1

Collecting spatio-temporal data

Jorge Mateu[1] **and Werner G. Müller**[2]

[1]*Department of Mathematics, University of Jaume I of Castellon, Spain*

[2]*Department of Applied Statistics, Johannes Kepler University Linz, Austria*

1.1 Introduction

In this volume we intend to provide a comprehensive state-of-the-art presentation combining both classical and modern treatments of network design and planning for spatial and spatio-temporal data acquisition. A common problem set is interwoven throughout the chapters, providing various perspectives to illustrate a complete insight to the problem at hand. Motivated by the high demand for statistical analysis of data that takes spatial and spatio-temporal information into account, this book incorporates ideas from the areas of time series, spatial statistics and stochastic processes, and combines them to discuss optimum spatio-temporal sampling design.

The past has seen, perhaps initiated by Gribik *et al*. (1976), a great number of statistical papers devoted to the purely spatial aspect of sampling design mainly in the context of monitoring networks. Other early papers include those by Caselton and Zidek (1984), Olea (1984) and Fedorov and Müller (1988); book-length treatments are given by Müller (1998, 2007) and de Gruijter *et al*.

Spatio-temporal design: Advances in efficient data acquisition, First Edition.
Edited by Jorge Mateu and Werner G. Müller.

(2006). An excellent recent overview over this literature is provided by Zidek and Zimmerman (2010) and we will take the liberty in this introduction to draw heavily from their structure and exposition, albeit complementing it with aspects that enter due to the additional temporal component. Another excellent review paper is that by Dobbie *et al.* (2008), who provide some material on spatio-temporal aspects. As those two texts are comprehensive we can restrict ourselves to a brief exposition with the goal and emphasis to lead the readers into the more substantial subsequent chapters.

1.2 Paradigms in spatio-temporal design

An important clarification that needs to be made before thinking about spati(o-tempor)al design is whether we assume the randomness of the observations to stem from stochastic disturbances or from the sampling process itself. This leads to the distinction of so-called model-based and design-based (rather than this common but confusing expression we prefer to call them probability-based) inferences and their respective design procedures.

The probability-based paradigm is rooted in classical sampling theory and assumes the ability of defining a population explicitly and a respective randomness in the design. These methods aim at restoring unobserved observations and more importantly general attributes of the spatial population, such as total means (cf. de Gruijter and ter Braak 1990) and variances (cf. Fewster 2011). Probability-based inferences of these attributes are bias-free and allow uncertainty assessments under mild assumptions. The corresponding design techniques reach from the benchmark random sampling to stratified, two-stage, cluster or sequential random sampling with a multitude of variants (Stehman 1999) that all lend themselves to straightforward extensions into incorporating a temporal dimension (Brus and de Gruijter 2011). An excellent account of the latter can be found in Part IV of de Gruijter *et al.* (2006), which can in general be considered the most definitive text for the probability-based design paradigm.

The model-based paradigm on the other hand requires a statistical model to describe the data-generating spatio-temporal process. Here we typically assume that observations stem from a random field generally given by

$$Z(x, \mathbf{s}, t) = \eta(x(\mathbf{s}, t), \mathbf{s}, t, \beta) + \epsilon(x, \mathbf{s}, t), \quad \mathbf{s} \in D, t \in T, \tag{1.1}$$

where $\mathbf{s}$ denotes a spatial location, t a time point, x some potentially space and time dependent regressors, and η a parametrized trend model (a nearly encyclopedic reference for these type of processes is Cressie and Wikle 2011). Note that the random element here is the error ϵ rather than the design mechanism. This allows to assign meaning to purely geometric designs, such as regular grids or space-filling lattices, that are common in applications. Another advantage here is that by borrowing inference strength from the model we can make meaningful inferences from rather small samples and for very specific aspects derived from the model parameters, such as threshold exceedances, times of trend-reversals, local outliers, etc.

We believe that unless a reasonable modeling is out of reach the model-based approach offers more flexibility and statistical power, which is why most of the contributions in this volume will fall into this category. However, this issue has been the subject of considerable debate in the literature and further details are provided in Papritz and Webster (1995), Brus and de Gruijter (1997), Stevens (2006), and an overview is given in Table 1.1 of Dobbie *et al.* (2008). A general discussion that goes beyond the spatio-temporal realm can be found in Thompson (2002). Recently, there also have been attempts to fuse the two paradigms, Brus and de Gruijter (2012) for instance employ probability-based sampling for the spatial coordinates, whereas they build a time series model and use a respective design for the temporal trend. How to include probability-based design in a hierarchical statistical modeling framework is surveyed in Cressie *et al.* (2009).

1.3 Paradigms in spatio-temporal modeling

Another dichotomy clearly shows when one examines the spati(o-tempor)al modeling literature. Historically, two schools have somewhat independently developed, one based on discrete time series model analogies and the other one derived from generalizations of stochastic process methodologies. The former is much used by geographers and economists particularly in the advent of what was termed ‘new economic geography’ and was consequently referred to as ‘spatial econometrics’ (for perhaps the earliest full exposition see Anselin 1988; a recent account of the history of the field thereafter by the same author can be found in Anselin 2010). The latter school stems from the theory of regionalized variables developed among mining scientists and geologists and has consequently been named ‘geostatistics’ (a book-length treatment is provided by Chilès and Delfiner 1999). Comparative discussions on these two paradigms can be found in the encyclopedic Cressie (1993) and more recently in Griffith and Paelinck (2007), Hae-Ryoung *et al.* (2008) and Haining *et al.* (2010).

Both of these modeling views are encompassed by the random field (1.1) and can be solely distinguished by the nature of the indexing variables s and t. In spatial econometrics spatial econometrics we usually assume the s’s to form a discrete geographic lattice and their relationships are usually described in the form of a so-called spatial weight or link matrix W. Various types of dependence structures can be modeled by assigning particular forms of η and covariances of ϵ employing W, such as the common simultaneous and conditionally autoregressive regression models (SAR and CAR), the latter being spatial manifestations of Gaussian Markov random fields (GMRF; see Rue and Held 2005 for a definitive text).

In geostatistics the locations $\mathbf{s}$ are assumed to vary continuously in D and again the implied models differ by the choice of η and the error dependence usually determined by the so-called variogram $\gamma(\mathbf{s}, \mathbf{s}') = E(|Z(x, \mathbf{s}, t) - Z(x, \mathbf{s}', t)|^2)$. Under normality assumptions for the errors these processes also come under the notion of Gaussian random fields (GRF) and they are also much in use in other contexts such as machine learning and computer simulation experiments (cf. Rasmussen and Williams 2005). Though more flexible the corresponding models are usually much more estimation intensive than those for GMRF, which can typically be employed for much larger spatial datasets.

Despite this divide there have lately been successful attempts to merge those two spatio-temporal modeling paradigms. While previously only results for regular sampling schemes were available (cf. Griffith and Csillag 1993), Lindgren *et al*. (2011) provide an explicit link for arbitrary lattices, thus opening the issue for the question of sampling design. In a discussion to this article Müller and Waldl (2011) indeed uncover relationships between the respective designs that will allow to exploit properties from both paradigms.

A great number of spatio-temporal extensions of these models exist particularly for GRF; see Cressie and Wikle (2011) for an extensive review and Baxevani *et al*. (2011) for a particular representation using velocity fields. GMRF are usually extended by modeling them in discrete time, so-called spatial panel models (see e.g., Elhorst 2012 for a recent survey); a continuous time extension of spatial panels is given in Oud *et al*. (2012).

1.4 Geostatistics and spatio-temporal random functions

Geostatistical research has typically analyzed random fields, in which every spatio-temporal location can be seen as a point on $\mathbb{R}^d \times \mathbb{R}$. While from a mathematical point of view $\mathbb{R}^d \times \mathbb{R} = \mathbb{R}^{d+1}$, from a physical perspective it would make no sense to consider spatial and temporal aspects in the same way, due to the significant differences between the two axes of coordinates. Therefore, while the time axis is ordered intrinsically (as it exists in the past, present and future), the same does not occur with the spatial coordinates.

Recalling (1.1), assume that observations stem from a random field (r.f.) given by $Z(x, \mathbf{s}, t) = \eta(x(\mathbf{s}, t), \mathbf{s}, t, \beta) + \epsilon(x, \mathbf{s}, t), \mathbf{s} \in D, t \in T$, where $\mathbf{s}$ denotes a spatial location, t a time point, x some potentially space and time dependent regressors, η a parametrized trend model, $D \subset \mathbb{R}^d$ (very often $d = 2$), and $T \subset \mathbb{R}$. For ease of notation, we remove the term in the covariates x, and write $Z(\mathbf{s}, t)$, assuming whenever necessary that any trend coming from a set of covariates has already been removed.

1.4.1 Relevant spatio-temporal concepts

A spatio-temporal r.f. $Z(\mathbf{s}, t)$ is said to be *Gaussian* if the random vector $\mathbf{Z} = (Z(\mathbf{s}_1, t_1), \ldots, Z(\mathbf{s}_n, t_n))'$ for any set of spatio-temporal locations

$\{(\mathbf{s}_1, t_1), ..., (\mathbf{s}_n, t_n)\}$ follows a multivariate normal distribution. When not stated explicitly, the indexes i and j will go from 1 to n.

The spatio-temporal r.f. $Z(\mathbf{s}, t)$ is said to have a *spatially stationary covariance* function if, for any two pairs $(\mathbf{s}_i, t_i)$ and $(\mathbf{s}_j, t_j)$ on $\mathbb{R}^d \times \mathbb{R}$, the covariance $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ only depends on the distance between the locations $\mathbf{s}_i$ and $\mathbf{s}_j$ and the times t_i and t_j. And the spatio-temporal r.f. $Z(\mathbf{s}, t)$ is said to have a *temporarily stationary covariance* function if, for any two pairs $(\mathbf{s}_i, t_i)$ and $(\mathbf{s}_j, t_j)$ on $\mathbb{R}^d \times \mathbb{R}$, the covariance $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ only depends on the distance between the times t_i and t_j and the spatial locations $\mathbf{s}_i$ and $\mathbf{s}_j$. If the spatio-temporal r.f. $Z(\mathbf{s}, t)$ has a *stationary covariance function* in both spatial and temporal terms, then it is said to have a stationary covariance function. In this case, the covariance function can be expressed as

$$C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = C(\mathbf{h}, u) \tag{1.2}$$

with $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ and $u = t_i - t_j$ the distances in space and time, respectively.

A spatio-temporal r.f. $Z(\mathbf{s},t)$ has a *separable covariance function* if there is a purely spatial covariance function $C_s(\mathbf{s}_i, \mathbf{s}_j)$ and a purely temporal covariance function $C_t(t_i, t_j)$ such that

$$C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = C_s(\mathbf{s}_i, \mathbf{s}_j)C_t(t_i, t_j) \tag{1.3}$$

for any pair of spatio-temporal locations $(\mathbf{s}_i, t_i)$ and $(\mathbf{s}_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$.

A spatio-temporal r.f. $Z(\mathbf{s},t)$ has a *fully symmetrical covariance function* if

$$C((\mathbf{s}_i, t_i), (\mathbf{s}_j,t_j)) = C_s(\mathbf{s}_i, t_j)C_t(\mathbf{s}_j, t_i) \tag{1.4}$$

for any pair of spatio-temporal locations $(\mathbf{s}_i, t_i)$ and $(\mathbf{s}_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$.

Separability is a particular case of complete symmetry and, as such, any test to verify complete symmetry can be used to reject separability. In the case of stationary spatio-temporal covariance functions, the condition of full symmetry reduces to

$$C(\mathbf{h}, u) = C(\mathbf{h}, -u) = C(-\mathbf{h}, u) = C(-\mathbf{h}, -u),\ \forall(\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R}. \tag{1.5}$$

A spatio-temporal r.f. has a *compactly supported covariance function* if, for any pair of spatio-temporal locations $(\mathbf{s}_i, t_i)$ and $(\mathbf{s}_j, t_j) \in \mathbb{R}^d \times \mathbb{R}$, the covariance function $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ tends towards zero when the spatial or temporal distance is sufficiently large.

If $C(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j)$ depends only on the distance between positions, that is, $\left(\|\mathbf{s}_i - \mathbf{s}_j\|, t_i - t_j\right)$, the r.f., apart from being stationary, is also *isotropic* in space and time. Note that if the covariance function of a stationary r.f. is isotropic in space and time, then it is fully symmetrical.

The spatio-temporal *variogram* is defined as the function

$$2\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = V(Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j)), \tag{1.6}$$

where V is the variance, and half this quantity is called a semivariogram.

In the case of a r.f. with a zero mean,

$$2\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = E[(Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j))^2]. \tag{1.7}$$

Whenever it is possible to define the covariance function and the variogram, they will be related by means of the following expression

$$2\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = V(Z(\mathbf{s}_i, t_i)) + V(Z(\mathbf{s}_j, t_j)) - 2C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)). \tag{1.8}$$

If the spatio-temporal r.f. $Z(\mathbf{s}, t)$ has an intrinsically stationary variogram in both space and time, then it is said to have an intrinsically stationary variogram. In this case, the variogram can be expressed as

$$2\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = 2\gamma(\mathbf{h}, u). \tag{1.9}$$

The marginals $2\gamma(\cdot, u)$ and $2\gamma(\mathbf{h}, \cdot)$ are called purely spatial and purely temporal variograms, respectively.

A r.f. $Z(\mathbf{s}, t)$ is *strictly stationary* if its probability distribution is translation invariant. Second-order stationarity is a less demanding condition than strict stationarity. A spatio-temporal r.f. $Z(\mathbf{s}, t)$ is *second-order stationary* in the broad sense or weakly stationary if it has a constant mean and the covariance function depends on $\mathbf{h}$ and u.

A spatio-temporal r.f. $Z(\mathbf{s}, t)$ is said to be intrinsically stationary if it has a constant mean and an intrinsically stationary variogram. Intrinsic stationarity is less restrictive than second-order stationarity. Another widely used function when modeling implicit spatio-temporal dependence in a stationary r.f. is the correlation function. Let $Z(\mathbf{s}, t)$ be a second-order stationary r.f. with a priori variance $\sigma^2 = C(\mathbf{0}, 0) > 0$. The autocorrelation function of this r.f. is defined as

$$\rho(\mathbf{h}, u) = \frac{C(\mathbf{h}, u)}{C(\mathbf{0}, 0)}. \tag{1.10}$$

If $\rho(\mathbf{h}, u)$ is a correlation function on $\mathbb{R}^d \times \mathbb{R}$, then its marginal functions $\rho(\mathbf{0}, u)$ and $\rho(\mathbf{h}, 0)$ will respectively be the spatial correlation function on $\mathbb{R}^d$ and the temporal correlation function on $\mathbb{R}$.

1.4.2 Properties of the spatio-temporal covariance and variogram functions

A function $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ of real values, defined on $\mathbb{R}^d \times \mathbb{R}$ is a *covariance function* if it is symmetrical, $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = C((\mathbf{s}_j, t_j), (\mathbf{s}_i, t_i))$ and positive-definite, that is,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) \geq 0 \tag{1.11}$$

for any $n \in \mathbb{N}$, $(\mathbf{s}_i, t_i) \in \mathbb{R}^d \times \mathbb{R}$, and $a_i \in \mathbb{R}$, $i = 1, ..., n$. The condition (1.11) is *sufficient* if the covariance function can take complex values. Similarly, one *necessary and sufficient* condition for a non-negative function of real values $\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ defined on $\mathbb{R}^d \times \mathbb{R}$ to be a *semivariogram* is that it is a symmetrical function and conditionally negative-definite, that is,

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) \leq 0 \tag{1.12}$$

with $\sum_{i=1}^{n} a_i = 0$.

Schoenberg (1938) proved the following theorem characterizing the spatio-temporal semivariogram. Let $\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ be a function defined on $\mathbb{R}^d \times \mathbb{R}$, with $\gamma((\mathbf{s}, t), (\mathbf{s}, t)) = 0$, $\forall (\mathbf{s}, t) \in \mathbb{R}^d \times \mathbb{R}$. Then the following statements are equivalent:

- $\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ is a semivariogram on $\mathbb{R}^d \times \mathbb{R}$.
- $\exp\big(-\theta\gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))\big)$ is a covariance function on $\mathbb{R}^d \times \mathbb{R}$, for any $\theta > 0$.
- $C((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j)) = \gamma((\mathbf{s}_i, t_i), (\mathbf{0}, 0)) + \gamma((\mathbf{s}_j, t_j), (\mathbf{0}, 0)) - \gamma((\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j))$ is a covariance function on $\mathbb{R}^d \times \mathbb{R}$.

In case of stationarity, the above results reduce to functions depending on spatial and temporal lags. Another seminal result that characterizes covariance functions is that given in Bochner (1933). A function $C(\mathbf{h}, u)$ defined on $\mathbb{R}^d \times \mathbb{R}$ is a stationary covariance function if, and only if, it has the following form

$$C(\mathbf{h}, u) = \int\int e^{i(\boldsymbol{\omega}'\mathbf{h}+\tau u)} dF(\boldsymbol{\omega}, \tau), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R} \tag{1.13}$$

where the function F is a non-negative distribution function with a finite mean defined on $\mathbb{R}^d \times \mathbb{R}$, which is known as a *spectral distribution function*. Therefore, the class of stationary spatio-temporal covariance functions on $\mathbb{R}^d \times \mathbb{R}$ is identical to the class of Fourier transforms of non-negative distribution functions with finite means on that domain. If the function C can also be integrated, then the spectral distribution function F is absolutely continuous and the representation (1.13) simplifies to

$$C(\mathbf{h}, u) = \int\int e^{i(\boldsymbol{\omega}'\mathbf{h}+\tau u)} f(\boldsymbol{\omega}, \tau) d\boldsymbol{\omega} d\tau, \ \ (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R} \tag{1.14}$$

where f is a non-negative, continuous and integrable function that is known as a *spectral density function*. The covariance function C and the spectral density function f then form a pair of Fourier transforms , and

$$f(\boldsymbol{\omega}, \tau) = (2\pi)^{-d-1} \int\int e^{-i(\boldsymbol{\omega}'\mathbf{h}+\tau u)} C(\mathbf{h}, u) d\mathbf{h} du \tag{1.15}$$

The decomposition (1.13) can be specialized for fully symmetrical covariance functions. Let $C(\cdot, \cdot)$ be a continuous function defined on $\mathbb{R}^d \times \mathbb{R}$, then $C(\cdot, \cdot)$ is a fully symmetrical stationary covariance function if, and only if, the following decomposition is possible

$$C(\mathbf{h}, u) = \int \int \cos(\boldsymbol{\omega}'\mathbf{h}) \cos(\tau u) dF(\boldsymbol{\omega}, \tau), \quad (\mathbf{h}, u) \in \mathbb{R}^d \times \mathbb{R} \tag{1.16}$$

where F is the non-negative and symmetrical spectral distribution function defined on $\mathbb{R}^d \times \mathbb{R}$.

Cressie and Huang (1999) provide a theorem for characterizing the class of stationary spatio-temporal covariance functions under the additional hypothesis of integrability. Let $C(\cdot, \cdot)$ be a continuous, bounded, symmetrical and integrable function defined on $\mathbb{R}^d \times \mathbb{R}$, then $C(\cdot, \cdot)$ is a stationary covariance function if, and only if, in view of $u \in \mathbb{R}$,

$$C_{\boldsymbol{\omega}}(u) = \int e^{-i\boldsymbol{\omega}'\mathbf{h}} C(\mathbf{h}, u) d\mathbf{h}, \tag{1.17}$$

is a covariance function for every $\boldsymbol{\omega} \in \mathbb{R}^d$ except, at the most, in a set with a null Lebesgue mean. Gneiting (2002) generalizes this result for C defined on $\mathbb{R}^d \times \mathbb{R}^l$, from which the previous statement is a particular case for $l = 1$.

Both the covariance function and the spectral density function are important tools for characterizing random stationary spatio-temporal fields. Mathematically speaking, both functions are closely related as a pair of Fourier transforms. Furthermore, the spectral density function is particularly useful in situations where there is no explicit expression of the covariance function. Stein (2005) shows the benefit of using smooth covariance functions far from the origin, which can be tested by verifying whether their spectral densities have derivatives of certain orders.

1.4.3 Spatio-temporal kriging

Kriging is aimed at predicting an unknown point value $Z(\mathbf{s}_0, t_0)$ at a point $(\mathbf{s}_0, t_0)$ that does not belong to the sample. To do so, all the information available about the regionalized variable is used, either at the points in the entire domain or in a subset of the domain called the *neighborhood*.

Assume that the value of the r.f. has been observed on a set of n spatio-temporal locations $\{Z(\mathbf{s}_1, t_1), ..., Z(\mathbf{s}_n, t_n)\}$. We now want to predict the value of the r.f. on a new spatio-temporal location $(\mathbf{s}_0, t_0)$, for which we use the linear predictor

$$Z^*(\mathbf{s}_0, t_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_\mathrm{i}, \mathrm{t}_\mathrm{i}) \tag{1.18}$$

constructed from the random variables $Z(\mathbf{s}_i, t_i)$. As in the spatial case, spatio-temporal kriging equations will depend on the degree of stationarity attributed to the r.f. that supposedly generates the observed realization. The most widely used kriging techniques in the spatio-temporal case are simple spatio-temporal kriging, ordinary spatio-temporal kriging, and universal spatio-temporal kriging. In the case of *simple spatio-temporal kriging* we assume that $Z(\mathbf{s}, t)$ is a second-order stationary spatio-temporal r.f., with a constant and known mean $\mu(\mathbf{s}, t)$, constant and known variance $C(\mathbf{0}, 0)$, and a known covariance function $C(\mathbf{h}, u)$. The kriging equations (n equations with n unknown elements) are of the form

$$\sum_{j=1}^{n} \lambda_j C\left(\mathbf{s}_i - \mathbf{s}_j, \mathrm{t}_i - \mathrm{t}_j\right) = C\left(\mathbf{s}_i - \mathbf{s}_0, \mathrm{t}_i - \mathrm{t}_0\right), \forall i = 1, ..., n \tag{1.19}$$

from which we obtain the values λ_i that minimize the prediction error variance, which is given by

$$V\left[Z^*(\mathbf{s}_0, t_0) - Z(\mathbf{s}_0, t_0)\right] = C(\mathbf{0}, 0) - \sum_{i=1}^{n} \lambda_i C(\mathbf{s}_i - \mathbf{s}_0, t_i - t_0) \tag{1.20}$$

In the case of *ordinary spatio-temporal kriging*, the constant mean $\mu(\mathbf{s}, t)$ is not known, and the covariance function $C(\mathbf{h}, \mathrm{u})$ is known, under second-order stationarity. In the case of an intrinsic r.f. the variance is unbounded. In these two cases, simple kriging cannot be performed as the mean cannot be subtracted. We must therefore impose a condition of unbiasedness. In these situations, ordinary spatio-temporal kriging equations can be expressed, in the first case, in terms of the covariance function, and in the second case, in terms of the semivariogram, as there is no covariance at the origin.

In the *universal kriging* approach, assume $Z(\mathbf{s}, t)$ is a r.f. with drift, and so the mean of the r.f. is not constant, but depends on the pairs $(\mathbf{s}, t)$. In this situation the so-called condition of unbiasedness is substantially affected. In this case, the r.f. can be disaggregated into two components: one deterministic $\mu(\mathbf{s}, \mathrm{t})$ and the other stochastic $e(\mathbf{s}, t)$ which can be treated as an intrinsically stationary r.f. with zero expectation, $E\left[e(\mathbf{s}, \mathrm{t})\right] = 0$

$$Z(\mathbf{s}, t) = \mu(\mathbf{s}, \mathrm{t}) + \mathrm{e}(\mathbf{s}, \mathrm{t}) \tag{1.21}$$

We can assume that the mean, even unknown, can be expressed locally by

$$\mu(\mathbf{s}, t) = \sum_{h=1}^{p} a_h f_h(\mathbf{s}, \mathrm{t}) \tag{1.22}$$

where $\left\{f_h(\mathbf{s}, \mathrm{t}),\ h = 1, ..., p\right\}$ are p known functions, a_h constant coefficients, and p the number of terms used in the approximation. It must be taken into account that this expression is only valid locally. In this case, the equations that yield the prediction of the weights are obtained from the prediction error conditions of zero expectation and minimum variance.

1.4.4 Spatio-temporal covariance models

One key stage in the spatio-temporal prediction procedure is choosing the covariance function (covariogram or semivariogram) that models the structure of the spatio-temporal dependence of the data. However, while the semivariogram is normally chosen for this purpose in the spatial case, in a spatio-temporal framework the covariance function is the most commonly chosen tool. By referring to a valid covariographic spatio-temporal model, we are implicitly stating that the covariance function must be positive-definite. The purely spatial and temporal covariance models have been widely studied and there is a long list of those which can be used to model spatial or spatio-temporal dependence that guarantee the (spatial or temporal) covariance function is positive-definite. However, this is not the case in the spatio-temporal scenario, in which constructing valid spatio-temporal covariance models is one of the main research activities. In addition, while it is difficult to demonstrate that a spatial or temporal function is positive-definite, it is even more so when seeking to determine valid spatio-temporal covariance models. For this reason, many authors began to study how to combine valid spatial and temporal models to obtain (valid) spatio-temporal covariance models.

By way of introduction, the first approximations to modeling spatio-temporal dependence using covariance functions were nothing more than generalizations of the stationary models used in the spatial scenario. In this sense, early studies often modeled the spatio-temporal covariance using metric models by defining a metric in space and time that allowed researchers to directly use isotropic models that are valid in the spatial case. Such metric models were characterized by being nonseparable, isotropic and stationary. The next step in this initial stage consisted of configuring spatio-temporal covariance functions by means of the sum or product of a spatial covariance and a temporal covariance, both of which were stationary, giving rise to separable, isotropic and stationary models. Later, realizing the limitations of the two procedures detailed above in terms of capturing the spatio-temporal dependence that really exists in the large majority of the phenomena studied, interest shifted towards including the interaction of space and time, in covariance models, giving rise to the so-called nonseparable models (while remaining isotropic and stationary). It is worth highlighting the nonseparable models developed by Jones and Zhang (1997), Cressie and Huang (1999), Brown *et al*. (2000), De Cesare *et al*. (2001a, b), De Iaco *et al*. (2001, 2002a, b, 2003), Gneiting (2002), Ma (2002, 2003a, c, 2005a, b, c), Fernández-Casal *et al*. (2003), Kolovos *et al*. (2004), and Stein (2005) among others.

Development continued with the search for nonseparable spatio-temporal, spatially anisotropic and/or temporally asymmetrical models such as those described in Fernández Casal *et al*. (2003), Porcu *et al*. (2006), and Mateu *et al*. (2007). Finally, we can cite some recent approaches to the problem of modeling nonstationary covariance functions, such as those made by Ma (2002, 2003b), Fuentes *et al*. (2005), Stein (2005), Chen *et al*. (2006), Porcu *et al*. (2006, 2007a, b, 2009), Mateu *et al*. (2007), Porcu and Mateu (2007) and Gregori *et al*. (2008).

1.4.5 Parametric estimation of spatio-temporal covariograms

The empirical determination of the covariance function or the variogram of a spatio-temporal process can be generalized naturally using the procedures for merely spatial processes. Let $Z(\cdot, \cdot)$ be an intrinsically stationary process observed on a set of n spatio-temporal pairs $\{(\mathbf{s}_1, t_1), ..., (\mathbf{s}_n, t_n)\}$. Two direct and popular alternatives to obtain an estimation of the variogram $2\gamma(\cdot, \cdot)$ [and its covariance function $C(\cdot, \cdot)$, if the process is also second-order stationary] are the classical estimator based on the method-of-moments (MoM), and the robust estimator proposed by Cressie and Hawkins (1980).

This MoM estimator for the variogram is given in its most general form by

$$2\hat{\gamma}(\mathbf{h}(l), u(k)) = \frac{1}{|N(\mathbf{h}(l), u(k))|} \sum_{(\mathbf{s}_i, t_i)(\mathbf{s}_j, t_j) \in N(\mathbf{h}(l), u(k))} \left(Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j)\right)^2, \tag{1.23}$$

where $N(\mathbf{h}(l), u(k)) = \{(\mathbf{s}_i, t_i), (\mathbf{s}_j, t_j) : \mathbf{s}_i - \mathbf{s}_j \in N(\mathbf{h}(l)), t_i - t_j \in T(u(k))\}$, $T(\mathbf{h}(l))$ being an area of tolerance in $\mathbb{R}^d$ around $\mathbf{h}(l)$, $T(u(k))$ is a region of tolerance in $\mathbb{R}$ around $u(k)$, and $|N(\mathbf{h}(l), u(k))|$ is the number of different elements in $N(\mathbf{h}(l), u(k))$, with $l = 1, ..., L$ and $k = 1, ..., K$.

Although the classical estimation method has the advantage of being easy to calculate, it also has some practical drawbacks, such as not being robust in the case of extreme values. In order to avoid this problem, and following and extending Cressie and Hawkins (1980) we have the following variogram estimator for the spatio-temporal case

$$2\hat{\gamma}(\mathbf{h}(l), u(k)) = \left(\frac{1}{|N(\mathbf{h}(l), u(k))|} \sum_{(\mathbf{s}_i, t_i)(\mathbf{s}_j, t_j) \in N(\mathbf{h}(l), u(k))} |\, Z(\mathbf{s}_i, t_i) - Z(\mathbf{s}_j, t_j)\, |^{\frac{1}{2}}\right)^4 \times \left(0.457 + \frac{0.494}{|N(\mathbf{h}(l), u(k))|}\right)^{-1}$$

As in the spatial case, these estimators of the covariance function or variogram of the process do not generally fulfill the condition of being positive-definite or conditionally negative definite, respectively. For this reason, in practice we select a parametric model of covariance or variogram that we already know is valid, and estimate the parameters of the covariance function or variogram that best fits the values of the empirical estimator.

Any of the procedures for *least squares* (LS) estimation used in the spatial case can be easily generalized to the spatio-temporal case. Let us assume that the process being analyzed has been decomposed such that either the original process or the process of residuals after modeling its mean, is intrinsically stationary. In this case, we can estimate the parameters that define the semivariogram model chosen using ordinary (OLS), generalized (GLS) or weighted (WLS) least squares. But the OLS procedure does not take into account the behavior of the semivariogram (or covariance function) near the origin and, above all, does not consider the possibility of the values of the semivariogram being correlated, which could have an adverse effect on the estimations of the vector of parameters as the amount of data increases. This last problem is normally solved by resorting to GLS, but it is also true that in order to determine the covariance matrix, which is normally quite large, we should work in a Gaussian context and proceed using iterative methods (cf. Müller 1999). A compromise between efficiency and computation is provided by the WLS method, which is the most popular LS estimation method. The WLS method consists of (iteratively) minimizing the expression

$$\sum_{i=1}^{n} \omega_i (\hat{\gamma}(\mathbf{h}_i, u_i) - \gamma(\mathbf{h}_i, u_i; \boldsymbol{\theta}))^2 = (\hat{\gamma} - \gamma(\boldsymbol{\theta}))' W(\boldsymbol{\theta})^{-1} (\hat{\gamma} - \gamma(\boldsymbol{\theta})) \tag{1.24}$$

where $W(\boldsymbol{\theta})$ is a diagonal matrix, the elements of which, ω_i, are approximately the variances of $\hat{\gamma}$.

It is a well-known fact that the *maximum likelihood* (ML) method requires knowing the distribution of the r.f. underlying the phenomenon under study. However, only the case of Gaussian random functions has been developed in the literature (Mardia and Marshall 1984). With this in mind, if the spatio-temporal process is Gaussian, then it is possible to use the ML method, which is also more efficient than the least squares procedures when estimating the parameters of the valid covariance function (or semivariogram) chosen. ML estimators have desirable properties for spatial analysis, such as consistency and asymptotic normality in the field of extendible domains (Mardia and Marshall 1984). However, they also have one significant limitation: even in the merely spatial case they entail high computation costs as the number of parameters the model includes is high. The most critical part of the estimation process is calculating the determinant and the inverse of the covariance matrix. An alternative to ML is given by the restricted maximum likelihood (RML), in which the estimators are obtained by applying the ML method to the so-called ‘error contrasts’.

As the problem with the ML procedure is computational, it is natural that research has focused on searching for approaches to the likelihood function that require less than $O\left(N^3\right)$ stages and which have desirable statistical properties. However, another line of research has resorted to exploiting the special structure of the covariance matrix in order to enhance computation. In this sense, Zimmerman (1989) indicates that in certain sampling procedures it is possible to compute ML more efficiently. Another way of simplifying the number of operations in the context of lattice data is to represent the process spectrally (cf. Whittle 1954). Kaufman *et al.* (2008) focus directly on the covariance matrix and propose calculating likelihood with a reduced version of the covariance matrix instead of the original matrix. Furrer *et al.* (2006) use this technique to interpolate observations rather than estimate the parameters.

The computational problem was once again addressed recently using likelihood approaches, such as composite likelihood. According to Varin and Vidoni (2005), there are two types of composite likelihood. The first includes the so-called 'subsetting methods' and is based on marginal functions or 'subset pieces' of full likelihood. The second is based on the so-called 'mission methods', obtaining composite likelihood by omitting components from the full likelihood. The latter intuitively yield computational benefits insofar as they do not consider certain 'pieces of likelihood' when computing composite likelihood. Three types of composite likelihood have been proposed in the literature to approach full likelihood when working with a large set of spatial data (see Vecchia 1988, Curriero and Lele 1999, Caragea and Smith 2006, and the modified version of Stein *et al.* 2004).

Vecchia (1988) proposes factorizing the full likelihood into a product of conditional densities to later reduce the size of the conditioning sets. Stein *et al.* (2004) suggest that it might be more efficient to carry out the above procedure in blocks. Stein (2005) extended this approach to the spatio-temporal field when the data are measured in monitoring stations at regular intervals of time.

The Weighted Composite Likelihood (WCL) technique (cf. Lindsay 1988) defines a general estimation method when working with large sets of observations and has been applied in a large number of fields of science in recent years (particularly in agronomy). It is easy to appreciate that the approximations in Vecchia (1988), Stein *et al.* (2004), Stein (2005), and Caragea and Smith (2006) are special cases of composite likelihood. This approach is the starting point of the procedure conceived by Bevilacqua *et al.* (2012). Using WCL to estimate the parameters of a valid covariographic or semivariographic model aims to obtain accurate estimates comparable with those obtained by ML, but with lower computation costs (cf. Bevilacqua *et al.* 2012). This issue is of vital importance in the spatio-temporal field, as there is usually a sufficient number of observations that make the ML prohibitive in terms of computation.

1.5 Types of design criteria and numerical optimization

When it comes to actually determining the spatio-temporal design we have a variety of choices and face a multitude of constraints. Perhaps it should be first stressed that many of the decisions we are going to make will have a considerable influence on the designs we yield but it is the decision to design in the first place, that will impact the quality of our inference the most. As Ver Hoef (2012) puts it: 'The main point [...] is that any search for a good design is better than relying on a randomly chosen one, [...]'. We (and he) argue for designs that are robust with respect to choice that need to be made like prior guesses of parameters and spatial correlations and insensitive to the constraints imposed, ideally even those stemming from cost considerations.

Nevertheless to make concrete progress it may be worthwhile to define clear-cut objectives that can be optimized by the sampling site determination, so-called design criteria. If we define the design as the finite collection of sites $\mathbf{s}_i$ and observation times t_i, the set $\xi = \{\mathbf{s}_1, t_1, ..., \mathbf{s}_n, t_n\}$ we can usually state the design problem as the minimization task

$$\xi^* = \arg\min_{\xi} \phi(\xi), \tag{1.25}$$

with ϕ being a suitable scalar function. The result ξ^* is referred to as a (ϕ)-optimal design. For spatio-temporal designs the time dimension is sometimes treated separately (cf. de Gruijter *et al*. 2006) since there are often specific restrictions as to whether and how temporal changes in the spatial configuration of measurement sites are allowed (see Wikle and Royle 1999 for an early description of the problem). Sometimes one even has to take into account measurement devices that move in space according to given trajectories (Ucinski and Chen 2005). In other cases one might want to exploit information gained from early stages of the data collecting process as time progresses leading to so-called sequential or adaptive designs (Berliner *et al* 1999).

The choice of a particular design criterion ϕ clearly reflects the purpose of a study. Is it (cf. Overton and Stehman 1996):

- exploratory in nature to gain a first understanding of the phenomenon under study?
- confirmatory aimed at estimating global or local characteristics or model parameters?
- solely used for most accurate prediction purposes?

For the first in this list one usually constructs designs that are filling the space (and sometimes period) available in some efficient way. One could now look for a design that is trying to get good coverage of every point in the space D. Let us for this purpose define a metric for the (inverse) distance between a given point $\mathbf{s}$ and a design ξ, i.e

$$d_p(\mathbf{s}, \xi) = \left(\sum_{i=1}^{n} \frac{1}{\|\mathbf{s} - \mathbf{s}_i\|^p} \right)^{\frac{1}{p}}, \quad p > 0,$$

which can be considered as a measure of (lack of) coverage of a design ξ for the point $\mathbf{s}$. Consequently we can use the L_q average over the space D

$$\phi_{[p,q]}(\xi) = \left(\int_{\mathbf{s} \in D} d_p^q(\mathbf{s}, \xi) d\mathbf{s} \right)^{\frac{1}{q}}, \quad q > 0,$$

which forms a design criterion to be minimized over ξ (cf. Royle and Nychka 1998; a probability-based concept with a similar purpose was devised in Stevens and Olsen 2004).

As a particular choice, it is natural to attempt to make the maximum distance from all the points in D to their closest points in ξ as small as possible. This is achieved by letting $p \to \infty$ and $q \to \infty$ giving

$$\phi_{[\infty,\infty]}(\xi) = \max_{\mathbf{s} \in D} \min_{\mathbf{s}_i} \|\mathbf{s} - \mathbf{s}_i\| = \phi_{mM}(\xi) \tag{1.26}$$

and we call the resulting design

$$\xi^*_{mM} = \arg\min_{\xi} \phi_{mM}(\xi)$$

a *minimax distance design*. If we instead seek for a design that wants to achieve a high spread solely amongst its support points, we can similarly to above define a criterion that L_q averages the (inverse) interdistances for maximization, i.e.

$$\phi^{-1}_{[q]}(\xi) = \left(\sum_{\mathbf{s}_i \neq \mathbf{s}_j \in D} \frac{1}{\|\mathbf{s}_i - \mathbf{s}_j\|^q} \right)^{\frac{1}{q}}.$$

Then a design that is seeking for a high spread amongst its support points within the design region must attempt to make the smallest distance between neighboring points in ξ as large as possible. That is ensured by the maximin design criterion which corresponds to letting $q \to \infty$,

$$\phi^{-1}_{[\infty]}(\xi) = \min_{\mathbf{s}_i \neq \mathbf{s}_j \in D} \|\mathbf{s}_i - \mathbf{s}_j\| = \phi^{-1}_{Mm}(\xi) \tag{1.27}$$

and we call the resulting design

$$\xi^*_{Mm} = \arg\max_{\xi} \phi^{-1}_{Mm}(\xi)$$

a *maximin distance design*, first introduced by Johnson *et al.* (1990). For more details on space-fillingness refer to Pronzato and Müller (2012).

For the second point in the above list, that is improving the quality of the estimation of the model parameters or derived characteristics it is customary to base the design criterion on the corresponding Fisher information matrix $M(\xi, \beta)$ for the parameters β. A large theory of optimal experimental design is built around those criteria (Atkinson *et al.* 2007), however most of it covers the case of independent errors ϵ, which is usually violated in spatial studies. Some attempts have been made to extend the theory into the correlated error case (cf. Fedorov 1996 and Müller and Pázman 2003), but the respective criteria are still popular for spatial design even without proper adaptations and despite lacking theoretical justification they seem to fare well. The most common choice for a criterion here is

$$\phi_D(\xi) = -\log\ \det M(\xi, \beta)$$

and resulting designs are termed *D-optimal*, although occasionally other scalar functions of $M(\xi, \beta)$ such as the trace (Ver Hoef 2012) are preferred for simplicity (see Nowak 2010 for a survey). Sometimes rather than estimating trend parameters the focus could be on fitting parametrized variograms $\gamma(\mathbf{s}, \mathbf{s}', \theta)$ and estimation of their parameters, which consequently leads to designs based upon information matrices $M'(\xi, \theta)$ (Müller and Zimmerman 1999). A compound version taking into account both purposes

$$\phi_{D\alpha} = -\alpha \log \det M(\xi, \beta) - (1 - \alpha) \log \det M'(\xi, \theta) \tag{1.28}$$

with a weighting factor $0 \leq \alpha \leq 1$ was proposed by Müller and Stehlík (2010). A Bayesian approach for the same problem was proposed in Diggle and Lophaven (2006) and extended in Nowak *et al.* (2010).

When it comes to providing most accurate prediction for unobserved locations and times it is naturally to base a design criterion on the so-called kriging variance $\mathrm{Var}(\hat{Z}(\mathbf{s}, t))$, the mean squared prediction error of the best linear unbiased predictor of the random field (see McBratney and Webster 1981 for some early suggestions and Gao *et al.* 1996 for an efficient formulation). Most commonly the concrete design criterion employed for prediction is

$$\phi_G(\xi) = \max_{D,T} \mathrm{Var}(\hat{Z}(\mathbf{s}, t)|\xi),$$

which amounts to minimizing the maximum prediction variance over the design space, which in design theory is often called *G-optimal*. Some authors prefer to use the average variance (cf. Brus and Heuvelink 2007) but all variants basically yield space-filling designs. It has, however, to be noted that the kriging variance undervalues the uncertainty of the prediction, when – as is common – the variogram parameters θ are estimated together with β from the same data (see also Marchant and Lark 2007 for a discussion of this issue). Therefore Zhu and Stein (2006) and Zimmerman (2006) instead calculate their optimal prediction designs based on the modified kriging variance

$$\phi_{\tilde{G}}(\xi) = \max_{D,T} \left\{ \mathrm{Var}(\hat{Z}(\mathbf{s}, t)|\xi) + \mathrm{tr}\left[M'^{-1}(\xi, \theta) \mathrm{Var}(\partial \hat{Z}(\mathbf{s}, t)/\partial \theta) \right] \right\} \tag{1.29}$$

resulting in much less space-filling and more patchy optimal designs. Note that designs for criteria of type (1.29) are computationally much more difficult to obtain than those of (1.28) since they require evaluations at all points of the design space. However, for uncorrelated regression there exist a certain duality between D- and G-optimality via the celebrated general equivalence theorem by Kiefer and Wolfowitz (1960). This formed the basis of the promising investigations relating criteria $\phi_{D\alpha}$ and $\phi_{\tilde{G}}$ given in Müller *et al.* (2012).

Finally, particularly from a Bayesian viewpoint, it may be desirable to reduce the uncertainty associated to an (a posteriori) distribution. This naturally leads to the so-called entropy criterion, which is given a comprehensive exposition in Le and Zidek (2006) and offers a quite general unified approach at least in the Gaussian case. Note that for most of the above methods the application in the spatio-temporal setting will require adjustments for specific purposes, such as allowing for adapting sampling (Marchant and Lark 2006), ensuring robustness from model assumptions (Wiens 2005) or knowledge of variograms (Spöeck and Pilz 2010).

Finding exact optimum designs is obviously a challenging computational problem and most of the mentioned criteria are nonconvex functions of the design. Thus the algorithms known from the classical design theory (cf. Wynn 1970 and Fedorov 1971) either require some straightforward heuristic adaptions (cf. Brimkulov *et al*. 1980) or taylor-cut numerical routines. Recently, stochastic search algorithms such as spatial simulated annealing (van Groeningen and Stein 1998) and genetic algorithms (Ruiz-Cárdenas *et al*. 2010) or hybrids (Guedes *et al*. 2011) have been considered; a comparison of some of these methods with classic procedures can be found in Baume *et al*. (2010). An approach based on Markov chain Monte Carlo (MCMC) calculations first proposed by Müller (1999) seems to be particularly suitable for finding dynamic spatio-temporal designs (cf. Wikle and Royle 2005).

Available software implementation of the above discussed algorithms and methods is scarce, and we can only find something in R. Some indications can be found in Bivand *et al*. (2008); concrete useful packages are 'fields' by Furrer *et al*. (2011) (model-based) and 'spsurvey' by Kincaid and Olsen (2012) (design-based).

1.6 The problem set: Upper Austria

It was a cornerstone of this project to make the different approaches to design assembled in this volume comparable on an identical problem set. A natural choice was the region of Oberösterreich (Upper Austria), which has the city of Linz, where the university of one of the authors is located, as its capital and was also featured in Müller (1998, 2007). Since we wanted to encourage the use of methods for both discretely and continuously indexed random fields two complementary datasets were provided to the authors.

1.6.1 Climatic data

This dataset contains climatic data measured at 37 stations irregularly placed over the region provided from http://www.zamg.ac.at/fix/klima/oe71-00/klima2000/klimadaten_oesterreich_1971_frame1.htm. Here, we have (incomplete) monthly data from 1994 to 2009 on average temperature and total rainfall. Due to some missing observations, however, not all of the stations could be effectively used. A map of the region with the respective locations of the measurement stations is displayed in Figure 1.1.

Both of these climatic indicators have been frequently analyzed by geostatistical methods in the past, a recent showcase analysis for rainfall is found in Grimes and Pardo-Igúzquiza (2010). A kriging variance based early study for the optimal design of a rain gauge network was undertaken by Papamichail and Metaxa (1996), whereas Zimmermann *et al*. (2010) provide a mixed probability and model-based approach for the related issue of throughfall

Figure 1.1 The sampling locations of the climatic dataset within the region of Upper Austria.

monitoring. For the current dataset (July 1994) the results of a straightforward kriging analysis employing an exponential semivariogram can be found in Figures 1.2 (temperature) and 1.3 (rainfall), respectively.

To get an impression of the temporal character of this dataset the rainfall measurements for Linz/Stadt from 1994 to 2009 are displayed in Figure 1.4. The seasonality is less pronounced here than in the temperature series, which require particular modeling.

1.6.2 Grassland usage

The second part of the dataset contains information on the grassland usage in the 444 municipal districts of Upper Austria over the years 1995, 1999, 2003, 2005, 2007 and 2008. Those data were processed from a set that was provided by the 'Abteilung Statistik' of the 'Amt der oberösterreichischen Landesregierung' (www.land-oberoesterreich.gv.at/statistik), a regionalized and extended version of what is provided by Statistics Austria (http://www.statistik.at/web_de/statistiken/land_und_forstwirtschaft/agrarstruktur_flaechen_ertraege/bodennutzung/index.html). The districts were represented by a lattice of local coordinates (easting and

Figure 1.2 Average temperature map for July 1994 in Upper Austria.

northing) of the respective centroids and the boundary information is contained in shape-file (http://doris.ooe.gv.at/index.asp?MenuID = 4; see Figure 1.5).

Specifically we have provided the authors with the following variables: longitude, latitude (in local coordinates), LBBGG (identification number of municipality), BEZNR (identification number for district), FLKM2 (area of the region in square kilometers), ALTITUDE (elevation of the district's capital), and R95,R99,R03,R05,R07,R08 [log (area of arable land +1) − log (area of grassland +1)], i.e. practically the log of the ratio of the two areas for the periods 1995 (Figure 1.6) until 2008, making it scalefree. The latter variables should give some indication of the characteristic of interest to environmental scientists and policy makers, as to whether grassland is increasingly substituted by arable land leading to numerous economic and ecological impacts (see Vellinga *et al*. 2004 and Carlier *et al*. 2005 for a statistical analysis). The indicator is thus well suited for a spatio-temporal analysis and a positive temporal trend as well as some short range spatial autocorrelations can be expected (Yang *et al*. 2006). Furthermore some scientists postulated some interactions between climate change and land use (Thornley and Cannell 1997) thus allowing speculation about integrating some of our climatic variables into a potential spatio-temporal grassland usage model.

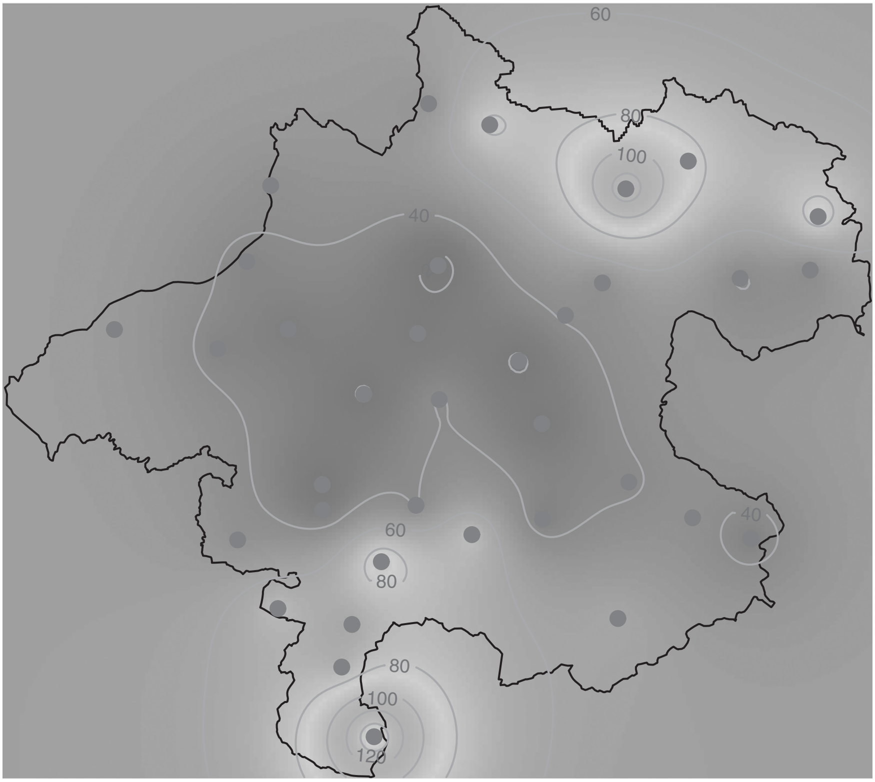

Figure 1.3 *Average rainfall map for July 1994 in Upper Austria.*

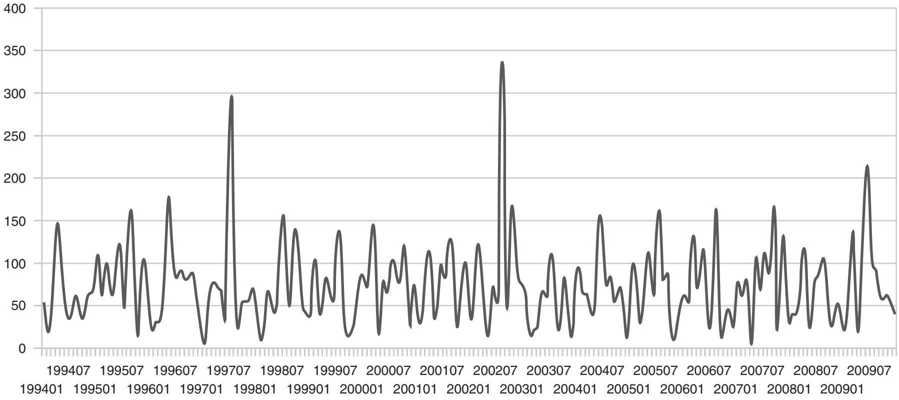

Figure 1.4 *Monthly rainfall in Linz/Stadt.*

Figure 1.5 The 444 municipal districts of Upper Austria.

The altitudes were not directly available, but were read off the respective Wikipedia pages (http://de.wikipedia.org/wiki/Liste_der_Gemeinden_in_Oberoe sterreich). Some authors have used a full digital elevation model such as provided by http://gdem.ersdac.jspacesystems.or.jp/; a corresponding plot can be found in Figure 1.7. As the local coordinate system (MGI Austria GK Central projection system based on Gauss-Krüger) was not sufficient for all purposes a transformation to the WGS84 geographical coordinates was performed (see http://spatial-analyst.net/book/uppera, where some more useful information on the case study can be found). A Google-Earth depiction of the area is presented in Figure 1.8.

Particularly in the environmental application areas we can naturally find most of the examples of spatio-temporal monitoring network design, see Vašát *et al.* (2009), Corwin *et al.* (2010), Brus and de Gruijter (2011), and Creelman and Risk (2011) for soil assessment, Reed and Minsker (2004) and Masoumi and Kerachian (2010) for groundwater and Murtojärvi *et al.* (2011) for sea water quality, Romary *et al.* (2011) and Wu and Bocquiet (2011) for air pollution, Abida *et al.* (2008) and Melles *et al.* (2011) for radioactive releases, Martínez *et al.* (2008) for solar radiation, Hooten *et al.* (2009) and Fewster (2011) for ecological surveys, Stehman (2009) for land cover evaluation and Barabesi *et al.* (2012) for forestry among the most recent. Note, however, that the problem of where and

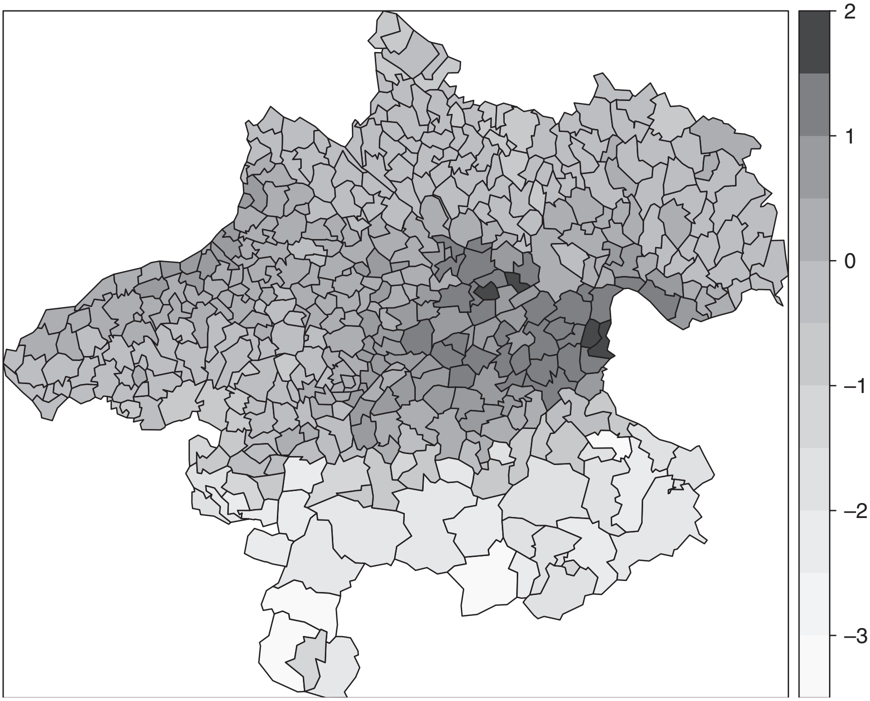

Figure 1.6 *Grassland usage: log (area of arable land +1) − log (area of grassland +1) in Upper Austria 1995.*

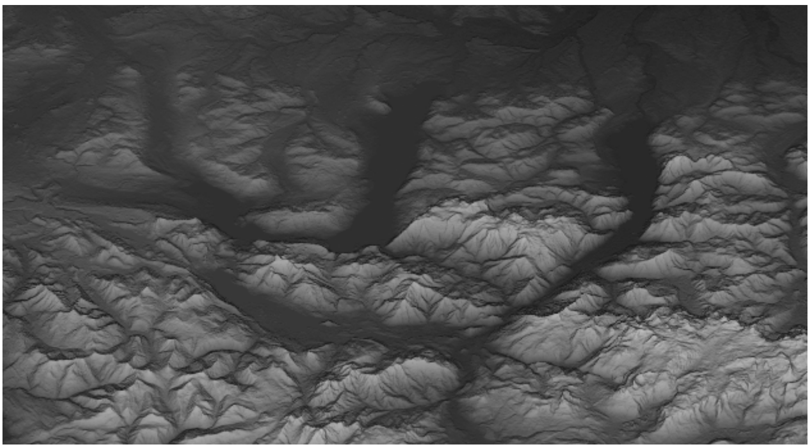

Figure 1.7 *A perspective contour plot of Upper Austria.*

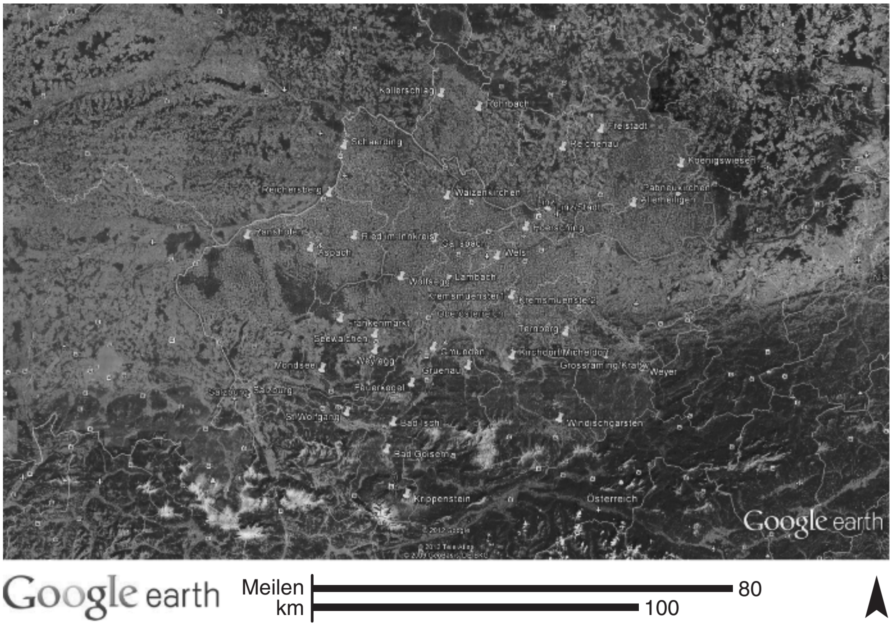

Figure 1.8 A Google-Earth view of observation regions and sites. (Please see plate section for color version of the figure.)

when to collect spatio-temporal data is also of relevance in many other fields, for example geophysics (Maurer *et al*. 2010), engineering (Ucinski and Patan 2010), industrial production (Borgoni *et al*. 2010), agriculture (Heuvelink and Egmond 2010), social sciences (Kumar 2007) and signal transmission (Rogerson *et al*. 2004). Furthermore the proposed techniques could even be fruitful for solving problems that are just loosely related to data collection (Guhaniyogi *et al*. (2011).

1.7 The chapters

Let us now give a quick overview of the contents of the book in terms of the individual chapters.

In the geostatistical context, the quality of inferences are affected substantially by the spatial configuration of the network of sites where measurements are taken. An extensive literature exists on spatial network design encompassing a variety of design objectives and statistical perspectives. Chapter 2 (*Model-based frequentist design for univariate and multivariate geostatistics*), written by Dale L. Zimmerman and Jie Li, considers spatial network design for four design objectives from a frequentist, parametric model-based perspective. The design objectives considered pertain to optimal estimation of model parameters and optimal spatial prediction, both for the univariate case (one spatially varying quantity of interest) and multivariate case (two or more such quantities).

Chapter 3 (*Model-based criteria heuristics for second-phase spatial sampling*), written by Eric M. Delmelle, discusses several objectives related to second-phase spatial sampling which is the process of collecting additional measurements of a spatial variable of interest. Different criteria exist to guide the location of these new measurements, and they generally are nonlinear. The author focuses on simulated annealing, a heuristic method which facilitates the process of finding a suitable sampling set among candidate locations.

As with any statistical prediction method, a general goal is to obtain as good predictions as possible for the complete area of investigation. One possibility to formalize this goal is to try to select the sampling locations in such a way that the sum of all kriging mean square errors of prediction becomes a minimum over the area of investigation. Notably, this is a very complicated optimization problem and becomes still more complicated by the fact that the covariance matrix enters the kriging mean square error of prediction in its inverse form. This design problem has been tackled in a number of recent research papers. Gunter Spöck and Jürgen Pilz in Chapter 4 (*Spatial sampling design by means of spectral approximations to the error process*) first consider the integrated kriging variance averaged over the area of investigation and then, as a combined design criterion, the averaged expected lengths of predictive intervals are considered. Finally it is shown how these design criteria can be generalized to spatial variables having skewed distributions. In such context a criterion for spatial sampling design with Box–Cox transformed spatial variables is proposed. Mathematically speaking, they approximate the investigated spatial random field by means of a large regression model consisting of cosine-sine Bessel surface harmonics with random amplitudes. This approximating regression model is a direct result of the polar spectral representation theorem for isotropic random fields. The authors show that kriging prediction in the original random field model is equivalent to trend prediction in this approximating Bayesian linear regression model.

Assessing the changes both spatially and temporally of data are highly important for monitoring purposes. Many spatio-temporal data are collected from monitoring stations. However, the selection of these stations is often influenced by administrative, political, and other pragmatic considerations. Thus it is important to evaluate if an existing station is statistically unnecessary. It is also crucial to find a location such that a new monitoring station upon it can greatly improve the data modeling. These are considered as environmental network design problems. Most of the approaches that tackle this type of problems can be classified into the following three categories: (1) geometry-based; (2) probability-based; (3) model-based. In Chapter 5 (*Entropy-based network design using hierarchical Bayesian kriging*), written by Baisuo Jin, Yuehua Wu, and Baiqi Miao, the attention is paid to the entropy-based design within Category (3). Although the authors cannot escape from the 'curse of dimensionality', they take advantage of recent increases in computational speed and numerical advances (e.g. MCMC) that allow the implementation of Bayesian space-time dynamical models in a hierarchical framework. Such specifications provide simple strategies for incorporating complicated space-time interactions at different stages of the model's hierarchy, and the models are feasible to implement in high dimensions.

In spatial statistics, the data locations and the measurement process are commonly stochastically dependent. In Chapter 6 (*Accounting for design in the analysis of spatial data*) Brian J. Reich and Montserrat Fuentes identify and distinguish between two such situations: informative sampling and informative missingness. Informative or preferential sampling occurs when measurement locations are selected in a way that depends on the underlying process. For example, air pollution monitors may be placed in locations with high air pollution levels. On the other hand, informative missingness occurs when the measurement locations are fixed, but observations are censored in a way that is informative about the underlying process. An approach that can be used in both situations is the shared variable model which introduces random effects that are shared between the data location and the measurement processes. Conditioned on the random effects, the data location and measurement processes are assumed to be independent. These methods can be naturally implemented in a Bayesian framework using MCMC methods.

Chapter 7 (*Spatial design for knot selection in knot-based dimension reduction models*), written by Alan E. Gelfand, Sudipto Banerjee and Andrew O. Finley, investigates the challenge of fitting hierarchical models for large spatial datasets. The authors propose the use of approximation through dimension reduction models, and work in the setting of spatial random effects specified through Gaussian processes. Such specification requires development of knot locations, whence the spatial design problem emerges. More precisely, they have a post-data collection, predate analysis design problem, and need to specify knots in order to fit desired spatial models. They propose an implementation of knot design based upon an average predictive variance criterion, and show how to implement the design and model fitting in a two-step process.

Efficient data acquisition requires some prior understanding of the process to be observed, ideally in the form of a spatio-temporal model. If this is not the case, the employed design must guarantee that an appropriate model can be identified. These designs, usually called exploratory designs, typically employ minimum assumptions which make random sampling a reasonable choice. However, random sampling can be very inefficient in a spatio-temporal context for the assessment of whether spatial or temporal dependence is present or not and if present at what intensity (and form). Chapter 8 (*Exploratory designs for assessing spatial dependence*), written by Agnes Fussl, Werner G. Müller and Juan Rodríguez-Díaz, addresses these questions and possible improvements over random sampling in coping with them.

Space-time geostatistics has received increasing attention over the past decades, both on the theoretical side with the publication of extensions to the set of valid space-time covariance structures, and on the practical side with the publication of many real-world applications. In this context, the use of space-time kriging will likely increase dramatically in the near future. Chapter 9 (*Sampling* *design optimization for space-time kriging*), written by Gerard B.M. Heuvelink, Daniel A. Griffith, Tomislav Hengl, and Stephanie J. Melles, deals with the choice of the number and configuration of the observation points in space and time. As the authors state, the main differences are that in space-time kriging, strong anisotropies must often be taken into account because variation in time can be quite different from variation in space, and in space-time there is a much richer set of sampling designs that one may choose from (e.g., static, synchronous and rotational designs). In their chapter, the authors restrict themselves to a criterion that simultaneously minimizes the estimation error of a linear trend and the average interpolation error of the kriging residual. This boils down to minimizing the average universal kriging variance. To obtain an optimal design, a solution algorithm that minimizes the defined criterion is necessary. They work with one particular numerical optimization technique known as spatial simulated annealing.

In many applications, spatio-temporal sampling procedures allow for temporal adaptability, in the sense that the set of spatial sampling locations can be (totally or partially) modified between different observation times. Prior or inferential knowledge on structural properties, jointly with real-time sample-path observed evolution, provide the basic information sources for dynamic modeling and prediction in this context. In Chapter 10 (*Space-time adaptive sampling and data transformations*), written by José M. Angulo, María C. Bueso and Francisco J. Alonso, some related approaches, mainly based on entropy criteria, are described and illustrated. The effect of variable transformations such as averages and/or maxima over given time periods or space regions, often considered in the formulations of critical indicators, is also evaluated in terms of entropy-based information measures. Optimality criteria based on entropy and related quantities such as mutual information have been adopted for the design of spatial and spatio-temporal sampling strategies, particularly in the context of environmental applications. Here, the authors consider the sampling network real-time adaptation based on the observations, and the quantification of the information obtained by transformation of the data.

As exposure assessment technology improves, it is likely that mobile monitoring networks will be used instead of, or in addition to, permanent networks, possibly in conjunction with alternative technologies such as leaf sampling. This gives rise to the adaptive or dynamic sampling design problem. In Chapter 11 (*Adaptive sampling design for spatio-temporal prediction*) Thomas R. Fanshawe and Peter J. Diggle consider model-based design, in which the optimal design problem requires two key features to be specified: (i) a statistical or mathematical model for the process under consideration; (ii) a criterion with respect to which the design is required to be optimized. In practice, a good choice of model may not be known until data collection has begun. This highlights another benefit of adaptive designs: model fitting can take place as data become available, and future design locations chosen accordingly.

Capturing this dynamical component of the spatial structure of many environmental processes, through effective modeling strategies, can lead to efficient and cost effective designs for spatial monitoring networks. Furthermore, environmental variables related to processes are often non-Gaussian. In many cases, there also exist important environmental factors that are related nonlinearly to the process mean (or location parameter). To accommodate these types of data complexities, Scott H. Holan and Christopher K. Wikle present in Chapter 12 (*Semiparametric dynamic design of monitoring networks for non-Gaussian spatio-temporal data*) a low-rank semiparametric simulation-based design approach for data distributions from the spatio-temporal exponential family. Fundamental to their approach is that the principal random component in the conditional mean function is a spatio-temporal dynamic process modeled with a low-rank basis function expansion and that other environmental factors are potentially modeled semiparametrically using low-rank basis function representations.

Chapter 13 (*Active learning for monitoring network optimization*), written by Devis Tuia, Alexei Pozdnoukhov, Loris Foresti and Mikhail Kanevski, presents the use of active learning to design monitoring networks that are optimal for environmental modeling with machine learning algorithms. In machine learning, active learning defines a family of algorithms aimed at the selection of new data points from a pool of available or potential measurements. In the case of monitoring networks optimization, these data are the locations of new measuring stations or measurement points. A general framework of active learning is presented in this chapter. The main attention is paid to the Support Vector algorithms which recently demonstrated high efficiency in data analysis and modeling tasks and gave rise to new interesting and important approaches in active learning.

Monitoring of pollution in air or water can be improved by considering the physical laws underlying the dispersion process of such substances. Pollutants usually spread from point sources in the shape of a plume. The randomness of these fields is mainly due to the uncertainty of flow parameters or weather forecasts and to the limited information about the source characteristics. Simulations of plumes are a feasible way to study these fields for sampling design optimization. They give a detailed estimation of what sensors could measure and how good the conclusions are that we draw from such measurements. Although these phenomena are dynamic, Chapter 14 (*Stationary sampling designs based on plume simulations*), written by Kristina B. Helle and Edzer Pebesma, focuses on optimizing networks with stationary sensors, meaning that the sensors do not move during plume passage. Optimizing sampling designs with the help of plume simulations is quite common for groundwater monitoring. It is a bit less common for the monitoring of air pollution because winds change faster than ground water fluxes and therefore the dispersion parameters of the fields are less well known in most applications. Their main concern is monitoring of pollution that spreads from a few potential sources by single plumes like poisonous or radioactive gases or aerosols. The random fields of such plumes are the concentrations or dose rates in the space around the source in the time after the beginning of the release. Uncertainty in these fields is due to the incomplete knowledge of the boundary conditions of the dispersion like the wind fields and sometimes also due to uncertainty about the source. Probabilities of these conditions are very complex as is their propagation through the dispersion. However, by adopting a dispersion model we can generate realizations to test proposed sampling designs.

Acknowledgments

Werner Müller would like to acknowledge the support of the projects OEAD-WTZ FR 11/2010 and FWF/ANR I-833 N18 and his colleagues Luc Pronzato, Joao Rendas, and Milan Stehlík; and particularly Helmut Waldl, who provided the R-code for producing the kriging variance maps that are displayed on the book cover. He also wants to thank attentive students from various courses at the JKU Linz, particularly Andreas Rappold, who provided Figure 1.7. The work on this book was made easier by the supportive environment at the Department of Applied Statistics; particular gratitude goes to Andreas Quatember. Gabriele Mack provided invaluable secretarial assistance.

Jorge Mateu acknowledges the support of the projects MTM2010-14961 and P11B2008-27.

Both authors would like to thank their international colleagues for showing interest in the project, in particular Noel Cressie for fruitful discussions and providing the Foreword. We are grateful to Tomislav Hengl for providing transformations of our data between different coordinate systems. We are also grateful to the Zentralanstalt für Meteorologie und Geodynamik, Vienna and Thomas Raferzeder from the Statistics department of the Amt der Oberösterreichischen Landesregierung, Linz for providing the data.

We are thankful for the stimulating professional environment Wiley has provided conducive to this book activity. The support of the Wiley team through the Production Editor, Project Manager and Copyeditor is acknowledged with appreciation for encouragement. They have done a magnificent job. Special gratitude goes to Ilaria Meliconi, who initiated the project, Richard Davies, Heather Kay, Prachi Sinha Sahay from Wiley and Jayashree Saishankar from Laserwords.

References

Abida R, Bocquet M, Vercauteren N and Isnard O 2008 Design of a monitoring network over France in case of a radiological accidental release. *Atmospheric Environment* **42**(21), 5205–5219.

Anselin L 1988 *Spatial Econometrics: Methods and Models (Studies in Operational Regional Science)*. Springer.

Anselin L 2010 Thirty years of spatial econometrics. *Papers in Regional Science* **89**(1), 3–25.

Atkinson A, Donev A and Tobias R 2007 *Optimum Experimental Designs, with SAS (Oxford Statistical Science Series)*. Oxford University Press.

Barabesi L, Franceschi S and Marcheselli M 2012 Properties of design-based estimation under stratified spatial sampling with application to canopy coverage estimation. *Annals of Applied Statistics* **6**(1), 210–228.

Baume OP, Gebhardt A, Gebhardt C, Heuvelink GBM and Pilz J 2010 Network optimization algorithms and scenarios in the context of automatic mapping. *Computers & Geosciences* **37**, 289–294.

Baxevani A, Podgórski K and Rychlik I 2011 Dynamically evolving Gaussian spatial fields. *Extremes* **14**(2), 223–251.

Berliner LM, Lu ZQ and Snyder C 1999 Statistical design for adaptive weather observations. *Journal of the Atmospheric Sciences* **56**(15), 2536–2552.

Bevilacqua M, Gaetan C, Mateu J and Porcu E 2012 Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. *Journal of the American Statistical Association* **107**(497), 268–280.

Bivand RS, Pebesma EJ and Gómez-Rubio V 2008 *Applied Spatial Data Analysis with R (Use R!)*. Springer.

Bochner S 1933 A theorem on fourier-stieltjes integrals. *Mathematische Annalen* **108**, 378–410.

Borgoni R, Radaelli L, Tritto V and Zappa D 2010 Optimal reduction of a monitoring grid for SiO_2 deposition surface over a wafer for semiconductor devices. *Atti della XLV Riunione Scientifica della Società ltaliana di Statistica*.

Brimkulov UN, Krug GK and Savanov VL 1980 Numerical construction of exact experimental designs when the measurements are correlated. *Zavodskaya Laboratoria* **36**, 435–442.

Brown P, Roberts G, KÅresen K and Tonellato S 2000 Blur-generated non-separable space–time models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **62**(4), 847–860.

Brus D and de Gruijter JJ 1997 Random sampling or geostatistical modelling? Choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma* **80**(1–2), 1–44.

Brus D and Heuvelink G 2007 Optimization of sample patterns for universal kriging of environmental variables. *Geoderma* **138**(1–2), 86–95.

Brus DJ and de Gruijter JJ 2011 Design-based Generalized Least Squares estimation of status and trend of soil properties from monitoring data. *Geoderma* **164**(3–4), 172–180.

Brus DJ and de Gruijter JJ 2012 A hybrid design-based and model-based sampling approach to estimate the temporal trend of spatial means. *Geoderma* **173–174**, 241–248.

Caragea P and Smith R 2006 Approximate likelihoods for spatial processes. *Proceedings of the American Statistical Association Preprint*.

Carlier L, De Vliegher A, Van Cleemput O and Boeckx P 2005 Importance and functions of European grasslands. *Communications in Agricultural and Applied Biological Sciences* **70**(1), 5–15.

Caselton WF and Zidek JV 1984 Optimal monitoring network designs. *Statistics & Probability Letters* **2**(4), 223–227.

Chen L, Fuentes M and Davis J 2006 Spatial–temporal statistical modeling and prediction of environmental processes. In *Hierarchical Modelling for the Environmental Sciences* (eds Clark JS and Gelfand A). Oxford University Press, pp. 121–144.

Chilès JP and Delfiner P 1999 *Geostatistics: Modeling Spatial Uncertainty (Wiley Series in Probability and Statistics)*. Wiley-Interscience.

Corwin DL, Lesch SM, Segal E, Skaggs TH and Bradford SA 2010 Comparison of sampling strategies for characterizing spatial variability with apparent soil electrical conductivity directed soil sampling. *Journal of Environmental & Engineering Geophysics* **15**(3), 147–162.

Creelman C and Risk D 2011 Network design for soil CO_2 monitoring of the northern North American region. *Ecological Modelling* **222**(18), 3421–3428.

Cressie N 1993 *Statistics for Spatial Data (Wiley Series in Probability and Statistics)* revised. Wiley-Interscience.

Cressie N and Hawkins D 1980 Robust estimation of the variogram: I. *Mathematical Geology* **12**(2), 115–125.

Cressie N and Huang H 1999 Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* **94**, 1330–1340.

Cressie N and Wikle CK 2011 *Statistics for Spatio-Temporal Data (Wiley Series in Probability and Statistics)*. John Wiley & Sons, Ltd.

Cressie N, Calder CA, Clark JS, Hoef JM and Wikle CK 2009 Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecological Applications* **19**(3), 553–570.

Curriero F and Lele S 1999 A composite likelihood approach to semivariogram estimation. *Journal of Agricultural, Biological, and Environmental Statistics* **4**(1), 9–28.

De Cesare L, Myers D and Posa D 2001a Estimating and modeling space–time correlation structures. *Statistics & Probability Letters* **51**(1), 9–14.

De Cesare L, Myers D and Posa D 2001b Product-sum covariance for space-time modeling: an environmental application. *Environmetrics* **12**(1), 11–23.

de Gruijter JJ and ter Braak CJF 1990 Model-free estimation from spatial samples: a reappraisal of classical sampling theory. *Mathematical Geology* **22**(4), 407–415.

de Gruijter JJ, Brus DJ, Bierkens MFP and Knotters M 2006 *Sampling for Natural Resource Monitoring*. Springer.

De Iaco S, Myers D and Posa D 2001 Space–time analysis using a general product–sum model. *Statistics & Probability Letters* **52**(1), 21–28.

De Iaco S, Myers D and Posa D 2002a Nonseparable space-time covariance models: some parametric families. *Mathematical Geology* **34**(1), 23–42.

De Iaco S, Myers D and Posa D 2002b Space–time variograms and a functional form for total air pollution measurements. *Computational Statistics & Data Analysis* **41**(2), 311–328.

De Iaco S, Myers D and Posa D 2003 The linear coregionalization model and the product–sum space–time variogram. *Mathematical Geology* **35**(1), 25–38.

Diggle P and Lophaven S 2006 Bayesian Geostatistical Design. *Scandinavian Journal of Statistics* **33**(1), 53–64.

Dobbie MJ, Henderson BL and Stevens DL 2008 Sparse sampling: spatial design for monitoring stream networks. *Statistical Survey* **2**, 113–153.

Elhorst JP 2012 Dynamic spatial panels: models, methods, and inferences. *Journal of Geographical Systems* **14**(1), 5–28.

Fedorov V 1996 Design of spatial experiments: model fitting and prediction. In *Handbook of Statistics* (eds Gosh S and Rao CR) vol. 13. Elsevier, pp. 515–553.

Fedorov VV 1971 The design of experiments in the multiresponse case. *Theory of Probability and its Applications* **16**(2), 323–332.

Fedorov VV and Müller WG 1988 Two approaches in optimization of observing networks. In *Optimal Design and Analysis of Experiments* (eds Dodge Y, Fedorov VV and Wynn HP). North Holland, pp. 239–256.

Fernández-Casal R, González-Manteiga W and Febrero-Bande M 2003 Flexible spatio-temporal stationary variogram models. *Statistics and Computing* **13**(2), 127–136.

Fewster RM 2011 Variance estimation for systematic designs in spatial surveys. *Biometrics* **67**(4), 1518–1531.

Fuentes M, Chen L, Davis J and Lackmann G 2005 Modeling and predicting complex space–time structures and patterns of coastal wind fields. *Environmetrics* **16**(5), 449–464.

Furrer R, Genton M and Nychka D 2006 Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics* **15**(3), 502–523.

Furrer R, Nychka D and Sain S 2011 R-package 'fields', version 6.6.3.

Gao H, Wang J and Zhao P 1996 The updated kriging variance and optimal sample design. *Mathematical Geology* **28**(3), 295–313.

Gneiting T 2002 Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* **97**(458), 590–600.

Gregori P, Porcu E, Mateu J and Sasvári Z 2008 On potentially negative space time covariances obtained as sum of products of marginal ones. *Annals of the Institute of Statistical Mathematics* **60**(4), 865–882.

Gribik PR, Kortanek KO and Sweigart JR 1976 Designing a regional air pollution monitoring network: An appraisal of a regression experimental design approach. In *Conference on Environmental Modeling and Simulation* (ed. Ott WR). EPA, pp. 86–91.

Griffith D and Paelinck J 2007 An equation by any other name is still the same: on spatial econometrics and spatial statistics. *The Annals of Regional Science* **41**(1), 209–227.

Griffith DA and Csillag F 1993 Exploring relationships between semi-variogram and spatial autoregressive models. *Papers in Regional Science* **72**(3), 283–295.

Grimes DIF and Pardo-Igúzquiza E 2010 Geostatistical analysis of rainfall. *Geographical Analysis* **42**(2), 136–160.

Guedes LPC, Ribeiro PJ, De Stefano SM and Uribe-Opazo MA 2011 Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing. *Chilean Journal of Statistics* **2**(2), 39–50.

Guhaniyogi R, Finley AO, Banerjee S and Gelfand AE 2011 Adaptive Gaussian predictive process models for large spatial datasets. *Environmetrics* **22**(8), 997–1007.

Hae-Ryoung S, Fuentes M and Ghosh S 2008 A comparative study of Gaussian geostatistical models and Gaussian Markov random field models. *Journal of Multivariate Analysis* **99**(8), 1681–1697.

Haining RP, Kerry R and Oliver MA 2010 Geography, spatial data analysis, and geostatistics: an overview. *Geographical Analysis* **42**(1), 7–31.

Heuvelink GBM and Egmond FM 2010 Space-time geostatistics for precision agriculture: a case study of NDVI mapping for a Dutch potato field. In *Geostatistical Applications for Precision Agriculture* (ed. Oliver MA). Springer, pp. 117–137.

Hooten MB, Wikle CK, Sheriff SL and Rushin JW 2009 Optimal spatio-temporal hybrid sampling designs for ecological monitoring. *Journal of Vegetation Science* **20**(4), 639–649.

Johnson ME, Moore LM and Ylvisaker D 1990 Minimax and maximin distance designs. *Journal of Statistical Planning and Inference* **26**(2), 131–148.

Jones R and Zhang Y 1997 Models for continuous stationary space-time processes. *Lecture Notes in Statistics* 289–298.

Kaufman C, Schervish M and Nychka D 2008 Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association* **103**(484), 1545–1555.

Kiefer J and Wolfowitz J 1960 The equivalence of two extremum problems. *Canadian Journal of Mathematics* **12**, 363–366.

Kincaid T and Olsen T 2012 R-package ‘spsurvey’, version 2.3.

Kolovos A, Christakos G, Hristopulos D and Serre M 2004 Methods for generating non-separable spatiotemporal covariance models with potential environmental applications. *Advances in Water Resources* **27**(8), 815–830.

Kumar N 2007 Spatial Sampling Design for a Demographic and Health Survey. *Population Research and Policy Review* **26**(5), 581–599.

Le ND and Zidek JV 2006 *Statistical Analysis of Environmental Space-Time Processes (Springer Series in Statistics).* Springer.

Lindgren F, Rue H and Lindström J 2011 An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(4), 423–498.

Lindsay B 1988 Composite likelihood methods. *Contemporary Mathematics* **80**(1), 221–239.

Ma C 2002 Spatio-temporal covariance functions generated by mixtures. *Mathematical Geology* **34**(8), 965–975.

Ma C 2003a Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference* **116**(2), 489–501.

Ma C 2003b Nonstationary covariance functions that model space–time interactions. *Statistics & Probability Letters* **61**(4), 411–419.

Ma C 2003c Spatio-temporal stationary covariance models. *Journal of Multivariate Analysis* **86**(1), 97–107.

Ma C 2005a Linear combinations of space-time covariance functions and variograms. *IEEE Transactions on Signal Processing* **53**(3), 857–864.

Ma C 2005b Semiparametric spatio-temporal covariance models with the arma temporal margin. *Annals of the Institute of Statistical Mathematics* **57**(2), 221–233.

Ma C 2005c Spatio-temporal variograms and covariance models. *Advances in Applied Probability* **37**(3), 706–725.

Marchant B and Lark R 2007 Optimized sample schemes for geostatistical surveys. *Mathematical Geosciences* **39**(1), 113–134.

Marchant BP and Lark RM 2006 Adaptive sampling and reconnaissance surveys for geostatistical mapping of the soil. *European Journal of Soil Science* **57**(6), 831–845.

Mardia K and Marshall R 1984 Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**(1), 135–146.

Martínez F, Mateu J, Montes F, Bodas-Salcedo A and López-Baeza E 2008 A comparative analysis of different spatial sampling schemes: modelling of SSRB data. *International Journal of Remote Sensing* **29**(6), 1635–1647.

Masoumi F and Kerachian R 2010 Optimal redesign of groundwater quality monitoring networks: a case study. *Environmental Monitoring and Assessment* **161**(1), 247–257.

Mateu J, Juan P and Porcu E 2007 Geostatistical analysis through spectral techniques: some words of caution. *Communications in Statistics - Simulation and Computation* **36**(5), 1035–1051.

Maurer H, Curtis A and Boerner DE 2010 Recent advances in optimized geophysical survey design. *Geophysics* **75**(5), 75A177–75A194.

McBratney A and Webster R 1981 The design of optimal sampling schemes for local estimation and mapping of regionalized variables. II Program and examples. *Computers & Geosciences* **7**(4), 335–365.

Melles SJ, Heuvelink GBM, Twenhöfel CJW, van Dijk A, Hiemstra PH, Baume O and Stöhlker U 2011 Optimizing the spatial pattern of networks for monitoring radioactive releases. *Computers & Geosciences* **37**(3), 280–288.

Müller WG 1998 *Collecting Spatial Data: Optimum Design of Experiments for Random Fields*. Springer.

Müller WG 1999 Least-squares fitting from the variogram cloud. *Statistics & Probability Letters* **43**(1), 93–98.

Müller WG 2007 *Collecting Spatial Data: Optimum Design of Experiments for Random Fields* 3rd revised and extend edn. Springer.

Müller WG and Pázman A 2003 Measures for designs in experiments with correlated errors. *Biometrika* **90**(2), 423–434.

Müller WG and Stehlík M 2010 Compound optimal spatial designs. *Environmetrics* **21**(3-4), 354–364.

Müller WG and Waldl H 2011 Discussion of Lindgren, Rue and Lindström 'An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach'. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(4), 483–485.

Müller WG and Zimmerman DL 1999 Optimal designs for variogram estimation. *Environmetrics* **10**(1), 23–37.

Müller WG, Pronzato L and Waldl H 2012 Relations between designs for prediction and estimation in random fields: an illustrative case. In *Advances and Challenges in Space-time Modelling of Natural Events* (ed. Porcu E, Montero J and Schlather M) vol. 207 of *Lecture Notes in Statistics*. Springer. pp. 125–139.

Murtojärvi M, Suominen T, Uusipaikka E and Nevalainen OS 2011 Optimising an observational water monitoring network for Archipelago Sea, South West Finland. *Computers & Geosciences* **37**(7), 844–854.

Nowak W 2010 Measures of parameter uncertainty in geostatistical estimation and geostatistical optimal design. *Mathematical Geosciences* **42**(2), 199–221.

Nowak W, de Barros FPJ and Rubin Y 2010 Bayesian geostatistical design: task-driven optimal site investigation when the geostatistical model is uncertain. *Water Resources Research* **46**(3), W03535+.

Olea RA 1984 Sampling design optimization for spatial functions. *Mathematical Geology* **16**(4), 369–392.

Oud JHL, Folmer H, Patuelli R and Nijkamp P 2012 Continuous-time modeling with spatial dependence. *Geographical Analysis* **44**(1), 29–46.

Overton WS and Stehman SV 1996 Desirable design characteristics for long-term monitoring of ecological variables. *Environmental and Ecological Statistics* **3**, 349–361.

Papamichail DM and Metaxa IG 1996 Geostatistical analysis of spatial variability of rainfall and optimal design of a rain gauge network. *Water Resources Management* **10**(2), 107–127.

Papritz A and Webster R 1995 Estimating temporal change in soil monitoring: I. Statistical theory. *European Journal of Soil Science* **46**(1), 1–12.

Porcu E and Mateu J 2007 Mixture-based modeling for space–time data. *Environmetrics* **18**(3), 285–302.

Porcu E, Gregori P and Mateu J 2006 Nonseparable stationary anisotropic space–time covariance functions. *Stochastic Environmental Research and Risk Assessment* **21**(2), 113–122.

Porcu E, Gregori P and Mateu J 2007a La descente et la montée étendues: the spatially d-anisotropic and the spatio-temporal case. *Stochastic Environmental Research and Risk Assessment* **21**(6), 683–693.

Porcu E, Mateu J and Bevilacqua M 2007b Covariance functions that are stationary or nonstationary in space and stationary in time. *Statistica Neerlandica* **61**(3), 358–382.

Porcu E, Mateu J and Christakos G 2009 Quasi-arithmetic means of covariance functions with potential applications to space-time data. *Journal of Multivariate Analysis* **100**(8), 1830–1844.

Pronzato L and Müller WG 2012 Design of computer experiments: space filling and beyond. *Statistics and Computing* **22**(3), 681–701.

Rasmussen CE and Williams CKI 2005 *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning series)*. The MIT Press.

Reed PM and Minsker BS 2004 Striking the balance: long-term groundwater monitoring design for conflicting objectives. *Journal of Water Resources Planning and Management* **130**(2), 140–149.

Rogerson P, Delmelle E, Batta R, Akella MR, Blatt A and Wilson G 2004 Optimal sampling design for variables with varying spatial importance. *Geographical Analysis* **36**(2), 177–194.

Romary T, de Fouquet C and Malherbe L 2011 Sampling design for air quality measurement surveys: an optimization approach. *Atmospheric Environment* **45**(21), 3613–3620.

Royle J and Nychka D 1998 An algorithm for the construction of spatial coverage designs with implementation in SPLUS. *Computers & Geosciences* **24**(5), 479–488.

Rue H and Held L 2005 *Gaussian Markov Random Fields: Theory and Applications (Chapman & Hall/CRC Monographs on Statistics & Applied Probability)* Chapman and Hall/CRC.

Ruiz-Cárdenas R, Ferreira MAR and Schmidt AM 2010 Stochastic search algorithms for optimal design of monitoring networks. *Environmetrics* **21**(1), 102–112.

Schoenberg I 1938 Metric spaces and positive definite functions. *Transactions of the American Mathematical Society* **44**(3), 522–536.

Spöck G and Pilz J 2010 Spatial sampling design and covariance-robust minimax prediction based on convex design ideas. *Stochastic Environmental Research and Risk Assessment* **24**(3), 463–482.

Stehman SV 1999 Basic probability sampling designs for thematic map accuracy assessment. *International Journal of Remote Sensing* **20**(12), 2423–2441.

Stehman SV 2009 Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing* **30**(20), 5243–5272.

Stein M 2005 Space-time covariance functions. *Journal of the American Statistical Association* **100**(469), 310–321.

Stein M, Chi Z and Welty L 2004 Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **66**(2), 275–296.

Stevens DL 2006 Spatial properties of design-based versus model-based approaches to environmental sampling. In *Proceedings of the 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, pp. 119–125.

Stevens DL and Olsen AR 2004 Spatially balanced sampling of natural resources. *Journal of the American Statistical Association* **99**(465), 262–278.

Thompson SK 2002 On sampling and experiments. *Environmetrics* **13**(5–6), 429–436.

Thornley JHM and Cannell MGR 1997 Temperate grassland responses to climate change: an analysis using the Hurley Pasture Model. *Annals of Botany* **80**(2), 205–221.

Ucinski D and Chen Y 2005 Time optimal path planning of moving sensors for parameter estimation of distributed systems. In *Proceedings of the 44th IEEE Conference on Decision and Control*. IEEE, pp. 5257–5262.

Ucinski D and Patan M 2010 Sensor network design for the estimation of spatially distributed processes. *International Journal of Applied Mathematics and Computer Science* **20**(3), 459–481.

van Groenigen JW and Stein A 1998 Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality* **27**(5), 1078–1086.

Varin C and Vidoni P 2005 A note on composite likelihood inference and model selection. *Biometrika* **92**(3), 519–528.

Vašát R, Heuvelink GBM and Borůvka L 2009 Sampling design optimization for multivariate soil mapping. *Geoderma* **155**(3–4), 147–153.

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

Vellinga T, van den Pol-van Dasselaar A and Kuikman PJ 2004 The impact of grassland ploughing on CO_2 and N_2O emissions in the Netherlands. *Nutrient Cycling in Agroecosystems* **70**(1), 33–45.

Ver Hoef JM 2012 Practical considerations for experimental designs of spatially autocorrelated data using computer intensive methods. *Statistical Methodology* **9**(1–2), 172–184.

Whittle P 1954 On stationary processes in the plane. *Biometrika* **41**, 434–449.

Wiens DP 2005 Robustness in spatial studies II: minimax design. *Environmetrics* **16**(2), 205–217.

Wikle CK and Royle JA 1999 Space-time dynamic design of environmental monitoring networks. *Journal of Agricultural, Biological, and Environmental Statistics* **4**(4), 489–507.

Wikle CK and Royle JA 2005 Dynamic design of ecological monitoring networks for non-Gaussian spatio-temporal data. *Environmetrics* **16**(5), 507–522.

Wu L and Bocquet M 2011 Optimal redistribution of the background ozone monitoring stations over France. *Atmospheric Environment* **45**(3), 772–783.

Wynn HP 1970 The sequential generation of D-optimum experimental designs. *The Annals of Mathematical Statistics* **41**(5), 1655–1664.

Yang Y, Feng Z, Liu D and Zhang J 2006 The Spatial-temporal change of grassland in Qinghai-Tibet Plateau. In *2006 IEEE International Symposium on Geoscience and Remote Sensing*. IEEE, pp. 3099–3102.

Zhu Z and Stein ML 2006 Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological, and Environmental Statistics* **11**(1), 24–44.

Zidek JV and Zimmerman DL 2010 *Monitoring Network Design*. CRC Press, pp. 131–148.

Zimmerman D 1989 Computationally efficient restricted maximum likelihood estimation of generalized covariance functions. *Mathematical Geology* **21**(7), 655–672.

Zimmerman DL 2006 Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. *Environmetrics* **17**(6), 635–652.

Zimmermann B, Zimmermann A, Lark RM and Elsenbeer H 2010 Sampling procedures for throughfall monitoring: a simulation study. *Water Resources Research* **46**(1), W01503+.