Stoch Environ Res Risk Assess
DOI 10.1007/s00477-017-1409-9

ORIGINAL PAPER

A simple two-step method for spatio-temporal design-based balanced sampling

Ramin Khavarzadeh[1] · **Mohsen Mohammadzadeh**[1] · **Jorge Mateu**[2]

Abstract We introduce a two-step method to perform spatio-temporal balanced sampling in a design-based approach. For populations with spatio-temporal trends and with anisotropic effects in the variable of interest, the prediction can be further improved by selecting samples that are well spread over the entire population in space and time. We control the spread of the sample over the population by using the volume of the corresponding three-dimensional Voronoi tessellation. Indeed, spatio-temporal design-based balanced sampling is even more efficient under the presence of a trend and anisotropic effects. We present an intensive simulation study comparing our method to other available methods for spatio-temporal sampling. Finally, we analyze real data by sampling from a population of temperature stations over six European countries.

Keywords Balanced sampling · Design-based sampling · Spatio-temporal sampling

1 Introduction

In data analysis, each one of the individuals in a large population of interest cannot generally be surveyed. Instead, we usually sample a subset of individuals, and use these observations to draw conclusions about the whole population. Ideally, the sample resembles the characteristics of the target population, and thus the conclusions obtained from the sample are likely applicable to the whole population. Sampling is concerned with choosing a group of individuals from an entire population to estimate particular characteristics of such a population. Examples of such characteristics could be the mean parameter of a random field (Haining 2003; Christakos 2005), the location of targets (Rogerson et al. 2004), or values at unsampled sites (Goovaerts 1997). In a classical survey sampling the population is usually finite, and some auxiliary information is assumed to be known about the whole population. Let U indicate the population of N units, and the indicator I_i represents whether the unit i is selected or not. Note that we use the usual terminology in sampling designs, where the individuals of a population are called units. We have $I_i = 1$ if unit i is included in the sample, and $I_i = 0$ otherwise. Each unit i has a positive inclusion probability $0 < \pi_i < 1$. The first-order inclusion probability, $\pi_i = \mathrm{E}(I_i) = \mathrm{P}(I_i = 1)$, is the probability that the unit i is selected. The second-order inclusion probability, $\pi_{ij} = \mathrm{E}(I_i I_j) = \mathrm{P}(I_i = 1, I_j = 1)$, is the probability that both units i and j are selected. When the sample size is equal to n and it is fixed, the inclusion probabilities should satisfy

$$n = \sum_{i=1}^{N} \pi_i. \tag{1}$$

Note that, for example, in simple random sampling (SRS) with replacement, the inclusion probability for each unit in each selection is $\frac{1}{N}$, so for a sample with size n, the inclusion probability is $\frac{n}{N}$, and we then have $\sum_{i=1}^{N} \pi_i = \frac{n}{N} + \cdots + \frac{n}{N} = n$. Suppose S_n is the set of all possible samples of size n drawn from the population U. Then S_n is a sampling design along with a function

✉ Mohsen Mohammadzadeh
Mohsen_m@modares.ac.ir

[1] Department of Statistics, Tarbiat Modares University, Tehran, Iran

[2] Department of Mathematics, University Jaume I, Castellón, Spain

Published online: 07 April 2017

$p(\cdot) > 0$, where $p(s)$ denotes the selection probability of a sample $s \in S_n$. Let y_i denote the value of the target variable for the population unit i, then for a sample $s \in S_n$, the total quantity

$$Y_i = \sum_{i=1}^{N} y_i, \tag{2}$$

can be estimated using the unbiased Horvitz–Thompson estimator (Horvitz and Thompson 1952), as follows

$$\hat{Y}_{\mathrm{HT}} = \sum_{i=1}^{N} \frac{y_i}{\pi_i} I_i. \tag{3}$$

Let $X_1, \ldots, X_J$ be a set of auxiliary variables, so that their values $x_{k1}, \ldots, x_{kJ}$ for the kth unit (kth individual of the population) are known. According to the definition given in Tillé (2006), a balanced sampling design is defined as follows

$$\sum_{k \in s} \frac{x_{kj}}{\pi_k} \approx \sum_{k \in U} x_{kj}, \quad j = 1, \ldots, J \tag{4}$$

Balanced sampling can be used in stratified sampling and in sampling with a fixed sample size. Indeed, a stratified design is balanced on the indicator variables of the strata, because the Horvitz–Thompson estimators of the sizes of the strata are proportional to the population sizes of such strata. In design-based inference, balanced sampling allows for a large improvement in the efficiency of the Horvitz–Thompson estimator when the auxiliary variables are correlated with the variable of interest (Deville and Tillé 2004). In model-based inference, the selection of balanced samples has often been considered to protect against misspecification of the model (Valliant et al. 2000).

This article presents a new method for sampling from a spatio-temporal population. Many populations under study are in fact distributed over space and time, but a wide number of sampling designs, such as SRS, do not accommodate the spatial and/or temporal aspects into the design. If nearby units (or locations in space and/or time) behave more similarly than units further apart, which is a very common feature, then it is useful to make sure that the sample is well spread over the population. A well-spread sample in space and time is said to be spatio-temporally balanced (Grafström and Tillé 2013). It is well established that spatio-temporally balanced sampling is efficient, and that is why difficult types of systematic designs are commonly used. With a systematic design it is a trouble to use unequal inclusion probabilities. In one dimension it is possible to use systematic sampling with unequal probabilities, also known as systematic π_{ps} sampling, to make sure that the sample is well spread over the population. Stevens and Olsen (2004) generalized this concept to two dimensions by introducing the generalized random-tessellation stratified (GRTS) method. The GRTS method uses a specific random mapping to map the two-dimensional locations into one dimension, while preserving some spatial order. Units close in the two-dimensional space tend to be close in the one-dimensional space after the mapping. The sample is then selected in one dimension using systematic π_{ps} sampling. This procedure assures that a sample that is well spread over the population is selected. Another sampling method that maps two dimensions into one, by using space-filling curves, is provided and assessed by Lister and Scott (2009). It is also a common approach to perform spatial stratification. This is often not straightforward and can be done in many different ways. To achieve a well-spread sample, it is required that the strata are quite small. When the units have unequal inclusion probabilities it is more difficult to split the population into smaller strata. To select a fixed number of units within each stratum, the inclusion probabilities within each stratum must sum to an integer. More simplicity is achieved if it is possible to avoid a spatial stratification. In this article, we define a new method based on balanced sampling by using the cube method to choose representative sampling from a spatio-temporal population. We use the spatial coordinates (x, y) and the temporal instants as balanced variables. This approach selects a sample that is well-spread over the population in space and time. Most of the spatial applications naturally concern populations spread in one, two, or three dimensions. For populations with auxiliary variables available (thus variables that provide useful information related to the problem at hand), the sample can be balanced in the auxiliary space or auxiliary time, which might consist of more than three dimensions.

The paper is structured as follows. In Sect. 2, we briefly discuss the history of spatial and spatio-temporal sampling. In Sect. 3, we provide short descriptions of the balanced sampling strategy and the cube method, and present our method that makes use of the cube method for spatio-temporal data. Spatio-temporal universal kriging is presented and discussed in Sect. 4, together with some accuracy measures. A simulation study and a real data example are presented in Sect. 5. Concluding remarks are given in Sect. 6.

2 Spatial sampling: a historical follow-up

Sampling is the process of selecting units from a target population so that the sample allows unknown quantities of the population to be estimated. Intuitive applications of the principles of sampling in science have taken place for a long time from very early human history in Egypt, China and other places throughout the world. The first known

attempt to make statements about a population using information from only a part of it was by the English merchant John Graunt (1620–1674). His famous tract describes a method for estimating the population of London on the basis of partial information. Since then, the sampling theory has developed separately from the mainstream of classical statistics (Neyman 1934) and has evolved into an extensive body of theory, methods, and operations that are used on a daily basis all over the world. "Sampling Techniques", a landmark book by Cochran (1977), is widely used in modern sampling practices. To deal with two-dimensional spatial sampling, the regionalized variable theory, often referred to as geostatistics, was well built and is widely applied in geosciences (Matheron 1971). This approach uses spatial autocorrelation to improve the sampling efficiency in terms of the estimator error variance in relation to the sample design and the sample size (Stein and Ettema 2003; Christakos 2005). Spatial stratified heterogeneity was considered to achieve more efficient spatial sampling and inference (Goovaerts 1997; Li et al. 2008; Wang et al. 2010).

The importance of (optimal) spatial sampling design for environmental applications and soil science has been shown in several papers and monographs (Cox 1999; van Groenigen et al. 1999; Brus and Heuvelink 2007; Dobbie et al. 2008; Delmelle and Goovaerts 2009). Published papers on spatial sampling design may be divided into several categories, although some of which are overlapping. We may differentiate between design criteria for spatial prediction and estimation of the covariance function, and criteria combining both objectives. Contributions falling into the category of criteria for prediction are provided by Müller (2005) and Brus and Heuvelink (2007). Criteria for the estimation of the covariance function are considered by Zimmerman and Homer (1991) and Müller and Zimmerman (1999). Combined criteria can be found in Zhu and Stein (2006), who considered the minimization of the average expected length of predictive intervals. Other papers falling into this category of combined criteria are Bayesian approaches specifying prior distributions over covariance functions such as those by Brown et al. (1994), Müller et al. (2004) and Fuentes et al. (2007). Indeed, Brown et al. (1994) and Fuentes et al. (2007) considered the covariance function to be nonstationary and deal with an entropy-based design criterion according to which the determinant of the covariance matrix between locations to be added to the design must be maximized. Both of them make use of simulated annealing algorithms to find optimal designs satisfying their criteria.

On the more computational side, we can distinguish between stochastic search algorithms such as simulated annealing (Aarts and Korst 1989), or evolutionary genetic algorithms, and deterministic algorithms for optimizing the investigated design criteria. With the exception of Müller (2005), and Spöck and Pilz (2010), almost all algorithms for spatial sampling design optimization use stochastic search algorithms to find optimal configurations of sampling locations $x_1, \ldots, x_n$. The term "spatial-simulated annealing" (SSA) finds its first appearance in the work of van Groenigen et al. (1999). Trujillo-Ventura and Ellis (1991) consider multi-objective sampling design optimization.

The local pivotal method (LPM) is another method that was introduced by Grafström et al. (2012). In this method, selecting spatially balanced samples with prescribed inclusion probabilities from a finite (large) population uses a sub-optimal implementation of the LPM. The local cube method (or doubly balanced sampling) selects doubly balanced samples with prescribed inclusion probabilities from a finite population. This method was introduced by Grafström and Tillé (2013). We note that none of these methods appeared in the literature considered balanced sampling in a spatio-temporal setting.

3 Balanced sampling for a spatio-temporal population

Balanced sampling can be used with two different inferential approaches, model-based (Royall and Herson 1973; Valliant et al. 2000) and design-based perspectives (Deville and Tillé 2004). In the model-based approach, inference is made on a statistical super population model and it may be performed by probability or non-probability samples. In this case a sample is balanced when the sample means of a set of auxiliary variables are equal to the known population means (Valliant et al. 2000). These auxiliary variables usually provide useful information related directly to the problem at hand, and they are often measured at particular latitude and longitude geographical coordinates along with time. The design-based method needs a sampling frame and uses a probability sample. In this case a sample is balanced when the Horvitz-Thompson sample estimates for the auxiliary variables are equal to their known population counterparts. The selection of a balanced sample generally improves the efficiency of the sampling estimates (Cochran 1977). There are several methods to select a balanced sample with a fixed sample size. The cube algorithm is the method for selecting balanced sampling with unequal inclusion probabilities (Deville and Tillé 2004).

Spatially-balanced sampling combines elements of simple random and systematic sampling. Locations are selected randomly, but are guaranteed to be spread over space in an attempt to maximize the spatial dependence among sample locations. Yates (1949) showed that a

sample of a response Y was balanced over an auxiliary variable Z, that is correlated with Y, if the values of Z (which are known in advance) are chosen so that the sample mean of the Z values is exactly equal to the true population mean of Z. Royall and Herson (1973) required the strict condition that the first several sample moments of Z exactly match the corresponding population moments. The intuition behind balancing is that by forcing the Z sample moments to match the population moments, we should get approximate balance over Y, and therefore a more precise sample. Royall and Herson (1973) showed that a balanced sample is optimal in some cases. They noted that an option between strict balancing and SRS is to partition the range of Z into quantiles, then pick one point in each quantile, and observe the corresponding Y. While such a sample will not be strictly balanced, it guarantees a good estimate of the distribution function of Z for every sample draw. Because of the correlation between Y and Z, one should also get a good estimate for Y.

Let the ancillary variable be located at a spatial site $s = (x, y) \in \mathrm{R}^2$ and time location $t \in \mathrm{R}$ in a spatio-temporal subregion $\mathrm{D} \subseteq \mathrm{R}^2 \times \mathrm{R}$. Then we define a sample to be spatio-temporally balanced if the spatial moments of the sample locations match the spatial moments of the population, and the temporal moments of the sample locations match the temporal moments of the population. The first two spatial moments are the center of gravity and the inertia. The center of gravity for a region D is given by the ordered triplet (μ_x, μ_y, μ_t), where μ_x is the central moment about the spatial y-axis and temporal t-axis given by

$$\mu_x = \int_{-\infty}^{\infty} x\vartheta_y(x)\tau_t(x)\mathrm{d}x, \tag{5}$$

where $\vartheta_y(x)$ and $\tau_t(x)$ are the extended cross-sections of D at the points y and t, respectively, given by

$$\vartheta_y(x) = \int_{-\infty}^{\infty} I_{D_x}(y)\mathrm{d}y, \tag{6}$$

$$\tau_t(x) = \int_{-\infty}^{\infty} I_{D_x}(t)\mathrm{d}t, \tag{7}$$

where $D_x = \{(x, ., .) \in D\}$ and $I_{D_x}(z)$ is an indicator function that is equal to 1 when z belongs to the domain D_x and it is 0 otherwise. Similarly, μ_y and μ_t can be derived in the same way. In particular,

$$\mu_t = \int_{-\infty}^{\infty} t\delta_x(t)\vartheta_y(t)\mathrm{d}t$$

with

$$\vartheta_y(t) = \int_{-\infty}^{\infty} I_{D_t}(y)\mathrm{d}y$$

and

$$\delta_x(t) = \int_{-\infty}^{\infty} I_{D_t}(x)\mathrm{d}x.$$

The second spatial moment is analogous to the covariance matrix, and measures the regularity of the shape of D, or of the point pattern formed by the sample points. Designs with some degree of spatio-temporal regularity or balance tend to be more efficient (i.e. yield responses that are less variable) for sampling natural resources than designs with no spatio-temporal structure. Spatio-temporal balance also ensures that there is minimal effect of spatial–temporal correlation on parameter estimates.

In this paper, we propose a two-step sampling method to obtain a spatio-temporally balanced sample from a target population. In the first step a stratified design that is balanced on the spatio-temporal variable of the strata is used. Indeed, a local cube method chooses the stratifications as samples that satisfy the balanced equations. In the second step a member of each stratification is chosen so that it has a maximum distance to the center of each stratified neighborhood in the three dimensions. So, we choose three near stratifications for each one, the first one from the spatial x-axis, the second one from the spatial y-axis, and the third one from the temporal axis. Note that for all three strata, the nearest stratum is considered as a neighbor stratum, and thus, for each selected stratum, a member of the stratum is selected having a maximum distance from all three nearest strata.

As an illustrative simple, one-dimensional example, the top plot of Fig. 1 shows such a stratification, with samples chosen as in the bottom plot.

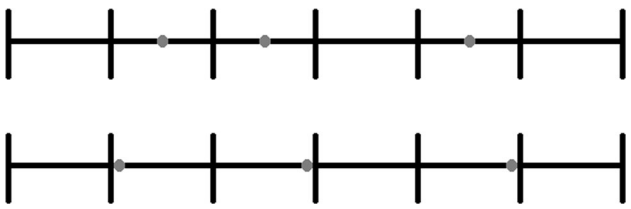

Fig. 1 Illustration of a two-step spatio-temporal design-based balanced sampling

4 Measuring the accuracy of the optimal sampling design

To measure the accuracy of the resulting sampling strategy, we used spatio-temporal kriging and accuracy measures to test the prediction coming from the selected sample.

4.1 Spatio-temporal kriging

Kriging is a generalized least-square regression technique that allows one to account for the spatio-temporal dependence between observations, as revealed by the covariogram, to perform spatio-temporal prediction.

Assume that the values of a random field $Z(.,.)$ have been observed on a set of nm spatio-temporal locations $\{(s_1,t_1),\ldots,(s_n,t_m)\}$, and that $Z(s,t)$ is a second-order stationary spatio-temporal random field, with a constant unknown mean $\mu(s,t)=\mu$, and a known covariance function $C(h,u)$. Let $\mathbf{C}=(C(s_0-s_1,t_0-t_1),\ldots,C(s_0-s_n,t_0-t_m))$ and $\boldsymbol{\Sigma}_{nm\times nm}=(C(s_i-s_j,t_i-t_j))$, then the prediction of the random field at a new spatio-temporal site (s_0,t_0) is given by the linear form

$$Z^*(s_0,t_0)=\sum_{i=1}^{n}\sum_{j=1}^{m}\lambda_{ij}Z(s_i,t_j), \tag{8}$$

where $\boldsymbol{\lambda}'=(\lambda_{11},\ldots,\lambda_{nm})=\left(\mathbf{C}+1\frac{(1-1'\boldsymbol{\Sigma}^{-1}\mathbf{C})}{1'\Sigma^{-1}1}\right)'\boldsymbol{\Sigma}^{-1}$, $\mathbf{1}$ is a vector of *nm* ones, and prime stands for the transpose of a vector (see Mateu and Müller 2013; Cressie and Wikle 2011).

4.2 Accuracy measures

To measure the accuracy of the resulting sampling strategy, we test the prediction coming from the selected sample by using ordinary kriging via the root-mean-squared error (RMSE) defined as

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{t=1}^{T}\sum_{s=1}^{S}(\hat{y}_{st}-y_{st})^2}{ST}}, \tag{9}$$

where the y_{st} and $\hat{y}_{st}$ are the true and predicted values at the *s*th location on the *t*th temporal instant, respectively. The smaller the RMSE, the more accurate the kriging prediction is. In terms of comparison under different setups and scenarios (as happens in the simulation study), we use a normalized root-mean-squared error (NRMSE), which is given by

$$\mathrm{NRMSE}=\sqrt{\frac{\sum_{t=1}^{T}\sum_{s=1}^{S}(\hat{y}_{st}-y_{st})^2}{\sum_{t=1}^{T}\sum_{s=1}^{S}(y_{st})^2}}. \tag{10}$$

As an alternative measure that evaluates the spread of the sample over space and time, we use the variance of the volume of the Voronoi tessellation. A Voronoi diagram is a partition of a plane into regions based on distances to points in a specific subset of the plane. That set of points (called seeds, sites, or generators) is specified beforehand, and for each seed there is a corresponding region consisting of all points closer to that seed than to any other. These regions are called Voronoi cells. The Voronoi diagram of a set of points is dual to its Delaunay triangulation. It is a diagram drawn by taking pairs of points that are close together and drawing a line that is equidistant between them and perpendicular to the line connecting them. That is, all points on the lines in the diagram are equidistant to the nearest two (or more) source points.

A 3d-Voronoi subdivision is not that hard to imagine. Consider two lonely points in a cube. A good way of dividing the cube is splitting it with the bisector plane. This plane is perpendicular to the line connecting the two points and it is placed exactly halfway between them (Fig. 2, left). There is no need to limit ourselves to two points. Once we can split the cube by a plane, we can repeat this as often as we like. Figure 2 (middle) shows and example with three points, and Fig. 2 (right) an example with ten points (Du and Wang 2005).

The volume of each Voronoi tessellation is approximated by using a Monte Carlo method, in which we generate a large number of random points in a cube. Smaller values of RMSE, of NRMSE, and of the variance of the volume of the Voronoi tessellation indicate a better performance of the sampling method in the spatio-temporal context.

5 Simulation study and real data analysis

We considered a variety of scenarios to show the performance of our proposed design-based balanced sampling in comparison with other sampling strategies. In the second part we considered recorded average daily temperatures at 617 stations across six European countries taken every day for a 1-year period.

5.1 Simulation study

We considered two cases for the population size ($N=4{,}000{,}000$, $N=125{,}000$) living in the rectangular space–time regions [0, 200] × [0, 200] × [0, 100] and [0, 50] × [0, 50] × [0, 50], respectively. We also considered isotropic and anisotropic cases in combination with homogenous (no present trend) and inhomogeneous (with a particular polynomial trend) cases. Finally, we chose a

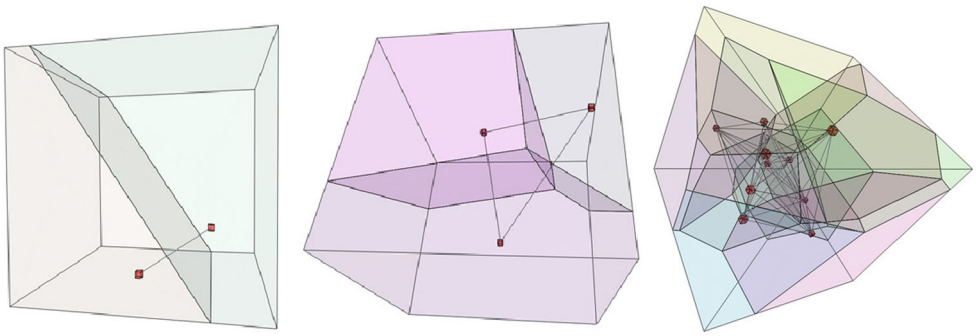

Fig. 2 Voronoi tessellation in a three-dimensional space

number of spatial and temporal covariance models giving rise to several separable spatio-temporal covariance models. A non-separable covariance model was also considered. The whole set of combined scenarios is shown in Table 1.

For completeness, we briefly comment on the covariance models used [further insight can be found in Christakos (2005)]. The stationary, isotropic exponential covariance model is a function that only depends on the spatial distance $r \geq 0$ between any two points, and is given by

$$\mathrm{C}(r) = \exp(-r). \tag{11}$$

The Matérn isotropic correlation function is defined by Handcock and Stein (1993) as follows

$$\mathrm{K}_\theta(r) = \frac{1}{2^{\theta_2-1}\Gamma(\theta_2)}\left(\frac{r}{\theta_1'}\right)^{\theta_2}\mathcal{K}_{\theta_2}\left(\frac{r}{\theta_1'}\right), \tag{12}$$

where r is the separation lag, $\theta_1' = \theta_1/(2\sqrt{\theta_2})$, $\mathcal{K}_{\theta_2}$ is a modified Bessel function of the second kind of order θ_2, Γ is the gamma function, $\theta_1 > 0$ is a scale parameter controlling the range of correlation, and $\theta_2 > 0$ is the smoothness parameter controlling the smoothness of the random field. The spatial isotropic covariance function is then, for $\sigma \geq 0$,

$$\mathrm{C}(r) = \sigma^2\mathrm{K}_\theta(r), \tag{13}$$

where σ^2 stands for the overall variance.

As temporal covariance functions we use the Cauchy model (Gneiting and Schlather 2004)

$$\mathrm{C}(t) = \sigma^2\left(1 + (\theta t)^\phi\right)^{-v}, \tag{14}$$

where $t > 0$, $\phi \in (0, 2]$, $v > 0$, $\theta > 0$, and the stable model with a covariance function of the form

$$\mathrm{C}(t) = \exp(-t^\alpha), \tag{15}$$

with $\alpha \in (0, 2]$.

We finally consider the non-separable spatio-temporal covariance function of the Gneiting class (Gneiting 2001) given by

$$\mathrm{C}(r,t) = \frac{\sigma^2}{\psi(t^2)^{d/2}}\mathrm{f}\left(\frac{r^2}{\psi(t^2)}\right), \tag{16}$$

where $\mathrm{f}(x)$ is a completely monotone function on $[0, \infty)$, d denotes the dimension of the random field (the models can be used for any dimension), $\psi(x)$ is positive with a

Table 1 Trend and covariance models for different scenarios

Scenario	Isotropy	Trend	Spatial cov.	Temporal cov.	Non-separable cov.
1	Yes	–	Exponential	Cauchy	
2	Yes	–	Matérn	Stable	
3	Yes	–			Gneiting
4	Yes	$1 + x - y + t$	Exponential	Cauchy	
5	Yes	$1 + x - y + t$	Matérn	Stable	
6	Yes	$1 + x - y + t$			Gneiting
7	No	–	Exponential	Cauchy	
8	No	–	Matérn	Stable	
9	No	–			Gneiting
10	No	$1 + x - y + t$	Exponential	Cauchy	
11	No	$1 + x - y + t$	Matérn	Stable	
12	No	$1 + x - y + t$			Gneiting

The names of the covariance functions are referred to in the text

completely monotone derivative on $[0, \infty)$, and $\sigma^2 > 0$ is the variance.

For the anisotropic cases, we used a 3×3 matrix

$$\begin{bmatrix} \cos(a)\cos(L) & \sin(a)\cos(L) & \sin(L) \\ -\sin(a) & \cos(a) & 0 \\ -\cos(a)\sin(L) & -\sin(a)\sin(L) & \cos(L) \end{bmatrix}$$

where a stands for the angle in space, and L for the angle in time. In particular, we fixed $a = \pi/4$ and $L = \pi/8$.

For the case of the population with size $N = 4{,}000{,}000$, we simulated a random field of this size using the corresponding covariance structure within the spatial–temporal cube $[0, 200] \times [0, 200] \times [0, 100]$, and proceeded as follows: (a) we first divided the total volume into 32,000 parts of small cubes of $5 \times 5 \times 5$ points; (b) we then selected a sample using the cube method from these 32,000 blocks; (c) for each selected block we chose a member that had the greatest distance to selected centers of other blocks.

Out of the $N = 4{,}000{,}000$, we selected four sample sizes, $n = 500$, 1000, 2000 and 4000, and used the sampled data to predict the observation at the locations of the rest of the population to obtain a value of NRMSE.

For sampling purposes, variables such as longitude, latitude and time have been used as ancillary variables, and the sampling has been conducted in a manner that the sampling result is balanced within these considered variables.

This procedure was repeated 100 times, and we averaged the NRMSE values. For each simulation, we also calculated the variance of the volume of the Voronoi tessellation, providing an average out of the 100 simulations. Figure 3 shows the average value of the variance of the volume of the Voronoi tessellation for each sample size and for each of the four compared methods. Our method

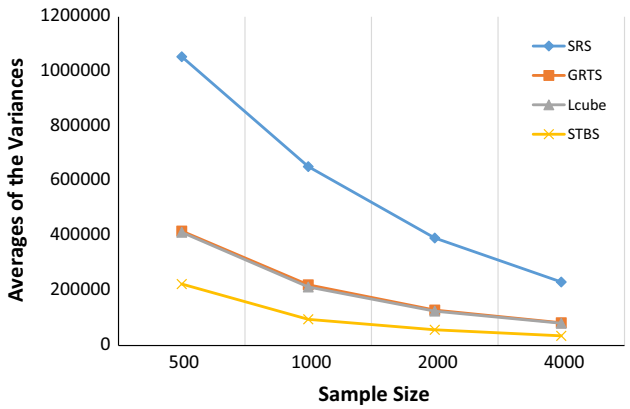

Fig. 3 Averages (over 100 simulations) of the variances of the Voronoi tessellation volume for each sample size and for each of the four compared sampling methods (i.e. *SRS* the simple random sampling, *GRTS* the generalized random-tessellation stratified, *Lcube* the local cube method; and *STBS* the spatio-temporal balanced sampling)

(named as STBS, spatio-temporal design-based balanced sampling) clearly provides a much larger reduction in the variance of the volume, indicating a better spread of the sample over the space and time domains, in addition to a better performance in the sampling strategy.

A similar procedure was followed in the case $N = 125{,}000$, for which we selected sample sizes of $n = 100$, 200, 400 and 800. Tables 2 and 3 show the averaged NRMSE values for the whole set of considered scenarios for the two cases of the population size. We clearly note that NRMSE decreases with the sample size, and in all cases, our method STBS provides the lowest values of NRMSE. In general, the largest values of NRMSE are obtained under anisotropic, and non-stationary cases. Comparing isotropic versus anisotropic cases (the first six vs the second six scenarios), we note that when using the SRS method SRS and the STBS there were not significant differences in terms of NRMSE values. However, values of NRMSE were larger under anisotropic cases when considering the GRTS method and the local cube method (Lcube). In terms of non-stationary cases with the presence of a trend, GRTS and Lcube methods are sensitive to such trend providing larger values of NRMSE. However, the performance of prediction under the STBS method is not affected by a trend as this method provides similar values of NRMSE for any of these cases. Indeed, spatio-temporal design-based balanced sampling is even more efficient under the presence of a trend and anisotropic effects. The effect of separability in the covariance structure of the spatio-temporal data does not seem to affect the performance of each method. We also highlight the following fact. We considered two population sizes $N = 4{,}000{,}000$ and $N = 125{,}000$. The largest sample size in the former case was $n = 4000$, while it was $n = 800$ in the latter. Thus the corresponding ratios of the largest sample size to the population size were 0.001 and 0.0064, six times largest in the $N = 125{,}000$ case. This notably results in a reduction of that magnitude of the averaged NRMSE in Table 3 compared to those values in Table 2.

5.2 Real data analysis

We considered recorded average daily temperatures at 617 stations across six European countries (France, Germany, Belgium, Luxembourg, the Netherlands and Switzerland) taken every day for a 1-year period (January, 2014–January, 2015). See the locations of the stations in Fig. 4. This data set is available on the web site "wunderground.com". The color of the stations indicates variances of the temperature over the studied period.

Table 2 Averaged NRMSE for the different scenarios considered and $N = 4{,}000{,}000$

n	Model	Scenario											
		1	2	3	4	5	6	7	8	9	10	11	12
500	SRS	0.564	0.567	0.550	0.570	0.561	0.553	0.573	0.568	0.585	0.563	0.569	0.558
	GRTS	0.312	0.325	0.322	0.314	0.323	0.328	0.320	0.325	0.556	0.331	0.313	0.324
	Lcube	0.287	0.303	0.291	0.279	0.293	0.293	0.294	0.296	0.316	0.301	0.302	0.291
	STBS	0.267	0.267	0.281	0.263	0.269	0.269	0.288	0.285	0.290	0.271	0.295	0.272
1000	SRS	0.305	0.315	0.305	0.308	0.302	0.294	0.306	0.311	0.228	0.293	0.295	0.303
	GRTS	0.234	0.246	0.236	0.234	0.242	0.246	0.243	0.242	0.310	0.250	0.248	0.242
	Lcube	0.231	0.227	0.232	0.235	0.237	0.241	0.234	0.235	0.242	0.232	0.238	0.240
	STBS	0.222	0.233	0.218	0.224	0.238	0.244	0.234	0.228	0.232	0.241	0.235	0.242
2000	SRS	0.265	0.272	0.266	0.261	0.268	0.262	0.272	0.269	0.204	0.273	0.271	0.266
	GRTS	0.211	0.211	0.207	0.203	0.212	0.214	0.217	0.216	0.264	0.211	0.218	0.216
	Lcube	0.218	0.215	0.218	0.217	0.219	0.219	0.219	0.218	0.217	0.219	0.224	0.225
	STBS	0.197	0.198	0.196	0.194	0.213	0.212	0.201	0.204	0.222	0.210	0.198	0.200
4000	SRS	0.215	0.213	0.212	0.218	0.214	0.214	0.213	0.212	0.155	0.215	0.212	0.216
	GRTS	0.164	0.165	0.160	0.165	0.165	0.166	0.167	0.167	0.213	0.169	0.169	0.170
	Lcube	0.161	0.163	0.157	0.157	0.161	0.164	0.162	0.164	0.167	0.163	0.164	0.165
	STBS	0.154	0.151	0.151	0.157	0.156	0.155	0.161	0.155	0.164	0.155	0.156	0.161

Table 3 Averaged NRMSE for the different scenarios considered and $N = 125{,}000$

n	Model	Scenario											
		1	2	3	4	5	6	7	8	9	10	11	12
100	SRS	0.092	0.095	0.093	0.092	0.096	0.092	0.094	0.096	0.092	0.095	0.094	0.092
	GRTS	0.051	0.053	0.055	0.056	0.053	0.054	0.055	0.054	0.053	0.053	0.050	0.051
	Lcube	0.048	0.050	0.049	0.050	0.050	0.049	0.050	0.049	0.049	0.046	0.047	0.048
	STBS	0.042	0.045	0.045	0.048	0.048	0.047	0.049	0.048	0.046	0.044	0.043	0.042
200	SRS	0.049	0.051	0.052	0.049	0.051	0.050	0.052	0.052	0.052	0.053	0.052	0.049
	GRTS	0.040	0.041	0.041	0.040	0.039	0.039	0.039	0.041	0.040	0.038	0.039	0.040
	Lcube	0.037	0.039	0.039	0.040	0.039	0.039	0.040	0.040	0.040	0.039	0.037	0.037
	STBS	0.036	0.040	0.040	0.041	0.038	0.040	0.039	0.038	0.041	0.039	0.036	0.036
400	SRS	0.044	0.045	0.045	0.043	0.045	0.045	0.043	0.044	0.045	0.044	0.046	0.044
	GRTS	0.035	0.037	0.037	0.036	0.036	0.036	0.037	0.035	0.036	0.034	0.035	0.035
	Lcube	0.036	0.038	0.037	0.038	0.037	0.037	0.037	0.037	0.038	0.036	0.036	0.036
	STBS	0.033	0.033	0.033	0.034	0.033	0.034	0.034	0.033	0.033	0.033	0.033	0.033
800	SRS	0.036	0.036	0.037	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036	0.036
	GRTS	0.028	0.028	0.028	0.028	0.028	0.028	0.027	0.027	0.028	0.027	0.028	0.028
	Lcube	0.027	0.028	0.028	0.027	0.027	0.027	0.027	0.027	0.027	0.027	0.028	0.027
	STBS	0.026	0.026	0.027	0.026	0.027	0.027	0.027	0.026	0.027	0.026	0.025	0.026

We fitted three covariance models: (a) a separable model "Exponential–Cauchy", (b) a separable "Matérn–Stable" model, and (c) a non-separable Gneiting model. We used a marginal likelihood procedure to estimate the parameters, with a 7-day temporal lag, and a 100 km spatial lag. The non-separable Gneiting model provided the best results in terms of prediction. Figure 5 shows the 3d spatio-temporal fitted Gneiting covariogram, and the

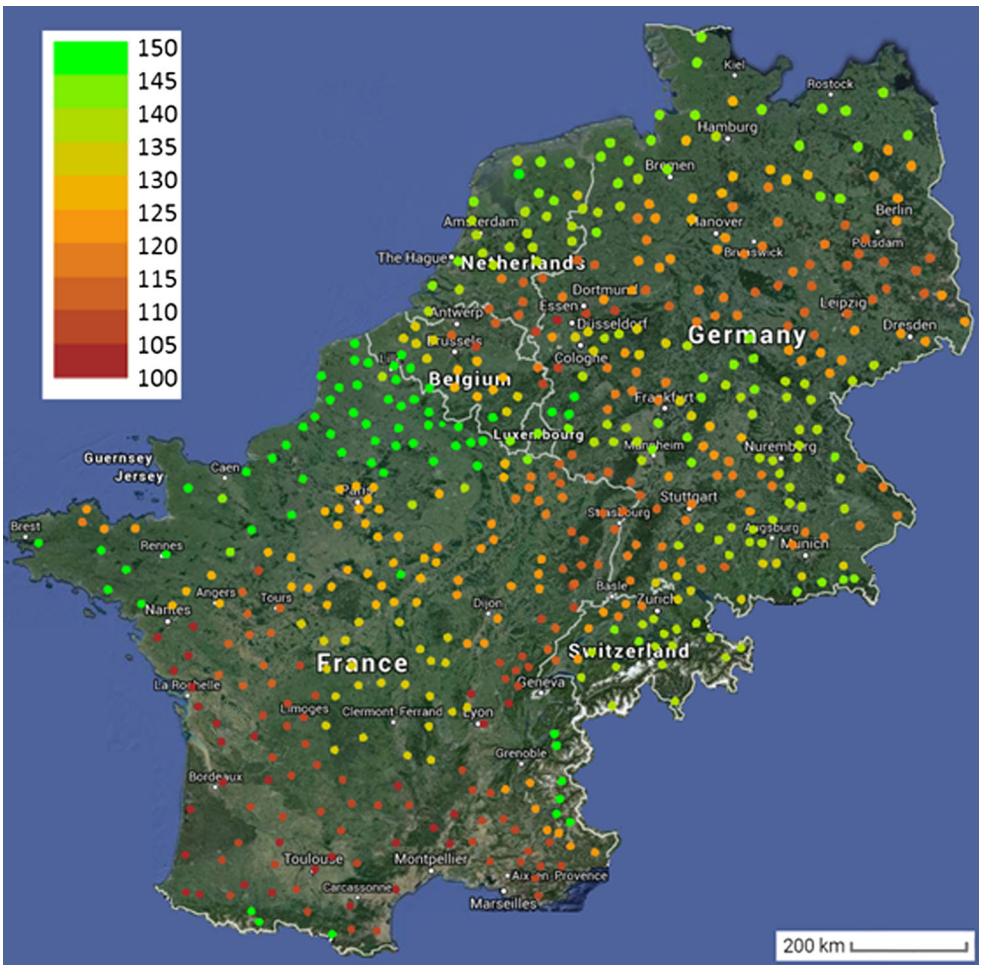

Fig. 4 Locations of the weather stations in six European countries

corresponding fitted variogram. The empirical spatio-temporal variogram is also shown.

Table 4 shows the fitted parameters of the spatio-temporal covariance model based on days as the temporal unit, and 10^{-2} km as the spatial one. We note that the separable parameter takes the value $\text{Sep} = 0.5763$, indicating a strong interaction between space and time.

Finally, we computed the NRMSE for several sample sizes, and we used the same four sampling procedures as in the simulation study. Note that we have 617 stations taking temperature values during 365 days per year. So we have 225,205 points for sampling. Out of these, we selected samples sizes of $n = 500$, 1000, 2000 and 4000. Using these samples, we used the non-separable spatio-temporal Gneiting model for kriging prediction. The results are shown in Table 5 which shows comparisons among the SRS, GRTS, Lcube, and our approach, the STBS method. The NRMSE values indicate that the spatio-temporal balanced sampling design has a better kriging performance for the average temperatures in these six countries compared to the other methods, confirming the outperformance of our proposed method.

6 Conclusions and discussion

We have introduced a simple two-step method that performs spatio-temporal balanced sampling in a design-based approach. The presence of spatio-temporal trends and/or anisotropic effects in the variable of interest makes our method even more competitive with respect to other existing methods. The spread of the sample over the population is controlled by using the volume of the corresponding three-dimensional Voronoi tessellation, and the spatio-temporal design-based balanced sampling strategy provides the best spread. So, we have presented a sampling strategy that outperforms any other adapted strategy for spatio-temporal data.

We performed a simulation study comparing the performance of our proposed method with other three sampling methods. It is shown that our method (the STBS method) outperformed its competitors in several fronts. It provided the lowest spatio-temporal prediction root mean square errors. But in addition, the STBS method provided balanced samples on the auxiliary variables, and located the samples homogeneously spread all over the region

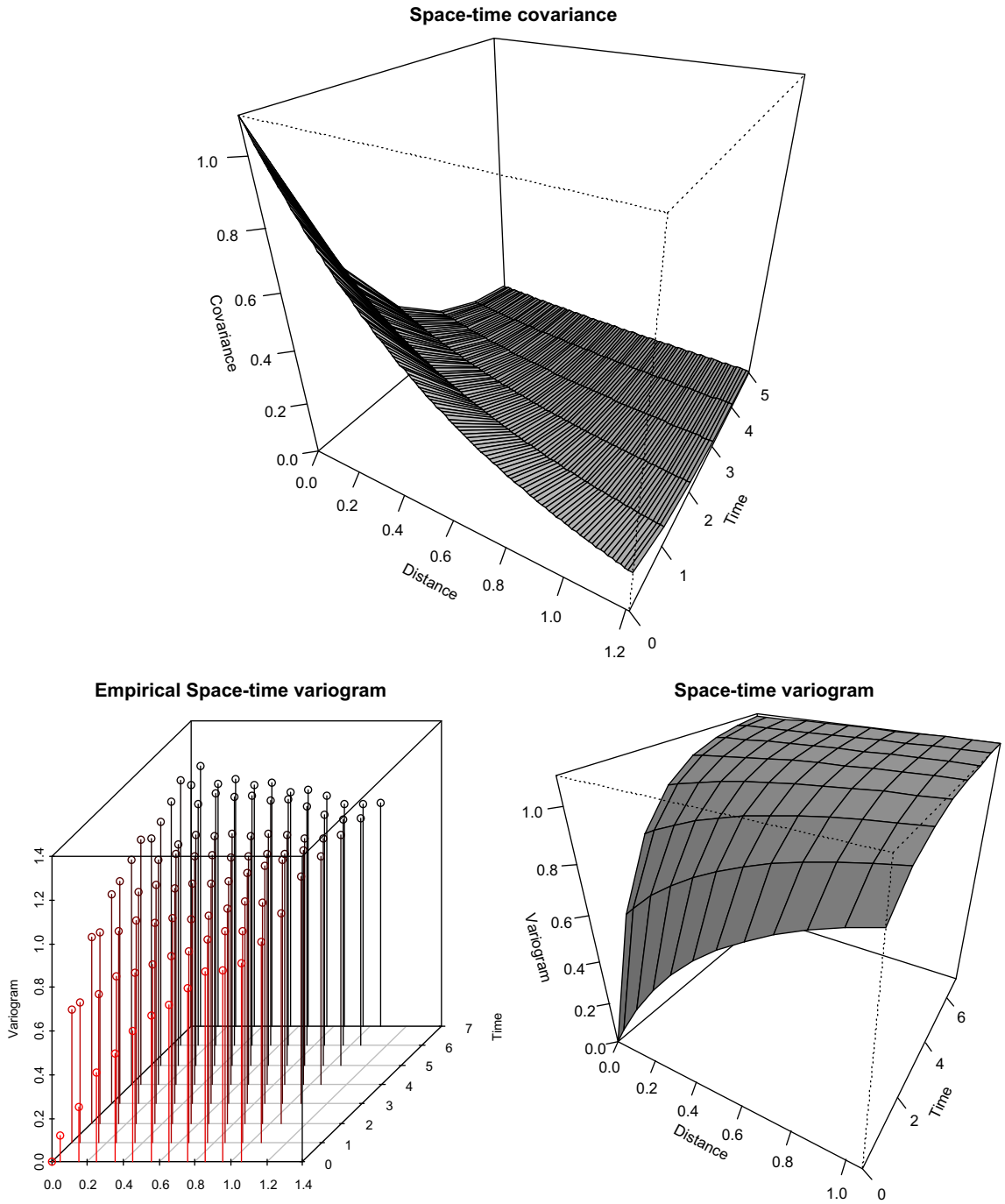

Fig. 5 Empirical spatio-temporal variogram (*bottom-left panel*) and the fitted non-separable Gneiting model (*top* and *bottom-right panels*). The temporal units are days, and the spatial units are shown with the proportion of 1 unit = 100 km)

Table 4 Parameter estimates of the spatio-temporal Gneiting covariance model

Nugget	Sill	Temporal scale	Spatial scale	Power in time	Power in space	Sep
0.1275	0.3576	0.3850	0.7245	0.3187	0.6374	0.5763

preserving the spatio-temporal structure of the data. Note that the SRS method provides samples that are not balanced on the auxiliary variables.

Globally speaking, the SRS method usually reports the largest NRMSE values. The reason can be that samples may be selected near each other, and it may reduce the

Table 5 Averaged temperatures in Europe: NRMSE values for several sample sizes and methods

n	SRS	GRTS	Lcube	STBS
500	12.532	8.643	8.274	8.145
1000	9.324	7.834	7.346	6.935
2000	7.245	6.013	5.854	5.634
4000	5.935	4.724	4.583	4.483

A non-separable spatio-temporal Gneiting model is used

gained information from the sample. Therefore, it might be a wrong method for spatio-temporal sampling. Among all presented methods, GRTS is an appropriate method to stratify various spatial and temporal parts. However, it is not completely successful to place samples in the optimal locations. Although the GRTS method reports lower NRMSE values compared to the SRS method, it is not better than the Lcube and STBS ones. Lcube method can spread samples suitably all over the region and it causes a reduction in NRMSE. An important issue is that the Lcube method still makes a large error in the marginal spatial and temporal dimensions, and it cannot locate samples in optimal regions. Finally, the STBS method can reduce NRMSE by applying balanced sampling and creates space in final sample locations. The reason can be based on sampling from spatial and temporal margins creating enough space between samples.

We note here that there has been a growing literature in the field of optimal spatial sampling designs for environmental applications and soil sciences (Brus and Heuvelink 2007; Dobbie et al. 2008; Delmelle and Goovaerts 2009). However these approaches have mainly focussed on the spatial structure of the data. Some of these approaches are more focussed on the computational side, providing stochastic search algorithms, evolutionary genetic algorithms, and deterministic ones for optimizing the investigated design criteria. Again all published papers deal with the spatial component. More recently, Grafström and Tillé (2013) introduce the local cube method (or doubly balanced sampling) which selects doubly balanced samples with prescribed inclusion probabilities from a finite population.

Our method is completely new as works for spatio-temporal data, and represents a step forward into the optimal design for spatio-temporal structures. We have provided some comparisons with the other existing methods in space, and have shown that our proposal is clearly competitive even only in the space against other existing methods, with the additional gain of dealing also with space–time data.

The STBS method depends on the inclusion probabilities which can be clearly refined in terms of the potential auxiliary variables. A further step not considered here could be calculating these probabilities in terms of such auxiliary variables. Another important point is that we are proposing a method for selection of sample points in a spatio-temporal context, but we give no answer to the optimal sample size. This question still remains open. Finally, we have assumed that the region of interest is a subregion of the plane. A latent open question here is how our method can be adapted to those situations where the support of the spatio-temporal data is not a continuous planar region but a network. Sampling in networks is again an open area for research.

Acknowledgements The authors are thankful to the referees for their many helpful comments that greatly improved this paper. We also wish to acknowledge for the support from Ordered and Spatial Data Center of Excellence of the Ferdowsi University of Mashhad.

References

Aarts EH, Korst J (1989) Simulated annealing and boltzman machines. Wiley, New York

Brown PJ, Le ND, Zidek JV (1994) Multivariate spatial interpolation and exposure to air pollutants. Can J Stat 2:489–509

Brus DJ, Heuvelink GBM (2007) Optimization of sample patterns for universal kriging of environmental variables. Geoderma 138:86–95

Christakos G (2005) Random field models in earth sciences. Dover, New York

Cochran WG (1977) Sampling techniques, 3rd edn. Wiley, New York

Cox LA Jr (1999) Adaptive spatial sampling of contaminated soils. Risk Anal 19:1059–1069

Cressie N, Wikle CK (2011) Statistics for spatio-temporal data. Wiley, Hoboken

Delmelle E, Goovaerts P (2009) Second-phase sampling designs for non-stationary spatial variables. Geoderma 153:205–216

Deville JC, Tillé Y (2004) Efficient balanced sampling: the cube method. Biometrika 91:893–912

Dobbie MJ, Henderson BL, Stevens DL Jr (2008) Sparse sampling: spatial design for monitoring stream networks. Stat Surv 2:113–153

Du Q, Wang D (2005) The optimal centroidal voronoi tessellations and the Gersho's conjecture in the three-dimensional space. Comput Ind Eng 49:1355–1373

Fuentes M, Chaudhuri A, Holland DM (2007) Bayesian entropy for spatial sampling design of environmental data. Environ Ecol Stat 14:323–340

Gneiting T (2001) Criteria of Pólya type for radial positive definite functions. Proc Am Math Soc 129:2309–2318

Gneiting T, Schlather M (2004) Stochastic models that separate fractal dimension and the Hurst effect. SIAM Rev 46:269–282

Goovaerts P (1997) Geostatistics for natural resources evaluation. Oxford University Press, Oxford

Grafström A, Tillé Y (2013) Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. Environmetrics 24:120–131

Grafström A, Lundström NLP, Schelin L (2012) Spatially balanced sampling through the pivotal method. Biometrics 68:514–520

Haining RP (2003) Spatial data analysis: theory and practice. Cambridge University Press, Cambridge

Handcock MS, Stein ML (1993) A Bayesian analysis of kriging. Technometrics 35:403–410

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47:663–685

Li LF, Wang JF, Cao ZD, Feng XL, Zhang LL, Zhong ES (2008) An information-fusion method to regionalize spatial heterogeneity for improving the accuracy of spatial sampling estimation. Stoch Environ Res Risk A 22:689–704

Lister AJ, Scott CT (2009) Use of space-filling curves to select sample locations in natural resource monitoring studies. Environ Model Assess 149:71–80

Mateu J, Müller WG (2013) Spatio-temporal design. Advances in efficient data acquisition. Wiley, Chichester

Matheron G (1971) The theory of regionalized variables and its application. Ecole Nationale Supérieure des Mines de Paris, France

Müller WG (2005) A comparison of spatial design methods for correlated observations. Environmetrics 16:495–505

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

Müller P, Sanso B, De Iorio M (2004) Optimal Bayesian design by inhomogeneous Markov chain simulation. J Am Stat Assoc 99:788–798

Neyman J (1934) On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. J R Stat Soc Ser B Stat Methodol 97:558–606

Rogerson PA, Delmelle E, Batta R, Akella M, Blatt A, Wilson G (2004) Optimal sampling design for variables with varying spatial importance. Geogr Anal 36:177–194

Royall RM, Herson J (1973) Robust estimation in finite populations I. J Am Stat Assoc 68:880–889

Spöck G, Pilz J (2010) Spatial sampling design and covariance-robust minimax prediction based on convex design ideas. Stoch Environ Res Risk A 24:463–482

Stein A, Ettema C (2003) An overview of spatial sampling procedures and experimental design of spatial studies for ecosystem comparisons. Agric Ecosyst Environ 94:31–47

Stevens DL Jr, Olsen AR (2004) Spatially-balanced sampling of natural resources. J Am Stat Assoc 99:262–278

Tillé Y (2006) Sampling algorithms. Spinger, New York

Trujillo-Ventura A, Ellis JH (1991) Multiobjective air pollution monitoring network design. Atmos Environ 25:469–479

Valliant R, Dorfman AH, Royall RM (2000) Finite population sampling and inference: a prediction approach. Wiley, New York

van Groenigen JW, Siderius W, Stein A (1999) Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma 87:239–259

Wang JF, Haining RP, Cao ZD (2010) Sample surveying to estimate the mean of a heterogeneous surface: reducing the error variance through zoning. Int J Geogr Inf Sci 24:523–543

Yates F (1949) Sampling methods for censuses and surveys. Griffin, London

Zhu Z, Stein ML (2006) Spatial sampling design for prediction with estimated parameters. J Agric Biol Environ Stat 11:24–49

Zimmerman DL, Homer KE (1991) A network design criterion for estimating selected attributes of the semivariogram. Environmetrics 2:425–441