134	0.052	0.007
90	0.102	0.053	0.019
100	0.110	0.046	0.012
110	0.095	0.045	0.013

Table 5.14: For the class in equation (5.16), estimated empirical error of the type I ($\hat{\alpha}$), and nominal value (α) for both tests, under $T = 40, 90, 100, 110$. The simulation parameter setting is the scenario (**b**).

Test 1			
T	$\hat{\alpha}$ $(\alpha = 0.1)$	$\hat{\alpha}$ $(\alpha = 0.05)$	$\hat{\alpha}$ $(\alpha = 0.01)$
40	0.067	0.026	0.005
90	0.075	0.053	0.010
100	0.095	0.039	0.007
110	0.103	0.045	0.008
Test 2			
T	$\hat{\alpha}$ $(\alpha = 0.1)$	$\hat{\alpha}$ $(\alpha = 0.05)$	$\hat{\alpha}$ $(\alpha = 0.01)$
40	0.067	0.028	0.005
90	0.078	0.053	0.013
100	0.097	0.040	0.007
110	0.105	0.052	0.008

Table 5.15: For the class in equation (5.16), empirical power of the tests $\hat{\pi}$, under $T = 40, 90, 100, 110$. The RF is simulated under $H_1 : \delta = 0.3$. The simulation parameter setting is the scenario (**a**).

Test 1 and $\delta = 0.3$			
T	$\hat{\pi}$ $(\alpha = 0.1)$	$\hat{\pi}$ $(\alpha = 0.05)$	$\hat{\pi}$ $(\alpha = 0.01)$
40	0.457	0.312	0.118
90	0.523	0.365	0.189
100	0.579	0.447	0.208
110	0.632	0.491	0.283
Test 2 and $\delta = 0.3$			
T	$\hat{\pi}$ $(\alpha = 0.1)$	$\hat{\pi}$ $(\alpha = 0.05)$	$\hat{\pi}$ $(\alpha = 0.01)$
40	0.475	0.322	0.136
90	0.546	0.407	0.207
100	0.604	0.472	0.227
110	0.652	0.531	0.291

5.4.5 Evaluation of the power of the test

To evaluate the power of the test, we simulate, under H_1, $K = 300$ weakly stationary zero-mean Gaussian random fields with covariance function belonging to the class (5.16) for the scenario (**a**) and scenario (**b**). Specifically we simulate with $\delta = 0.3$ and $\delta = 0.5$. To compute the empirical power we compare the quantile of α order of the estimates with the quantile computed in the previous section for both tests. Tables 5.15, 5.16, 5.17, 5.18 present the results. As it can be seen, the power increases with the time for both tests and tend to converge near 90 or 100 times. Moreover, as we expect, the power increases when moving from $\delta = 0.3$ to $\delta = 0.5$. Note that the second test always seems to perform slightly better.

5.4.6 Conclusion and discussion

It is of interest to test the separability of space-time covariance function. Here we work under parametric assumption and we propose two parametric montecarlo tests to verify the separability of a covariance model exploiting WCL estimation method. The choice of the computational approach comes out for two main reasons. First, the asymptotic distribution of the tests proposed in

Table 5.16: For the class in equation (5.16), empirical power of the tests $\hat{\pi}$, under $T = 40, 90, 100, 110$. The RF is simulated under $H_1 : \delta = 0.3$. The simulation parameter setting is the scenario (**b**).

Test 1 and $\delta = 0.3$			
T	$\hat{\pi}$ $(\alpha = 0.1)$	$\hat{\pi}$ $(\alpha = 0.05)$	$\hat{\pi}$ $(\alpha = 0.01)$
40	0.413	0.258	0.13
90	0.627	0.457	0.157
100	0.704	0.498	0.242
110	0.728	0.532	0.265
Test 2 and $\delta = 0.3$			
T	$\hat{\pi}$ $(\alpha = 0.1)$	$\hat{\pi}$ $(\alpha = 0.05)$	$\hat{\pi}$ $(\alpha = 0.01)$
40	0.466	0.322	0.175
90	0.682	0.533	0.228
110	0.754	0.598	0.344
100	0.775	0.632	0.352

Table 5.17: For the class in equation (5.16), empirical power of the tests $\hat{\pi}$, under $T = 40, 90, 100, 110$. The RF is simulated under $H_1 : \delta = 0.5$. The simulation parameter setting is the scenario (**a**).

Test 1 and $\delta = 0.5$			
T	$\hat{\pi}$ $(\alpha = 0.1)$	$\hat{\pi}$ $(\alpha = 0.05)$	$\hat{\pi}$ $(\alpha = 0.01)$
40	0.616	0.485	0.230
90	0.856	0.785	0.580
100	0.896	0.818	0.620
110	0.912	0.886	0.643
Test 2 and $\delta = 0.5$			
T	$\hat{\pi}$ $(\alpha = 0.1)$	$\hat{\pi}$ $(\alpha = 0.05)$	$\hat{\pi}$ $(\alpha = 0.01)$
40	0.646	0.495	0.260
90	0.860	0.801	0.616
100	0.910	0.833	0.650
110	0.920	0.904	0.680

Table 5.18: For the class in equation (5.16), empirical power of the tests $\hat{\pi}$, under $T = 40, 90, 100, 110$. The RF is simulated under $H_1 : \delta = 0.5$. The simulation parameter setting is the scenario (**b**).

Test 1 and $\delta = 0.5$			
T	$\hat{\pi}$ $(\alpha = 0.1)$	$\hat{\pi}$ $(\alpha = 0.05)$	$\hat{\pi}$ $(\alpha = 0.01)$
40	0.583	0.423	0.201
90	0.903	0.826	0.523
100	0.953	0.840	0.572
110	0.965	0.900	0.701
Test 2 and $\delta = 0.5$			
T	$\hat{\pi}$ $(\alpha = 0.1)$	$\hat{\pi}$ $(\alpha = 0.05)$	$\hat{\pi}$ $(\alpha = 0.01)$
40	0.682	0.510	0.342
90	0.936	0.890	0.660
110	0.973	0.893	0.755
100	0.975	0.950	0.803

literature in the spatial and spatio-temporal case depends on which kind of asymptotics one consider, second testing for separability of a space-time covariance model often means works on the bound of the parameter space. This is a non standard problem for likelihood and approximations of the likelihood.

We perform a simulation study on a particular class which presents this kind of problem. Simulation results seems promising for both tests even if the test based on WCLRT seems to work slightly better.

Chapter 6

Appendix

6.1. PROOF OF PROPOSITION 1

Proof. Firstly, suppose that ψ is a Bernstein function. Any bivariate Laplace transform $\mathcal{L}$ yields the integral representation

$$\mathcal{L}(\theta_1, \theta_2) = \int_0^\infty \int_0^\infty e^{-\theta_1 w_1 - \theta_2 w_2} \mathrm{d}F(w_1, w_2), \tag{6.1}$$

where F is some distribution function. Indeed, (3.13) can be written as

$$\int_0^\infty \int_0^\infty \prod_{i=1}^{d} e^{-\psi_i(|h_i|)w_1} e^{-\psi_t(|u|)w_2} \mathrm{d}F(w_1, w_2). \tag{6.2}$$

Now, observe that for Criterion 2 in (W. (1966) p. 494), if ψ is a Bernstein function and φ a completely monotone function, then $\varphi \circ \psi$ is completely monotone, and for well known results, any completely monotone function of the type $\varphi(|t|)$ is positive definite. Finally, note that the function $\exp(-|x|)$, x real, is completely monotone. Thus, $e^{-\psi_i(|h_i|)w_1}$ is a covariance function for any $i = 1, \ldots, d$ and $w_1 > 0$. Also observe that the product of covariance functions is still a covariance function. To end up, note that (6.2) is a positive mixture of covariance functions, indeed a valid covariance function in $\mathbb{R}^d \times \mathbb{R}$.

Secondly, if ψ is a continuous, increasing and concave function, then it is negative definite on $\mathbb{R}$, according to a Pólya type criterion (Berg and Forst (1975), Proposition 10.6). So is the sum of continuous, increasing and concave functions. Thus, for the same arguments as before, the proof is completed. □

6.2. PROOF OF PROPOSITION 2

Proof. Recall that $\mathcal{L}$ admits the integral representation in equation (6.1), and that for construction F yields condition (3.18). Now, in order to apply the operator $D_k\mathcal{L}$, observe that the negative exponential $\exp(-w_1\theta_1 - w_2\theta_2)$, which is the integrand in (6.1), has bounded derivatives, and following condition (3.18) we can differentiate under the integral sign. Thus

$$D_k\mathcal{L}(\theta_1,\theta_2) = \int_0^\infty \int_0^\infty w_1^k w_2^k \exp(-w_1\theta_1 - w_2\theta_2)\mathrm{d}F(w_1,w_2). \tag{6.3}$$

Now, for condition (3.18) it is possible to apply the dominated convergence theorem, so that we can substitute in accordance to (3.19) under the integral sign and obtain that

$$\nu(\mathbf{h},u) = \frac{\int_0^\infty \int_0^\infty \exp(-w_1\gamma_s(\mathbf{h}) - w_2\gamma_t(u))w_1^k w_2^k \mathrm{d}F(w_1,w_2)}{\int_0^\infty \int_0^\infty w_1^k w_2^k \mathrm{d}F(w_1,w_2)}. \tag{6.4}$$

Note that Schoenberg Theorem applies to $\exp(-\gamma)$ for every positive w_1, w_2, while the denominator in (6.4) normalises the measure in the numerator, so that (6.4) is a positive mixture of space-time covariance functions, indeed a covariance function. And this fact completes the proof. □

6.3. PROOF OF PROPOSITION 3

Proof. For abuse of notation, the class $\mathcal{L}^*$ admits the representation

$$\mathcal{L}^*(x_1,\ldots,x_d,x_t) = \int_0^\infty \int_0^\infty e^{-w_1\sum_{i=1}^d \psi_i(x_i) - w_2\psi_t(x_t)}\mathrm{d}F(w_1,w_2). \tag{6.5}$$

Now, as the derivatives of the integrand are bounded, we can differentiate under the integral sign in the representation above, so that

$$\begin{aligned} D\mathcal{L}^*(x_1,\ldots,x_d,x_t) &= (-1)^{d+1}\prod_{i=1}^d \psi_i'(x_i)\psi_t'(x_t) \\ &\times \int_0^\infty \int_0^\infty w_1^d w_2 e^{-w_1\sum_{i=1}^d \psi_i(x_i) - w_2\psi_t(x_t)}\mathrm{d}F(w_1,w_2). \end{aligned} \tag{6.6}$$

In order to obtain the result, let us focus on the first factor, *i.e.* $\prod_{i=1}^d \psi_i'(x_i)\psi_t'(x_t)$. Note that the first derivative of a Bernstein function is a completely monotone function, and for similar arguments to the previously shown results, we have that $\psi' \circ \gamma$, where γ is a variogram, is a covariance function. As the tensorial product of covariances is still a covariance, we have that

$$\xi(h_1,\ldots,h_d,u) = \prod_{i=1}^d \psi_i'(\gamma_i(h_i))\psi_t'(\gamma_t(u))$$

is a covariance function.

Let us now focus on the second factor. Applying formula (3.22) we get

$$\nu^*(h_1,\ldots,h_d,u) = \frac{\xi(h_1,\ldots,h_d,u)\int_0^\infty\int_0^\infty w_1^d w_2 e^{-w_1\sum_{i=1}^d \psi_i(\gamma_i(h_i))-w_2\psi_t(\gamma_t(u))}\mathrm{d}F(w_1,w_2)}{k_\xi\int_0^\infty\int_0^\infty w_1^d v F(w_1,w_2)}, \tag{6.7}$$

where:

- $k_\xi = \prod_{i=1}^d \psi_i'(\gamma_i(0))\psi_{d+1}'(\gamma_t(0))$ is a finite positive constant (due to direct theorem's construction), as $\gamma_i(0) = 0$ and $\psi_i'(0)$, $i = 1,\ldots,d,t$, exist and are finite;
- following the dominated convergence theorem, we can evaluate the integrand at $\mathbf{0} \in \mathbb{R}^2$ and obtain the above expression. As the denominator normalises the measure in the numerator, the second factor results in a positive mixture of covariance functions, indeed a covariance function.

These facts complete the proof. □

6.4. TECHNICAL DETAILS FOR THE PERMISSIBILITY OF THE CLASS IN EQUATION (5.16).

We recall that a function $f :]0,\infty[\to \mathbb{R}$ is called completely monotonic, if it is C^∞ and

$$(-1)^n f^{(n)}(x) \geq 0 \text{ for } x > 0, n = 0, 1, \ldots. \tag{6.8}$$

By the celebrated theorem of Bernstein the set $\mathcal{C}$ of completely monotonic functions coincides with the set of Laplace transforms of positive measures μ on $[0,\infty[$, cf. (Widder, 1949), i.e.

$$f(x) = \int_0^\infty e^{-xt}\, d\mu(t), \tag{6.9}$$

where the only requirement on μ is that e^{-xt} is μ-integrable for any $x > 0$. The set $\mathcal{C}$ is closed under addition, multiplication and pointwise convergence.

A function $f :]0,\infty[\to [0,\infty[$ is called a *Bernstein function*, if it is C^∞ and $f' \in \mathcal{C}$.

The set of Bernstein functions is denoted $\mathcal{B}$ and is a convex cone closed under pointwise convergence.

Since a Bernstein function is non-negative and increasing, it has a non-negative limit $f(0+)$. Integrating the Bernstein representation of the completely monotonic function f' gives the following integral representation of $f \in \mathcal{B}$

$$f(x) = \alpha x + \beta + \int_0^\infty (1 - e^{-xt})\, d\nu(t), \tag{6.10}$$

where $\alpha, \beta \geq 0$ and ν, called the *Lévy measure*, is a positive measure on $]0, \infty[$ satisfying

$$\int_0^\infty \frac{t}{1+t}\, d\nu(t) < \infty.$$

The following composition result is useful, see Berg (2007):

Let X be any of the sets $\mathcal{B}, \mathcal{C}, \mathcal{L}$. Then

$$f \in X, g \in \mathcal{B} \Rightarrow f \circ g \in X.$$

We are now able to show the permissibility conditions for the class (5.16). Recall that, under the constraint $\beta < 1/2\delta$, the function $t \mapsto (1 + t^\gamma)^{2\beta\delta}$, for t positive argument and $\gamma \in (0, 2]$, is a Bernstein function, and thus is evaluation in either the arguments $\|\mathbf{x}\|$, $\mathbf{x} \in \mathbb{R}^d$, and $|x|$, $x \in \mathbb{R}$, is a Bernstein function and thus negative definite respectively on $\mathbb{R}^d$ and $\mathbb{R}$. So is the sum that will be negative definite on $\mathbb{R}^d \times \mathbb{R}$. Finally, direct application of Schoenberg (1993) theorem completes the proof.

Bibliography

Abromowitz, M., and Stegun, I. (1967), *Handbook of Mathematical Functions*, : U.S. Goverment Printing Office.

Aho, A. V., Hopcroft, J. E., and Ullman, J. D. (1974), *The Design and Analysis of Computer Algorithms*, Reading, Massachusset: Addison-Wesley.

Akaike, H. (2003), "Block Toeplitz matrix inversion," *SIAM Journal of Applied Mathematics*, 14, 234–241.

Banerjee, S. (2003), "On geodetic distance computations in spatial modelling," *Biometrics*, 61, 617–625.

Bellio, R., and Varin, C. (2005), "A pairwise likelihood approach to generalized linear models with crossed random effects," *Statistical Modelling*, 5, 217–227.

Benjamin, B., and Pollard, J. (1991), *The Analysis of Mortality and Other Actuarial Statistics*, London: Butterworth-Heinemann.

Berg, C., and Forst, G. (1975), *Potential Theory on Locally Compact Abelian Groups*, New York: Springer.

Bogaert, P., and Christakos, G. (1997), "Stochastic analysis of spatio-temporal solute content measurements using a regression model.," *Stoch. Hydrol. Hydraul.*, 11, 267–295.

Booth, H. (2006), "Demographic forecasting: 1980 to 2005 in review.," *International Journal of Forecasting*, 22(3), 547–582.

Booth, H., Maindonald, J., and Smith, L. (2002), "Applying Lee-Carter under conditions of variable mortality decline.," *Population Studies*, 56(3), 325–336.

Brouhns, N., D. M., and Vermunt, J. (2002), "A Poisson log-bilinear regression approach to the construction of projected lifetables.," *Insurance: Mathematics and Economics*, 31(3), 373–393.

Caragea, P., and Smith, R. (2006), Approximate likelihoods for spatial processes,, Technical report, Department of Statistics, Iowa State University.

Carlstein, E. (1986), "The use of subseries values for estimating the variance of a general statistic from a stationary sequence," *The Annals of Statistics*, 14(3), 1171–1179.

Carroll, S., and Cressie, N. (2002), "Spatial modeling of snow water equivalent using covariances estimated from spatial and geomorphic attributes.," *Journal of Hydrology*, 190, 42–59.

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 Hawkins, D. (1980), "Robust estimation of the variogram I," *Mathematical Geology*, 12, 115–125.

Cressie, N., and Huang, H. (1999), "Classes of nonseparable, spatiotemporal stationary covariance functions," *Econometric Theory*, 2, 305–330.

Cressie, N., and Lahiri, S. (1996), "Asymptotics for REML estimation of spatial covariance parameters," *Journal of Statistical Planning and Inference*, 50, 327–341.

Cressie, N., and Lahiri, S. (2002), "On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters," *J. Statist. Plann. Inference*, 21, 65–85.

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*a*), "Product-sum covariance for space-time modeling: an environmental application," *Environmetrics*, 12, 11–23.

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

De Cesare, L., Myers, D., and Posa, D. (2002), "Nonseparable space-time covariance models: some parametric families," *Mathematical Geology,*, 34, 23–42.

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

Debón, A., Martínez-Ruiz, F., and Montes, F. (2004), "Dynamic life tables: a geostatistical approach.," *International Congress on Insurance: Mathematics and Economics,Rome, Italy*, .

Debón, A., Montes, F., and Puig, F. (2006), "Modelling and forecasting mortality in Spain.," *European Journal of Operation Research*, .

Debón, A., Montes, F., and Sala, R. (2005), "A comparison of parametric models for mortality graduation. application to mortality data of the Valencia region (Spain).," *Statistics and Operations Research Transactions*, 29(2), 269–287.

Debón, A., Montes, F., and Sala, R. (2006*a*), "A comparison of models for dinamical life tables. application to mortality data of the Valencia region (Spain).," *Lifetime Data Analysis*, 12(2), 223–244.

Debón, A., Montes, F., and Sala, R. (2006*b*), "A comparison of nonparametric methods in the graduation of mortality: application to data from the Valencia region (Spain).," *International Statistical Review*, 74(2), 215–233.

Diggle, P. J., Moyeed, R. A., and Tawn, J. A. (1998), "Model-based geostatistics," *Applied Statistics*, 47(299-350).

Dimitrakopoulos, R., and Lou, X. (1994), *Spatiotemporal Modeling: covariances and ordinary kriging system, Geostatistics for the Next Century*, : Kluwer Academic.

Felipe, A., M., G., and Pérez-Marín, A. (2002), "Recent mortality trends in the Spanish population.," *British Actuarial Journal*, 8(4), 757–786.

Fernández-Casal, R. (2003), "Geostadistica Espacio-Temporal. Modelos Flexibles de Variogramas Anisotropicos no Separables.," *Unpublished PhD. Thesis, University of Santiago de Compostela. In spanish.*, .

Forfar, D., McCutcheon, J., and Wilkie, A. (2002), "On graduation by mathematical formula.," *Journal of the Institute of Actuaries*, 459, 1–149.

Fuentes, M. (2002), "Spectral methods for nonstationary spatial processes," *Biometrika*, 89(197-210).

Fuentes, M. (2005), "Testing separability of spatio-temporal covariance functions," *Journal of Statistical Planning and Inference*, 136(447-466).

Fuentes, M. (2007), "Approximate likelihood for large irregularly spaced spatial data," *Journal of the American Statistical Association*, 102(321-331).

Furrer, R., Genton, M., and Nychka, D. (2007), "Covariance tapering for interpolation of large spatial datasets," *J. Comput. Graph. Statist.*, 15(502-523).

Gavin, J., Haberman, S., and Verrall, R. (1993), "Moving weighted average graduation using kernel estimation.," *Insurance: Mathematics & Economics*, 12(2), 113–126.

Gavin, J., Haberman, S., and Verrall, R. (1994), "On the choice of bandwidth for kernel graduation.," *Journal of the Institute of Actuaries*, 121, 119–134.

Gavin, J., Haberman, S., and Verrall, R. (1995), "Graduation by kernel and adaptive kernel methods with a boundary correction.," *Transactions. Society of Actuaries*, XLVII, 173–209.

Genest, C. (1987), "Frank's family of bivariate distributions," *Biometrika*, 74, 549–555.

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 and Hall/CRC, pp. 151–175.

Gneiting, T., and Schlather, M. (2004), "Stochastic models that separate fractal dimension and the Hurst effect," *SIAM Review*, 46, 69–282.

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

Gregori, P., Porcu, E. Mateu, J., and Sasvári, Z. (2007), "On potentially negative space time covariances obtained as sum of products of marginal ones," *Annals of the Institute of Statistical Mathematics*, to appear.

Guillen, M., and Vidiella, A. (2005), "Forecasting Spanish natural life expectancy.," *Risk Analysis*, 101, 1161–1170.

Guyon, X. (1995), *Random fields on a network. Modeling, statistics and applications*, New York: Springer.

Handcock, M. S., and Wallis, J. R. (1994), "An approach to statistical spatial-temporal modeling of metereological fields," *Journal of the American Statistical Association*, 89, 368–378.

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.

Heyde, C. (1997), *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*, New York: Springer.

Higdon, D. (1998), "A process-convolution approach to modeling temperatures in the North Atlantic Ocean," *Journal of Environmental and Ecological Statistics*, 5, 173–190.

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

Huang, H.-C., and Cressie, N. (1996), "Spatio-temporal prediction of the snow water equivalent using the Kalman filter.," *Computational Statistics and Data Analysis*, 22, 159–175.

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.

Joe, H. (1997), *Multivariate Models and Dependence Concept*, : Chapman & Hall.

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.

Journel, A. G., and Huijbregts, C. J. (1978), *Mining geostatistics*, revised edn, London: Academic Press.

Kaufmann, C., Schervish, M., and Nychka, D. (2007), Covariance tapering for likelihood-based estimation in large spatial datasets,, Technical report, Statistical and Applied Mathematical Sciences Institute,Research Triangle Park, North Carolina.

Koissi, M., Shapiro, A., and Högnäs, G. (2006), "Evaluating and extending the Lee-Carter model for mortality forecasting confidence interval.," *Insurance: Mathematics and Economics*, 38(1), 1–1.

Kyriakidis, P., and Journel, A. (1999), "Geostatistical Space-Time Models: A Review," *Mathematical Geology*, 31, 651–684.

Lee, R. (2000), "The Lee-Carter method for forecasting mortality, with various extensions and applications.," *North American Actuarial Journal*, 4(1), 80–91.

Lee, R., and Carter, L. (1992), "Modelling and forecasting U. S. mortality.," *Journal of the American Statistical Association*, 87(419), 659–671.

Li, B., G., G., and Sherman, M. (2007), "A nonparametric assessment of properties of space-time covariance functions," *Journal of the American Statistical Association*, 102, 736–744.

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.

Ma, C. (2005), "Spatio-temporal variograms and covariance models," *Advances in Applied Probability*, 37, 706–725.

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.

Paciorek, C. J., and Schervish, M. J. (2006), "Spatial modelling using a new class of nonstationary covariance functions," *Environmetics*, 17, 483–506.

Pedroza, C. (2006), "A bayesian forecasting model: predicting U.S. male mortality.," *Biostatistics*, 7(4), 530–550.

Perrin, O., and Senoussi, R. (1999), "Reducing non-stationary random fields to stationarity and isotropy using a space deformation.," *Statistics and Probability Letters*, 48(1), 23–32.

Pitacco, E. (2004), "Survival models in dynamic context: a survey.," *Insurance: Mathematics & Economics*, 35(2), 279–298.

Porcu, E., Gregori, P., and Mateu, J. (2007), "La descente et la montée étendues: the spatially d-anisotropic and the spatio-temporal case," *Stochastic Environmental Research and Risk Assessment*, 21, 683–693.

Porcu, E., Mateu, J., and Christakos, G. (2007), Quasi-arithmetic means of covariance functions with potential applications to space-time data,, Technical report, Dept Maths, Universitat Jaume I.

Porcu, E., Saura, F., and Mateu, J. (2005), "New classes of covariance and spectral density functions for spatio-temporal modelling," *Stochastic Environmental Research and Risk Assessment*, To appear.

Renshaw, A. (1991), "Actuarial graduation practice and generalised linear models.," *Journal of the Institute of Actuaries*, 118, 295–312.

Renshaw, A., and Haberman, S. (2003*a*), "Actuarial Research paper No. 153," *City University, London*, .

Renshaw, A., and Haberman, S. (2003*b*), "Lee-Carter mortality forecasting: a parallel generalized linear modelling aproach for England and Wales mortality projections.," *Journal of the Royal Statistical Society C*, 52(1), 119–137.

Renshaw, A., and Haberman, S. (2003*c*), "Lee-Carter mortality forecasting with age specific enhancement.," *Mathematics and Economics*, 255-272.

Rouhani, S., and Hall, T. (1989), "Space-time Kriging of Groundwater Data," *Geostatistics, Armstrong, M. (Ed.), Kluwer Acaemic Publishers*, 2.

Rouhani, S., and Wackernagel, H. (1990), "Multivariate geostatistical approach to space-time data analysis," *Water Resour. Res.*, 26.

Scaccia, L., and Martin, R. (2005), "Testing axial symmetry and separability of lattice processes," *Journal of Statistical Planning and Inference*, 131.

Schoenberg, I. (1938), "Metric spaces and completely monotone functions," *Annals of Mathematics*, 39.

Shapiro, A., and Botha, J. (1991), "Variogram fitting with a conditional class of conditionally nonnegative definite functions," *Computational Statistics and Data Analysis*, 11.

Shumway, R., and Stoffer, D. (2006), "Time Series Analysis and its Applications with R Examples," *Springer, New York*, .

Smith, R. (2001), *Environmental Statistics*, http://www.unc.edu/depts/statistics/postscript/rs/envnotes.pdf: Under revision.

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.

Tabeau, E., v. d. B. J. A. (2001), *A Review of Demographic Forecasting Models for Mortality.Forecasting in Developed Countries: From description to explanation*, : Kluwer Academic.

Thanassoulas, C., and Tsokas, G. (1990), "A geothermal field model based on geophysical and thermal prospectings in Nea Kessani (NE Greece)," *Geothermics*, 19, 77–85.

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

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

W., F. (1966), *An intruduction to Probability Theory and Its Applications. Vol II*, New York: WIley.

Whittle, P. (1954), "On stationary processes in the plane.," *Biometrika*, 49, 305–314.

Wilmoth, J. (1993), "Computational methods for fitting and extrapolating the Lee-Carter model of mortality change.," *Technical report, Departament of Demography, University of California, Berkeley*, .

Wong-Fupuy, C., and Haberman, S. (2004), "Projecting mortality trens: Recent developents in the United Kingdom and the United States," *North American Actuarial Journal*, 8(2), 56–83.

Yaglom, A. (1987), *Correlation Theory of Stationary and Related Random Functions*, New York: Springer.

Ying, Z. (1993), "Maximum likelihood estimation of parameters under a spatial sampling scheme," *Annals of Statistics*, 21, 1567–1590.

Zhang, H. (2004), "Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics," *Journal of the American Statistical Association*, 99, 250–261.

Zhang, H., and Zimmerman, D. (2005), "Towards reconciling two asymptotic frameworks in spatial statistics," *Biometrika*, 92, 921–936.

Zimmerman, D. (1989), "Computationally exploitable structure of covariance matrices and generalized covariance matrices in spatial models," *Journal of Statistical Computation and Simulation*, 32, 1–15.