Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche
Dottorato di ricerca in

Metodologia Statistica per la Ricerca Scientifica
XXV ciclo

Employing Distances in Design-based Spatial Estimation

Alessandro Vagheggini

Dipartimento di Scienze Statistiche "Paolo Fortunati"
Gennaio 2013

Alma Mater Studiorum - Università di Bologna

Scuola di Dottorato in Scienze Economiche e Statistiche
Dottorato di ricerca in

Metodologia Statistica per la Ricerca Scientifica
XXV ciclo

Employing Distances in Design-based Spatial Estimation

Alessandro Vagheggini

Coordinatore:
Prof.ssa Angela Montanari

Tutor:
Prof.ssa Daniela Cocchi

Co-Tutor:
Dott.ssa Francesca Bruno

Settore Disciplinare: SECS-S/01
Settore Concorsuale: 13/D1

Dipartimento di Scienze Statistiche "Paolo Fortunati"
Gennaio 2013

There are three kinds of lies:
lies, damned lies, and statistics.

attributed to Benjamin Disraeli

Abstract

In the last couple of decades we assisted to a reappraisal of spatial design-based techniques. Usually the spatial information regarding the spatial location of the individuals of a population has been used to develop efficient sampling designs.

This thesis aims at offering a new technique for both inference on individual values and global population values able to employ the spatial information available before sampling at estimation level by rewriting a deterministic interpolator under a design-based framework. The achieved point estimator of the individual values is treated both in the case of finite spatial populations and continuous spatial domains, while the theory on the estimator of the population global value covers the finite population case only.

A fairly broad simulation study compares the results of the point estimator with the simple random sampling without replacement estimator in predictive form and the kriging, which is the benchmark technique for inference on spatial data. The Monte Carlo experiment is carried out on populations generated according to different superpopulation methods in order to manage different aspects of the spatial structure. The simulation outcomes point out that the proposed point estimator has almost the same behaviour as the kriging predictor regardless of the parameters adopted for generating the populations, especially for low sampling fractions. Moreover, the use of the spatial information improves substantially design-based spatial inference on individual values.

Keywords: spatial estimation, design-based inference, spatial sampling, ratio estimator.

Acknowledgements

This thesis is the synthesis of almost five years of work. During this period many people, with their involvement, have helped me to achieve this result.

I would like to give my most sincere thanks to the two people who have supported and pushed me to give my best: my supervisor Professor Daniela Cocchi and my co-supervisor Doctor Francesca Bruno. Without their guidance and advices it would have been harder to come to the end of this work and of the three years of my PhD. I also would like to thank the whole working group I have been part of: Professor Carlo Trivisano, Doctor Fedele Greco, Professor Michele Scagliarini and all my fellow students who along with me have taken part at the meetings.

I would also like to thank all the students that have shared with me this adventure.

An important part of this dissertation is the result of the period I spent at the JPSM department of the University of Maryland, College Park. I would like to thank Professor Partha Lahiri for inviting me and giving me the opportunity of share my thoughts with him.

During my early years as a university student Professor Lucio Barabesi has played an important role in my education. I also thank him for giving me the opportunity to work on this topic which has revealed itself interesting beyond all expectations.

In this years an often underestimated role has been played by all my friends, the ones with whom I grew up, the ones I met later and the ones from my Contrada, my second family.

A great thank goes to all the people I met on and off the rugby field,

everyone of them has taught me something. Without this sport and the people involved, I definitely would have had a different approach to life.

But most of all, I feel the need to thank my parents and my family: they have always been supporting and advising me. I am who I am because you have always been there when I needed you.

Sincerely,
Alessandro

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the last couple of decades, spatial statistics has experienced a rapid growth and has become one of the most interesting research field. The object of the inference is often represented by the individual values at unsampled locations. This problem has often been addressed through a model-based approach. Indeed, it was thought that design-based techniques were inappropriate since they could not capture any spatial dependence. However in the last few years we witnessed a rediscovery of design-based techniques for inference on individual values.

1.1 Motivations

From the review of the literature on spatial design-based techniques it emerged that spatial information is usually used in order to obtain more efficient sampling designs but almost never to infer on individual values. This aspect is strictly related to the fact that the main objective of the design-approach is the inference on global values (i.e. population total or mean). However, in literature (see for example Bolfarine and Zacks, 1992) can be found predictive formulation of the usual global estimators.

Brus and de Gruijter (1997) offer a deep insight on the misconceptions that have always accompanied design-based spatial estimation of values at unsampled locations. Moreover, they propose an estimator for the value of any subarea the domain has been divided in: the sample mean of the values

belonging to same subarea is assigned to all the points in it. Thus, their estimator is not able to replicate the changes at different locations typical of any spatial dataset nor to replicate the observed values at sampled locations.

The development of an individual design-based spatial estimator arose from the lack of those in the literature. By observing that labels in a spatial domain corresponds to the locations coordinates, we propose to use this information at estimation level in order to arrange pre-sampling weights according to the design-based paradigm. In this way the use of spatial information is merged with inference on individual values under a design-based approach. Inference on population global values is treated as the sum of the observed values plus the sum of the estimated values.

This thesis tries to answer whether the design-based estimator is suitable for inference both *(1)* on individual values and *(2)* on population global values. Moreover, through an extensive simulation study, *(3)* we investigate the conditions favourable to their application. In order to make the theory complete, several variance estimators are proposed whose characteristics are studied with Monte Carlo experiments.

1.2 Structure of the thesis

The thesis is organized as follows. Chapter 2 provides a historical excursus on the statistical spatial techniques and gives an insight on the basic definitions that will be needed in the following chapters.

Chapter 3 offers a review of the kriging: a model-based technique widely used for inference on individual values belonging to a spatial domain. At the beginning of the chapter the superpopulation model is introduced and its properties formalised. Then the parametric functions managing the spatial dependence and the estimation of their parameters are presented. The prediction of the individual values through the use of these estimated functions is treated in the following section. Finally, the practical applications of the kriging are explained.

Chapter 4 covers the estimator we propose and its properties. After the section addressing the main ideas used for its development the point

estimator is presented in the case of a finite spatial population. Then, it is extended to a continuous spatial domain. Finally, the estimation of the population total is considered as the sum of the estimated values.

A chapter providing the results of individual and global estimations on spatial populations generated according to different superpopulation models completes the thesis.

Chapter 2

Analysis of spatial data

In the last few years, the statistical analysis of spatial data has known a quick growth. An increasing number of application fields need a powerful and reliable tool upon which other analyses can be carried out. Despite nowadays its wide application is generally reknown, spatial statistics has struggled to be fully understood and widely accepted as a statistical methodology.

2.1 A historical excursus of of spatial analysis

Following Webster and Oliver (2007) the first records of what can be defined spatial statistics began to appear in the early 1910s, although probabilistic problems which can be related to this topic had been raised before (see Gelfand et al., 2010). In their 1911 paper, Mercer and Hall studied the uniformity of the yield of small crop plots at the Rothamsted Experimental Station (now known as Rothamsted Research) and noted that "plots similarly treated [..] yield considerably different results, even when the soil appears to be uniform and the conditions under which the experiment is conducted are carefully designed to reduce errors in weighing and measurement". Gosset (aka Sudent) wrote an appendix to the same paper where he recognized that both autocorrelation and a complete random effect were accountable for the plots variability. In a 1910 letter to Karl Pearson (Pearson, 1990), the very Student wrote that it would have been interesting to "work out the law according to which the correlation is likely to weaken with increase of unit".

Both the paper by Mercer and Hall and the letter by Student showed several embryonic ideas which nowadays are basic to modern geostatistics: spatial dependence, modellization and correlation range.

In his years at the Rothamsted Experimental Station (1919-1933), Fisher mostly dealt with the development of statistical inference for agricultural field trials data. He also detected a sort of spatial variation in his data; however, he decided to remove it through the use of blocking (i.e. design of experiment). The variation effects were then estimated by the analysis of variance (Fisher, 1966). A quite common criticism to Fisher's works is that proposing the use of the design of experiment techniques he might have delayed the understanding of spatial variation. For a deeper look at Student's and Fisher's ideas, see Gelfand et al. (2010).

In his work about how to describe the weather variation and its forecasting, Kolomogorov (1941b) achieved some interesting results. Among other things, he detected spatial correlation and explicitly wrote an analytic function able to represent such dependence. Furthermore he developed a method to interpolate such function using optimal weights (Cressie, 1990). Unfortunately his work never achieved much fame and was forgotten.

Likely, the most important breakthroughs in modern geostatistics were made in the 1950s and the early 1960s, when works in different discipline gave it a substantial boost. In his doctoral dissertation about forestry efficient sampling, Matérn (1960, reprinted as Matérn, 1986) detected the effects of spatial correlation. According to Guttorp and Gneiting (2006), he can be considered the first to write down a fully parametric class of models for describing spatial correlation. This family of functions, that now brings his name thanks to Handcock and Stein (1993), has some appealing properties that has made it the most used model for spatial correlation (see Subsection 3.2.2; Cressie, 1993; Diggle and Ribeiro, 2007).

Another fundamental boost was the one given by South African mining engineer Krige (1951). He figured out how to improve estimation of ore grades in block mining by taking into account the grades in the neighbouring blocks. Through the use of empirical methods he developed a way to determine ore grades distribution starting from a sample and using the spatial autocorrelation between its elements. Krige's method had become widely

used in gold mines.

Even if Krige's work is considered a corner stone in geostatistics, the formulation of a statistical spatial predictor was derived at the Paris School of Mines in Fontainebleau. There, in his doctoral dissertation, Matheron (1962) proposed a theory for optimal spatial linear prediction which takes into account the spatial correlation, given a finite sample of locations where ore are extracted. Matheron named his technique *kriging*, proving how important Krige's work had been. As Matheron was working on mining engineering, Gandin (1963) independently developed quite the same ideas for meteorology forecasting in the Soviet Union. In the 1980s, it became clear that these techniques are able to capture spatial correlation and to use it in order to predict the value of the variable under study at unobserved locations using a model-based approach.

Design-based techniques have been considered not suitable for spatial inference for quite a long time. However, in the last twenty years a reappraisal of the methods of sampling theory has been promoted (Cox et al., 1997; Stehman, 2000). The reasons of this presumed inadequacy have to be sought in some misunderstandings. First of all, under a design-based approach, the spatial dependency of the sampled elements can be avoided since the sampling design randomizes the locations, not the value of the variable under study (de Gruijter and ter Braak, 1990). It follows directly from probability theory (Parzen, 1960) that choosing a sampling design which independently select the locations lead to the independence of the data. In this sense the design-based approach "creates independence through randomization" (Brus and de Gruijter, 1997). As a result of this reappraisal some new design-based spatial techniques for inference on individual values have been developed. Among those we recommend the ones recently proposed by Cicchitelli and Montanari (2012) and Ghosh et al. (2012).

2.2 Some definitions

Let $\mathcal{D}$ be a fixed spatial subset of the d-dimensional Euclidean space $\mathbb{R}^d$, and let $\mathbf{u}$ be a generic location in it having coordinates $(u_1, \ldots, u_d)$. More-

over, let $z(\mathbf{u})$ be the value of the variable under study gathered at $\mathbf{u}$. The datum $z(\mathbf{u})$ can be thought of as the realization of a not necessarily known function of the spatial location $\mathbf{u}$. No assumption on the nature of the function $z(\cdot)$ has been made since the model- and the design-based framework need different definitions of the function $z(\cdot)$, probabilistic and deterministic respectively.

The nature of spatial data is somewhat complicated: a datasets consists of the values of the variable under study, $z(\mathbf{u})$, and the information about the locations where those are observed at, $\mathbf{u}$-s. Therefore, datasets have to be seen in a matrix form, but this doesn't mean that the phenomenon is necessarily multivariate. If the spatial information is not available, then inference can be carried out only using simple random sampling.

The forecasting of the unknown values at unobserved locations is usually carried out by means of interpolation. The large amount of methods that has been developed can be classified according to various criteria. An interpolator is said *global* when it predicts the values using a single function for the whole region. These methods are often computationally complex and adding or deleting a data point requires a new computation. On the contrary a *local* interpolator predicts the value at an unobserved location using multiple different functions whose parameters are optimized for a neighbourhood (Franke, 1982; Shepard, 1968; Mitas and Mitasova, 1999).

Interpolators assuming a superpopulation probabilistic model are able to exploit the randomness peculiar to the phenomenon and therefore are said *stochastic*. For the values predicted with these methods it is possible to obtain uncertainty measures and to assess their probabilistic properties. *Deterministic* interpolators fail to capture the random nature of the phenomenon since they are based only on geometric properties. It is therefore impossible to assess any probabilistic property (Webster and Oliver, 2007; Li and Heap, 2008).

Interpolators which honour the data locations passing through the observed values are defined *exact* and are opposed to the *approximate* ones which fail in doing so. The latter produce a very slowly varying interpolating function and are used when there is some uncertainty about a given spatial configuration since they reduce the effect of measurement errors.

Moreover, interpolating methods can be classified on the basis of the interpolated values depending on whether they produce *gradual* or *abrupt* changes in the interpolated surface.

Finally *point* interpolation is used for those data that can be collected at location point level while *areal* interpolation is used when the domain is divided in subareas.

Regardless of the method used, the interpolating function must follow the Tobler's first law of geography (Tobler, 1970)

> *Everything is related to everything else, but near things are more related than distant things.*

This assumption indicates a certain spatial relationship between the points in the domain. The relationship is strong for close points and decreases as the the distances between them increase. Tobler's law can be stated in a mathematical form so that it would be possible to achieve formal results.

The interpolated value at a generic location is a weighted mean of the observed values

$$\hat{z}(\mathbf{u}) = \sum_{i=1}^{n} \lambda_i \, z(\mathbf{u}_i),$$

with weights λ_i suitably chosen. As it will be seen in the following chapter the choice of the interpolation methods will affect the definition of the weighting system.

2.3 Spatial sampling

Let us suppose we are interested in the spatial pattern of a variable under study in a given domain $\mathcal{D}$. Collecting the data for a continuous surface is almost impossible given the cost and the time involved. Therefore spatial sampling seems the only reasonable choice.

When the object of inference is a spatial domain, sampling means selecting n locations where to gather the data. An arbitrary sample of locations is defined as

$$S = \{\mathbf{u}_i : \mathbf{u}_i \in \mathcal{D}, 1 \leq i \leq n\}.$$

Then the observed set of values at the sampled locations is defined as

$$\{z(\mathbf{u}_i) : \mathbf{u}_i \in S, 1 \leq i \leq n\}.$$

Many different spatial sampling designs have been proposed each of those suitable for the problem under study. In this dissertation we will deal with finite and continuous population sampling.

According to Cressie (1993) the most commonly used probabilistic sampling designs are the following three:

- in *simple random sampling*, the sampled locations are chosen according to a uniform distribution over the region $\mathcal{D}$;
- in *stratified random sampling*, the domain is divided in non-overlapping strata and within any of those a simple random sample is drawn;
- in *systematic random sample*, a location is randomly chosen within $\mathcal{D}$ and the remaining $(n-1)$ are chosen according to a predefined pattern.

Simple random sampling design is used when little is known about the population and when there is not any restraint on the inclusion of close locations in the sample. The stratified random sampling design is used when the domain can be divided in subregion and can lead to more efficient inference when the the strata are homogeneous (Gelfand et al., 2010). The systematic random sampling design can easily be implemented and allow to have well defined directional and equally distant classes of sampled locations (Haining, 2003).

A way to add more complex design is to consider the variable probability ones (Gelfand et al., 2010) where the locations are sampled with *varying probabilities*. In this way it is possible to manage inclusion in or exclusion from the sample for the locations close to the ones already sampled (e.g. adaptive spatial sampling).

For what regards a two-dimensional domain, it may be divided in subregions by a superimposed grid and each subregion becomes one of the population elements to be sampled (*finite population sampling*). On the contrary if no grid is superimposed on the domain we will sample from a *continuous population*. Grids can be regular (i.e. triangular, squared or hexagonal) or irregular (e.g., Voronoi tessellation).

2.4 Model-based and design-based spatial inference

The randomness involved in (spatial) sampling comes from the combination of a sampling design with a superpopulation model (Brus and de Gruijter, 1997). The resulting combination leads to the four strategies highlighted in Table 2.1. As previously mentioned in the introduction, this dissertation will focus on the duality model- versus design-based spatial inference. Fully deterministic strategies involve the use of non-stochastic interpolators whereas fully random strategies can be considered as a blend of the two different statistical approaches (model- and design-based).

Table 2.1: Sampling strategies involving two sources of randomness (Brus and de Gruijter, 1997)

		Values at given locations	
		Fixed	Random
Sample locations	Fixed	Fully deterministic	Model-based
	Random	Design-based	Fully random

Model-based inference assumes that the observed values are the outcome of a superpopulation model generating the data (see Chapter 3). In spatial analysis the data generating random process is called random field. The function defined in Section 2.2 in this case is a random function $Z(\cdot)$ depending on the spatial location $\mathbf{u}$ belonging to the domain $\mathcal{D}$ and on a random event ω of the σ-algebra $\mathcal{A}$ of subsets of the sample space Ω. The function can be explicitly written as $Z(\mathbf{u}, \omega)$. Once the realization of the random event, ω', is observed, the function $z(\mathbf{u}, \omega')$ is no longer random and depends only on the spatial location $\mathbf{u}$. On the contrary, if the spatial location $\mathbf{u}'$ is fixed, $Z(\omega, \mathbf{u}')$ is the probability distribution of the variable in that location. In what follows the dependence on the random event will be implicit whenever there will be no risk of misunderstanding. Formally speaking the spatial random process generating the data is defined as

$$\{Z(\mathbf{u}, \omega) : \mathbf{u} \in \mathcal{D}, \omega \in \mathcal{A}\} = \{Z(\mathbf{u}) : \mathbf{u} \in \mathcal{D}\}.$$

Once a set of locations is sampled, observing the values represents the realization of the random event involved. Thus, the set of observed values is defined as

$$\{z(\mathbf{u}, \omega') : \mathbf{u} \in S\} = \{z(\mathbf{u}) : \mathbf{u} \in S\}.$$

The object of inference is to model the random function $Z(\cdot)$ in order to predict the values at unsampled locations. *Prediction* involves the process of forecasting the random outcome of the random field through the *super-population model* whose parameters are estimated according to the observed values at the set of sampled locations. Predictors' properties are derived according to the joint distribution ξ of the N random variables $Z(\cdot)$.

Since in the model-based approach the sampling design is assumed to be non-random, sampling from a finite or a continuous population is quite the same. When the population is finite either a grid is superimposed over the domain $\mathcal{D}$ or we are interested in only a finite number of locations in it.

Probability sampling designs are the core of *design-based* inference where the only source of randomness is represented by the sampling design itself (see Chapter 4). Indeed, the population values are considered fixed, but unknown. Without loss of generality, these can be seen as the outcome of a not known function of the spatial locations $z(\mathbf{u})$.

The object of inference can either be the *estimation* of a summary value of the population (e.g. total) over the domain $\mathcal{D}$ or the individual values z at any location (*point estimation*). The properties of the resulting estimator or point estimator, respectively, are derived according to their distribution over the σ-algebra $\mathcal{S}$ of all the possible samples of the sample space $\mathfrak{S}$. The inclusion probabilities (i.e. the probability that subsets of elements of the population enters the sample) play a fundamental role.

Spatial design-based inference has been developed in the context of finite populations, however it is possible to extend its application to the case of infinite population by substituting the inclusion probabilities with the inclusion density functions (for an example in spatial case, see Cordy, 1993).

From now on, following Cressie (1993), the model-based forecasting of the value of the variable under study at an unsampled location is called *prediction*, while the design-based one is said *point prediction*.

Chapter 3

Model-based spatial statistics

When dealing with spatial data, the forecasting of the value of the variable under study at an unsampled location is commonly addressed by using model-based techniques. Those are based on the assumption of the existence of an underlying superpopulation random model generating the data. The values at the sampled locations are the realization of a random event of the σ-algebra $\mathcal{A}$ of the sample space Ω. Moreover, they are function of the spatial location $\mathbf{u}$, having coordinates $(u_1, \ldots, u_d)$, in the domain $\mathcal{D} \in \mathbb{R}^d$.

The object of inference is the superpopulation model parameter estimated according to the observed values at the sampled locations. Hence, the estimated model can be used in order to predict the values at the unsampled locations. In order to infer the random process, sampling is still essential for convenience and cost reasons. However, it is not necessary for the sampling to be random since the randomness is induced by the superpopulation model itself (see Table 2.1).

In this chapter a review of the main geostatistical technique (*kriging*) will be presented, mostly following the book by Cressie (1993). In Sections 3.1 the underlying superpopulation model is presented. Section 3.2 concerns the structure of spatial dependence. Section 3.3 regards the estimation of the superpopulation parameters. The stochastic interpolator used for prediction is introduced in Section 3.4. A short section highlighting the practical applications of the kriging ends the chapter.

3.1 The superpopulation model

3.1.1 Some important definitions

In geostatistics, the observed values are the realizations of the spatial random process

$$\{Z(\mathbf{u}) : \mathbf{u} \in \mathcal{D}\},$$

where the $\mathcal{D} \in \mathbb{R}^d$ is the spatial domain with spatial positive volume. In the previous equation the dependence on the random event ω has been omitted for sake of ease; whenever the emphasis has to be put on the source of randomness, it can be either indicated as $\{Z(\mathbf{u}, \omega) : \mathbf{u} \in \mathcal{D}, \omega \in \mathcal{A}\}$, where $(\Omega, \mathcal{A}, \Pr)$ is a probability space.

Random processes are usually defined through the use of their *finite-dimensional* distribution

$$F_{\mathbf{u}_1,\ldots,\mathbf{u}_k}(z_1, \ldots, z_k) = \Pr\{Z(\mathbf{u}_1) \leq z_1, \ldots, Z(\mathbf{u}_k) \leq z_k\},$$

where $k \geq 1$. It must meet the Kolmogorov consistency conditions:

(i) for any permutation π of the set of labels $\{1, \ldots, k\}$, F must remain the same

$$F_{\mathbf{u}_{\pi(1)},\ldots,\mathbf{u}_{\pi(k)}}(z_{\pi(1)}, \ldots, z_{\pi(k)}) = F_{\mathbf{u}_1,\ldots,\mathbf{u}_k}(z_1, \ldots, z_k);$$

(ii) for any $m \in \mathbb{N}$, F must satisfy

$$F_{\mathbf{u}_1,\ldots,\mathbf{u}_k,\mathbf{u}_{k+1},\ldots,\mathbf{u}_{k+m}}(z_1, \ldots, z_k, \infty, \ldots, \infty) = F_{\mathbf{u}_1,\ldots,\mathbf{u}_k}(z_1, \ldots, z_k).$$

The random field is said to be real valued if $z_j \in \mathbb{R}$ for any j.

For any location $\mathbf{u}$ belonging to the domain $\mathcal{D}$ we define the random field expected value

$$\mathrm{E}[Z(\mathbf{u})] = \mu(\mathbf{u}),$$

which is said *trend* or *drift*, and its variance

$$\mathrm{V}[Z(\mathbf{u})] = \mathrm{E}[(Z(\mathbf{u}) - \mu(\mathbf{u}))^2].$$

For any couple of locations $\mathbf{u}$ and $\mathbf{u}'$ belonging to the domain $\mathcal{D}$ we define the random field autocovariance

$$\mathrm{Cov}(Z(\mathbf{u}), Z(\mathbf{u}')) = \mathrm{E}[(Z(\mathbf{u}) - \mu(\mathbf{u}))(Z(\mathbf{u}') - \mu(\mathbf{u}'))].$$

3.1.2 Stationarity, isotropy and ergodicity

A random field is said to be *strong stationary* if for any $k \geq 1$ and for any spatial lag $\mathbf{h} \in \mathbb{R}^d$ its finite-dimensional distribution is invariant to space translations

$$F_{\mathbf{u}_1+\mathbf{h},\ldots,\mathbf{u}_k+\mathbf{h}}(z_1,\ldots,z_k) = F_{\mathbf{u}_1,\ldots,\mathbf{u}_k}(z_1,\ldots,z_k).$$

Strong stationarity requires a rather strict condition; therefore, a random field is said *weakly stationary* or *second-order stationary* if the following conditions hold:

(i) its expected value does not depend on the spatial location

$$\mathrm{E}[Z(\mathbf{u})] = \mu; \tag{3.1.1}$$

(ii) its covariance depends only on the spatial lag between the two locations

$$\mathrm{Cov}(Z(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})) = C(\mathbf{h})$$

and is said *covariogram*. If condition *(ii)* holds, then the random field variance does not depend on the location $\mathbf{u}$

$$\mathrm{V}[Z(\mathbf{u})] = \mathrm{Cov}(Z(\mathbf{u}), Z(\mathbf{u})) = C(\mathbf{0}) = \sigma^2. \tag{3.1.2}$$

In other words stationarity is the property a random field has to replicate itself in a way that the absolute coordinates lack their importance. Furthermore, if the covariogram is function only of the distance $h = \|\mathbf{h}\|$, then it is said *isotropic*, meaning that the orientation (angle) of the coordinates does not matter.

We define a process Z *intrinsic stationary* if the second-order stationarity conditions hold for the increments $Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})$:

(i) the expected value does not depend on the spatial location

$$\mathrm{E}[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})] = 0;$$

(ii) its variance depends only on the spatial lag between the two locations

$$\frac{1}{2}\mathrm{V}[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})] = \frac{1}{2}\mathrm{E}[(Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h}))^2] = \gamma(\mathbf{h}). \tag{3.1.3}$$

Function $\gamma(\cdot)$ is known as the *semivariogram* of $Z(\mathbf{u})$. Note that intrinsic stationarity does not requires finite variance σ^2. Again, if the semivariogram is function only of the distance, then it is said *isotropic*. The semivariogram plays a fundamental role in geostatistics; prediction is often based on its estimation because the assumptions on the underlying process are the weakest, i.e. intrinsic stationarity implies weak stationarity.

Among the weak stationary random fields there is a subset possessing a fundamental property named *ergodicity* allowing to estimate the expectations over the event space Ω through spatial averages. At the time Cressie (1993) was writing, no one had yet generalized Birkhoff's ergodic theorem (Kallenberg, 1997) to the case of a finite spatial sample having continuous values. However, Gaetan and Guyon (2010) extends it to the multivariate domain case and states the relationships between ergodicity and the weak and strong laws of large numbers respectively. A common belief is that statisticians should make the weaker ergodic assumption of the L_2-convergence of the spatial mean $\bar{Z}$ and of the covariogram (semivariogram) estimator $\hat{C}(h)$ ($\hat{\gamma}(h)$) as the domain bounds expand towards infinity in all directions. This property has more theoretical relevance than practical, since geostatistics always deals with bounded domains and no replication is possible. For a deeper coverage of the ergodicity extension to random fields one can have a look at the book by Cressie (1993), Chilès and Delfiner (1999) and Gaetan and Guyon (2010).

Finally we define *Gaussian* a random field if all its finite dimensional distributions are multivariate Gaussian. Then, if a Gaussian random field is weak stationary, it is also strong stationary because the Gaussian distribution is uniquely characterized by its expectation and its covariance function. Moreover, $C(\mathbf{h}) \to 0$, as $\|\mathbf{h}\| \to \infty$, is a sufficient condition for ergodicity (Adler, 1981).

3.1.3 The geostatistical model

So far we have stated the stochastic properties of random fields; however, the forecasting at unsampled locations involves the modellization of the process itself. Cressie (1993) proposes to decompose the variability of

the process into various sources so that it will be possible to manage all the aspects that can induce undesired variability. He models the random process $Z(\mathbf{u})$ as

$$Z(\mathbf{u}) = \mu(\mathbf{u}) + W(\mathbf{u}) + \eta(\mathbf{u}) + \epsilon(\mathbf{u}), \tag{3.1.4}$$

where each component represents a source of variability, endogenous or exogenous, of the process. The quantity $\mu(\mathbf{u}) = \mathrm{E}[Z(\mathbf{u})]$ captures the *large-scale variation* of the process and is expressed by its deterministic mean. As a result all the other components must have null expectation. It can either be function only of the spatial coordinates or even of some auxiliary variables. $W(\mathbf{u})$ is an intrinsic stationary random process having semivariogram $\gamma_W(\mathbf{u})$ whose range is larger than the minimum observed lag between the sampled locations. This component is known as *smooth small-scale variation*. Component $\eta(\mathbf{u})$ is said *micro-scale variation*. It is an intrinsic stationary process independent of $W(\mathbf{u})$ having a semivariogram range smaller than the minimum observed lag. Given that, its semivariogram, $\gamma_\eta(\mathbf{u})$, can not be modelled. Finally, $\epsilon(\mathbf{u})$ is the *measurement error* and is a white noise process independent of both $W(\mathbf{u})$ and $\eta(\mathbf{u})$. The $\epsilon(\mathbf{u})$-s are *i.i.d.* with variance $\mathrm{V}[\epsilon(\mathbf{u})] = \sigma^2_\epsilon$.

According to decomposition (3.1.4) and given the characteristics of each of its components, it results that the semivariogram of the second-order stationary process $Z(\mathbf{u})$ satisfies the decomposition

$$\gamma(\cdot) = \gamma_W(\cdot) + \gamma_\eta(\cdot) + \sigma^2_\epsilon.$$

By combining the first three components $S(\mathbf{u}) = \mu(\mathbf{u}) + W(\mathbf{u}) + \eta(\mathbf{u})$ of decomposition (3.1.4), we obtain the signal model

$$Z(\mathbf{u}) = S(\mathbf{u}) + \epsilon(\mathbf{u}).$$

The signal model is largely used for assessing the stochastic behaviour of the random field in order to predict $Z(\cdot)$ or $S(\cdot)$ whether the measurement error is null or not. The trend $\mu(\mathbf{u})$ is not of interest in this analysis with the exception that if it depends on the location, then a spatial structure for it is needed since otherwise the process would be non-stationary.

An alternative model collects the smooth small-scale variation, the micro-scale variation and the measurement error in the random component $\delta(\mathbf{u}) = W(\mathbf{u}) + \eta(\mathbf{u}) + \epsilon(\mathbf{u})$ highlighting the large-scale variation

$$Z(\mathbf{u}) = \mu(\mathbf{u}) + \delta(\mathbf{u}). \tag{3.1.5}$$

In this way the emphasis is posed on the expected value of the process allowing to analyse the spatial trend (see Subsection 3.4.3)

Cressie (1993) detects more decomposition for $Z(\mathbf{u})$.

3.2 The structure of spatial dependence

Semivariogram and covariogram estimation plays a critical role in the prediction of the value of the variable under study at an unsampled location. Indeed they are the parameters of the random field $\{Z(\mathbf{u}) : \mathbf{u} \in \mathcal{D}\}$ that needs to be estimated (see Section 3.3). In this section we will have a deeper look at their properties and empirical estimation.

In Subsection 3.1.2 the semivariogram function $\gamma(\mathbf{h})$ has been defined as the variance of the increment $Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})$ of the intrinsic stationary process $Z(\mathbf{u})$. Moreover if the process is second-order stationary, then it is possible to obtain $\gamma(\cdot)$ from $C(\cdot)$ and vice versa by using the following relationship:

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h}). \tag{3.2.1}$$

This is why kriging predictors can either be written in semivariogram or covariogram form. Second-order stationarity is a necessary and sufficient condition for deriving $\gamma(\mathbf{h})$ and $C(\mathbf{h})$ one from another. If the process is just intrinsic stationary, then $C(\mathbf{h})$ can still be obtained as $C(\mathbf{0}) - \gamma(\mathbf{h})$, but it is not a superpopulation parameter anymore.

3.2.1 Properties of the covariance functions

The validity of a covariogram holds if certain properties are satisfied. Let us consider a second-order stationary random field $\{Z(\mathbf{u}) : \mathbf{u} \in \mathcal{D}\}$. First of all, its covariogram must be even, $C(\mathbf{h}) = C(-\mathbf{h})$, since covariance is a symmetric operator, $\mathrm{Cov}(Z(\mathbf{u}), Z(\mathbf{u}+\mathbf{h})) = \mathrm{Cov}(Z(\mathbf{u}+\mathbf{h}), Z(\mathbf{u}))$. Moreover, in

order to guarantee non-negative prediction variances, the covariogram must be *positive semidefinite*

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j C(\mathbf{u}_j-\mathbf{u}_i)\geq 0,$$

for any real number $\alpha_1,\ldots,\alpha_m$ and any spatial location. This requirement means that the model must not have negative variance

$$\mathrm{V}\left[\sum_{i=1}^{m}\alpha_i Z(\mathbf{u})\right]=\sum_{i=1}^{m}\sum_{j=1}^{n}\alpha_i\alpha_j C(\mathbf{u}_j-\mathbf{u}_i).$$

This condition ensures that the covariogram has monotonically decreasing positive values since, by the Cauchy-Schwartz inequality, it can be shown that $0\leq|C(\mathbf{h})|\leq C(\mathbf{0})$, where variance (3.1.2) of the observations is non-negative.

The properties of the semivariogram of a second-order stationary random field $\{Z(\mathbf{u}):\mathbf{u}\in\mathcal{D}\}$ arise from the covariogram's ones. Semivariograms must be even functions, $\gamma(\mathbf{h})=\gamma(-\mathbf{h})$, and pass through the origin, $\gamma(\mathbf{0})=0$, since $\mathrm{V}[Z(\mathbf{u})-Z(\mathbf{u}+\mathbf{0})]=\mathrm{V}[0]=0$. Moreover, semivariograms must satisfy the *conditional negative semidefiniteness*

$$\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\gamma(\mathbf{u}_j-\mathbf{u}_i)\leq 0,$$

for any real number $\alpha_1,\ldots,\alpha_m$ such that $\sum_{i=1}^{m}\alpha_i=0$ and any spatial location (Cressie, 1993). In order to explain this requirement, let us consider the intrinsically stationary process $Z(\cdot)$, then it results

$$\left(\sum_{i=1}^{m}\alpha_i Z(\mathbf{u}_i)\right)^2=-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j(Z(\mathbf{u}_j)-Z(\mathbf{u}_j))^2,$$

since $\sum_{i=1}^{m}\alpha_i=0$. Taking expectations we obtain

$$0\leq\mathrm{V}\left[\sum_{i=1}^{m}\alpha_i Z(\mathbf{u})\right]=-\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j\gamma(\mathbf{u}_j-\mathbf{u}_i).$$

A necessary condition for the validity of a semivariogram is that it must satisfy the *intrinsic hypothesis*, i.e. it must grow more slowly than the squared distance, $\|\mathbf{h}\|^2$,

$$\lim_{\|\mathbf{h}\|\to\infty} \frac{\gamma(\mathbf{h})}{\|\mathbf{h}\|^2} = 0. \tag{3.2.2}$$

For more conditions on the validity of covariograms and semivariograms one can have a look at Subsections 2.5.1 and 2.5.2 of the book by Cressie (1993).

The isotropic covariogram $C(\|\mathbf{h}\|)$ of a second-order stationary random field $Z(\mathbf{u})$ is monotonically decreasing and approaches the x-axis as the distance $\|\mathbf{h}\|$ increases. Since property (3.2.1) holds, $\gamma(\|\mathbf{h}\|)$ approaches the upper asymptote, corresponding to $C(\mathbf{0})$, as $\|\mathbf{h}\| \to \infty$; it is called *sill* of the semivariogram, σ^2. The distance at which the semivariogram reaches its sill is said *range* of the semivariogram, α. When the semivariogram approaches the sill only asymptotically, the *practical range*, $\phi = 0.95 \times C(\mathbf{0})$, is used instead.

It has been said that the semivariogram has to pass through the origin; however, it is more a theoretical property rather than an empirical one. In applications it may often happen that $\gamma(\mathbf{h}) \to \tau^2 > 0$, as $\|\mathbf{h}\| \to 0$. There are two possible explanations to this phenomenon, one endogenous and one exogenous. The former depends on the fact that no distance smaller than $\min\{\|\mathbf{u}_j - \mathbf{u}_i\|, \mathbf{u}_i, \mathbf{u}_j \in S\}$ can be observed from the sample. Hence, the random process $\eta(\mathbf{u})$, having sill σ^2_η, operates at a micro-scale level. The exogenous phenomenon depends on measurement errors; any repeated measure at location $\mathbf{u}$ can not be gathered without error leading to variability represented by the variance σ^2_ϵ. The combination of the two phenomena leads to the discontinuity at the origin known as *nugget effect* whose magnitude is given by

$$\tau^2 = \sigma^2_\eta + \sigma^2_\epsilon.$$

In presence of a nugget effect, the quantity $\sigma^2 = C(\mathbf{0}) - \tau^2$ is called *partial sill* of the semivariogram. Hence, the practical range is defined as the lag at which the semivariogram reaches $\phi = \tau^2 + 0.95 \times C(\mathbf{0})$.

3.2.2 Valid parametric isotropic semivariogram models

Matheron (1971) states that the behaviour of the semivariogram near the origin is expression of the continuity and regularity in the space of the process Z. First of all, let us define the L_2-*continuity* or *mean square continuity* of the process Z at the point $\mathbf{u}$, as

$$\lim_{\mathbf{u}' \to \mathbf{u}} \mathrm{E}[(Z(\mathbf{u}') - Z(\mathbf{u}))^2] = 0.$$

Furthermore, if Z is weakly stationary having semivariogram

$$\frac{1}{2}\mathrm{E}[(Z(\mathbf{u}') - Z(\mathbf{u}))^2] = C(\mathbf{0}) - C(\mathbf{u} - \mathbf{u}')$$

according to (3.2.1), then Z is L_2-*continuous* at $\mathbf{u}$ if and only if $C(\cdot)$ is continuous at the origin. Moreover, a process Z is said L_2-*differentiable* or *mean square differentiable* at the generic point $\mathbf{u}$ if the increment

$$\frac{Z(\mathbf{u} + \mathbf{h}) - Z(\mathbf{u})}{\mathbf{h}}$$

converge in L_2 as $\mathbf{h} \to \mathbf{0}$. If such a limit exists, we call it $Z'(\mathbf{u})$. Furthermore, if Z is weakly stationary and the second differentials

$$\frac{\partial^2}{\partial h_1^2}\gamma(\mathbf{h}), \ldots, \frac{\partial^2}{\partial h_d^2}\gamma(\mathbf{h})$$

exist and are finite for $\mathbf{h} \to \mathbf{0}$, where $\mathbf{h} = [h_1, \ldots, h_d]^\top$, then Z is mean square differentiable for any $\mathbf{u} \in \mathbb{R}^d$. For more mathematical details regarding L_2-continuity and L_2-differentiability one can have a look at Sections 2.4 and 2.6 of Stein (1999).

Then, following Matheron (1971), we classify the semivariogram models according to their behaviour near the origin as follows:

(i) $\gamma(\mathbf{h})$ L_2-differentiable is at the origin, so that $Z(\cdot)$ is L_2-differentiable itself and the semivariogram is regular (parabolic behaviour);

(ii) $\gamma(\mathbf{h})$ is L_2-continuous, but not L_2-differentiable at the origin, then $Z(\cdot)$ is mean square continuous and the semivariogram is less regular (linear trend);

(iii) $\gamma(\mathbf{h})$ presents a discontinuity at the origin (i.e. it does not approach 0 as $\mathbf{h}$ approaches the origin), then $Z(\cdot)$ is not even L_2-continuous and the semivariogram is highly irregular (*nugget effect*);

(iv) $\gamma(\mathbf{h})$ is null at the origin and a positive constant elsewhere, then $Z(\mathbf{u})$ and $Z(\mathbf{u}')$ are independent for any $\mathbf{u} \neq \mathbf{u}'$ despite their proximity (Z is called *white noise*).

Literature is rich of semivariogram models satisfying the previous conditions; however, just a few of them are of practical use. In the following part of this subsection we introduce the key models, while for a wider range of them one can have a look at the books by Journel and Huijbregts (1978), Cressie (1993), Stein (1999) and Chilès and Delfiner (1999). The presented models are parametric families, $\gamma(\mathbf{h};\boldsymbol{\theta})$, and the parameter vector can be composed, among other parameters, by the nugget τ^2, the partial sill σ^2 or the range ϕ or all of them together. All the model that will be presented are valid in $\mathbb{R}^2$ and some of them even for higher dimensional domains. Since this dissertation deals with two-dimensional random fields, where no different will be stated we will assume $\mathcal{D} \in \mathbb{R}^2$.

Nugget-only model

The semivariogram (Figure 3.1) of a *white noise* process, case *(iv)* of the classification given by Matheron (1971), corresponds only to the sill, σ^2, of the process. This means that the process has the same expectation and variance all over the domain with no correlation reflecting the void of spatial structure. Clearly, it is second-order stationary and satisfy all the validity conditions. The nugget-only model mathematical formulation is

$$\gamma(\mathbf{h};\sigma^2) = \begin{cases} 0, & \text{if } \mathbf{h} = \mathbf{0}; \\ \sigma^2, & \text{if } \mathbf{h} \neq \mathbf{0}. \end{cases}$$

Linear model

The linear semivariogram model (Figure 3.2) corresponds to a process whose covariances change linearly over a large range. However, this assump-

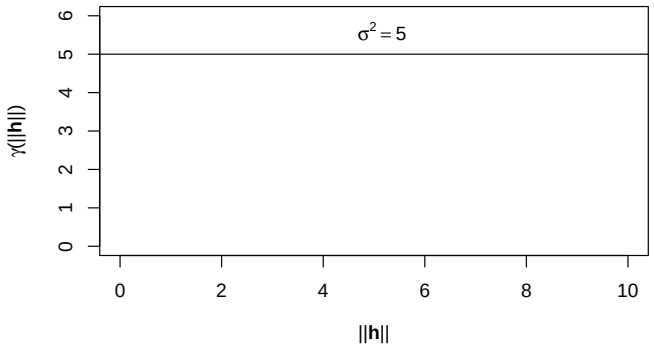

Figure 3.1: Nugget-only semivariogram model

tion is hardly verified in practice; more realistically it may represent the initial increase of a second-order stationary model having a linear trend near the origin. This means that the set of sampled locations does not allow to infer the range and the sill of the semivariogram because no large enough lag can be observed. The linear model is intrinsically stationary and its mathematical formulation is

$$\gamma(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } \mathbf{h} = \mathbf{0}; \\ \tau^2 + \beta\|\mathbf{h}\|, & \text{if } \mathbf{h} \neq \mathbf{0}. \end{cases}$$

The parameter vector $\boldsymbol{\theta} = [\tau^2, \beta]^\top$ is constituted by the nugget parameter and the slope parameter, both non-negative.

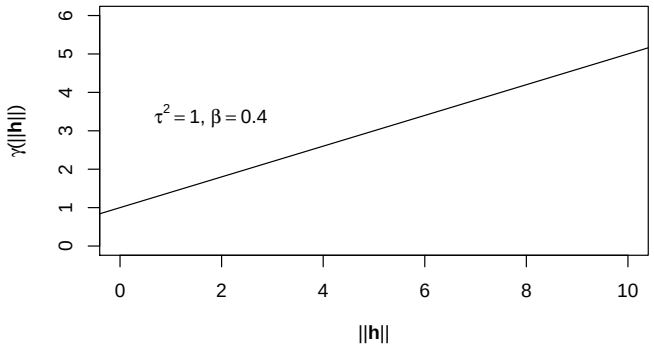

Figure 3.2: Linear semivariogram model

Spherical model

The spherical semivariogram model (Figure 3.3) is widely used when modelling a second-order stationary random field. It has a linear behavior near the origin, then it grows monotonically until it reaches the sill beyond some distance. Stein (1999) strongly criticizes the use of this model, arguing that adopting such a semivariogram may lead to problems when using likelihood-based methods because $\gamma(\mathbf{h};\boldsymbol{\theta})$ is only once differentiable at the lag where the semivariogram reaches the sill. Its mathematical formulation is

$$\gamma(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } \mathbf{h} = \mathbf{0}; \\ \tau^2 + \sigma^2 \left(\frac{3}{2}\frac{\|\mathbf{h}\|}{\alpha} - \frac{1}{2}\left(\frac{\|\mathbf{h}\|}{\alpha}\right)^3 \right), & \text{if } 0 < \|\mathbf{h}\| \leq \alpha; \\ \tau^2 + \sigma^2, & \text{if } \|\mathbf{h}\| > \alpha. \end{cases}$$

The model parameters are the nugget, τ^2, the sill, σ^2, and the range, α, which must be non-negative.

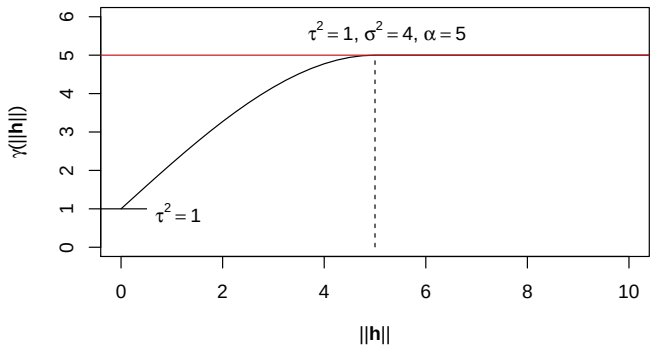

Figure 3.3: Spherical semivariogram model

Exponential model

The exponential model (Figure 3.4) grows monotonically from the origin and, then, approaches the sill asymptotically as $\|\mathbf{h}\| \to \infty$. This model is associated with a second-order stationary random field and has been found to fit well spatial data coming from different applications. According to Webster

and Oliver (2007), its corresponding autocovariance function has even been used in order to arrange efficient sampling design. Hence, it can be basically thought of as the main model of randomness in space. Its mathematical formulation is

$$\gamma(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } \mathbf{h} = \mathbf{0}; \\ \tau^2 + \sigma^2 \left(1 - \exp\left\{-\dfrac{\|\mathbf{h}\|}{\alpha}\right\}\right), & \text{if } \mathbf{h} \neq \mathbf{0}. \end{cases}$$

The nugget, τ^2, the sill, σ^2, and the range, α, parameters must be non-negative. The practical range ϕ of the exponential model is usually taken at $\alpha/3$.

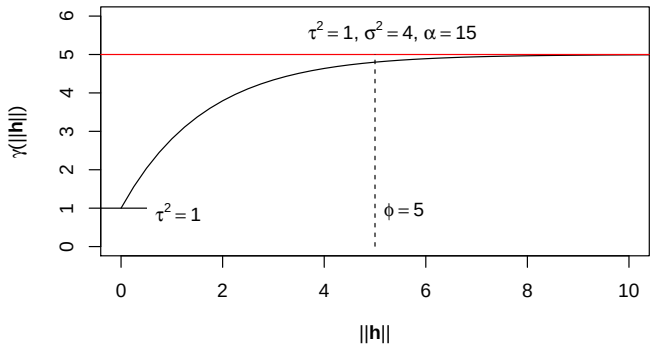

Figure 3.4: Exponential semivariogram model

Power model

The power semivariogram model (Figure 3.5) is intrinsically stationary and presents a convex monotone growth. Its formula is

$$\gamma(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } \mathbf{h} = \mathbf{0}; \\ \tau^2 + \beta\|\mathbf{h}\|^{\lambda}, & \text{if } \mathbf{h} \neq \mathbf{0}. \end{cases}$$

The nugget parameter, τ^2, and the slope parameter, β, must be non-negative. The exponent has to be non negative, but smaller than 2, $0 \leq \lambda < 2$, because otherwise the intrinsic hypothesis (3.2.2) would not be satisfied. As particular cases of the power model we have the nugget-only model, for $\tau^2 = 0$ and

$\lambda = 0$, and the linear model, for $\lambda = 1$. Moreover, the power model is the only one among the presented to be scale invariant, i.e. $\forall a > 0, \gamma(a\mathbf{h}) = a^{\lambda}\gamma(\mathbf{h})$ (Gaetan and Guyon, 2010).

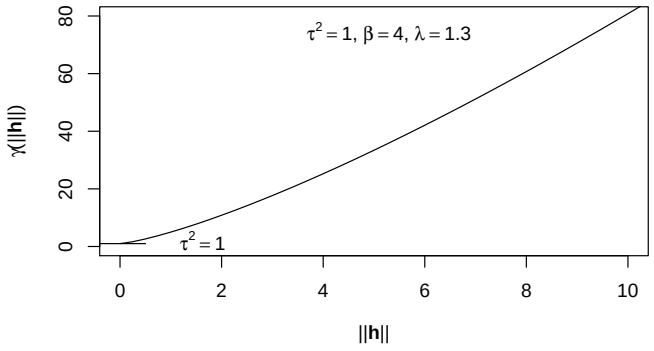

Figure 3.5: Power semivariogram model

Gaussian model

The Gaussian semivariogram model (Figure 3.6) exhibits a convex curvature near the origin and, then, a concave one before reaching the upper asymptote. Its formula is

$$\gamma(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} 0, & \text{if } \mathbf{h} = \mathbf{0}; \\ \tau^2 + \sigma^2 \left(1 - \exp\left\{-\left(\frac{\|\mathbf{h}\|}{\alpha}\right)^2\right\}\right), & \text{if } \mathbf{h} \neq \mathbf{0}. \end{cases}$$

Its only difference from the exponential model is the square in the exponential function; the parameter constraints are the same too. Relationship between the range parameter, α, and the practical range, ϕ, is $\phi = \alpha/\sqrt{3}$. The Gaussian model approaches the origin with a zero gradient which may lead to unstable kriging equations (Webster and Oliver, 2007).

Matérn model

The Matèrn model (Figure 3.7) is of great interest because it generalizes several other models (Gaetan and Guyon, 2010). Its mathematical formula-

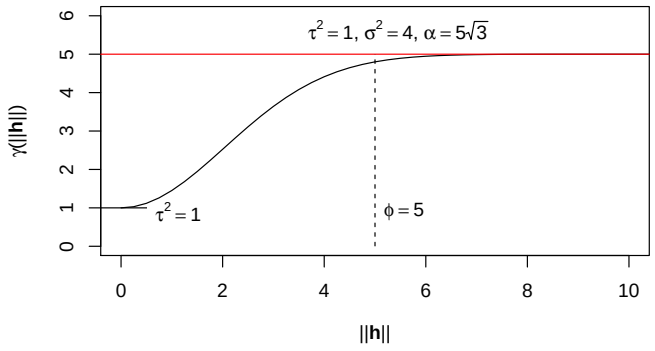

Figure 3.6: Gaussian semivariogram model

tion is

$$\gamma(\mathbf{h};\boldsymbol{\theta}) = \begin{cases} \tau^2 + \sigma^2, & \text{if } \mathbf{h} = \mathbf{0}; \\ \sigma^2 \left(1 - \dfrac{2^{1-\nu}}{\Gamma(\nu)} \left(\dfrac{\mathbf{h}}{\alpha}\right)^{\nu} K_\nu \left(\dfrac{\mathbf{h}}{\alpha}\right)\right), & \text{if } \mathbf{h} \neq \mathbf{0}, \end{cases}$$

where the Γ function and νth-order Bessel function $K_\nu(\cdot)$ are involved. All the parameters must be non-negative; ν is the smoothness parameter. The exponential semivariogram model is obtained as a particular case when $\nu = 1/2$,while the Gaussian when $\nu \to \infty$.

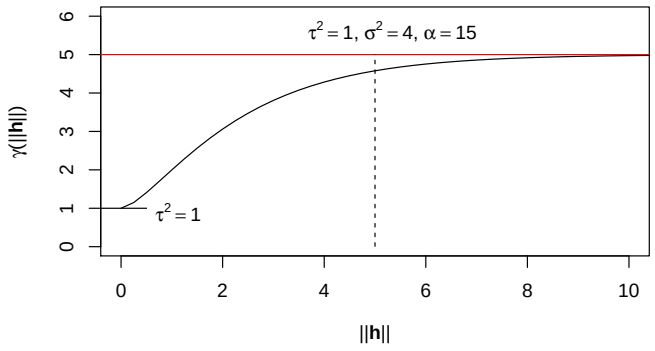

Figure 3.7: Matèrn semivariogram model

Complex models

More complex semivariogram models can be obtained by combining the ones presented since any combination of conditional negative semidefinite semivariogram functions is still conditional negative semidefinite. For a brief coverage of the *combined models* one can have a look at Section 5.3 of the book by Webster and Oliver (2007).

In some occasions the semivariogram may present a periodic fluctuation. In this case a semivariogram model involving trigonometric functions (e.g. *wave model*) is the most suitable choice (Cressie, 1993; Webster and Oliver, 2007).

Finally, an anisotropic semivariogram is function of both the distance between locations and the direction of it. Hence, it can not be represented as a function of the distance anymore. To deal with this problem one can consider a linear transformation of the lag vector $\mathbf{h}$. Let us consider a $d \times d$ matrix $\mathbf{A}$, then if

$$\gamma(\mathbf{h}) = \gamma^{\#}(\|\mathbf{Ah}\|)$$

is a real valued function, the semivariogram $\gamma^{\#}(\cdot)$ is said *geometrically anisotropic* (Cressie, 1993; Webster and Oliver, 2007). The linear transformation $\mathbf{A}$ of the Euclidean space is necessary to measure distance between locations. More complex anisotropic processes can be dealt with hierarchical methods (Cressie, 1993) or zonal anisotropic models where the components of vector $\mathbf{h}$ are treated separately (Webster and Oliver, 2007).

3.3 Estimating the superpopulation parameters

Semivariogram estimation can be achieved by using *empirical estimators* or *parametric models*. The former allow to obtain a semivariogram estimate at those lags obtainable from the set of sampled locations. These methods are useful to obtain a rough idea of $\gamma(\mathbf{h})$. The latter consists in fitting a semivariogram model to the data in order to estimate the model parameters. Thus, the estimated function is used to infer at the unobservable spatial lags.

3.3.1 Empirical semivariogram estimators

Let us suppose that the data $Z(\mathbf{u}_1), \ldots, Z(\mathbf{u}_n)$ can be modelled as an intrinsically stationary process; that is its increment process are weak stationary. Hence, the expected values $\mathrm{E}[(Z(\mathbf{u})-Z(\mathbf{u}+\mathbf{h}))^2]$ appearing in (3.1.3) can be estimated by the averages of the squared differences of the observed values at the sampled locations having the same lag, $(Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h}))^2$, once the weak ergodicity assumption has been put forward. Therefore, we define the *empirical estimator* of the semivariogram at lag $\mathbf{h}$

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{|N(\mathbf{h})|} (Z(\mathbf{u}_i) - Z(\mathbf{u}_j))^2, \tag{3.3.1}$$

where $|N(h)|$ is the number of distinct couples of locations having spatial lag $\mathbf{h}$ between them

$$N(\mathbf{h}) = \{(\mathbf{u}_i, \mathbf{u}_j) : \mathbf{u}_j - \mathbf{u}_i = \mathbf{h}, i, j = 1, \ldots, n\}.$$

Estimator (3.3.1) was introduced by Matheron in 1962 and that is why it is sometimes called *Matheron's method-of-moments estimator*. If the underlying process is isotropic, then the lag vector $\mathbf{h}$ is replaced by its length $\|\mathbf{h}\|$.

Estimator (3.3.1) is even, $\hat{\gamma}(\mathbf{h}) = \hat{\gamma}(-\mathbf{h})$, and pass through the origin, $\hat{\gamma}(\mathbf{0}) = 0$. Furthermore, it is an unbiased estimator of $\gamma(\mathbf{h})$ unless the random field is not at least intrinsically stationary (Cressie, 1993) and the data have a skewed distribution (Lloyd, 2011).

In the Gaussian case, the variance of estimator (3.3.1) can be computed

$$\mathrm{V}[\hat{\gamma}(\mathbf{h})] \simeq \frac{2\gamma^2(\mathbf{h})}{|N(\mathbf{h})|}. \tag{3.3.2}$$

It is function of the true semivariogram and, therefore, increasing values of the semivariogram lead to increasing estimated variances. For those interested in the asymptotic behaviour of $\hat{\gamma}(\mathbf{h})$, we recommend the books by Cressie (1993) and Gaetan and Guyon (2010).

The advantage in estimating the semivariogram in respect of the covariogram is that $\gamma(\mathbf{h})$ is defined in more cases than $C(\mathbf{h})$ since the latter relies on the stronger assumption of second-order stationarity. For weak stationary

random fields the empirical covariogram is defined as

$$\hat{C}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{|N(\mathbf{h})|} (Z(\mathbf{u}_i) - \bar{Z})(Z(\mathbf{u}_j) - \bar{Z}), \tag{3.3.3}$$

where $\bar{Z}$ is the sample mean

$$\bar{Z} = \frac{1}{n} \sum_{i=1}^{n} Z(\mathbf{u}_i)$$

estimating the process expected value μ. It is not possible to obtain $\hat{\gamma}(\mathbf{h})$ from $\hat{C}(\mathbf{h})$, and vice versa, since relation (3.2.1) does not hold for estimators (3.3.1) and (3.3.3). However, if the ratio $|N(h)|/n$ is close to one, then the difference will be small.

Estimator (3.3.1) is sensitive to outliers; Cressie and Hawkins (1980) proposed two robust estimators of the semivariogram able to safeguard inference from small independent contamination of a Gaussian process. The former is basically a correction of estimator (3.3.1) and employs scaled observation in order to make them more Gaussian like

$$\bar{\gamma}(\mathbf{h}) = \frac{1}{2} \left(0.457 + \frac{0.494}{|N(\mathbf{h})|} \right)^{-1} \left(\frac{1}{|N(\mathbf{h})|} \sum_{|N(\mathbf{h})|} |Z(\mathbf{u}_i) - Z(\mathbf{u}_j)|^{1/2} \right)^4 . \tag{3.3.4}$$

The latter involves the median of the square root of the absolute value of the increments having lag $\mathbf{h}$

$$\tilde{\gamma}(\mathbf{h}) = \frac{\left(\text{med}\{|Z(\mathbf{u}_i) - Z(\mathbf{u}_j)|^{1/2} : (\mathbf{u}_i, \mathbf{u}_j) \in N(\mathbf{h})\} \right)^4}{2B(\mathbf{h})},$$

where $B(\mathbf{h})$ is a bias correction factor which is asymptotically equal to 0.457.

3.3.2 Semivariogram fitting

Empirical estimators can be useful in getting a rough idea of the semivarioram; however, they can not be employed to predict the values at unsampled locations. In order to predict over the whole domain, continuous functions are necessary: the fitting of a semivariogram parametric model to the data is, therefore, crucial.

Schabenberger and Pierce (2002) divide the fitting methods between the *indirect* and the *direct* ones. The former use transformation of the data as the responses for the estimation of the model parameters. The latter rely on rough data for the estimation process.

Least squares methods

Let $\Theta \subset \mathbb{R}^p$ be the parameter space within which the semivariogram model is defined (see Subsection 3.2.2). Moreover, let k be the number of classes, $N(\mathbf{h_i})$, obtainable form the sampled locations, for any of which the empirical estimate $\gamma^*(\mathbf{h}_i)$ has been computed. Then, the *ordinary least squares* (OLS) estimator of the parameters vector is the value

$$\hat{\boldsymbol{\theta}}_{OLS} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{k} (\gamma^*(\mathbf{h}_i) - \gamma(\mathbf{h}_i;\boldsymbol{\theta}))^2 \right\},$$

where $\gamma^*(\cdot)$ is any empirical estimator and $\gamma(\mathbf{h}_i;\boldsymbol{\theta})$ is any valid parametric semivariogram model.

As in regression, OLS estimators require uncorrelated and homoscedastic data; both assumptions are not satisfied. Therefore, we introduce the *generalized least squares* (GLS) criterion to preserve the appealing geometric property of the least square estimation and introduce the heteroscedasticity. Let us suppose that $\boldsymbol{\Gamma}^* = [\Gamma^*(\mathbf{h}_1), \ldots, \Gamma^*(\mathbf{h}_k)]^\top$ is the vector of the k random variable of the k computable empirical semivariogram, then it has variance $\mathbf{V}_{\boldsymbol{\Gamma}} = \mathrm{V}[\boldsymbol{\Gamma}^*]$. The GLS estimator is defined as

$$\hat{\boldsymbol{\theta}}_{GLS} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \big\{ (\boldsymbol{\gamma}^* - \boldsymbol{\gamma}(\boldsymbol{\theta}))^\top \mathbf{V}_{\boldsymbol{\Gamma}}^{-1} (\hat{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}(\boldsymbol{\theta})) \big\},$$

where $\boldsymbol{\gamma}^* = [\gamma^*(\mathbf{h}_1), \ldots, \gamma^*(\mathbf{h}_k)]^\top$ is the vector collecting the empirical semivariogram for any of the k classes $N(\mathbf{h_i})$, $\boldsymbol{\gamma}(\boldsymbol{\theta}) = [\gamma(\mathbf{h}_1;\boldsymbol{\theta}), \ldots, \gamma(\mathbf{h}_k;\boldsymbol{\theta})]^\top$ is the vector collecting the corresponding values of the semivariogram computed using a valid model.

Unfortunately the covariances of matrix $\mathbf{V}_{\boldsymbol{\Gamma}}$ are hard to compute. However, Cressie (1985) derived an approximated formula for the variance of the empirical estimator (3.3.1) and its robust version (3.3.4). Therefore, a widely

used compromise is to employ the *weighted least squares* (WLS) estimation where approximated variances (3.3.2) are collected in the diagonal matrix $\mathbf{W}$

$$\hat{\boldsymbol{\theta}}_{WLS} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}}\big\{(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}(\boldsymbol{\theta}))^\top \mathbf{W}^{-1}(\hat{\boldsymbol{\gamma}}^* - \boldsymbol{\gamma}(\boldsymbol{\theta}))\big\}.$$

Cressie (1993) shows that in the case of estimator (3.3.1) the WLS minimization problem can be reformulated as

$$\hat{\boldsymbol{\theta}}_{WLS} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}}\left\{\sum_{i=1}^{k}\left(|N(\mathbf{h})|\frac{\hat{\gamma}(\mathbf{h}_i)}{\gamma(\mathbf{h}_i;\boldsymbol{\theta})} - 1\right)^2\right\}.$$

He also provides an analogous formulation for the robust estimator (3.3.4).

Estimator $\hat{\boldsymbol{\theta}}_{WLS}$ is a rough approximation of the real values, since $\mathbf{W}$ is a poor approximation of $\mathbf{V}$ and variances (3.3.2) are themselves approximations of the real variances.

The empirical estimator involved in the least squares estimation techniques is a transformation of the data. Therefore, they belong to the indirect methods.

Likelihood methods

Likelihood estimation methods, introduced in spatial applications by Mardia and Marshall (1984), are said direct since they rely directly on the data; however, some assumptions on their distribution need to be made. Firstly, suppose that the underlying process is second-order stationary. Then, suppose that the vector collecting the observed values at the sampled locations is multivariate Gaussian $\mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\theta}))$, where $\mathbf{X}$ is a $n \times q$ matrix having rank $q < n$ and $\boldsymbol{\beta}$ a q-dimensional vector of coefficients. The variance-covariance matrix

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} C(\mathbf{0};\boldsymbol{\theta}) & C(\mathbf{u}_1 - \mathbf{u}_2;\boldsymbol{\theta}) & \cdots & C(\mathbf{u}_1 - \mathbf{u}_N;\boldsymbol{\theta}) \\ C(\mathbf{u}_1 - \mathbf{u}_2;\boldsymbol{\theta}) & C(\mathbf{0};\boldsymbol{\theta}) & \cdots & C(\mathbf{u}_2 - \mathbf{u}_N;\boldsymbol{\theta}) \\ \vdots & \vdots & \ddots & \vdots \\ C(\mathbf{u}_1 - \mathbf{u}_N;\boldsymbol{\theta}) & C(\mathbf{u}_2 - \mathbf{u}_N;\boldsymbol{\theta}) & \cdots & C(\mathbf{0};\boldsymbol{\theta}) \end{bmatrix}$$

is linked to $\gamma(\mathbf{h};\boldsymbol{\theta})$ through relation (3.2.1) and depends on $\boldsymbol{\theta}$ through the same parameter space Θ. Then, the log-likelihood is

$$\ell(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{z}) = c - \frac{1}{2}\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta}),$$

where $\mathbf{z}$ is the vector of the observed values at the sampled locations. As a function of $\boldsymbol{\beta}$, it is maximized by

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{z},$$

corresponding to the GLS estimator of $\boldsymbol{\beta}$ in the case $\mathbf{V} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$. This leads to the profile log-likelihood

$$\ell(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}; \mathbf{z}) = c - \frac{1}{2}\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2}\mathbf{z}^\top \mathbf{M}(\boldsymbol{\theta})\mathbf{z}, \tag{3.3.5}$$

where $\mathbf{M}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}(\mathbf{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$. Maximizing $\ell(\hat{\boldsymbol{\beta}}, \boldsymbol{\theta}; \mathbf{z})$ leads to the maximum likelihood (ML) estimator $\hat{\boldsymbol{\theta}}_{ML}$.

Maximization of (3.3.5) is a constrained problem for $\boldsymbol{\theta} \in \Theta$ which can not be solved analytically; iterative methods are necessary, the most known of which is the Newton-Raphson algorithm.

The ML estimator of the spatial dependence parameter is highly biased (see Subsection 2.6.3 of Cressie, 1993). In order to avoid this inconvenience Kitanidis and Vomvoris (1983) proposed to use the restricted maximum likelihood (REML) estimation, originally proposed by Patterson and Thompson (1971), for the estimation in spatial data. It consists in estimating the model parameters on $n - p$ linearly independent combinations of the original data known as error contrasts. A linear combination $\mathbf{a}^\top \mathbf{Z}$ is defined an error contrast if $\mathrm{E}[\mathbf{a}^\top \mathbf{Z}] = 0$ for all $\boldsymbol{\beta} \in \mathbb{R}^q$ and all $\boldsymbol{\theta} \in \Theta$; hence, $\mathbf{a}^\top \mathbf{Z}$ is an error contrast if and only if $\mathbf{a}^\top \mathbf{X} = 0$. Those are chosen as $n - p$ linearly independent elements among those of $(\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X})\mathbf{z}$. The log-likelihood becomes (Gelfand et al., 2010)

$$\ell_R(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{w}) = c - \frac{1}{2}\ln|\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2}\ln|\mathbf{X}^\top \boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{X}| - \frac{1}{2}\mathbf{z}^\top \mathbf{M}(\boldsymbol{\theta})\mathbf{z}.$$

The REML estimator of the model parameters is obtained by maximizing $\ell_R(\boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{w})$. Once the estimate, $\tilde{\boldsymbol{\theta}}$, is available, the estimator of the coefficient vector is obtained via its GLS estimator, $\tilde{\boldsymbol{\beta}} = (\mathbf{X}^\top \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}})^{-1}\mathbf{X})^{-1}\mathbf{X}^\top \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}})^{-1}\mathbf{z}$. The REML estimates need iterative procedures too.

3.4 Spatial prediction using the kriging

With the word *kriging*, geostatisticians indicate the set of techniques allowing to predict the values of a variable of interest over a spatial domain employing a function expressing the spatial correlation. Each technique relies on some assumptions made on the geostatistical model and on the nature of the data to predict. These techniques can be summarized in Table 3.1 (Schabenberger and Pierce, 2002). This subsection only regards the *simple kriging*, *ordinary kriging* and *universal kriging* which are the most used and known among the others. According to the assumption made, different methods can be mixed, e.g. universal block kriging. *Block kriging* will later be used in Chapter 5; it is about the estimation of the value of the variable

Table 3.1: Kriging methods according to the assumption made on the underlying process (Schabenberger and Pierce, 2002)

Assumptions		Method
Target size	Points	(Point) Kriging
	Areas	Block kriging
Known quantities	μ known	Simple kriging
	μ unknown, but constant	Ordinary kriging
	$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, $\boldsymbol{\beta}$ unknown	Universal kriging
Distribution	$Z(\mathbf{u}) \sim \mathcal{N}$	Kriging
	$\log Z(\mathbf{u}) \sim \mathcal{N}$	Lognormal kriging
	$\phi(Z(\mathbf{u})) \sim \mathcal{N}$	Transgaussian kriging
	$Z(\mathbf{u})$ is an indicator	Indicator kriging
	Gaussian with outliers	Robust kriging
	Unknown	Median polish kriging
	$\mu(\mathbf{u})$ is a random process	Bayesian kriging
Dimension	One-dimensional	Kriging
	d-dimensional	Cokriging
Linearity	Linear in $Z(\mathbf{u})$	Kriging
	Linear in functions of $Z(\mathbf{u})$	Disjunctive kriging

under study at a subarea of the domain, but the mathematics involved does not differ from the one of the techniques presented. For a comprehensive coverage of the other kriging techniques one can have a look at the books by Cressie (1993), Webster and Oliver (2007), Gaetan and Guyon (2010) and Gelfand et al. (2010).

3.4.1 Simple kriging

Let Z be a second-order stationary random field whose expected value function, $\mu(\mathbf{u})$, and variance structure, $\text{Cov}(Z(\mathbf{u}), Z(\mathbf{u}')) = C(\mathbf{u} - \mathbf{u}')$, are known for any location of the domain. Moreover, let us suppose that we are interested in predict the value Z at any location of the domain by employing the observed values, $\mathbf{Z} = [Z(\mathbf{u}_1), \ldots, Z(\mathbf{u}_n)]^\top$, of a random field.

Then, for any generic $\mathbf{u}_0 \in \mathcal{D}$, we define the predictor $p(\mathbf{Z}; \mathbf{u}_0)$ *mean squared (prediction) error* (MSPE)

$$\text{MSPE}(p(\mathbf{Z}; \mathbf{u}_0)) = \text{E}[(Z(\mathbf{u}_0) - p(\mathbf{Z}; \mathbf{u}_0))^2], \tag{3.4.1}$$

which is widely used in prediction error problems. Given that a sample of n locations has been drawn, the MSPE is minimized by the conditional expectation

$$p(\mathbf{Z}; \mathbf{u}_0) = \text{E}[Z(\mathbf{u}_0)|\mathbf{Z}], \tag{3.4.2}$$

which is known to be the optimal predictor.

Obtaining predictor (3.4.2) can result in a daunting task even if the true law of the random field Z is known. Hence, the use of linear predictors

$$p(\mathbf{Z}; \mathbf{u}_0) = \lambda_0 + \sum_{i=1}^{n} \lambda_i Z(\mathbf{u}_i) = \lambda_0 + \boldsymbol{\lambda}^\top \mathbf{Z}, \tag{3.4.3}$$

is a very common answer. In this case MSPE (3.4.1) is the squared mean of the prediction error plus its variance

$$\begin{aligned} \text{E}[(Z(\mathbf{u}_0) - \lambda_0 - \boldsymbol{\lambda}^\top \mathbf{Z})^2] &= (\mu(\mathbf{u}_0) - \lambda_0 - \boldsymbol{\lambda}^\top \boldsymbol{\mu})^2 + \text{V}[(Z(\mathbf{u}_0) - \lambda_0 - \boldsymbol{\lambda}^\top \mathbf{Z})^2] \\ &= (\mu(\mathbf{u}_0) - \lambda_0 - \boldsymbol{\lambda}^\top \boldsymbol{\mu})^2 + \sigma_0^2 - 2\boldsymbol{\lambda}^\top \mathbf{c}_0 + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda}, \end{aligned} \tag{3.4.4}$$

where $\mathrm{E}[\mathbf{Z}] = \boldsymbol{\mu}$, $\mathrm{Cov}(\mathbf{Z}, Z(\mathbf{u}_0)) = \mathbf{c}_0$ and $\mathrm{V}[\mathbf{Z}] = \boldsymbol{\Sigma}$ and the expectation and variance at point $\mathbf{u}_0$ are $\mathrm{E}[Z(\mathbf{u}_0)] = \mu(\mathbf{u}_0)$ and $\mathrm{V}[Z(\mathbf{u}_0)] = \sigma_0^2$, respectively. Once the the squared mean term of (3.4.4) is made null, in order to find the *best linear predictor* (BLP), we need to find the values of λ_0 and $\boldsymbol{\lambda}$ minimizing the variance term. These conditions are met by

$$\lambda_0 = \mu(\mathbf{u}_0) - \boldsymbol{\lambda}^\top \boldsymbol{\mu} \tag{3.4.5}$$

and

$$\boldsymbol{\lambda} = \boldsymbol{\Sigma}^{-1}\mathbf{c}_0, \tag{3.4.6}$$

provided $\boldsymbol{\Sigma}$ is invertible. Equations (3.4.5) and (3.4.6) together are the weighting system that, substituted in linear predictor (3.4.3), leads to the *simple kriging* predictor

$$\hat{Z}_{\mathrm{sk}}(\mathbf{u}_0) = \mathbf{c}_0^\top \boldsymbol{\Sigma}^{-1}(\mathbf{Z} - \boldsymbol{\mu}) + \mu(\mathbf{u}_0).$$

The minimized prediction error

$$\sigma^2_{\mathrm{sk}}(\mathbf{u}_0) = \sigma_0^2 - \mathbf{c}_0^\top \boldsymbol{\Sigma}^{-1}\mathbf{c}_0$$

is also known as *kriging variance*. The simple kriging takes his name from the fact that the assumption of knowing the spatial mean function, $\mu(\cdot)$, is needed. By relaxing this assumption we obtain the ordinary kriging and universal kriging (see Table 3.1).

Stein (1999) states that if the process is Gaussian, then under mean squared prediction error the simple kriging predictor is the best predictor since the BLP is equal to the conditional expectation. Moreover, still in the case of a Gaussian process, he adds that the conditional distribution of $\hat{Z}_{sk}(\mathbf{u}_0)$ given the data, $\mathbf{Z} = \mathbf{z}$, is Gaussian $\mathcal{N}(\lambda_0 + \boldsymbol{\lambda}^\top \mathbf{z}, \sigma_0^2 - \mathbf{c}_0^\top \boldsymbol{\Sigma}^{-1}\mathbf{c}_0)$.

3.4.2 Ordinary kriging

By supposing the mean structure of the process unknown but fixed all over the domain (see Table 3.1), the predictor we obtain by minimizing the MSPE is the *ordinary kriging* predictor which is known to be optimal. The

resulting model comes from (3.1.5) for which a fixed mean has been assumed

$$Z(\mathbf{u}) = \mu + \delta(\mathbf{u}), \tag{3.4.7}$$

where $\delta(\mathbf{u})$ is a zero-mean intrinsic stationary process with known variance structure, $\boldsymbol{\Gamma} = [\gamma(\mathbf{u}-\mathbf{u}')]$. The predictor at the generic location $\mathbf{u}_0$ is formally equal to (3.4.3), where $\lambda_0 = 0$ has been set,

$$p(\mathbf{Z}; \mathbf{u}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{u}_i) = \boldsymbol{\lambda}^\top \mathbf{Z} \tag{3.4.8}$$

and the following condition on the λ coefficients must hold

$$\sum_{i=1}^{n} \lambda_i = \mathbf{1}^\top \boldsymbol{\lambda} = 1. \tag{3.4.9}$$

The condition on the coefficients guarantees that linear predictor (3.4.8) is unbiased, $\mathrm{E}[p(\mathbf{Z}; \mathbf{u}_0)] = \mu = \mathrm{E}[Z(\mathbf{u}_0)]$, and once the MSPE is minimized leads to the *best linear unbiased predictor* (BLUP).

Let us define the vector collecting the squared differences between values at the location $\mathbf{u}_0$ to predict and the ones the sampled locations,

$$\mathbf{y}_0 = [(Z(\mathbf{u}_0) - Z(\mathbf{u}_1))^2, \ldots, (Z(\mathbf{u}_0) - Z(\mathbf{u}_1))^2]^\top,$$

and the matrix of the squared differences between the values at sampled locations, $\mathbf{Y} = [\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top]^\top$. Then, under condition (3.4.8), the MSPE of predictor (3.4.8) is

$$\begin{aligned} \mathrm{E}[(Z(\mathbf{u}_0) - \boldsymbol{\lambda}^\top \mathbf{Z})^2] - m(\mathbf{1}^\top \boldsymbol{\lambda} - 1) &= \boldsymbol{\lambda}^\top \mathrm{E}[\mathbf{y}_0] - \frac{1}{2} \boldsymbol{\lambda}^\top \mathrm{E}[\mathbf{Y}] \boldsymbol{\lambda} - m(\mathbf{1}^\top \boldsymbol{\lambda} - 1) \\ &= 2\boldsymbol{\lambda}^\top \boldsymbol{\gamma}_0 - \boldsymbol{\lambda}^\top \boldsymbol{\Gamma} \boldsymbol{\lambda} - m(\mathbf{1}^\top \boldsymbol{\lambda} - 1), \end{aligned} \tag{3.4.10}$$

where $\boldsymbol{\gamma}_0 = \mathrm{E}[\mathbf{y}_0]$, $\boldsymbol{\Gamma} = \mathrm{E}[\mathbf{Y}]$ and m is a Lagrange multiplier ensuring that the condition on the coefficients is satisfied. Equation (3.4.10) is minimized by the following coefficient vector and value of the Lagrange multiplier, respectively:

$$\boldsymbol{\lambda} = \boldsymbol{\Gamma}^{-1} \left(\boldsymbol{\gamma}_0 + \mathbf{1} \frac{1 - \mathbf{1}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}_0}{\mathbf{1}^\top \boldsymbol{\Gamma}^{-1} \mathbf{1}} \right)$$

and

$$m = -\frac{1 - \mathbf{1}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}_0}{\mathbf{1}^\top \boldsymbol{\Gamma}^{-1} \mathbf{1}},$$

where m is the estimator of the mean, provided $\boldsymbol{\Gamma}$ is invertible. The ordinary kriging predictor results

$$\hat{Z}_{\text{ok}}(\mathbf{u}_0) = \left(\boldsymbol{\gamma}_0 + \mathbf{1} \frac{1 - \mathbf{1}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}_0}{\mathbf{1}^\top \boldsymbol{\Gamma}^{-1} \mathbf{1}} \right)^\top \boldsymbol{\Gamma}^{-1} \mathbf{Z}$$

having kriging variance

$$\sigma^2_{\text{ok}}(\mathbf{u}_0) = \boldsymbol{\gamma}_0^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}_0 - \frac{(\mathbf{1}^\top \boldsymbol{\Gamma}^{-1} \boldsymbol{\gamma}_0 - 1)^2}{\mathbf{1}^\top \boldsymbol{\Gamma}^{-1} \mathbf{1}}$$

A stronger hypothesis need to be posed in order to reformulate the ordinary kriging weighting system in terms of the covariogram function. Assuming model (3.4.7) still holds, component $\delta(\mathbf{u})$ needs to be a second-order stationary random field with known variance structure, $\boldsymbol{\Sigma}$. In this way MSPE (3.4.10) becomes

$$C(\mathbf{0}) + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \mathbf{c}_0 - m(\mathbf{1}^\top \boldsymbol{\lambda} - 1),$$

leading to the corresponding ordinary kriging weighting system

$$\boldsymbol{\lambda} = \boldsymbol{\Sigma}^{-1} \left(\mathbf{c}_0 + \mathbf{1} \frac{1 - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{c}_0}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}} \right)$$

and

$$m = -\frac{1 - \mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{c}_0}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}.$$

The corresponding kriging variance is

$$\sigma^2_{\text{ok}}(\mathbf{u}_0) = C(\mathbf{0}) - \mathbf{c}_0^\top \boldsymbol{\Sigma}^{-1} \mathbf{c}_0 - \frac{(\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{c}_0 - 1)^2}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}}.$$

3.4.3 Universal kriging

Let us recall that $\mu(\mathbf{u}) = \mathrm{E}[Z(\mathbf{u})]$ is the large-scale variation of the process at location $\mathbf{u}$, depending on it indirectly from auxiliary functions known over the domain. Assuming that the mean structure is function of the spatial location, i.e. it varies over the domain, corresponds to adopt the mean model

(3.1.5). Moreover, if the function $\mu(\cdot)$ is unknown, the underlying hypothesis is a further relaxation of the one at the basis of the ordinary kriging.

Let us assume that $\delta(\cdot)$ is a zero-mean intrinsic stationary random field with semivariogram $\gamma(\cdot)$. Moreover, let $\mu(\mathbf{u})$ be an unknown combination of known functions $\{f_0(\mathbf{u}), \ldots, f_p(\mathbf{u})\}$. Then, the mean model can be rewritten as

$$Z(\mathbf{u}) = \mathbf{x}(\mathbf{u})^\top \boldsymbol{\beta} + \delta(\mathbf{u}), \tag{3.4.11}$$

where $\mathbf{x}(\mathbf{u}) = [f_0(\mathbf{u}), \ldots, f_p(\mathbf{u})]^\top$ is the $(p+1)$-dimensional vector collecting the auxiliary functions $f_i(\mathbf{u})$ and $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$ is a vector of unknown parameters. Hence, the observed data can be written as

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta}, \tag{3.4.12}$$

where $\mathbf{X} = [\mathbf{x}(\mathbf{u}_1)^\top, \ldots, \mathbf{x}(\mathbf{u}_n)^\top]^\top$ is the $n \times (p+1)$ matrix collecting the values of the auxiliary functions $\{f_0(\cdot), \ldots, f_p(\cdot)\}$ computed at the sampled locations $\{\mathbf{u}_i \in S, i = 1, \ldots, n\}$.

The linear predictor

$$p(\mathbf{Z}; \mathbf{u}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{u}_i) = \boldsymbol{\lambda}^\top \mathbf{Z} \tag{3.4.13}$$

at a generic location $\mathbf{u}_0$ must meet the condition on the coefficients

$$\boldsymbol{\lambda}^\top \mathbf{X} = \mathbf{x}_0^\top. \tag{3.4.14}$$

where $\mathbf{x}_0 = [f_0(\mathbf{u}_0), \ldots, f_p(\mathbf{u}_0)]^\top$. This condition is necessary and sufficient to guarantee uniform unbiasedness of the predictor, $\mathrm{E}[p(\mathbf{Z}; \mathbf{u}_0)] = \mathrm{E}[\boldsymbol{\lambda}^\top \mathbf{Z}] = \boldsymbol{\lambda}^\top \mathbf{X}\boldsymbol{\beta}$, which is equal to $\mathbf{x}_0^\top \boldsymbol{\beta} = \mathrm{E}[Z(\mathbf{u}_0)]$, for all $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$. Once the MSPE is minimized, predictor (3.4.13) is the BLUP. By posing $\mathbf{1}^\top \boldsymbol{\lambda} = 1$ and $f_0 = 1$, we obtain the ordinary kriging assumption (3.4.9) on the predictor.

The MSPE of the *universal kriging* predictor (3.4.13) is

$$\mathrm{E}[(Z(\mathbf{u}_0) - \boldsymbol{\lambda}^\top \mathbf{Z})^2] - (\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\lambda})^\top \mathbf{m},$$

where $\mathbf{m} = [m_0, \ldots, m_p]^\top$ is a vector of $p+1$ Lagrange multipliers ensuring that condition (3.4.9) is satisfied. By assuming $f_0 = 1$ we obtain one of the unbiasedness conditions, $\mathbf{1}^\top \boldsymbol{\lambda} = 1$, which along with the use of model (3.4.11)

and the corresponding data formulation (3.4.12) under condition (3.4.14) lead to the mean squared term of the MSPE

$$\begin{aligned}\mathrm{E}[(Z(\mathbf{u}_0) - \boldsymbol{\lambda}^\top \mathbf{Z})^2] &= \mathrm{E}[(\mathbf{x}_0^\top + \delta(\mathbf{u}_0) - \boldsymbol{\lambda}^\top \mathbf{X}\boldsymbol{\beta} - \boldsymbol{\lambda}^\top \boldsymbol{\delta})^2] \\ &= \mathrm{E}[(\delta(\mathbf{u}_0) - \boldsymbol{\lambda}^\top \boldsymbol{\delta})^2] \\ &= \boldsymbol{\lambda}^\top \mathrm{E}[\mathbf{d}_0] - \frac{1}{2}\boldsymbol{\lambda}^\top \mathrm{E}[\mathbf{D}]\boldsymbol{\lambda} \\ &= 2\boldsymbol{\lambda}^\top \boldsymbol{\gamma}_0 - \boldsymbol{\lambda}^\top \boldsymbol{\Gamma}\boldsymbol{\lambda},\end{aligned}$$

where the vector collecting the squared differences of process $\delta(\cdot)$ between values at the location $\mathbf{u}_0$ to predict and the ones the sampled locations,

$$\mathbf{d}_0 = [(\delta(\mathbf{u}_0) - \delta(\mathbf{u}_1))^2, \ldots, (\delta(\mathbf{u}_0) - \delta(\mathbf{u}_1))^2]^\top,$$

and the matrix of the squared differences between the values at sampled locations, $\mathbf{D} = [\mathbf{d}_1^\top, \ldots, \mathbf{d}_n^\top]^\top$ are involved. Hence, the resulting MSPE to be minimized is

$$2\boldsymbol{\lambda}^\top \boldsymbol{\gamma}_0 - \boldsymbol{\lambda}^\top \boldsymbol{\Gamma}\boldsymbol{\lambda} - (\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\lambda})^\top \mathbf{m}. \tag{3.4.15}$$

Equation (3.4.15) is minimized for the following values of the coefficient vector and Lagrange multiplier vector, respectively:

$$\boldsymbol{\lambda} = \boldsymbol{\Gamma}^{-1}(\boldsymbol{\gamma}_0 + \mathbf{X}(\mathbf{X}^\top \boldsymbol{\Gamma}^{-1}\mathbf{X})^{-1}(\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0))$$

and

$$\mathbf{m} = -(\mathbf{X}^\top \boldsymbol{\Gamma}^{-1}\mathbf{X})^{-1}(\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0),$$

provided all the matrices requiring inverse are invertible. The resulting universal kriging predictor is defined as

$$\hat{Z}_{\mathrm{uk}}(\mathbf{u}_0) = (\boldsymbol{\gamma}_0 + \mathbf{X}(\mathbf{X}^\top \boldsymbol{\Gamma}^{-1}\mathbf{X})^{-1}(\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0))^\top \boldsymbol{\Gamma}^{-1}\mathbf{Z}$$

with kriging variance

$$\sigma^2_{\mathrm{uk}}(\mathbf{u}_0) = \boldsymbol{\gamma}_0^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0 - (\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0))^\top (\mathbf{X}^\top \boldsymbol{\Gamma}^{-1}\mathbf{X})^{-1}(\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma}_0)).$$

Universal kriging predictor (3.4.3) relies on the assumption that process Z is intrinsically stationary; let us strengthen this hypothesis by considering

it second-order stationary. This allow us to rewrite MSPE (3.4.15) in terms of the covariogram

$$C(\mathbf{0}) + \boldsymbol{\lambda}^\top \boldsymbol{\Sigma} \boldsymbol{\lambda} - 2\boldsymbol{\lambda}^\top \mathbf{c}_0 - (\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\lambda})^\top \mathbf{m},$$

which is minimized by vectors

$$\boldsymbol{\lambda} = \boldsymbol{\Sigma}^{-1}(\mathbf{c}_0 + \mathbf{X}(\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}(\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{c}_0))$$

and

$$\mathbf{m} = -(\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1}(\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{c}_0).$$

The corresponding kriging variance is

$$\sigma^2_{\text{uk}}(\mathbf{u}_0) = C(\mathbf{0}) - \mathbf{c}_0^\top \boldsymbol{\Sigma}^{-1} \mathbf{c}_0 + (\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{c}_0))^\top (\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} (\mathbf{x}_0 - \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{c}_0)).$$

3.4.4 Practical applications of the kriging predictor

The theoretical results of this section are far from being employable in practical applications. All the kriging methods, here presented or not, relies on the assumption that the process variance structure (covariogram or semivariogram) is known. Actually, it has to be estimated from the observed values using the techniques of Section 3.3, once a valid semivariogram (covariogram) model has been chosen. Simply put, prediction is based on the same data used for the estimation of the semivariogram parameters leading to coefficients that are no longer linear since they depend on the very data (Stein, 1999). Therefore, the resulting kriging predictors differs from optimal kriging weights (BLUP) derived in the previous section. Moreover, the corresponding kriging variances are not equal to the real variances of the subotptimal kriging predictors employed. This may lead to several problems in applications.

Chilès and Delfiner (1999) state that a semivariogram model whose behaviour near the origin is correctly represented leads to good estimates even if the model is seriously misspecified everywhere else (remember that the semivariogram behaviour near the origin completely characterizes the continuity and regularity of the process in the space, as stated in Subsection 3.2.2).

Cressie (1993) extends analyses on the effects of the estimation of the parameter of a semivariogram model. Let us firstly consider the effect of the nugget parameter, defined as $\gamma(\mathbf{h}) \to \tau^2 > 0$, as $\|\mathbf{h}\| \to 0$, on the semivariogram. Recall from Section 3.2 that its magnitude is given by

$$\tau^2 = \sigma_\eta^2 + \sigma_\epsilon^2,$$

where σ_η^2 and σ_ϵ^2 are the micro-scale process $\eta(\mathbf{u})$ sill and the measurement error $\epsilon(\mathbf{u})$ variance, respectively. In practice it is hard to determine the micro-scale process variance structure because the shorter observed lag is rarely short enough to model the semivariogram behaviour near the origin. Theoretically, the measurement errors may come from repeated measures or laboratory errors; however, in practice, it is not possible to assess its extent. These aspects combined lead to an estimation of the nugget effect through interpolation of semivariograms estimates near the origin. To safeguard inference from this inconvenience one should predict the *measurement-error-free* process $S(\mathbf{u}) = \mu(\mathbf{u}) + W(\mathbf{u}) + \eta(\mathbf{u})$ known as signal model. The resulting (ordinary or universal) kriging predictor, $\hat{S}(\mathbf{u})$, is obtained using the corresponding kriging equations where the ith element of vector $\boldsymbol{\gamma}_0$ is posed equal to the measurement error process variance, σ_ϵ^2, whenever $\mathbf{u} \in S$. Therefore, the kriging variance is modified by subtracting σ_ϵ^2 from the corresponding equation. Analogous modifications can be obtained for the simple kriging or the kriging equations expressed in terms of the covariogram. The signal predictor $\hat{S}(\mathbf{u})$ turns out to be non-exact leading to more smoothed values as larger as σ_ϵ^2 is. The most extreme scenario happens when the process has a fixed trend and its randomness is completely due to measurement errors $Z(\cdot) = \mu + \epsilon(\cdot)$. The resulting predictor is given by

$$\hat{Z}(\mathbf{u}) = \begin{cases} \bar{Z}, & \text{if } \mathbf{u} \notin S; \\ Z(u_i), & \text{if } \mathbf{u} = \mathbf{u}_i \in S, \end{cases}$$

where $\bar{Z}$ is the sample mean. Furthermore, it results that $\hat{S} = \bar{Z}$ (Cressie, 1993), that is the signal predictor corresponds to the SRSWoR estimator in predictive form (Bolfarine and Zacks, 1992). Then, one must pay attention when predicting outside the set of sampled locations since the kriging equa-

tions are the same for the the process $Z(\cdot)$ and $S(\cdot)$, but the kriging variances differ in the measurement error variance σ^2_ϵ.

Let us now focus on the sill parameter of the process Z. When existing, it has been defined as $\lim_{\|\mathbf{h}\|\to\infty} \gamma(\|\mathbf{h}\|)$ and its value is given by the sum of the sills of the large- and micro-scale processes plus the measurement error variance

$$\sigma^2 = \sigma^2_W + \sigma^2_\eta + \sigma^2_\epsilon,$$

where σ^2_W is the sill of the large-scale process W, when $\mu(\mathbf{u}) = \mu$. It is worth noticing that the micro-scale sill, σ^2_η, is larger than the estimated one and, therefore, the sill of the process, σ^2, is larger than the sum of the large-scale sill and the estimated nugget effect. This is due to the fact that the estimated nugget effect is obtained by extrapolating an experimental semivariogram at lags close to zero.

The range parameter has been defined as the maximum lag at which the process shows spatial dependence. It may be inaccurately thought to use the range estimate in order to provide a kriging neighbourhood for the location to predict and consider only the locations less distant. However, this approach could lead to some problems since the nugget effect would not be considered.

Cressie (1993) concludes stating that all the three parameters estimates are important and have a considerable effect on the kriging estimates. Moreover, since the kriging predictor is suboptimal, the kriging variance is smaller than the effective one due to the semivariogram parameters estimation

$$\sigma^2_{\mathrm{k}} \leq \mathrm{E}[(\hat{\hat{p}}(\mathbf{Z};\mathbf{u}) - Z(\mathbf{u}))^2],$$

where the second hat in the predictor emphasises the fact that the semivariogram parameters are estimated.

For a deeper insight into (mathematical and statistical) stability of the kriging predictor one can have a look at the book by Cressie (1993).

Likelihood semivariogram fitting methods (Subsection 3.3.2) strongly rely on the Gaussianity assumption of the random field; however, assuming this is highly restrictive. Maximum likelihood estimation methods are very sensitive to anomalous observations. Least squares methods (Subsection 3.3.2) can be thought of as a valid fitting alternative; nevertheless, they are still sensitive to

aberrant values even if less so (Stein, 1999). Diggle et al. (1998) empirically proved that Monte Carlo Markov Chain methods are a reliable prediction tool for non-Gaussian random fields.

The nugget parameter estimation is not the only one affected by lag problems. Indeed semivariogram fitting does not perform well when the sample is small and just few lags are observed for a sufficiently large domain. Therefore, the search for sampling designs, not necessarily probabilistic, that may lead to good estimates is still an open topic and, perhaps, it may vary case by case. In order to improve the estimation of the parameter of the Matèrn model, Stein (1999) suggest to add a few closely packed observations to an evenly spaced design; conceptually, this idea can be extended to any other semivariogram model.

Chapter 4

Design-based spatial estimation revised

In design-based spatial inference the spatial information has usually been used in order to obtain efficient sampling designs. According to Webster and Oliver (2007) several authors have used the exponential semivariogram as the basis for their theoretical studies on sampling startegies efficiency (e.g. Cochran, 1946; Yates, 1948; Quenouille, 1949; Matérn, 1960).

Since the 1990s we assisted to a reappraisal of the design-based techniques for inference on spatial data: de Gruijter and ter Braak (1990) were among the firsts to assess that these can be usefully applied to spatial datasets provided a random sampling is practicable. Brus and de Gruijter (1997) gave a first, and probably the only one, individual estimator which assigns the mean of the values observed at the set of sampled values belonging to a subarea to all the points in it. A part from them, researchers focused on the estimation of population quantities (e.g. mean and total). Among the last works regarding this topic we point out the papers by Cicchitelli and Montanari (2012) and Ghosh et al. (2012).

In this chapter a new technique for design-based spatial inference is developed. We propose an estimator able to use the spatial information known at population level for estimating, firstly, the individual values (*point estimation*) and, then, a summary population value (i.e. mean or total). In the following section the basic concepts behind this idea are presented. Sec-

tion 4.2 covers the achieved point estimator along with its properties and variance estimation in a finite population case. In Section 4.3 the point estimator is extended to a continuous domain. The population synthetic value estimation along with its properties is discussed in Section 4.4.

4.1 Basic concepts

The basic idea in design-based spatial inference is to use the spatial information known before sampling in order to obtain sampling designs more efficient than simple random sampling (SRS), which for cost and estimation reasons may not be the most suitable one. In this chapter we introduce and develop the idea of using the spatial information at estimation level rather than at sampling design level.

The starting point is a deterministic interpolator which can not be associated to any uncertainty measure. When the set of known locations is thought of as a realization of a probabilistic sampling design, then the deterministic interpolator itself can be randomized.

The weights of the deterministic interpolator chosen as the starting point depend on the spatial lag between the locations of the whole domain which are assumed to be known. Therefore, as typical of the design-based inference paradigm, the weighting system involved in the estimator can be constructed for the whole population before sampling.

The underlying concept is that the locations labels defined as the spatial coordinates may be regarded as population information. Viewing the labels as informative, simple random sampling without replacement (SRSWoR) seems the most suitable sampling design.

4.2 Estimating individual quantities in design-based finite spatial population

Let us consider a bounded finite spatial domain belonging to the two-dimensional Euclidean space $\mathcal{D} \subset \mathbb{R}^2$, and the function $z(\mathbf{u})$ belonging to the set $C^1(\mathcal{D})$. Moreover, let $L = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be a set of n locations, each

denoted by its coordinates $\mathbf{u}_i = (x_i, y_i)$, where the value of $z(\cdot)$ is known. Shepard (1968) proposed a smooth (continuous and once differentiable) function able to approximate $z(\mathbf{u})$ on $\mathcal{D}$ by means of interpolation, named inverse distance weighted (IDW) interpolator.

Let $\mathbf{w}(\mathbf{u})$ be a n-dimensional standardized weighting vector whose elements are a monotonic decreasing function of the Euclidean distance between each location belonging to L and an arbitrary one $\mathbf{u} = (x, y) \in \mathcal{D}$

$$w_i(\mathbf{u}) = \frac{\|\mathbf{u}_i - \mathbf{u}\|^{-\alpha}}{\sum_{j=1}^{n} \|\mathbf{u}_j - \mathbf{u}\|^{-\alpha}} = \frac{d_i^{-\alpha}}{\sum_{j=1}^{n} d_j^{-\alpha}}, \tag{4.2.1}$$

where $\alpha \in \mathbb{R}^+$ and $d_i = \|\mathbf{u}_i - \mathbf{u}\|$ is the Euclidean distance between two points. The proposed weighting system is constructed such that the influence of values at $\mathbf{u}_1, \ldots, \mathbf{u}_n$ decreases as the distance from the arbitrary point $\mathbf{u}$ increases, according to Tobler's law (Tobler, 1970).

The resulting interpolated value at an arbitrary location is

$$\hat{z}(\mathbf{u}) = \begin{cases} \mathbf{z}^\top \mathbf{w}(\mathbf{u}), & \text{if } \mathbf{u} \notin L; \\ z(\mathbf{u}), & \text{otherwise;} \end{cases} \tag{4.2.2}$$

where $\mathbf{z} = [z(\mathbf{u}_1), \ldots, z(\mathbf{u}_n)]^\top$ is the vector of the observed values at the point belonging to L. Interpolation involves evaluating $z(\cdot)$ at an arbitrary location $\mathbf{u}$ not belonging to the set L and reproducing the observed value when such location belongs to the set. Therefore, the IDW interpolator is exact.

The continuity assumption for the interpolating function holds:

$$\lim_{\mathbf{u} \to \mathbf{u}_i} \hat{z}(\mathbf{u}) = z(\mathbf{u}),$$

since as $\mathbf{u}$ approaches $\mathbf{u}_i$, the Euclidean distance $\|\mathbf{u}_i - \mathbf{u}\|$ approaches 0 so that the numerator and denominator of the ith term in (4.2.1) diverge while the others remain bounded.

First differentiability holds for any point in $\mathcal{D}$:

$$\frac{\partial \hat{z}(\mathbf{u})}{\partial x} = -\alpha \frac{\sum_{i=1}^{n} \sum_{j \neq i} d_i^{-2-\alpha} d_j^{-\alpha} (x - x_i) z(\mathbf{u}_i)(z(\mathbf{u}_i) - z(\mathbf{u}_j))}{\left(\sum_{i=1}^{n} d_i^{-\alpha} \right)^2},$$

where substituting $(y - y_i)$ to $(x - x_i)$ gives the expression for $\partial \hat{z}(\mathbf{u})/\partial y$.

Shepard (1968) suggested that a choice of $\alpha = 2$ for the exponent is the best, leading to easier computation and more satisfactory empirical results. Without loss of generality from now on, we assume $\alpha = 2$.

4.2.1 The IDW point estimator

Let us suppose that a (regular) grid has been superimposed over the domain $\mathcal{D}$ or that we are interested in estimating the value of the variable under study at a finite number of locations in the domain $\mathcal{D}$. In the second case we suppose that the coordinates are known for each location before the sampling. In the first case, without loss of generality, we suppose to know the coordinates of the centroid $\mathbf{u}_i = (x_i, y_i)$ of each sub-area. Both the cases are linked to the idea of finite population inference since we can think of a population $\mathcal{P}$ of fixed size N.

In spatial datasets the labels can be considered as informative since they corresponds to the locations' coordinates. In this thesis we propose a way to implement this information at the estimation level; therefore, in order not to use the same information at sampling level as well, simple random sampling sampling without replacement (SRSWoR) seems the most suitable design. Furthermore, let us suppose that the values $z(\mathbf{u}_1), \ldots, z(\mathbf{u}_N)$ come from a fixed unknown deterministic function, $z : \mathbb{R}^2 \to \mathbb{R}$, evaluated at the sampled location, $\mathbf{u}_1, \ldots, \mathbf{u}_n$.

Under the finite population design-based framework, the known locations, $\mathbf{u}_1, \ldots, \mathbf{u}_n$, belonging to the set L of equation (4.2.2) are considered as the realization of a probabilistic random sampling, denoted as s. Formally speaking s is an element of the the σ-algebra $\mathcal{S}$ of all the possible samples of the

sample space $\mathfrak{S}$. From now on we will use S as an arbitrary element belonging to $\mathcal{S}$, and s as its realization. Inclusion in the generic sample S is managed by the Bernoulli random variable $Q_i = \mathcal{I}_{(i \in S)}$ assuming value 1 if the location $\mathbf{u}_i$ belongs to the arbitrary sample S. On the contrary exclusion from the sample is managed by complement to 1 of the Bernoulli random variable managing inclusion, $1 - \mathcal{I}_{(i \in S)}$, since inclusion in and exclusion from the sample are two mutually exclusive events.

It is now possible to rewrite interpolator (4.2.2) for the generic ith population element as a design-based point estimator:

$$\begin{aligned}\hat{z}(\mathbf{u}_i) &= \frac{\mathcal{I}_{(i \in S)} z(\mathbf{u}_i) + (1 - \mathcal{I}_{(i \in S)}) \sum_{j \neq i} \phi_{ij} z(\mathbf{u}_j) \mathcal{I}_{(j \in S)}}{\mathcal{I}_{(i \in S)} + (1 - \mathcal{I}_{(i \in S)}) \sum_{j \neq i} \phi_{ij} \mathcal{I}_{(j \in S)}} \\ &= \frac{\mathcal{I}_{(i \in S)} z(\mathbf{u}_i) + \sum_{j \neq i} \phi_{ij} z(\mathbf{u}_j) \mathcal{I}_{(j \in S)} - \mathcal{I}_{(i \in S)} \sum_{j \neq i} \phi_{ij} z(\mathbf{u}_j) \mathcal{I}_{(j \in S)}}{\mathcal{I}_{(i \in S)} + \sum_{j \neq i} \phi_{ij} \mathcal{I}_{(j \in S)} - \mathcal{I}_{(i \in S)} \sum_{j \neq i} \phi_{ij} \mathcal{I}_{(j \in S)}}, \quad (4.2.3)\end{aligned}$$

where $\phi_{ij} = d_{ij}^{-2} = \|\mathbf{u}_j - \mathbf{u}_i\|^{-2}$ is the inverse squared Euclidean distance between the ith and the jth population locations. As the IDW interpolator, the achieved IDW point estimator attributes the observed value to the location ith if it is sampled, and otherwise evaluates $z(\cdot)$. The resulting IDW point estimator is a ratio of sums of linear combinations of random quantities and, therefore, its properties will be analytically obtained only in approximate form.

In the design-based finite population framework the information concerning the arbitrary sample S is usually summarized in the N-dimensional random vector $\boldsymbol{\mathcal{Q}} = [Q_1, \ldots, Q_N]^\top$ of Bernoulli random variables managing inclusion in the sample, i.e. $Q_i = \mathcal{I}_{(i \in S)}$. Once a sample s is drawn from the population, the realization of random vector $\boldsymbol{\mathcal{Q}}$ is the N-dimensional vector $\mathbf{q}$ containing n unit values in correspondence of the sampled locations and zero otherwise.

Let us define the $N \times N$ symmetric matrix containing the inverse squared Euclidean distances between the population locations and having null diag-

onal

$$\mathbf{\Phi} = \begin{bmatrix} 0 & \|\mathbf{u}_2 - \mathbf{u}_1\|^{-2} & \cdots & \|\mathbf{u}_N - \mathbf{u}_1\|^{-2} \\ \|\mathbf{u}_1 - \mathbf{u}_2\|^{-2} & 0 & \cdots & \|\mathbf{u}_N - \mathbf{u}_2\|^{-2} \\ \vdots & \vdots & \ddots & \vdots \\ \|\mathbf{u}_1 - \mathbf{u}_N\|^{-2} & \|\mathbf{u}_2 - \mathbf{u}_N\|^{-2} & \cdots & 0 \end{bmatrix}.$$

By using matrix $\mathbf{\Phi}$ and the random vector $\boldsymbol{\mathcal{Q}}$, we define the weighting vector

$$\begin{aligned} \mathbf{h}_i &= \boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i^* - Q_i \boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i \\ &= Q_i \mathbf{e}_i + (1 - Q_i) \boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i, \end{aligned} \tag{4.2.4}$$

where $\boldsymbol{\phi}_i = \mathbf{\Phi}\mathbf{e}_i$ is the ith column of matrix $\mathbf{\Phi}$, $\boldsymbol{\phi}_i^* = \boldsymbol{\phi}_i + \mathbf{e}_i$, $\mathbf{e}_i$ is the ith N-dimensional canonical basis and $\circ$ is the Hadamard product. The IDW point estimator (4.2.3) can be rewritten in matrix form as

$$\hat{z}(\mathbf{u}_i) = \mathbf{h}_i^{*\top} \mathbf{z} = (\mathbf{h}_i^\top \mathbf{1}_N)^{-1} \mathbf{h}_i^\top \mathbf{z}, \tag{4.2.5}$$

where $\mathbf{z} = [z(\mathbf{u}_1), \ldots, z(\mathbf{u}_n)]^\top$ is the N-dimensional vector collecting the observed values.

Estimation for all the population locations can be performed at the same time by using the $N \times N$ matrix $\mathbf{H} = [\mathbf{h}_1^\top, \ldots, \mathbf{h}_N^\top]^\top$ collecting row vectors $\mathbf{h}_i^\top$:

$$\begin{aligned} \mathbf{H} &= (\mathbf{1}_N \boldsymbol{\mathcal{Q}}^\top) \circ \mathbf{\Phi}^* - \operatorname{diag}(\boldsymbol{\mathcal{Q}})(\mathbf{1}_N \boldsymbol{\mathcal{Q}}^\top) \circ \mathbf{\Phi} \\ &= \operatorname{diag}(\boldsymbol{\mathcal{Q}}) + \operatorname{diag}(\mathbf{1}_N - \boldsymbol{\mathcal{Q}})(\mathbf{1}_N \boldsymbol{\mathcal{Q}}^\top) \circ \mathbf{\Phi}, \end{aligned}$$

where $\mathbf{\Phi}^* = \mathbf{\Phi} + \mathbf{I_N}$ and $\circ$ is the Hadamard product. The vector of the estimated values is defined as

$$\hat{\mathbf{z}} = \mathbf{H}^* \mathbf{z} = \operatorname{diag}(\mathbf{H}\mathbf{1}_N)^{-1} \mathbf{H}\mathbf{z}, \tag{4.2.6}$$

where $\mathbf{H}^* = [\mathbf{h}_1^{*\top}, \ldots, \mathbf{h}_N^\top]^{*\top}$ is the matrix collecting row vectors $\mathbf{h}_i^{*\top}$.

4.2.2 Expected value

Following Stuart and Ord (1987), the expected value of estimator (4.2.5) is obtained in approximate form using the delta method (Appendix A) since the IDW point estimator is the ratio of sums of linear combinations of random quantities.

Theorem 4.2.1. *The usual delta method first order approximation (A.2) of the expected value of estimator (4.2.5) is*

$$E[\hat{z}(\mathbf{u}_i)] = \frac{z(\mathbf{u}_i) + at_{1i}}{1 + at_{2i}} + O_p(n^{-1}), \tag{4.2.7}$$

where constant

$$a = \frac{N-n}{N-1} \tag{4.2.8}$$

is defined and the following combinations of the population values and function of the Euclidean distances are used:

$$t_{1i} = \sum_{j \neq i} \phi_{ij} z(\mathbf{u}_j) = \boldsymbol{\phi}_i^\top \mathbf{z} \tag{4.2.9}$$

and

$$t_{2i} = \sum_{j \neq i} \phi_{ij} = \boldsymbol{\phi}_i^\top \mathbf{1}_N. \tag{4.2.10}$$

Proof. In (4.2.7), the expected values of the numerator and denominator are easily obtained using the first-order inclusion probabilities

$$\mathrm{E}[Q_i] = \mathrm{E}[\mathcal{I}_{(i \in S)}] = \frac{n}{N} \tag{4.2.11}$$

and second-order inclusion probabilities

$$\mathrm{E}[Q_i Q_j] = \mathrm{E}[\mathcal{I}_{(i \in S)} \mathcal{I}_{(j \in S)}] = \mathrm{E}[\mathcal{I}_{(i \in S)}] \mathrm{E}[\mathcal{I}_{(j \in S | i \in S)}] = \frac{n}{N} \frac{n-1}{N-1}. \tag{4.2.12}$$

Combining these basic results from sampling theory with the delta method first-order approximation, we obtain the IDW point estimator's approximate expected value:

$$
\begin{aligned}
\mathrm{E}[\hat{z}(\mathbf{u}_i)] &= \frac{\mathrm{E}\left[\mathcal{I}_{(i\in S)}z(\mathbf{u}_i) + (1-\mathcal{I}_{(i\in S)})\sum_{j\neq i}\phi_{ij}z(\mathbf{u}_j)\mathcal{I}_{(j\in S)}\right]}{\mathrm{E}\left[\mathcal{I}_{(i\in S)} + (1-\mathcal{I}_{(i\in S)})\sum_{j\neq i}\phi_{ij}\mathcal{I}_{(j\in S)}\right]} + O_p(n^{-1}) \\
&= \frac{\mathrm{E}[\mathcal{I}_{(i\in S)}z(\mathbf{u}_i)] + \sum_{j\neq i}\phi_{ij}z(\mathbf{u}_j)\mathrm{E}[\mathcal{I}_{(j\in S)}] - \sum_{j\neq i}\phi_{ij}z(\mathbf{u}_j)\mathrm{E}[\mathcal{I}_{(i\in S)}\mathcal{I}_{(j\in S)}]}{\mathrm{E}[\mathcal{I}_{(i\in S)}] + \sum_{j\neq i}\phi_{ij}\mathrm{E}[\mathcal{I}_{(j\in S)}] - \sum_{j\neq i}\phi_{ij}\mathrm{E}[\mathcal{I}_{(i\in S)}\mathcal{I}_{(j\in S)}]} + O_p(n^{-1}) \\
&= \frac{\frac{n}{N}z(\mathbf{u}_i) + \frac{n}{N}\left(\sum_{j\neq i}\phi_{ij}z(\mathbf{u}_j) - \frac{n-1}{N-1}\sum_{j\neq i}\phi_{ij}z(\mathbf{u}_j)\right)}{\frac{n}{N} + \frac{n}{N}\left(\sum_{j\neq i}\phi_{ij} - \frac{n-1}{N-1}\sum_{j\neq i}\phi_{ij}\right)} + O_p(n^{-1}) \\
&= \frac{z(\mathbf{u}_i) + at_{1i}}{1+at_{2i}} + O_p(n^{-1}).
\end{aligned}
$$

Under the matrix formulation (4.2.5) of the IDW point estimator, we get the expected value of the weighting vector (4.2.4):

$$\mathrm{E}[\mathbf{h}_i] = \frac{n}{N}(\mathbf{e}_i + a\boldsymbol{\phi}_i),$$

where constant (4.2.8) is retrieved.

The first-order inclusion probabilities defined in (4.2.11) can be collected in the N-dimensional vector

$$\mathrm{E}[\boldsymbol{\mathcal{Q}}] = \frac{n}{N}\mathbf{1}_N \tag{4.2.13}$$

and the second-order inclusion probabilities (4.2.12) in the N-dimensional vector

$$\mathrm{E}[Q_i\boldsymbol{\mathcal{Q}}] = \frac{n}{N}\left(\frac{n-1}{N-1}\mathbf{1}_N + \frac{N-n}{N-1}\mathbf{e}_i\right). \tag{4.2.14}$$

In the previous vector, the second term in brackets manages the second-order inclusion probability of the ith individual in SRSWoR which corresponds to the first-order inclusion probability:

$$\mathrm{E}[Q_i^2] = \mathrm{E}[Q_i] = \frac{n}{N}. \tag{4.2.15}$$

Using vector (4.2.14), we obtain the expected value of vector $\mathbf{h}_i$:

$$\begin{aligned}
\mathrm{E}[\mathbf{h}_i] &= \mathrm{E}[Q_i]\mathbf{e}_i + \mathrm{E}[(1-Q_i)\boldsymbol{\mathcal{Q}}] \circ \boldsymbol{\phi}_i \\
&= \mathrm{E}[Q_i]\mathbf{e}_i + \mathrm{E}[\boldsymbol{\mathcal{Q}}] \circ \boldsymbol{\phi}_i - \mathrm{E}[Q_i\boldsymbol{\mathcal{Q}}] \circ \boldsymbol{\phi}_i \\
&= \frac{n}{N}\mathbf{e}_i + \frac{n}{N}\boldsymbol{\phi}_i - \frac{n}{N}\frac{n-1}{N-1}\boldsymbol{\phi}_i \\
&= \frac{n}{N}\left(\mathbf{e}_i + \frac{N-n}{N-1}\boldsymbol{\phi}_i\right) \\
&= \frac{n}{N}(\mathbf{e}_i + a\boldsymbol{\phi}_i). \qquad (4.2.16)
\end{aligned}$$

The second term in brackets of the vector of the second-order inclusion probabilities (4.2.14) does not appear since the Hadamard multiplication for $\boldsymbol{\phi}_i$ of $\mathbf{e}_i$ produces a null vector because of the iith null element of $\boldsymbol{\phi}_i$ (Appendix B).

Using expected value (4.2.16) and the usual delta method first-order approximation, the approximated expected value (4.2.7) can be rewritten in vector form as

$$\begin{aligned}
\mathrm{E}[\hat{z}(\mathbf{u}_i)] &= \mathrm{E}[\mathbf{h}_i^*]^\top \mathbf{z} = (\mathrm{E}[\mathbf{h}_i]^\top \mathbf{1}_N)^{-1}\mathrm{E}[\mathbf{h}_i]^\top \mathbf{z} + O_p(n^{-1}) \\
&= \frac{z(\mathbf{u}_i) + at_{1i}}{1 + at_{2i}} + O_p(n^{-1}), \qquad (4.2.17)
\end{aligned}$$

where $t_{1i} = \boldsymbol{\phi}_i^\top \mathbf{z}$ and $t_{2i} = \boldsymbol{\phi}_i^\top \mathbf{1}_N$.

The approximated expectation of the vector of estimated values (4.2.6) is vector

$$\mathrm{E}[\hat{\mathbf{z}}] = \mathrm{diag}(\mathbf{1}_N + a\mathbf{t}_2)^{-1}(\mathbf{z} + a\mathbf{t}_1) + O_p(n^{-1}), \qquad (4.2.18)$$

where the N-dimensional vectors $\mathbf{t}_1 = \boldsymbol{\Phi}\mathbf{z} = [t_{11}, \ldots, t_{1n}]^\top$ and $\mathbf{t}_2 = \boldsymbol{\Phi}\mathbf{1}_N = [t_{21}, \ldots, t_{2n}]^\top$ are involved. The previous expectation comes directly form the one of the weighting matrix $\mathbf{H}$:

$$\begin{aligned}
\mathrm{E}[\mathbf{H}] &= \mathrm{diag}(\mathrm{E}[\boldsymbol{\mathcal{Q}}]) + (\mathbf{1}_N\mathrm{E}[\boldsymbol{\mathcal{Q}}]^\top) \circ \boldsymbol{\Phi} - \mathrm{E}[\mathrm{diag}(\boldsymbol{\mathcal{Q}})(\mathbf{1}_N\boldsymbol{\mathcal{Q}}^\top)] \circ \boldsymbol{\Phi} \\
&= \frac{n}{N}\mathbf{I}_N + \frac{n}{N}\boldsymbol{\Phi} - \frac{n}{N}\frac{n-1}{N-1}\boldsymbol{\Phi} = \frac{n}{N}\left(\mathbf{I}_N + \frac{N-n}{N-1}\boldsymbol{\Phi}\right) \\
&= \frac{n}{N}(\mathbf{I}_N + a\boldsymbol{\Phi}),
\end{aligned}$$

where first- and second-order inclusion probabilities vectors, (4.2.13) and (4.2.14) respectively, are used. It is straightforward to obtain approximated

expected value (4.2.18) as follows

$$\begin{aligned}\mathrm{E}[\hat{\mathbf{z}}] &= \mathrm{E}[\mathrm{diag}(\mathbf{H}\mathbf{1}_N)^{-1}\mathbf{H}]\mathbf{z} = \mathrm{diag}(\mathrm{E}[\mathbf{H}]\mathbf{1}_N)^{-1}\mathrm{E}[\mathbf{H}]\mathbf{z} + O_p(n^{-1}) \\ &= \mathrm{diag}(\mathbf{1}_N + a\mathbf{t}_2)^{-1}(\mathbf{z} + a\mathbf{t}_1) + O_p(n^{-1}). \qquad \square\end{aligned}$$

4.2.3 Variance

For the same reasons as for the expected value, the variance of the IDW point estimator is obtained in approximate form using the delta method (Appendix A).

Theorem 4.2.2. *Following the standard first-order delta method approximation of the variance of the ratio of random quantities, the approximated variance of estimator (4.2.5) is*

$$V[\hat{z}(\mathbf{u}_i)] = \mathbf{k}_i^\top E[\mathbf{h}_i\mathbf{h}_i^\top]\mathbf{k}_i + O_p(n^{-2}), \tag{4.2.19}$$

where

$$\mathbf{k}_i = \frac{N}{n}\frac{(1+at_{2i})\mathbf{z} - (z(\mathbf{u}_i) + at_{1i})\mathbf{1}_N}{(1+at_{2i})^2}. \tag{4.2.20}$$

Proof. Starting from equation (A.3) the resulting first-order delta method approximation of the variance of the IDW point estimator is

$$\begin{aligned}\mathrm{V}[\hat{z}(\mathbf{u}_i)] = &\frac{\mathrm{E}[(\mathbf{h}_i^\top\mathbf{z})^2]\,(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^2}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} - 2\frac{\mathrm{E}[\mathbf{h}_i^\top\mathbf{z}\mathbf{h}_i^\top\mathbf{1}_N]\,\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}\,\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} \\ &+ \frac{\mathrm{E}[(\mathbf{h}_i^\top\mathbf{1}_N)^2]\,(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z})^2}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} + O_p(n^{-2}).\end{aligned} \tag{4.2.21}$$

By the algebraic manipulation of the second term of the previous equation, we obtain

$$2\mathrm{E}[\mathbf{h}_i^\top\mathbf{z}\mathbf{h}_i^\top\mathbf{1}_N] = \mathbf{z}^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top]\mathbf{1}_N + \mathbf{1}_N^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top]\mathbf{z},$$

leading to the following rewriting of equation (4.2.21):

$$\begin{aligned}\mathrm{V}[\hat{z}(\mathbf{u}_i)] = &\frac{\mathrm{E}[(\mathbf{h}_i^\top\mathbf{z})^2]\,(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^2}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} - \frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N\,\mathbf{z}^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top]\mathbf{1}_N\,\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} \\ &- \frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}\,\mathbf{1}_N^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top]\mathbf{z}\,\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} + \frac{\mathrm{E}[(\mathbf{h}_i^\top\mathbf{1}_N)^2]\,(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z})^2}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} + O_p(n^{-2})\end{aligned} \tag{4.2.22}$$

First of all, we need to compute the second moment of weighting vector (4.2.4). Using idempotence property (4.2.4) of the second-order inclusion probabilities of the same individual in SRSWoR, vector $\mathbf{h}_i$'s second moment results

$$\begin{aligned}
\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top] &= \mathrm{E}[(Q_i\mathbf{e}_i + \boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i - Q_i\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)(Q_i\mathbf{e}_i + \boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i - Q_i\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top] \\
&= \mathrm{E}[Q_i^2\mathbf{e}_i\mathbf{e}_i^\top + Q_i\mathbf{e}_i(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top - Q_i^2\mathbf{e}_i(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top \\
&\quad + Q_i(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)\mathbf{e}_i^\top + (\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top - Q_i(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top \\
&\quad - Q_i^2(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)\mathbf{e}_i^\top - Q_i(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top + Q_i^2(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top] \\
&= \mathrm{E}[Q_i\mathbf{e}_i\mathbf{e}_i^\top] + \mathrm{E}[(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top] - \mathrm{E}[Q_i(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top].
\end{aligned} \tag{4.2.23}$$

Starting from the last line of the previous equation, the first expected value's computation is straightforward

$$\mathrm{E}[Q_i\mathbf{e}_i\mathbf{e}_i^\top] = \mathrm{E}[Q_i]\mathbf{e}_i\mathbf{e}_i^\top = \frac{n}{N}\mathbf{e}_i\mathbf{e}_i^\top$$

and comes directly from first-order inclusion probability (4.2.11) in SRSWoR. Using the square matrix collecting the second-order inclusion probabilities proposed by Dol et al. (1996), the second expected value of equation (4.2.23) results

$$\begin{aligned}
\mathrm{E}[(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top] &= \mathrm{E}[\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{Q}}^\top] \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top \\
&= \frac{n}{N}\left(\frac{N-n}{N-1}\mathbf{I}_N + \frac{n-1}{N-1}\mathbf{1}_N\mathbf{1}_N^\top\right) \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top \\
&= \frac{n}{N}\left(\frac{N-n}{N-1}\mathrm{diag}(\boldsymbol{\phi}_i)^2 + \frac{n-1}{N-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top\right),
\end{aligned} \tag{4.2.24}$$

where the distributive property of the Hadamard product has been used. Using a matrix analogous to (4.2.24) collecting the the third-order inclusion probabilities, the third expectation in (4.2.23) is

$$\begin{aligned}
\mathrm{E}[Q_i(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}} \circ \boldsymbol{\phi}_i)^\top] &= \mathrm{E}[Q_i\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{Q}}^\top] \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top \\
&= \frac{n}{N}\left\{\frac{n-1}{N-1}\left(\frac{N-n}{N-2}\mathbf{I}_N + \frac{n-2}{N-2}\mathbf{1}_N\mathbf{1}_N^\top\right)\right. \\
&\quad \left. + \frac{N-n}{N-1}\mathbf{e}_i\mathbf{e}_i^\top\right\} \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top \\
&= \frac{n}{N}\frac{n-1}{N-1}\left(\frac{N-n}{N-2}\mathrm{diag}(\boldsymbol{\phi}_i)^2 + \frac{n-2}{N-2}\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top\right).
\end{aligned}$$

The Hadamard products between the quantities involved in the calculus of the previous equation are collected in Appendix B.

Finally, we obtain vector $\mathbf{h}_i$'s second moment

$$\begin{aligned}\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top] &= \frac{n}{N}\left(\mathbf{e}_i\mathbf{e}_i^\top + \frac{N-n}{N-1}\left(1-\frac{n-1}{N-2}\right)\mathrm{diag}(\boldsymbol{\phi}_i)^2\right.\\ &\quad\left.+\frac{n-1}{N-1}\left(1-\frac{n-2}{N-2}\right)\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top\right)\\ &= \frac{n}{N}(\mathbf{e}_i\mathbf{e}_i^\top + b\,\mathrm{diag}(\boldsymbol{\phi}_i)^2 + c\,\boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top),\end{aligned}$$

where the following constatnts are defined:

$$b = \frac{N-n}{N-1}\left(1-\frac{n-1}{N-2}\right) = \frac{N-n}{N-1}\frac{N-n-1}{N-2}$$

and

$$c = \frac{n-1}{N-1}\left(1-\frac{n-2}{N-2}\right) = \frac{n-1}{N-1}\frac{N-n}{N-2}.$$

Starting from equation (4.2.22), we factorize the approximated variance of the predictor as

$$\begin{aligned}\mathrm{V}[\hat{z}(\mathbf{u}_i)] &= \frac{\mathrm{E}[(\mathbf{h}_i^\top\mathbf{z})^2]\,(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^2}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} - \frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N\,\mathbf{z}^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top]\mathbf{1}_N\,\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4}\\ &\quad - \frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}\,\mathbf{1}_N^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top]\mathbf{z}\,\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} + \frac{\mathrm{E}[(\mathbf{h}_i^\top\mathbf{1}_N)^2]\,(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z})^2}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} + O_p(n^{-2})\\ &= \frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N\,\mathbf{z}^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top](\mathbf{z}\,\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N - \mathbf{1}_N\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z})}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4}\\ &\quad - \frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}\,\mathbf{1}^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top](\mathbf{z}\,\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N - \mathbf{1}_N\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z})}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} + O_p(n^{-2})\\ &= \frac{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N\,\mathbf{z}^\top - \mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}\,\mathbf{1}_N^\top)\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top](\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N\,\mathbf{z} - \mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}\,\mathbf{1}_N)}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^4} + O_p(n^{-2})\\ &= \mathbf{k}_i^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_i^\top]\mathbf{k}_i + O_p(n^{-2}) \qquad (4.2.25)\end{aligned}$$

and we define the following vector:

$$\begin{aligned}\mathbf{k}_i &= \frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N\,\mathbf{z} - \mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}\,\mathbf{1}_N}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^2} \\ &= \frac{N}{n}\frac{(1+at_{2i})\mathbf{z} - (z(\mathbf{u}_i)+at_{1i})\mathbf{1}_N}{(1+at_{2i})^2},\end{aligned} \tag{4.2.26}$$

where expectation (4.2.16) is retrieved. □

4.2.4 Covariance

The covariance between two IDW point estimators is obtained in approximated form through the delta method (Appendix A) since estimator (4.2.5) is the ratio of sums of linear combinations of random quantities.

Theorem 4.2.3. *Using the first-order delta method approximation of the covariance between two ratios of random quantities, the approximated covariance between two IDW point estimators is*

$$Cov(\hat{z}(\mathbf{u}_i), \hat{z}(\mathbf{u}_j)) = \mathbf{k}_i^\top E[\mathbf{h}_i\mathbf{h}_j^\top]\mathbf{k}_j + O_p(n^{-2}), \tag{4.2.27}$$

where vectors $\mathbf{k}_i$ *and* $\mathbf{k}_j$ *are defined in equation (4.2.20).*

Proof. Starting from equation (A.4) the resulting first-order delta method approximation of the covariance between two IDW point estimators is

$$\begin{aligned}\mathrm{Cov}(\hat{z}(\mathbf{u}_i), \hat{z}(\mathbf{u}_j)) = &\frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N\,\mathbf{z}^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_j^\top]\mathbf{z}\,\mathrm{E}[\mathbf{h}_j^\top]\mathbf{1}_N}{(\mathrm{E}[\mathbf{h}_i]^\top\mathbf{1}_N)^2(\mathrm{E}[\mathbf{h}_j^\top]\mathbf{1}_N)^2} \\ &- \frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N\,\mathbf{z}^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_j^\top]\mathbf{1}_N\,\mathrm{E}[\mathbf{h}_j^\top]\mathbf{z}}{(\mathrm{E}[\mathbf{h}_i]^\top\mathbf{1}_N)^2(\mathrm{E}[\mathbf{h}_j^\top]\mathbf{1}_N)^2} \\ &- \frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}\,\mathbf{1}_N^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_j^\top]\mathbf{z}\,\mathrm{E}[\mathbf{h}_j^\top]\mathbf{1}_N}{(\mathrm{E}[\mathbf{h}_i]^\top\mathbf{1}_N)^2(\mathrm{E}[\mathbf{h}_j^\top]\mathbf{1}_N)^2} \\ &+ \frac{\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}\,\mathbf{1}_N^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_j^\top]\mathbf{1}_N\,\mathrm{E}[\mathbf{h}_j^\top]\mathbf{z}}{(\mathrm{E}[\mathbf{h}_i]^\top\mathbf{1}_N)^2(\mathrm{E}[\mathbf{h}_j^\top]\mathbf{1}_N)^2} + O_p(n^{-2}).\end{aligned} \tag{4.2.28}$$

Analogously to what has been done for the variance, we need to compute the mixed first moment of two IDW point estimators' non-standardized

weighting vectors

$$\begin{aligned}
\mathrm{E}[\mathbf{h}_i\mathbf{h}_j^\top] &= \mathrm{E}[(Q_i\mathbf{e}_i + \boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i - Q_i\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)(Q_j\mathbf{e}_j + \boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j - Q_j\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top] \\
&= \mathrm{E}[Q_iQ_j\mathbf{e}_i\mathbf{e}_j^\top + Q_i\mathbf{e}_i(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top - Q_iQ_j\mathbf{e}_i(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top \\
&\quad + Q_j(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)\mathbf{e}_j^\top + (\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top - Q_j(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top \\
&\quad - Q_iQ_j(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)\mathbf{e}_j^\top - Q_i(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top \\
&\quad + Q_iQ_j(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top] \\
&= \mathrm{E}[Q_iQ_j\mathbf{e}_i\mathbf{e}_j^\top] + \mathrm{E}[Q_i\mathbf{e}_i(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top] - \mathrm{E}[Q_iQ_j\mathbf{e}_i(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top] \\
&\quad + \mathrm{E}[Q_j(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)\mathbf{e}_j^\top] + \mathrm{E}[(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top] \\
&\quad - \mathrm{E}[Q_j(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top] - \mathrm{E}[Q_iQ_j(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)\mathbf{e}_j^\top] \\
&\quad - \mathrm{E}[Q_i(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top] + \mathrm{E}[Q_iQ_j(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_i)(\boldsymbol{\mathcal{Q}}\circ\boldsymbol{\phi}_j)^\top].
\end{aligned} \tag{4.2.29}$$

We start by computing the first expected value involved in the previous equation by using the second-order inclusion probability:

$$\mathrm{E}[Q_iQ_j]\mathbf{e}_i\mathbf{e}_j^\top = \frac{n}{N}\frac{n-1}{N-1}\mathbf{e}_i\mathbf{e}_j^\top.$$

Using a formulation analogous to the one proposed by Dol et al. (1996) we compute the second expectation of equation (4.2.29) as

$$\begin{aligned}
\mathbf{e}_i(\mathrm{E}[Q_i\boldsymbol{\mathcal{Q}}]\circ\boldsymbol{\phi}_j)^\top &= \frac{n}{N}\mathbf{e}_i\left\{\left(\frac{N-n}{N-1}\mathbf{e}_i + \frac{n-1}{N-1}\mathbf{1}_N\right)^\top \circ \boldsymbol{\phi}_j^\top\right\} \\
&= \frac{n}{N}\left(\frac{N-n}{N-1}\phi_{ij}\mathbf{e}_i\mathbf{e}_i^\top + \frac{n-1}{N-1}\mathbf{e}_i\boldsymbol{\phi}_j^\top\right).
\end{aligned} \tag{4.2.30}$$

We continue by computing the expectation of the third term of equation (4.2.29)

$$\begin{aligned}
\mathbf{e}_i(\mathrm{E}[Q_iQ_j\boldsymbol{\mathcal{Q}}]\circ\boldsymbol{\phi}_j)^\top &= \frac{n}{N}\frac{n-1}{N-1}\mathbf{e}_i\left[\left\{\frac{N-n}{N-2}(\mathbf{e}_i+\mathbf{e}_j) + \frac{n-2}{N-2}\mathbf{1}_N\right\}^\top \circ \boldsymbol{\phi}_j^\top\right] \\
&= \frac{n}{N}\frac{n-1}{N-1}\left(\frac{N-n}{N-2}\phi_{ij}\mathbf{e}_i\mathbf{e}_i^\top + \frac{n-2}{N-2}\mathbf{e}_i\boldsymbol{\phi}_j^\top\right).
\end{aligned} \tag{4.2.31}$$

Similarly to (4.2.30), the fourth expected value is

$$(\mathrm{E}[Q_j\boldsymbol{\mathcal{Q}}]\circ\boldsymbol{\phi}_i)\mathbf{e}_j^\top = \frac{n}{N}\left(\frac{N-n}{N-1}\phi_{ij}\mathbf{e}_j\mathbf{e}_j^\top + \frac{n-1}{N-1}\boldsymbol{\phi}_i\mathbf{e}_j^\top\right). \tag{4.2.32}$$

The fifth expectation of equation (4.2.29) is similar to (4.2.24):

$$\begin{aligned}\mathrm{E}[\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{Q}}^\top]\circ\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \frac{n}{N}\left(\frac{N-n}{N-1}\mathbf{I}_N + \frac{n-1}{N-1}\mathbf{1}_N\mathbf{1}_N^\top\right)\circ\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top \\ &= \frac{n}{N}\left(\frac{N-n}{N-1}\mathrm{diag}(\boldsymbol{\phi}_i\circ\boldsymbol{\phi}_j) + \frac{n-1}{N-1}\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top\right).\end{aligned}$$

The sixth expectation in equation (4.2.29) is

$$\begin{aligned}\mathrm{E}[Q_j\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{Q}}^\top]\circ\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \frac{n}{N}\left[\frac{N-n}{N-1}\frac{N+n-3}{N-2}\mathbf{e}_j\mathbf{e}_j^\top + \frac{n-1}{N-1}\frac{n-2}{N-2}\mathbf{1}_N\mathbf{1}_N^\top\right. \\ &\quad + \frac{n-1}{N-1}\frac{N-n}{N-2}\left\{\mathbf{e}_j(\mathbf{1}_N-\mathbf{e}_j)^\top + (\mathbf{1}_N-\mathbf{e}_j)\mathbf{e}_j^\top\right. \\ &\quad \left.\left. + \mathrm{diag}(\mathbf{1}_N-\mathbf{e}_j)\right\}\right]\circ\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top \\ &= \frac{n}{N}\frac{n-1}{N-1}\left[\frac{N-n}{N-2}\left\{\phi_{ij}\mathbf{e}_j\boldsymbol{\phi}_j^\top + \mathrm{diag}(\boldsymbol{\phi}_i\circ\boldsymbol{\phi}_j)\right\}\right. \\ &\quad \left. + \frac{n-2}{N-2}\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top\right].\end{aligned} \tag{4.2.33}$$

The seventh expectation is similar to (4.2.31)

$$(\mathrm{E}[Q_iQ_j\boldsymbol{\mathcal{Q}}]\circ\boldsymbol{\phi}_i)\mathbf{e}_j^\top = \frac{n}{N}\frac{n-1}{N-1}\left(\frac{N-n}{N-2}\phi_{ij}\mathbf{e}_j\mathbf{e}_j^\top + \frac{n-2}{N-2}\boldsymbol{\phi}_i\mathbf{e}_j^\top\right).$$

The eighth expected value is computed in a way analogous to (4.2.33)

$$\begin{aligned}\mathrm{E}[Q_i\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{Q}}^\top]\circ\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \frac{n}{N}\frac{n-1}{N-1}\left[\frac{N-n}{N-2}\left\{\phi_{ij}\boldsymbol{\phi}_i\mathbf{e}_i^\top + \mathrm{diag}(\boldsymbol{\phi}_i\circ\boldsymbol{\phi}_j)\right\}\right. \\ &\quad \left. + \frac{n-2}{N-2}\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top\right].\end{aligned}$$

The last expectation involves fourth-order inclusion probabilities

$$
\begin{aligned}
\mathrm{E}[Q_iQ_j\boldsymbol{\mathcal{Q}}\boldsymbol{\mathcal{Q}}^\top]\circ\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top = \frac{n}{N}\frac{n-1}{N-1}\Bigg[&\frac{N-n}{N-2}\frac{N+n-5}{N-3}\left\{\mathbf{e}_i\mathbf{e}_i^\top+\mathbf{e}_j\mathbf{e}_j^\top\right.\\
&\left.+\,\mathbf{e}_i\mathbf{e}_j^\top+\mathbf{e}_j\mathbf{e}_i^\top\right\}+\frac{n-2}{N-2}\frac{N-n}{N-3}\left\{\mathbf{e}_i(\mathbf{1}_N-\mathbf{e}_i-\mathbf{e}_j)^\top\right.\\
&+\mathbf{e}_j(\mathbf{1}_N-\mathbf{e}_i-\mathbf{e}_j)^\top+(\mathbf{1}_N-\mathbf{e}_i-\mathbf{e}_j)\mathbf{e}_i^\top\\
&\left.+(\mathbf{1}_N-\mathbf{e}_i-\mathbf{e}_j)\mathbf{e}_j^\top+\mathrm{diag}(\mathbf{1}_N-\mathbf{e}_i-\mathbf{e}_j)\right\}\\
&+\frac{n-2}{N-2}\frac{n-3}{N-3}\mathbf{1}_N\mathbf{1}_N^\top\Bigg]\circ\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top\\
=\frac{n}{N}\frac{n-1}{N-1}\Bigg[&\frac{N-n}{N-2}\left\{\frac{N+n-5}{N-3}\phi_{ij}^2\mathbf{e}_j\mathbf{e}_i^\top\right.\\
&+\frac{n-2}{N-3}\Big(\phi_{ij}\mathbf{e}_j\left\{(\mathbf{1}_N-\mathbf{e}_i)\circ\boldsymbol{\phi}_j\right\}^\top\\
&+\phi_{ij}\left\{(\mathbf{1}_N-\mathbf{e}_j)\circ\boldsymbol{\phi}_i\right\}\mathbf{e}_i^\top\\
&\left.+\;\mathrm{diag}(\boldsymbol{\phi_i}\circ\boldsymbol{\phi}_j)\Big)\right\}+\frac{n-2}{N-2}\frac{n-3}{N-3}\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top\Bigg].
\end{aligned}
$$

The relations collected in Appendix B are involved in the calculus of the approximated covariance between two IDW point estimators.

The equation of the mixed first moment (4.2.29) becomes

$$
\begin{aligned}
\mathrm{E}[\mathbf{h}_i\mathbf{h}_j^\top] = \frac{n}{N}\Bigg\{&\frac{n-1}{N-1}\mathbf{e}_i\mathbf{e}_j^\top+b\phi_{ij}(\mathbf{e}_i\mathbf{e}_i^\top+\mathbf{e}_j\mathbf{e}_j^\top)+c(\mathbf{e}_i\boldsymbol{\phi}_j^\top+\boldsymbol{\phi}_i\mathbf{e}_j^\top)\\
&+\frac{N-n-2}{N-3}b\mathrm{diag}(\boldsymbol{\phi_i}\circ\boldsymbol{\phi}_j)+\frac{N-n-1}{N-3}c\boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top\\
&-c\phi_{ij}(\mathbf{e}_j\boldsymbol{\phi}_j^\top+\boldsymbol{\phi}_i\mathbf{e}_i^\top)+\frac{N+n-5}{N-3}c\phi_{ij}^2\mathbf{e}_j\mathbf{e}_i^\top\\
&+\frac{n-2}{N-3}c\phi_{ij}(\mathbf{e}_j\{(\mathbf{1}_N-\mathbf{e}_i)\circ\boldsymbol{\phi}_j\}^\top+\{(\mathbf{1}_N-\mathbf{e}_j)\circ\boldsymbol{\phi}_i\}\mathbf{e}_i^\top)\Bigg\}.
\end{aligned}
$$

Using the vector defined in equation (4.2.26), we factorize the approximated covariance in a way similar to equation (4.2.25)

$$
\begin{aligned}
\mathrm{Cov}(\hat{z}(\mathbf{u}_i),\hat{z}(\mathbf{u}_j)) &= \frac{\left(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N\,\mathbf{z}^\top-\mathrm{E}[\mathbf{h}_i^\top]\mathbf{z}\,\mathbf{1}_N^\top\right)\mathrm{E}[\mathbf{h}_i\mathbf{h}_j^\top]\left(\mathbf{z}\,\mathrm{E}[\mathbf{h}_j^\top]\mathbf{1}_N-\mathbf{1}_N\,\mathrm{E}[\mathbf{h}_j^\top]\mathbf{z}\right)}{(\mathrm{E}[\mathbf{h}_i^\top]\mathbf{1}_N)^2(\mathrm{E}[\mathbf{h}_j^\top]\mathbf{1}_N)^2}\\
&\quad+O_p(n^{-1})\\
&=\mathbf{k}_i^\top\mathrm{E}[\mathbf{h}_i\mathbf{h}_j^\top]\mathbf{k}_j+O_p(n^{-2}). \qquad (4.2.34)
\end{aligned}
$$

where vectors $\mathbf{k}_i$ and $\mathbf{k}_j$ are defined in equation (4.2.26). □

Variances (4.2.19) and covariances (4.2.27) of the IDW point estimators can be collected in the $N \times N$ symmetric matrix

$$\begin{aligned}\boldsymbol{\Sigma} &= \begin{bmatrix} \mathrm{V}[\hat{z}(\mathbf{u}_1)] & \mathrm{Cov}(\hat{z}(\mathbf{u}_1), \hat{z}(\mathbf{u}_2)) & \cdots & \mathrm{Cov}(\hat{z}(\mathbf{u}_1), \hat{z}(\mathbf{u}_2)) \\ \mathrm{Cov}(\hat{z}(\mathbf{u}_2), \hat{z}(\mathbf{u}_1)) & \mathrm{V}[\hat{z}(\mathbf{u}_2)] & \cdots & \mathrm{Cov}(\hat{z}(\mathbf{u}_2), \hat{z}(\mathbf{u}_N)) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\hat{z}(\mathbf{u}_N), \hat{z}(\mathbf{u}_1)) & \mathrm{Cov}(\hat{z}(\mathbf{u}_N), \hat{z}(\mathbf{u}_2)) & \cdots & \mathrm{V}[\hat{z}(\mathbf{u}_N)] \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{k}_1^\top \mathrm{E}[\mathbf{h}_1 \mathbf{h}_1^\top] \mathbf{k}_1 & \mathbf{k}_1^\top \mathrm{E}[\mathbf{h}_1 \mathbf{h}_2^\top] \mathbf{k}_2 & \cdots & \mathbf{k}_1^\top \mathrm{E}[\mathbf{h}_1 \mathbf{h}_N^\top] \mathbf{k}_N \\ \mathbf{k}_2^\top \mathrm{E}[\mathbf{h}_2 \mathbf{h}_1^\top] \mathbf{k}_1 & \mathbf{k}_2^\top \mathrm{E}[\mathbf{h}_2 \mathbf{h}_2^\top] \mathbf{k}_2 & \cdots & \mathbf{k}_2^\top \mathrm{E}[\mathbf{h}_2 \mathbf{h}_N^\top] \mathbf{k}_N \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{k}_N^\top \mathrm{E}[\mathbf{h}_N \mathbf{h}_1^\top] \mathbf{k}_1 & \mathbf{k}_N^\top \mathrm{E}[\mathbf{h}_N \mathbf{h}_2^\top] \mathbf{k}_2 & \cdots & \mathbf{k}_N^\top \mathrm{E}[\mathbf{h}_N \mathbf{h}_N^\top] \mathbf{k}_N \end{bmatrix} + O_p(n^{-2}). \end{aligned} \tag{4.2.35}$$

4.2.5 Asymptotic properties

The IDW point estimator turns out to be *finite population consistent* according to definition (c) of page 168 of Särndal et al. (1992). Let us consider only a fixed finite population $\mathcal{P}$ of N locations in the domain $\mathcal{D}$ for which we have an increasing sample size n, then

$$\lim_{n \to N} \hat{z}(\mathbf{u}_i) = z(\mathbf{u}_i).$$

It can be easily be proved by noticing that if the sample coincides to the whole population, then the event of being sampled is sure and point estimator (4.2.5) will be the observed value itself since $\hat{z}(\mathbf{u}_i)$ reproduces the observed value at sampled locations.

4.2.6 Variance estimation

In this subsection three different variance estimators for the variance of estimator (4.2.5) are presented: a naïve one and two jackknife ones.

Firstly, we propose to estimate the vector $\mathbf{k}_i$ of equation (4.2.20) using the vector of the estimated values (4.2.6). Vector $\mathbf{k}_i$ estimator is obtained by

substituting $\hat{\mathbf{z}}$ for the unknown quantities involved in equation (4.2.26):

$$\hat{\mathbf{k}}_i = \frac{N}{n} \frac{(1 + at_{2i})\hat{\mathbf{z}} - (z(\mathbf{u}_i) + a\hat{t}_{1i})\mathbf{1}_N}{(1 + at_{2i})^2}. \tag{4.2.36}$$

Analogously, linear combination (4.2.9) is estimated by

$$\hat{t}_1 = \sum_{j \neq i} \phi_{ij} \hat{z}(\mathbf{u}_j) = \boldsymbol{\phi}_i^\top \hat{\mathbf{z}},$$

while combination (4.2.10) does not need to be estimated because the squared inverse Euclidean distances are known before sampling.

The resulting naïve estimator of variance (4.2.19) is

$$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)] = \hat{\mathbf{k}}_i^\top \mathrm{E}[\mathbf{h}_i \mathbf{h}_i^\top] \hat{\mathbf{k}}_i.$$

Following Wolter (2007), we suggest the use of two jackknife estimators, one of which is known to be conservative. We define the first variance estimator as

$$\hat{\mathrm{V}}_1[\hat{\bar{z}}(\mathbf{u}_i)] = \frac{1}{n(n-1)} \sum_{k=1}^{n} (\hat{z}(\mathbf{u}_i)_k - \hat{\bar{z}}(\mathbf{u}_i))^2, \tag{4.2.37}$$

where $\hat{\bar{z}}(\mathbf{u}_i)$ is the mean of the n pseudo-values $\hat{z}(\mathbf{u}_i)_k$ of the IDW point estimator for the ith location defined as

$$\hat{z}(\mathbf{u}_i)_k = n\hat{z}(\mathbf{u}_i) - (n-1)\hat{z}(\mathbf{u}_i)_{(k)},$$

where $\hat{z}(\mathbf{u}_i)_{(k)}$ is the IDW point estimator computed by excluding the k-th sampled location. An alternative estimator, which is known to be conservative, is defined by substituting the mean of the pseudo-values $\hat{\bar{z}}(\mathbf{u}_i)$ for the estimated value $\hat{z}(\mathbf{u}_i)$:

$$\hat{\mathrm{V}}_2[\hat{\bar{z}}(\mathbf{u}_i)] = \frac{1}{n(n-1)} \sum_{k=1}^{n} (\hat{z}(\mathbf{u}_i)_k - \hat{z}(\mathbf{u}_i))^2. \tag{4.2.38}$$

Variance estimators (4.2.37) and (4.2.38) are both used when the sampling fraction $f = n/N$ is negligible, otherwise the modification

$$\hat{z}(\mathbf{u}_i)^*_{(k)} = \hat{z}(\mathbf{u}_i) + (1-f)^{1/2}(\hat{z}(\mathbf{u}_i)_{(k)} - \hat{z}(\mathbf{u}_i)) \tag{4.2.39}$$

might usefully be applied instead of $\hat{z}(\mathbf{u}_i)_{(k)}$.

For a deeper compendium see Appendix C.

4.3 Estimating individual quantities in design-based continuous spatial population

One of the most peculiar characteristics of spatial data is that they usually are gathered in a continuous domain. As we have seen in the previous chapter geostatistics can be employed for predicting unobserved data either in the spatial finite population case or the continuous domain case. In order to do so in design-based framework we need to pass from inclusion probabilities to inclusion density functions.

4.3.1 From inclusion probabilities to inclusion density functions

The main idea for using the design-based spatial inference in a continuous spatial domain is to replace the usual inclusion probabilities used for the finite population case with the inclusion density function defined by Cordy (1993).

Let us now suppose that the spatial domain is a bounded open set, $\mathcal{D} \subset \mathbb{R}^2$, and a fixed size sample design is chosen for the sake of simplicity. Following Cordy (1993), we define a sampling design as a probability measure P on the σ-algebra $\mathcal{S}_n$ of all the possible samples of size n, satisfying the following regularity conditions:

(i) the joint probability density function f defined on $\mathcal{S}_n$, of the n random variables denoting the locations $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ to sample, satisfies

$$P(E) = \int_E f \, \mathrm{d}\lambda_n$$

for each measurable set $E \subset \mathcal{S}_n$, where λ_n denotes Lebesgue measure on $\mathcal{D}^n$;

(ii) the inclusion density function

$$\pi(\mathbf{u}) = \sum_{i=1}^{n} f_i(\mathbf{u}), \tag{4.3.1}$$

where f_i denotes the marginal density of $\boldsymbol{U}_i$, is positive for each $\mathbf{u} \in \mathcal{D}$.

Requirement *(ii)* can be thought of as the continuous version of the statement that in the finite population case assesses that every unit in the population has a positive probability of being sampled.

Analogously we define the pairwise inclusion density function as

$$\pi(\mathbf{u}, \mathbf{u}') = \sum_{i=1}^{n} \sum_{i \neq j} f_{ij}(\mathbf{u}, \mathbf{u}'),$$

and so on.

For practical purposes one can consider inclusion density function (4.3.1) as the number of locations to be sampled per unit area of the domain.

For what regards SRSWoR the sampling locations are selected at random leading to independence among them. Therefore, the density function f of the n random variables denoting the locations $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n$ to sample can be expressed as the product of the marginal density functions f_i:

$$f(\mathbf{u}_1, \ldots, \mathbf{u}_n) = \prod_{i=1}^{n} f_i(\mathbf{u}_i).$$

The marginal density function managing the inclusion of a location in the sample is a Bernoulli random variable, leading to the first-order inclusion density function in SRSWoR from a continuous domain

$$\pi(\mathbf{u}) = \frac{n}{|\mathcal{D}|},$$

where $|\mathcal{D}| = \int_{\mathcal{D}} \mathrm{d}\mathbf{u}$ denotes the area of the domain $\mathcal{D}$. Analogously the pairwise inclusion density function of two spatial locations is defined as

$$\pi(\mathbf{u}, \mathbf{u}') = \frac{n(n-1)}{|\mathcal{D}|^2},$$

and so on.

4.3.2 The spatial point estimator in the continuous population case

Formally, the expression of the IDW point estimator for a continuous domain is the same as the one of the finite population case: (4.2.3) and, therefore, (4.2.5).

Its expectation

$$\mathrm{E}[\hat{z}(\mathbf{u}_i)] = \frac{z(\mathbf{u}_i) + at_{1i}}{1 + at_{2i}} + O_p(n^{-1}),$$

differs form (4.2.7) only in the constant

$$a = 1 - \frac{n-1}{|\mathcal{D}|},$$

now defined according to the inclusion density function of SRSWoR.

The same holds for the variance:

$$\mathrm{V}[\hat{z}(\mathbf{u}_i)] = \mathbf{k}_i^\top \mathrm{E}[\mathbf{h}_i \mathbf{h}_i^\top] \mathbf{k}_i + O_p(n^{-2}),$$

where vector $\mathbf{k}_i$ and the second moment $\mathrm{E}[\mathbf{h}_i \mathbf{h}_i^\top]$ become

$$\mathbf{k}_i = \frac{|\mathcal{D}|}{n} \frac{(1 + at_{2i})\mathbf{z} - (z(\mathbf{u}_i) + at_{1i})\mathbf{1}_N}{(1 + at_{2i})^2}$$

and

$$\mathrm{E}[\mathbf{h}_i \mathbf{h}_i^\top] = \frac{n}{|\mathcal{D}|}(\mathbf{e}_i \mathbf{e}_i^\top + b\,\mathrm{diag}(\boldsymbol{\phi}_i)^2 + c\,\boldsymbol{\phi}_i \boldsymbol{\phi}_i^\top),$$

respectively. As for the expectation, the involved constants need to be defined according the higher order inclusion density function:

$$b = \left(1 - \frac{n-1}{|\mathcal{D}|}\right)^2$$

and

$$c = \left(1 - \frac{n-1}{|\mathcal{D}|}\right) \frac{n-1}{|\mathcal{D}|}.$$

Following the same reasoning as above, we obtain the covariance between two IDW point estimators in a continuous spatial domain

$$\mathrm{Cov}(\hat{z}(\mathbf{u}_i), \hat{z}(\mathbf{u}_j)) = \mathbf{k}_i^\top \mathrm{E}[\mathbf{h}_i \mathbf{h}_j^\top] \mathbf{k}_j + O_p(n^{-2}),$$

where the mixed first moment $\mathrm{E}[\mathbf{h}_i \mathbf{h}_j^\top]$ becomes

$$\begin{aligned}
\mathrm{E}[\mathbf{h}_i \mathbf{h}_j^\top] = \frac{n}{|\mathcal{D}|} \Bigg\{ & \frac{n-1}{|\mathcal{D}|} \mathbf{e}_i \mathbf{e}_j^\top + b\phi_{ij}(\mathbf{e}_i \mathbf{e}_i^\top + \mathbf{e}_j \mathbf{e}_j^\top) + c(\mathbf{e}_i \boldsymbol{\phi}_j^\top + \boldsymbol{\phi}_i \mathbf{e}_j^\top) \\
& + b\mathrm{diag}(\boldsymbol{\phi}_i \circ \boldsymbol{\phi}_j) + c\boldsymbol{\phi}_i \boldsymbol{\phi}_j^\top - c\phi_{ij}(\mathbf{e}_j \boldsymbol{\phi}_j^\top + \boldsymbol{\phi}_i \mathbf{e}_i^\top) + c\phi_{ij}^2 \mathbf{e}_j \mathbf{e}_i^\top \\
& + c\frac{n-2}{|\mathcal{D}|}\phi_{ij}(\mathbf{e}_j\{(\mathbf{1}_N - \mathbf{e}_i) \circ \boldsymbol{\phi}_j\}^\top + \{(\mathbf{1}_N - \mathbf{e}_j) \circ \boldsymbol{\phi}_i\}\mathbf{e}_i^\top) \Bigg\}.
\end{aligned}$$

4.4 Estimating population quantities in design-based finite spatial population

In the previous sections point estimation has been treated either in a finite population or in a continuous domain. It represents a novelty in the context of design-based inference since it normally deals with the estimation of population synthetic quantities (i.e. totals or means). In this section the usual spatial design-based finite population inference is enriched by the use of the results of Section 4.2. For the sake of simplicity, in the following subsections only the population total estimator will be treated, since the population mean estimator is easily obtained by dividing by N.

4.4.1 Estimation of the population total

Given that the IDW interpolator is exact and that the resulting point estimator holds the same property, the population total estimator can be expressed as the sum of the values contained in vector (4.2.6) as follows:

$$\hat{t}(z) = \mathbf{1}_N^\top \hat{\mathbf{z}} = \sum_{i=1}^{N} \hat{z}(\mathbf{u}_i) = \sum_{i \in S} z(\mathbf{u}_i) + \sum_{j \notin S} \hat{z}(\mathbf{u}_j), \tag{4.4.1}$$

where the $\hat{z}(\mathbf{u}_j)$-s are the IDW point estimators evaluated in the unsampled locations. Previous equation's last equality is the usual estimator of the population total in predictive form of design-based inference (Bolfarine and Zacks, 1992).

4.4.2 Statistical properties

Since estimator of the population total is the sum of the estimated values, its expectation is the sum of the expectations of the IDW point estimators:

$$\begin{aligned} \mathrm{E}[\hat{t}(z)] &= \mathrm{E}[\mathbf{1}_N^\top \hat{\mathbf{z}}] = \mathbf{1}_N^\top \mathrm{E}[\hat{\mathbf{z}}] \\ &= \mathbf{1}_N^\top \mathrm{diag}(\mathbf{1}_N + a\mathbf{t}_2)^{-1}(\mathbf{z} + a\mathbf{t}_1) + O_p(n^{-1}) \\ &= \sum_{i=1}^{N} \frac{z(\mathbf{u}_i) + at_{1i}}{1 + at_{2i}} + O_p(n^{-1}), \end{aligned}$$

where expected values (4.2.18) and (4.2.17) are retrieved.

Estimator (4.4.1) variance is defined as

$$\begin{aligned} \mathrm{V}[\hat{t}(z)] &= \sum_{i=1}^{N} \mathrm{V}[\hat{z}(\mathbf{u}_i)] + \sum_{i=1}^{N} \sum_{j \neq i} \mathrm{Cov}(\hat{z}(\mathbf{u}_i), \hat{z}(\mathbf{u}_j)) \\ &= \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbf{k}_i^{\top} \mathrm{E}[\mathbf{h}_i \mathbf{h}_j^{\top}] \mathbf{k}_j + O_p(n^{-2}) = \mathbf{1}_N^{\top} \boldsymbol{\Sigma} \mathbf{1}_N, \end{aligned} \tag{4.4.2}$$

where variances (4.2.19) and covariances (4.2.27) are retrieved.

Using notation of Subsection 4.2.5, we define the infinite sequence of nested increasing size populations $\{\mathcal{P}_\nu\}_{\nu \in \mathbb{N}}$ belonging to the very spatial domain $\mathcal{D}$. From each of them we draw a sample S_ν generating the increasing sequence of sample sizes $\{n_\nu\}_{\nu \in \mathbb{N}}$. Both the population size N_ν and the sample size n_ν diverge as $\nu \to \infty$ (i.e. $N_\nu \to \infty$, $n_\nu \to N_\nu$). Hence, the estimator of the population total is asymptotically p-unbiased

$$\lim_{\nu \to \infty} \mathrm{E}[\hat{t}_\nu(z) - t_\nu(z)] = 0,$$

where $t_\nu(z)$ is the population total f population $\mathcal{P}_\nu$ to estimate (Särndal et al., 1992).

Estimator (4.4.1) of the population total is finite population consistent (Särndal et al., 1992); the proof is straightforward once one has noticed the convergence to zero of constant a of the individual IDW point estimator as $n \to N$.

4.4.3 Variance estimation

A naïve estimator of the variance can be obtained following Subsection 4.2.6 by using $\hat{\mathbf{k}}_i$ in place of $\mathbf{k}_i$ in equation (4.4.2):

$$\hat{\mathrm{V}}_n[\hat{t}(z)] = \sum_{i=1}^{N} \hat{\mathbf{k}}_i^{\top} \mathrm{E}[\mathbf{h}_i \mathbf{h}_i^{\top}] \hat{\mathbf{k}}_i + \sum_{i=1}^{N} \sum_{j \neq i} \hat{\mathbf{k}}_i^{\top} \mathrm{E}[\mathbf{h}_i \mathbf{h}_j^{\top}] \hat{\mathbf{k}}_j,$$

where estimator $\hat{\mathbf{k}}_i$ is defined in equation (4.2.36).

Analogously to what has been done for the IDW point estimator of individual quantities, two jackknife variance estimators are proposed. Let us

define the pseudo-values

$$\hat{t}(z)_k = n\hat{t}(z) - (n-1)\hat{t}(z)_{(k)},$$

where

$$\hat{t}(z)_{(k)} = \sum_{i \in S} z(\mathbf{u}_i) + \sum_{j \notin S} \hat{z}(\mathbf{u}_j) - z(\mathbf{u}_k) + \hat{z}(\mathbf{u}_k) \tag{4.4.3}$$

is the total estimator calculated on the subsample of $n-1$ locations where the kth sampled one has been omitted. The first jackknife variance estimator is defined as

$$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)] = \frac{1}{n(n-1)} \sum_{k=1}^{n} (\hat{t}(z)_k - \hat{\bar{t}}(z))^2.$$

The alternative conservative estimator is obtained by substituting the mean of the pseudo-values $\hat{\bar{t}}(z)$ for the estimated value $\hat{t}(z)$:

$$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)] = \frac{1}{n(n-1)} \sum_{k=1}^{n} (\hat{t}(z)_k - \hat{t}(z))^2.$$

If the sampling fraction $f = n/N$ is not negligible, values (4.4.3) might usefully be substituted for

$$\hat{t}(z)^*_{(k)} = \hat{t}(z) + (1-f)^{1/2}(\hat{t}(z)_{(k)} - \hat{t}(z)). \tag{4.4.4}$$

4.4.4 A GREG-like estimator

Let us start by defining the geographically weighted regression (GWR) proposed by Brunsdon et al. (1998). In a way similar to the IDW interpolator, the GWR employs the information close to an unobserved location to produce an estimate of the variable under study in that location.

Consider the general regression model for a generic location

$$z_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i,$$

whose estimate is given by

$$\hat{z}_i = \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + \varepsilon_i,$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{z}$ is the usual ordinary least squares (OLS) estimator. The regression coefficient vector $\boldsymbol{\beta}$ is global in the sense that it employs the

information concerning all individuals. Estimator $\hat{\boldsymbol{\beta}}$ is very helpful in a standard context, but in a spatial domain considering the spatial relation related to each location usually produces better results. GWR exploits this idea by modifying the usual regression model introducing a different coefficient vector which is able to use the information of each location's neighbourhood:

$$z(\mathbf{u}_i) = \mathbf{x}_i^\top \boldsymbol{\beta}(\mathbf{u}_i) + \varepsilon(\mathbf{u}_i).$$

The coefficient vector estimator is obtained as a weighted least square (WLS)

$$\hat{\boldsymbol{\beta}}(\mathbf{u}_i) = (\mathbf{X}^\top \mathbf{W}(\mathbf{u}_i)\mathbf{X})^{-1}\mathbf{X}^\top \mathbf{W}(\mathbf{u}_i)\mathbf{z}, \tag{4.4.5}$$

where the diagonal $N \times N$ matrix containing the reciprocal of the variances is substituted by the diagonal matrix containing a decreasing function of the Euclidean distance:

$$\mathbf{W}(\mathbf{u}_i) = \begin{bmatrix} w_1(\mathbf{u}_i) & 0 & \cdots & 0 \\ 0 & w_2(\mathbf{u}_i) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_N(\mathbf{u}_i) \end{bmatrix}. \tag{4.4.6}$$

Brunsdon et al. (1998) firstly suggest to use as function of the distances a step function

$$w_j(\mathbf{u}_i) = \begin{cases} 1, & \text{if } d_{ij} \leq r; \\ 0, & \text{otherwise;} \end{cases}$$

where r is the chosen radius of the neighbourhood and $d_{ij} = \|\mathbf{u}_j - \mathbf{u}_i\|$. They also suggest the use of spatially adaptive kernels leading, however, to the problem of the choice of the bandwidth parameter, h. For a deeper insight into GWR one can have a look at Fotheringham et al. (2000).

Consider now the weighting system defined in equation (4.2.1), it is a decreasing function of the Euclidean distance which, as such, is well suitable for being used as the diagonal elements of matrix (4.4.6). In this case no radius has been chosen, meaning that all the locations in the domain are taken into account. Note that the IDW interpolator weighting system can be easily modified in order to deal with the presence of a radius (Shepard,

1968). In this way the weights involved in matrix (4.4.6) are defined as

$$w_j(\mathbf{u}_i) = \begin{cases} d_{ij}^{-2}, & \text{if } i \neq j; \\ 0, & \text{otherwise.} \end{cases}$$

The proposed estimator of the total (4.4.1) is such that no auxiliary information is needed in order to obtain an estimate. In the GWR it means that matrix $\mathbf{X}$ reduces to a vector having only unit values, $\mathbf{1}_N$. As for the IDW point estimator, the information regarding the sample is summarized in vector $\boldsymbol{\mathcal{Q}}$ collecting the Bernoulli random variables managing inclusion in the sample defined in Subsection 4.2.1. Estimator (4.4.1) can, therefore, be rewritten as

$$\begin{aligned} \hat{t}(z) &= \sum_{i \in \mathcal{D}} \hat{z}(\mathbf{u}_i) = \sum_{i \in S} z(\mathbf{u}_i) + \sum_{i \notin S} \hat{z}(\mathbf{u}_i) \\ &= \left\{ \mathbf{1}_n + \sum_{i \notin S} \frac{\mathbf{1}_n^\top \mathbf{W}(\mathbf{u}_i)^{-1}}{\mathbf{1}_n^\top \mathbf{W}(\mathbf{u}_i)^{-1} \mathbf{1}_n} \right\}^\top \mathbf{z}_n, \end{aligned} \tag{4.4.7}$$

where $\mathbf{z}_n = [z(\mathbf{u}_1), \ldots, z(\mathbf{u}_n)]^\top$ is the vector of the observed values in the sampled locations.

Formulation (4.4.7) of the total estimator has the appealing look of the generalized regression (GREG) estimator (Särndal et al., 1992):

$$\hat{t}(y) = \left\{ \mathbf{1}_n + \sum_{i \notin S} \mathbf{x_i} (\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \right\}^\top \mathbf{y}^*,$$

where $\boldsymbol{\Sigma} = \text{diag}(\sigma_i^2)$ is the diagonal matrix containing the variances of each sampled individual and $\mathbf{y}^* = [w_1 y_1, \ldots, w_n y_n]$ is the vector of the observed values suitably weighted according to the sampling design. In the case without auxiliary variable it becomes an estimator where the unobserved values are estimated using a weighting system based on the inverse of the variances σ_i^2:

$$\hat{t}(y) = \left\{ \mathbf{1}_n + \sum_{i \notin S} \frac{\mathbf{1}_n^\top \boldsymbol{\Sigma}^{-1}}{\mathbf{1}_n^\top \boldsymbol{\Sigma}^{-1} \mathbf{1}_n} \right\} \mathbf{y}^*.$$

Chapter 5

Simulation study

In order to evaluate the performances of the IDW point estimator, a Monte Carlo experiment has been performed. The data used in the simulations are generated according to different random superpopulation models (Section 5.1). For each dataset the Monte Carlo experiment consists in drawing 1000 samples using SRSWoR for four different sampling fraction ($f = 0.05, 0.10, 0.15, 0.20$). Estimator (4.2.5) results are compared with the kriging predictor and the SRWoR estimator in predictive form (Appendix D and Bolfarine and Zacks, 1992) through some indicators.

The results of the simulation are obtained using a routine written for the statistical software `R`

5.1 Generating the populations

In order to fully appreciate the strengths and weaknesses of the proposed IDW point estimator in respect to the kriging predictor and the SRSWoR estimator in predictive form, we analyse different populations. Those are the realization of a random field having an exponential semivariogram with different parameters configuration, allowing us to control the reliance of the proposed technique on the change of the parameters of the random field generating the data. The superpopulation models generating the populations analysed in the simulations are collected in Table 5.1.

The spatial domain, $\mathcal{D}$, is a 20×20 square with a superimposed regu-

Table 5.1: Random superpoluation models generating the populations analysed in the simulations

	Semivariogram	μ	σ^2	ϕ	τ^2
A	exponential	2	4	15	0.00
B	exponential	2	4	6	0.00
C	exponential	2	4	45	0.00
D	exponential	2	4	90	0.00
E	exponential	2	1	15	0.00
F	exponential	2	1	45	0.00
G	exponential	2	8	15	0.00
H	exponential	2	8	45	0.00
I	exponential	2	4	15	0.25
J	exponential	2	4	45	0.25
K	exponential	2	4	15	1.00
L	exponential	2	4	45	1.00

lar grid. All the populations are generated as a realization of a Gaussian random field, using the function `grf` of the `R`-package `geoR` (Ribeiro Jr and Diggle, 2001). Table 5.1 summarizes the parameters of the exponential semivariogram of the random field generating the data as the result of an unconditional simulation (Diggle and Ribeiro, 2007). Suppose the random field $\{Z(\mathbf{u}) : \mathbf{u} \in \mathcal{D}\}$ has mean

$$\mu(\mathbf{u}) = \mathrm{E}[Z(\mathbf{u})], \quad \mathbf{u} \in \mathcal{D},$$

and covariance

$$C(\mathbf{u}, \mathbf{u}') = \mathrm{Cov}(Z(\mathbf{u}), Z(\mathbf{u}')), \quad \mathbf{u}, \mathbf{u}' \in \mathcal{D}.$$

Then, the vector of the population values $\mathbf{Z} = [Z(\mathbf{u}_1), \ldots, Z(\mathbf{u}_N)]^\top$ has mean $\mathrm{E}[\mathbf{Z}] = \boldsymbol{\mu}$ and covariance matrix $\mathrm{V}[\mathbf{Z}] = \boldsymbol{\Sigma}$ whose ijth element is the covariogram $C(\mathbf{u}_i, \mathbf{u}_j)$. Let $\boldsymbol{\varepsilon} = [\varepsilon(\mathbf{u}_1), \ldots, \varepsilon(\mathbf{u}_N)]^\top$ be a vector of N *i.i.d.* Gaussian random variables, $\mathcal{N}(0, 1)$, then the population values are obtained through the relation

$$\mathbf{Z} = \boldsymbol{\mu} + \mathbf{L}\boldsymbol{\varepsilon}.$$

In the previous equation $\mathbf{L}$ is a lower triangular matrix having strictly positive diagonal elements obtained through the Cholesky decomposition (Rubinstein and Kroese, 2007) of the matrix collecting the covariogram values of the process at the N locations

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^{\top}.$$

The main characteristics of each population of Table 5.1 are presented in Appendix E. A perspective plot and a tile plot shows the population spatial configuration. The random field semivariogram, having parameters as in Table 5.1, is plotted against the population one. A table summarizing the main descriptive statistics completes each section.

5.2 Results

In this section we present the results of the Monte Carlo experiments for the IDW point estimators (4.2.5) and (4.4.1) both for a finite spatial population. The simulation studies regard the comparison of the estimators' performances with those of the kriging predictor and the SRSWoR in predictive form through the following indicators computed on the basis of the Monte Carlo simulation:

- bias;
- root mean squared error (RMSE);
- coverage of the 95% confidence interval.

In order to evaluate the sampling fraction's effect on the estimation (prediction) of the unobserved values at unsampled locations, we firstly analyse each population at a time. We compare the IDW point estimator's performances with those of the kriging predictor and of the SRSWoR estimator in predictive form. Estimation of individual quantities and of the population total are considered. Finally, we focus on the variance estimation by comparing the performances of the naïve variance estimator and of the jackknife ones; in the case of the estimation of the population total we will consider only the jackknife ones for computational reasons.

Then, the analysis will focus on how varying each parameter of the super-population model generating the data affect estimation in comparison with the kriging and the SRSWoR estimator in predictive form. The comparison regards estimation of both individual quantities and the population total.

For what regards the kriging predictors, in the simulations we used the ordinary block kriging method. The choice of the ordinary kriging (Subsection 3.4.2) is motivated by a preliminary comparison with the results provided by the universal kriging (Subsection 3.4.3) where no appreciable difference has been found. Therefore, the computationally less demanding ordinary kriging has been adopted and implemented in the routine through the function `krige.control` of the `geoR` package (Ribeiro Jr and Diggle, 2001). The block kriging (see Table 3.1) is adopted when the location is not a spatial point but a subarea of the domain. The `geoR` package makes no difference between kriging and block kriging provided the coordinates of the point of balance of each subarea are used. The semivariogram model used for prediction is the same as the one used for generating the populations; thus, the model misspecification is not taken into account. Similarly, the recursive algorithm employs the parameters values used for generating the data as its initial values. Both the models and the parameter values are collected in Table 5.1.

5.2.1 The effect of the sampling fraction

Population A

In this subsection we concentrate on the analysis of the performances of the IDW point estimator both for individual and global estimationin the case of population A whose spatial configuration and main descriptive statistics are collected in Appendix E.1. Let us first focus on the IDW point estimator and its overall spatial behaviour as the sampling fraction increases compared to the kriging predictor and the SRSWoR estimator in predictive form. The results of the Monte Carlo experiments regarding their overall distributions of the bias, the RMSE and the coverage of the 95% confidence interval for all the locations at the same time are collected in Table 5.2. For all the three techniques the mean and median values of the bias distributions are

Table 5.2: Population A: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-3.045	-2.842	-2.615	-2.409	0.448	0.318	0.249	0.200	0.061	0.113	0.158	0.213
	1st Qu.	-0.915	-0.835	-0.753	-0.704	0.733	0.615	0.547	0.497	0.448	0.438	0.454	0.499
Point	median	-0.008	-0.039	-0.034	-0.028	1.088	0.947	0.880	0.804	0.716	0.654	0.640	0.655
Estimator	mean	-0.049	-0.042	-0.025	-0.019	1.208	1.058	0.967	0.903	0.661	0.636	0.636	0.652
	3rd Qu.	0.844	0.744	0.695	0.653	1.524	1.351	1.271	1.209	0.881	0.850	0.820	0.804
	max	3.172	2.879	2.693	2.506	3.359	3.118	2.982	2.865	0.963	0.964	0.964	0.959
	min	-3.042	-2.719	-2.420	-2.210	0.531	0.414	0.363	0.321	0.058	0.130	0.222	0.341
	1st Qu.	-0.899	-0.763	-0.635	-0.521	0.780	0.678	0.610	0.552	0.498	0.573	0.631	0.671
Kriging	median	0.041	-0.014	0.014	0.027	1.102	0.961	0.848	0.753	0.737	0.770	0.782	0.796
Predictor	mean	-0.022	-0.021	-0.006	-0.001	1.245	1.073	0.961	0.877	0.681	0.719	0.748	0.771
	3rd Qu.	0.867	0.735	0.624	0.548	1.559	1.369	1.216	1.123	0.887	0.890	0.888	0.892
	max	3.304	2.944	2.617	2.308	3.478	3.203	2.930	2.678	0.969	0.958	0.963	0.958
	min	-4.265	-4.068	-3.775	-3.554	0.366	0.245	0.181	0.142	0.037	0.079	0.122	0.171
	1st Qu.	-1.038	-0.990	-0.920	-0.859	0.665	0.588	0.550	0.523	0.062	0.106	0.156	0.211
SRSWoR	median	0.116	0.104	0.108	0.100	1.195	1.131	1.087	1.047	0.264	0.203	0.229	0.291
Estimator	mean	-0.007	-0.013	-0.001	0.001	1.413	1.334	1.281	1.234	0.384	0.351	0.362	0.394
	3rd Qu.	1.163	1.099	1.044	0.995	2.008	1.930	1.863	1.804	0.724	0.588	0.544	0.534
	max	3.359	3.136	2.944	2.767	4.385	4.277	4.111	3.986	0.952	0.945	0.939	0.939

sufficiently close to zero. In the case of the IDW, point estimator both the mean and median are negative indicating a possible slightly underestimation problem. In the kriging and SRSWoR cases the median assumes a negative sign, in contrast with a positive mean. Moreover, the three overall bias distributions shrink as the sampling fraction increases. The IDW point estimator and the kriging predictor are nearly identical and both seem reasonably symmetric around zero. The overall bias distribution of the SRSWoR estimator in predictive form shrinks less than the other two as the sampling fraction increases; moreover, the minimum values are much larger than those observed for the overall bias distributions of the other methods.

Almost the same can be said for the overall RMSE distribution: the IDW point estimator and the kriging predictor seem to globally outperform the SRSWoR estimator in predictive form; however, it shows smaller minimum values maybe due to the fact that values close to the population mean are less affected by estimation errors. The IDW point estimator's overall RMSE distribution shows smaller minimum than the kriging predictor ones and almost comparable maximum values.

For what regards the overall distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) values, the SRSWoR estimator in predictive form shows a poor behaviour in comparison with the other methods. The IDW point estimator overall distribution of the coverage has a peculiar characteristic: its mean and median decrease at increasing sampling fractions. Indeed at a sampling fraction of $f = 0.05$ the point estimator overall coverage distribution is comparable to the kriging's one.

In order to better understand the spatial behaviour of the proposed methods, the coverage of the 95% confidence interval centred in the estimated (predicted) values is plotted in Figure 5.1 for all locations. In the first row, corresponding to the IDW point estimator, the coverage tends to conform over the domain as the sampling fraction increases. This leads to the decreasing central tendency indices' values at growing sampling fractions of Table 5.2. By comparing all the plots, it appears even clearer how much both the IDW point estimator and the kriging predictor outperform the SRSRWoR estimator in predictive form in terms of coverage. Moreover, at small sampling fractions it seems that the IDW point estimator and the kriging

predictor coverages have very similar spatial distributions.

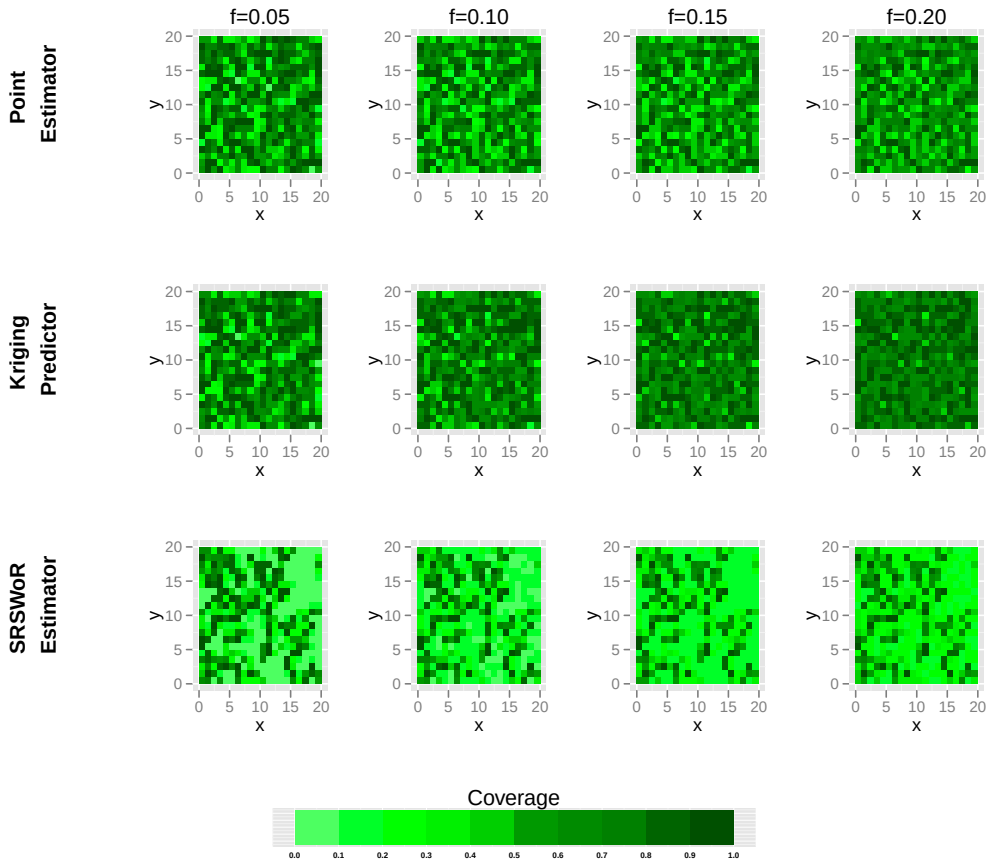

Figure 5.1: Population A: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

Table 5.3 collects the bias, RMSE and coverage of the estimator (predictor) of the population total computed through the Monte Carlo experiments. Let us first consider the bias, the IDW point estimator is the most biased of the three methods analysed: it tends to underestimate the population total, $t = 391.722$, by a quantity decreasing with the increase of the sampling fraction. Not very surprisingly the expansion estimator is the one performing better given the standard situation of population A. In terms of RMSE, the total predictor based on the individual kriging predictors performs slightly better than the IDW point estimator and the expansion estimator showing

Table 5.3: Population A: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	-19.687	-16.924	-9.927	-7.736	129.873	86.188	64.064	51.115	0.945	0.953	0.951	0.949
Kriging Predictor	-8.647	-8.519	-2.467	-0.322	126.093	81.123	58.189	44.790	0.947	0.955	0.953	0.949
Expansion Estimator	-2.810	-5.118	-0.401	0.538	152.485	104.808	80.859	66.171	0.953	0.946	0.949	0.952

faster decreasing values as the sampling fraction increases; the IDW point estimator is the second best. The coverage values of the three methods fluctuate around the nominal value of 95% and show a very similar behaviour.

Table 5.4: Population A: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-1.188	-1.226	-1.450	-1.739
	1st Qu.	-0.427	-0.351	-0.317	-0.322
	median	-0.314	-0.233	-0.190	-0.167
	mean	-0.371	-0.284	-0.254	-0.248
	3rd Qu.	-0.240	-0.154	-0.111	-0.087
	max	-0.124	-0.053	-0.033	-0.022
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-0.523	-0.978	-1.333	-1.722
	1st Qu.	-0.051	-0.138	-0.208	-0.257
	median	0.003	-0.049	-0.076	-0.098
	mean	-0.019	-0.094	-0.142	-0.183
	3rd Qu.	0.029	-0.002	-0.010	-0.024
	max	0.204	0.084	0.073	0.040
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.514	-0.976	-1.332	-1.722
	1st Qu.	-0.046	-0.136	-0.208	-0.258
	median	0.007	-0.047	-0.076	-0.098
	mean	-0.013	-0.093	-0.142	-0.183
	3rd Qu.	0.033	-0.001	-0.010	-0.024
	max	0.232	0.090	0.074	0.040

Finally, for population A, the kriging predictor and the corresponding predictor of the population total perform slightly better than the iDW point estimator and IDW point estimator of the total. In individual estimation (prediction), both these techniques outperform the SRSWoR estimator in predictive form. However, this is not true when the object of inference is the population total since population A is fair enough.

Consider now the variance estimation, we analyse the three estimators

proposed in Subsection 4.2.6 for the IDW point estimator. The naïve estimator bias is compared with the jackknife variance estimators (4.2.37) and (4.2.38) where modification (4.2.39) is used since in any of this cases the sampling fraction, f, is not negligible. The distribution of the overall bias of the variance estimators of the Monte Carlo experiments are collected in Table 5.4. The overall bias distribution of the two jackknife estimators is quite the same. The bias of the estimators of the variance of the IDW point estimator is less shrunk around its median as the sampling fraction increases; the same can be said for the jackknife estimators. At small sampling fractions, it seems that both the jackknife estimators performs slightly better having a smaller difference between the maximum and minimum values. However, the maximum of their overall bias distribution is always positive. At higher sampling fractions, the variance estimators perform quite in the same way. However, the naïve estimator bias is negative for each location. Estimator $\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$ seems the more conservative variance estimator.

Table 5.5: Population A: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	-278.204	-380.318	-9.446	125.877
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	-250.543	-375.690	-7.883	126.629

Analogously we consider the variance estimation of the IDW point estimator of the population total. The analysis focuses only on the two jackknife estimators of Subsection 4.4.3 where modification (4.4.4) is applied. The naïve estimator is not consider due to its computational intractability. The results of the Monte Carlo simulation are collected in Table 5.5. Even in the case of the estimator of the variance of the IDW point estimator of the population total both the jackknife estimators show a very similar behaviour.

Population B

Appendix E.2 collects the main descriptive statistics of population B. The overall distributions of the bias, the RMSE and the coverage of the 95%

Table 5.6: Population B: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-4.158	-3.926	-3.638	-3.367	0.561	0.413	0.331	0.277	0.048	0.093	0.153	0.207
	1st Qu.	-1.158	-1.077	-0.982	-0.910	0.990	0.840	0.741	0.661	0.405	0.405	0.430	0.482
Point	median	0.066	0.057	0.061	0.076	1.399	1.302	1.208	1.107	0.691	0.641	0.625	0.634
Estimator	mean	-0.021	-0.020	-0.003	0.001	1.590	1.425	1.316	1.235	0.640	0.619	0.624	0.642
	3rd Qu.	1.143	1.065	0.983	0.924	2.050	1.887	1.792	1.703	0.889	0.866	0.847	0.830
	max	3.990	3.709	3.496	3.313	4.358	4.176	3.976	3.794	0.972	0.960	0.957	0.952
	min	-4.378	-3.930	-3.578	-3.241	0.520	0.447	0.402	0.378	0.044	0.090	0.145	0.240
	1st Qu.	-1.252	-1.123	-0.988	-0.833	0.913	0.877	0.794	0.740	0.293	0.417	0.527	0.602
Kriging	median	0.063	0.073	0.074	0.065	1.430	1.300	1.205	1.107	0.629	0.662	0.702	0.738
Predictor	mean	-0.019	-0.023	-0.004	0.001	1.591	1.456	1.346	1.252	0.569	0.628	0.682	0.725
	3rd Qu.	1.140	1.064	0.948	0.859	2.075	1.904	1.777	1.660	0.855	0.865	0.871	0.877
	max	4.035	3.718	3.452	3.223	4.559	4.173	3.924	3.678	0.958	0.961	0.963	0.958
	min	-5.042	-4.807	-4.510	-4.22	0.406	0.271	0.205	0.166	0.037	0.079	0.122	0.166
	1st Qu.	-1.360	-1.295	-1.204	-1.122	0.768	0.687	0.653	0.626	0.064	0.106	0.156	0.209
SRSWoR	median	0.065	0.057	0.065	0.063	1.419	1.354	1.298	1.252	0.213	0.158	0.201	0.266
Estimator	mean	-0.014	-0.018	-0.003	0.000	1.599	1.513	1.454	1.402	0.364	0.333	0.350	0.385
	3rd Qu.	1.283	1.220	1.160	1.083	2.164	2.092	2.023	1.962	0.680	0.550	0.494	0.495
	max	4.105	3.880	3.684	3.491	5.154	5.024	4.859	4.702	0.950	0.949	0.944	0.944

confidence interval centred in the estimated (predicted) values are collected in Table 5.6. The dispersion of the distributions of the overall bias for the analysed methods shrinks as the sampling fraction increases. For all the techniques the mean and median overall bias fluctuate around the zero value suggesting that in the case of population B the estimators and predictor are generally unbiased. The IDW point estimator and the kriging predictor show a similar overall bias distribution and both outperform the SRSWoR estimator in predictive form.

In terms of the overall distribution of the RMSE, the SRSWoR in predictive form shows slightly lower minimums and greater maximums, but similar means and medians. The IDW point estimator and kriging predictor again show very similar distributions having lower values.

The overall coverage distribution of the SRSWoR estimator in predictive form has a worse behaviour than those of the IDW point estimator and kriging predictor: the mean and median values are lower and the the third quartile has a decreasing tendency as the sampling fraction increases. The IDW point estimator seems to perform slightly better than the kriging predictor when the sampling fraction is small, $f = 0.05$. At a sampling fraction of 10% the overall distributions of the coverage of the IDW point estimator and of the kriging predictor are still similar, but at the higher fractions the latter has a more desirable behaviour.

In order to better appreciate the coverages of the 95% confidence interval centred in the estimated (predicted) value for the analysed techniques we turn to their spatial distributions. Again, both the IDW point estimator and the kriging predictor outperform the SRSWoR estimator in predictive form (Figure 5.2). Moreover, the coverage maps show that the IDW point estimator has a better behaviour than the kriging predictor at the smallest sampling fraction, $f = 0.05$. As the sampling fraction increases, the comparison of the plots in the first and second rows of Figure 5.2 shows that the coverage of the kriging predictor gets little by little higher than the one of the IDW point estimator.

The values regarding the bias, the RMSE and the coverage of the estimator of the population total based on the IDW point estimator, the kriging predictor of the population total and the expansion estimator are collected

in Table 5.7. All the proposed methods of inference on the population total seem to underestimate the true value at almost all the sampling fractions analysed, being negative biased. Among those, the expansion estimator is the less biased. However, along with the kriging predictor its bias is not monotonously decreasing. The estimator of the population total based on the IDW point estimator has a slightly nicer behaviour than the one of the kriging predictor in terms of bias.

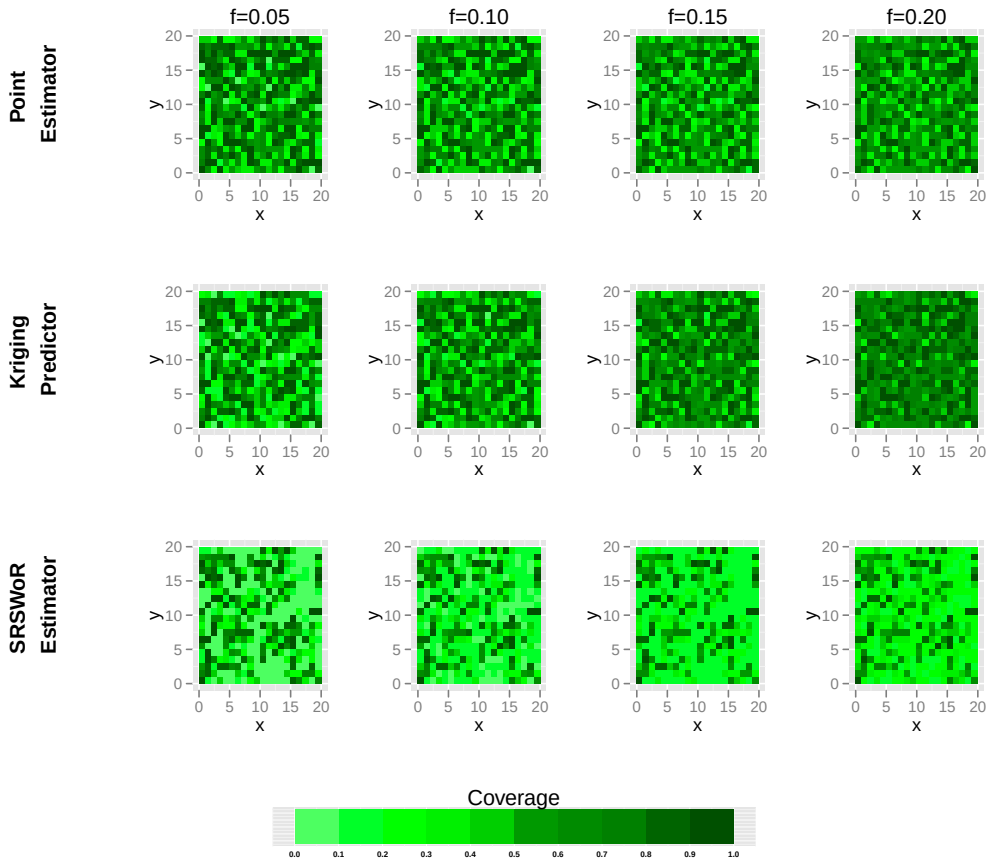

Figure 5.2: Population B: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

The RMSE of the total estimator based on the IDW point estimator is quite the same as the one of the kriging predictor. Both of them have values lower than the ones showed by the expansion estimator.

Table 5.7: Population B: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	-8.439	-8.036	-1.123	0.489	157.050	106.183	81.154	65.484	0.943	0.951	0.947	0.958
Kriging Predictor	-7.552	-9.330	-1.774	0.490	159.617	107.530	80.411	62.896	0.943	0.953	0.953	0.952
Expansion Estimator	-5.708	-7.242	-1.178	0.146	167.097	114.850	89.413	73.843	0.949	0.954	0.952	0.951

No substantial difference can be noticed in the coverages of the 95% confidence interval centred in the estimated (predicted) values, since they are close to the nominal level. Moreover, there is not any clear trend in the coverage as the sampling fraction increases.

Table 5.8: Population B: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-1.485	-1.837	-2.390	-2.862
	1st Qu.	-0.729	-0.624	-0.611	-0.639
	median	-0.584	-0.443	-0.357	-0.329
	mean	-0.613	-0.505	-0.471	-0.468
	3rd Qu.	-0.444	-0.299	-0.223	-0.176
	max	-0.225	-0.111	-0.069	-0.044
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-0.897	-1.543	-2.170	-2.778
	1st Qu.	-0.105	-0.286	-0.427	-0.533
	median	-0.005	-0.089	-0.147	-0.190
	mean	-0.050	-0.182	-0.271	-0.348
	3rd Qu.	0.043	-0.003	-0.019	-0.038
	max	0.223	0.098	0.105	0.066
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.885	-1.540	-2.169	-2.778
	1st Qu.	-0.098	-0.283	-0.427	-0.533
	median	0.001	-0.088	-0.146	-0.190
	mean	-0.041	-0.180	-0.270	-0.348
	3rd Qu.	0.050	-0.002	-0.019	-0.038
	max	0.249	0.101	0.106	0.067

Finally, for population B the IDW point estimator and the kriging predictor outperform the SRSWoR estimator in predictive regardless of the dimension of the sample. At the smallest sampling fraction, $f = 0.05$, the IDW point estimator is the most suitable for inference on individual quantities. This is not true any more at a sampling fraction of $f = 0.10$ when the two techniques are practically equivalent and, moreover, the relationship is

reversed at higher sampling fractions. For what regards inference on the population total, estimator (4.4.1) and the kriging predictor are just slightly more biased than the expansion estimator; however, they have smaller RMSE. The coverage values are similar and fluctuate around the nominal level.

Consider now variance estimation for the IDW point estimator, Table 5.8 collects the bias distribution of the naïve variance estimator and of the two jackknife estimators, where modification (4.2.39) has been adopted. Estimator $\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$ seems the more conservative presenting, however, negative values.

Table 5.9: Population B: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	438.450	-228.476	109.671	281.262
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	466.997	-224.075	111.084	281.930

We analyse the bias distribution of the jackknife estimators of the variance of the IDW point estimator of the population total, where the correction for the sampling fraction has been adopted. The results of the Monte Carlo experiment for population B are in Table 5.9. The two proposed estimators show little difference in the bias distribution, which has not a monotonous trend.

Population C

Appendix E.3 collects the main descriptive statistics of population C along with the perspective, tile and semivariogram plots. Table 5.10 collects the main descriptive quantities of the overall bias, RMSE and coverage distributions of the IDW point estimator, the kriging predictor and the SRSWoR estimator in predictive form. The kriging predictor overall bias distribution has generally lower values than those of the SRSWoR estimator and of the IDW point estimator; however, the differences with the latter are substantially smaller. The spatial median and mean of the three bias distri-

Table 5.10: Population C: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-2.211	-1.917	-1.741	-1.612	0.252	0.165	0.130	0.105	0.043	0.091	0.146	0.200
	1st Qu.	-0.444	-0.435	-0.384	-0.349	0.466	0.372	0.328	0.293	0.541	0.495	0.506	0.524
Point	median	0.017	0.033	0.038	0.035	0.633	0.548	0.501	0.454	0.760	0.703	0.683	0.679
Estimator	mean	0.013	0.012	0.010	0.010	0.751	0.642	0.585	0.544	0.696	0.660	0.658	0.668
	3rd Qu.	0.482	0.440	0.390	0.358	0.959	0.843	0.759	0.710	0.896	0.866	0.842	0.820
	max	2.055	1.840	1.690	1.571	2.367	2.069	1.919	1.826	0.969	0.968	0.963	0.957
	min	-2.035	-1.577	-1.397	-1.250	0.304	0.246	0.205	0.177	0.063	0.105	0.161	0.224
	1st Qu.	-0.456	-0.392	-0.370	-0.339	0.490	0.420	0.371	0.330	0.625	0.627	0.657	0.678
Kriging	median	0.030	0.040	0.014	0.004	0.647	0.546	0.495	0.457	0.807	0.792	0.802	0.809
Predictor	mean	0.003	0.005	0.003	0.001	0.743	0.629	0.564	0.514	0.735	0.739	0.757	0.774
	3rd Qu.	0.459	0.371	0.320	0.321	0.896	0.785	0.702	0.645	0.907	0.904	0.902	0.896
	max	1.906	1.582	1.455	1.316	2.199	1.712	1.592	1.490	0.968	0.966	0.969	0.966
	min	-3.583	-3.401	-3.207	-3.010	0.275	0.182	0.138	0.111	0.038	0.081	0.123	0.172
	1st Qu.	-0.845	-0.798	-0.746	-0.703	0.507	0.449	0.422	0.404	0.061	0.106	0.158	0.212
SRSWoR	median	0.079	0.076	0.071	0.070	0.916	0.865	0.833	0.804	0.249	0.188	0.235	0.303
Estimator	mean	-0.005	-0.005	-0.005	-0.003	1.053	0.994	0.955	0.921	0.384	0.345	0.362	0.398
	3rd Qu.	0.865	0.810	0.759	0.712	1.457	1.404	1.363	1.322	0.718	0.571	0.529	0.536
	max	2.966	2.815	2.651	2.515	3.680	3.578	3.473	3.363	0.947	0.942	0.937	0.944

butions are close enough to zero and the difference between the maximum and minimum decreases as the sampling fraction increases.

The RMSE overall distribution of the IDW point estimator and of the kriging predictor present basically the same mean and median values, whereas the corresponding SRSWoR estimator's ones are substantially larger. The IDW point estimator present lower minimum and larger maximum RMSE values than those of the kriging predictor, while both outperform the SRSWoR estimator in predictive form.

In terms of coverage the overall mean and median values of the kriging predictor are closer to the nominal level than those of the other methods. However, the difference with the IDW point estimator is smaller than the one with the SRSWoR estimator. Moreover, the latter has overall maximum values which fail in reach the nominal level at any sampling fraction. Looking at the trend at increasing sampling fractions, the mean and the median are not increasing monotonous for any of the three methods; furthermore, the IDW point estimator's median is monotonously decreasing. A deeper analysis on the spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) values follows. The IDW point estimator's coverage (Figure 5.3) tends to become more uniform as the sample size increases, by levelling peaks and troughs. As the sampling fraction increases, the kriging predictor basically maintains the same coverage level for those locations having a value close to the nominal level and improves coverage for the locations affected by lower values. The SRSWoR estimator in predictive form has a spatial coverage distribution which, consistently to its formulation, provides a good coverage only for those locations having a population value close to the population mean. Finally, the comparison of the plots in Figure 5.3 shows that both the IDW point estimator and the kriging predictor have a much better behaviour than the SRSWoR estimator in predictive form in terms of coverage. Moreover, at the smallest sampling fraction, $f = 0.05$, the spatial distribution of the coverage is almost the same to the kriging predictor's one. However, at increasing sampling fractions the latter removes the troughs more efficiently.

Let us now consider the inference on the population total. Table 5.11 collects the bias, RMSE and coverage of the 95% confidence interval centred

in the estimated (predicted) values for the three methods in the case of population C. Estimator (4.4.1) of the population total has the largest bias and tends to overestimate the true value. The expansion estimator has a slightly better behaviour and underestimate the population total. Finally, the predictor of the population total based on the kriging has the smallest bias among the analysed methods. Moreover, at the highest sampling fraction, $f = 0.20$ its bias is close to zero.

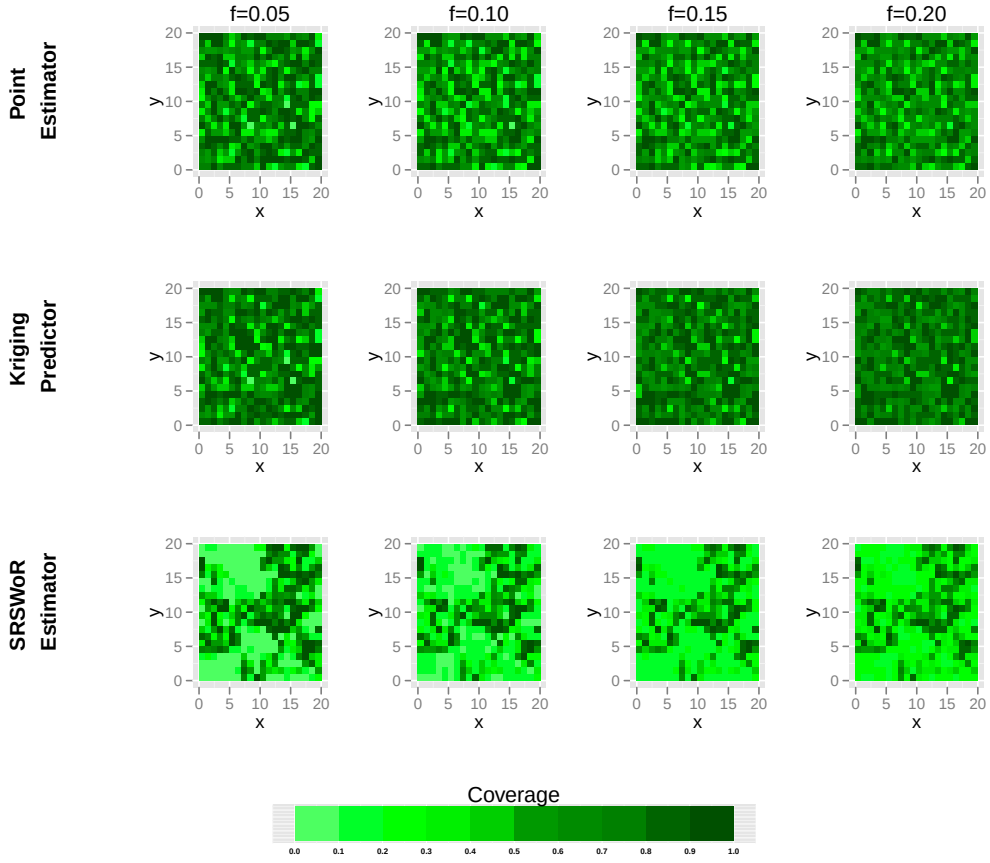

Figure 5.3: Population C: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

In terms of RMSE of the proposed techniques, the IDW point estimator of the population total and the kriging predictor of the population total have quite similar values; the former has slightly lower values. The expansion

Table 5.11: Population C: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	5.033	4.793	4.026	3.856	86.495	53.808	40.741	33.315	0.959	0.947	0.947	0.943
Kriging Predictor	1.341	1.859	1.121	0.485	75.832	44.325	32.077	25.273	0.942	0.950	0.958	0.956
Expansion Estimator	-2.017	-1.857	-2.070	-1.039	113.602	76.533	60.455	50.882	0.949	0.945	0.945	0.949

estimator's RMSE is substantially larger than the ones of the other techniques at all the sampling fractions.

Apart from the smallest sampling fraction, $f = 0.05$, the IDW point estimator of the population total fails to reach the nominal level. It happens just the opposite for the kriging predictor of the population total; it exceeds the nominal 95% level at $f = 0.15$ and $f = 0.20$. The expansion estimator of the population total has coverages lower than the nominal level at any sampling fraction.

Table 5.12: Population C: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-0.558	-0.496	-0.565	-0.670
	1st Qu.	-0.183	-0.134	-0.118	-0.117
	median	-0.128	-0.082	-0.064	-0.054
	mean	-0.150	-0.107	-0.096	-0.093
	3rd Qu.	-0.086	-0.053	-0.040	-0.029
	max	-0.045	-0.018	-0.010	-0.006
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-0.197	-0.373	-0.516	-0.658
	1st Qu.	-0.010	-0.042	-0.069	-0.091
	median	0.008	-0.009	-0.022	-0.030
	mean	0.001	-0.031	-0.053	-0.069
	3rd Qu.	0.018	0.002	-0.003	-0.006
	max	0.109	0.067	0.030	0.021
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.194	-0.373	-0.515	-0.658
	1st Qu.	-0.008	-0.042	-0.069	-0.091
	median	0.010	-0.009	-0.022	-0.030
	mean	0.003	-0.030	-0.053	-0.069
	3rd Qu.	0.020	0.003	-0.003	-0.006
	max	0.117	0.069	0.031	0.021

The overall bias distribution of the estimators of the variance of the IDW point estimator proposed in Subsection 4.2.6 are summarized in Table 5.12.

Jackknife estimator (4.2.38) with modification (4.2.39) for not negligible sampling fractions is the more conservative, despite even assuming negative values.

Table 5.13: Population C: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	-32.730	71.167	41.302	16.482
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	-15.975	73.510	41.997	16.812

The results on the bias of the jackknife estimators of the variance of estimator (4.4.1) of the population total adopting modification (4.4.4) are collected in Table 5.9. Both of them are far from being unbiased; however, estimator $\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$ has a more conservative behaviour and, therefore, is to be preferred.

Population D

Table 5.14 summarizes the overall distribution of the bias, the RMSE and the coverage of the 95% confidence interval of the three methods for population D (Appendix E.4). The IDW point estimator's overall bias distribution has maximum values larger than the kriging predictor's ones despite showing slightly lower minimum values; median and mean values are similar. In terms of the overall bias distribution, the SRSWoR estimator in predictive form performs worse than the other two methods; moreover, it is less concentrated around its central tendency values.

In terms of the overall RMSE distribution, both the IDW point estimator and the kriging predictor outperform the SRSWoR estimator. By comparing the distributions of the first two techniques, it results that the IDW point estimator has lower minimum and larger maximum values than the kriging distribution's ones; medians and means gets more different as the sampling fraction increases, while first quartile values are pretty similar.

The overall coverage distribution of the kriging predictor shows higher mean and median values than those of the other techniques in spite of the

Table 5.14: Population D: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-1.846	-1.605	-1.449	-1.315	0.176	0.114	0.086	0.068	0.043	0.092	0.152	0.201
	1st Qu.	-0.350	-0.304	-0.276	-0.255	0.362	0.278	0.244	0.219	0.549	0.511	0.528	0.556
Point	median	-0.001	0.006	-0.001	-0.004	0.517	0.428	0.381	0.350	0.768	0.718	0.702	0.701
Estimator	mean	0.029	0.021	0.018	0.016	0.608	0.506	0.457	0.423	0.710	0.675	0.675	0.684
	3rd Qu.	0.439	0.369	0.326	0.301	0.756	0.629	0.578	0.538	0.899	0.865	0.851	0.836
	max	2.253	2.013	1.864	1.731	2.373	2.169	2.043	1.940	0.980	0.967	0.955	0.952
	min	-1.747	-1.580	-1.506	-1.394	0.226	0.148	0.118	0.099	0.037	0.076	0.122	0.166
	1st Qu.	-0.317	-0.265	-0.232	-0.206	0.368	0.283	0.242	0.215	0.668	0.660	0.676	0.693
Kriging	median	-0.001	0.014	-0.008	-0.016	0.503	0.392	0.339	0.310	0.839	0.828	0.832	0.822
Predictor	mean	0.007	0.000	-0.001	0.000	0.567	0.450	0.397	0.361	0.770	0.767	0.777	0.788
	3rd Qu.	0.337	0.255	0.224	0.214	0.658	0.548	0.491	0.451	0.917	0.918	0.913	0.909
	max	1.954	1.567	1.332	1.144	2.067	1.715	1.653	1.573	0.962	0.965	0.959	0.958
	min	-2.758	-2.625	-2.485	-2.336	0.226	0.148	0.118	0.099	0.037	0.076	0.122	0.166
	1st Qu.	-0.666	-0.638	-0.598	-0.560	0.452	0.403	0.387	0.370	0.063	0.106	0.159	0.212
SRSWoR	median	-0.136	-0.132	-0.123	-0.117	0.782	0.740	0.719	0.693	0.223	0.159	0.212	0.291
Estimator	mean	0.004	-0.001	0.002	0.000	0.918	0.868	0.837	0.808	0.359	0.320	0.348	0.390
	3rd Qu.	0.851	0.802	0.763	0.719	1.269	1.216	1.179	1.140	0.654	0.502	0.482	0.518
	max	3.082	2.915	2.796	2.651	3.152	3.059	2.997	2.917	0.950	0.949	0.942	0.943

sampling fraction; however, the IDW point estimator's ones are not as dissimilar as those of the SRSWoR in predictive form. The maximum values of the overall coverage distribution of the IDW point estimator are similar to the kriging ones and both exceed the nominal level. The maximums of the SRSWoR estimator overall coverage distribution barely reach the nominal level of 95%.

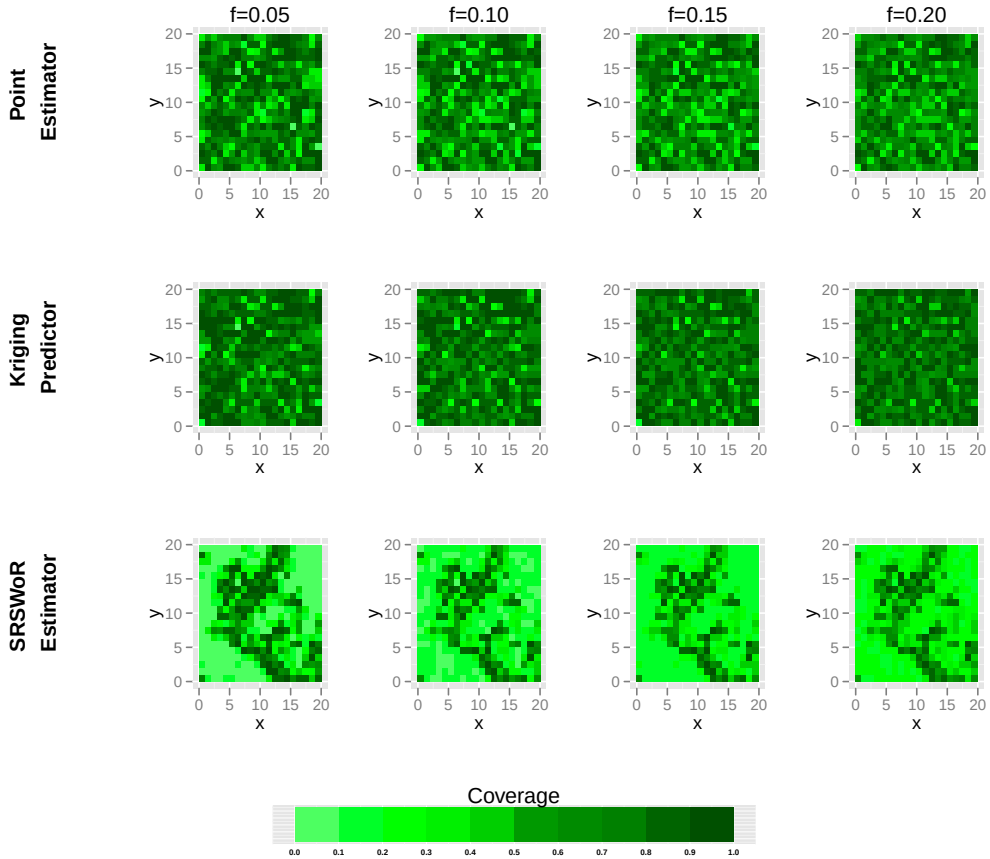

Figure 5.4: Population D: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

In order to better understand the spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) values, the coverage maps relative to each technique at the different sampling fractions are collected in Figure 5.4. It can be seen that as the sampling fraction

Table 5.15: Population D: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	11.800	8.256	7.360	6.445	68.487	42.113	33.480	27.420	0.948	0.948	0.957	0.944
Kriging Predictor	2.826	-0.169	-0.336	0.017	59.473	32.746	24.452	19.300	0.951	0.947	0.951	0.940
Expansion Estimator	1.410	-0.223	0.625	0.196	94.193	62.977	51.851	44.347	0.951	0.945	0.950	0.955

increases both the coverage of the IDW point estimator and kriging predictor gets higher; the same can not be said for the SRSWoR estimator in predictive form since the values of the coverage seems to remain quite the same, perhaps they even tend to decrease. As already pointed out for the overall coverage distribution, it seems that the IDW point estimator works almost as well as the kriging predictor for the lowest sampling fraction; as it grows, the latter shows higher and higher coverages.

Table 5.16: Population D: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-0.420	-0.535	-0.596	-0.665
	1st Qu.	-0.137	-0.084	-0.072	-0.066
	median	-0.081	-0.047	-0.037	-0.031
	mean	-0.102	-0.069	-0.060	-0.057
	3rd Qu.	-0.050	-0.029	-0.020	-0.016
	max	-0.019	-0.007	-0.003	-0.002
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-0.153	-0.362	-0.492	-0.618
	1st Qu.	-0.002	-0.023	-0.037	-0.047
	median	0.007	-0.003	-0.012	-0.017
	mean	0.004	-0.017	-0.032	-0.042
	3rd Qu.	0.016	0.005	-0.001	-0.003
	max	0.079	0.045	0.029	0.026
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.143	-0.360	-0.492	-0.618
	1st Qu.	-0.001	-0.023	-0.037	-0.047
	median	0.008	-0.003	-0.012	-0.017
	mean	0.006	-0.016	-0.032	-0.042
	3rd Qu.	0.018	0.005	-0.001	-0.003
	max	0.086	0.046	0.029	0.026

The indicators regarding the results of the Monte Carlo experiment on the techniques for inference on the population total are summarized in Table 5.15. In terms of bias the kriging predictor and the expansion estimator

have similar behaviours; it is worth noticing that at a sampling fraction $f = 20\%$ the bias is almost null. Both techniques perform better than the IDW point estimator showing lower bias.

In terms of RMSE the IDW point estimator shows values close to the kriging predictor's ones, despite being the most biased. Both outperform the expansion estimator showing substantially lower RMSE.

The coverages of three methods for inference on the population total are close to the nominal value of 95% regardless of the sampling fraction.

Table 5.16 shows the overall bias distributions of the three variance estimator for the individual IDW point estimator. The jackknife conservative estimator $\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$ seems the more conservative despite presenting negative values.

Table 5.17: Population D: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	441.242	245.552	30.417	6.374
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	453.126	247.404	31.060	6.681

Analogously, the conservative jackknife estimator of the variance of the IDW point estimator of the population total is the one to be preferred compared to the non conservative one.

Population E

Population E (Appendix E.5) presents a range value slightly larger than the mean distance observed in the domain and a low sill value. The main indicators of the overall distributions of the bias, the RMSE and of the coverage for the inference on individual quantities at different sampling fractions are shown in Table 5.18. In terms of bias, the three techniques presents quite the same median and mean values; however, the overall distribution of the IDW point estimator and the kriging predictor are more concentrated around their central tendency values. Moreover, at low sampling fraction (5% and 10%) these two methods have quite the same distribution. As the sampling

Table 5.18: Population E: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-1.523	-1.421	-1.308	-1.205	0.224	0.159	0.124	0.100	0.061	0.113	0.158	0.213
	1st Qu.	-0.458	-0.417	-0.376	-0.352	0.367	0.308	0.273	0.249	0.448	0.438	0.454	0.499
Point	median	-0.004	-0.019	-0.017	-0.014	0.544	0.474	0.440	0.402	0.716	0.655	0.640	0.656
Estimator	mean	-0.025	-0.021	-0.012	-0.010	0.604	0.529	0.484	0.452	0.661	0.636	0.636	0.652
	3rd Qu.	0.422	0.372	0.348	0.327	0.762	0.676	0.636	0.604	0.881	0.850	0.820	0.804
	max	1.586	1.440	1.347	1.253	1.680	1.559	1.491	1.433	0.963	0.964	0.964	0.959
	min	-1.521	-1.360	-1.210	-1.105	0.183	0.122	0.091	0.071	0.037	0.079	0.122	0.171
	1st Qu.	-0.450	-0.382	-0.318	-0.261	0.390	0.339	0.305	0.276	0.498	0.574	0.631	0.671
Kriging	median	0.021	-0.007	0.007	0.013	0.551	0.481	0.424	0.376	0.738	0.770	0.782	0.796
Predictor	mean	-0.011	-0.011	-0.003	0.000	0.623	0.536	0.481	0.439	0.681	0.719	0.748	0.772
	3rd Qu.	0.434	0.368	0.312	0.274	0.780	0.685	0.608	0.562	0.887	0.890	0.888	0.892
	max	1.652	1.472	1.308	1.154	1.739	1.602	1.465	1.339	0.969	0.958	0.963	0.958
	min	-2.133	-2.034	-1.888	-1.777	0.183	0.122	0.091	0.071	0.037	0.079	0.122	0.171
	1st Qu.	-0.519	-0.495	-0.460	-0.429	0.333	0.294	0.275	0.262	0.062	0.106	0.156	0.211
SRSWoR	median	0.058	0.052	0.054	0.050	0.597	0.565	0.544	0.524	0.264	0.204	0.229	0.291
Estimator	mean	-0.004	-0.006	-0.001	0.001	0.707	0.667	0.640	0.617	0.384	0.351	0.362	0.394
	3rd Qu.	0.582	0.550	0.522	0.498	1.004	0.965	0.931	0.902	0.724	0.588	0.544	0.534
	max	1.679	1.568	1.472	1.383	2.193	2.138	2.056	1.993	0.952	0.945	0.939	0.939

fraction increases the kriging predictor's overall distribution becomes more shrunk around its median.

The RMSE overall distributions of the IDW point estimator and of the kriging predictor have substantially lower values than those of the distribution of the SRSWoR. The IDW point estimator seems to perform slightly better at the lowest sampling fraction since its maximum is lower than the one of the overall RMSE distribution of the kriging predictor; nevertheless, the former has a slightly higher minimum value. As the sampling fraction increases this relation is reverted.

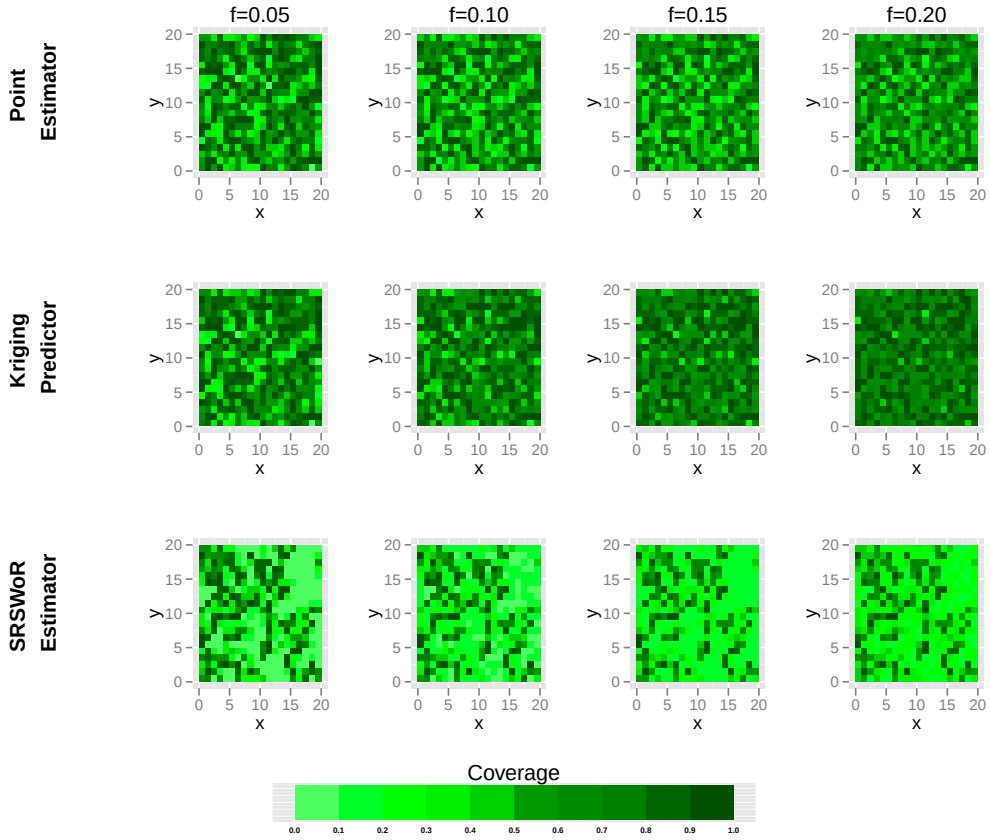

Figure 5.5: Population E: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

At a sampling fracion of $f = 0.05$ there is quite no difference between the overall coverage distribution of the IDW point estimator and the kriging

Table 5.19: Population E: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	-9.844	-8.462	-4.963	-3.868	64.936	43.094	32.032	25.557	0.945	0.953	0.951	0.949
Kriging Predictor	-4.324	-4.259	-1.233	-0.161	63.047	40.562	29.095	22.395	0.947	0.955	0.953	0.949
Expansion Estimator	-1.405	-2.559	-0.200	0.269	76.243	52.404	40.429	33.086	0.953	0.946	0.949	0.952

predictor. However, for higher sampling fractions it seems that the latter has better performances. The SRSWoR estimator in predictive form has a less appealing overall coverage distribution compared to the other ones.

In Figure 5.5 the maps of the coverage of the different methods over the domain are presented in order to give its spatial distribution. At the lowest sampling fraction, the maps of the coverage of the IDW point estimator and the kriging predictor are practically equal. As the sampling fraction increases the difference becomes relevant as the latter shows higher and higher coverages. The SRSWoR estimator has much lower coverages than the other techniques for inference on individual quantities.

Table 5.20: Population E: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-0.297	-0.307	-0.363	-0.435
	1st Qu.	-0.107	-0.088	-0.079	-0.081
	median	-0.079	-0.058	-0.047	-0.042
	mean	-0.093	-0.071	-0.064	-0.062
	3rd Qu.	-0.060	-0.039	-0.028	-0.022
	max	-0.031	-0.013	-0.008	-0.005
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-0.131	-0.245	-0.333	-0.430
	1st Qu.	-0.013	-0.035	-0.052	-0.064
	median	0.001	-0.012	-0.019	-0.025
	mean	-0.005	-0.024	-0.035	-0.046
	3rd Qu.	0.007	0.000	-0.003	-0.006
	max	0.051	0.021	0.018	0.010
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.129	-0.244	-0.333	-0.430
	1st Qu.	-0.012	-0.034	-0.052	-0.064
	median	0.002	-0.012	-0.019	-0.025
	mean	-0.003	-0.023	-0.035	-0.046
	3rd Qu.	0.008	0.000	-0.002	-0.006
	max	0.058	0.022	0.018	0.010

Let us now consider inference on the population total: Table 5.19 summarizes the indicators used for comparing the three techniques. In terms of bias the expansion estimator shows the best results; however, it is not dissimilar to the kriging predictor's one. The IDW point estimator of the population total seems to underestimate the true value slightly more than the other methods.

The RMSE of the IDW point estimator and of the kriging predictor are quite the same, especially at a sampling fraction of 5%. Despite being less biased than the other techniques, the expansion estimator has higher RMSE suggesting that it has higher variance.

The coverage of the 95% confidence interval of the three methods are pretty similar, and they lay close to the nominal level regardless of the sampling fraction.

Table 5.20 summarizes the overall bias distribution of the variance estimators of the IDW point estimator at different sampling fractions. As pointed out in the literature (Wolter, 2007), estimator (4.2.37) involving correction (4.2.39) for not negligible sampling fractions is the most conservative among the jackknife ones. The results prove that it is the best among those proposed in Subsection 4.2.6 too.

Table 5.21: Population E: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	-69.551	-95.080	-2.362	31.469
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	-62.636	-93.922	-1.971	31.657

Table 5.21 shows the behaviour of the two jackknife estimator of the variance of the IDW point estimator of the population total. Analogously to the estimator of the variance of the individual estimator, estimator $\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$ shows a more conservative behaviour.

Population F

Population F (Appendix E.6) is the realization of a stationary random field having exponential semivariogram with a low sill value and a relatively high range parameter. Table 5.22 shows the main indicators for the overall bias, RMSE and coverage distributions of the three techniques for inference on the individual values. The mean and median of the overall bias distribution of the IDW point estimator is close to zero as well as those of the kriging predictor and the SRSWoR estimator in predictive form; however, it is less concentrated around its central tendency values regardless of the sampling fraction. For what regards the IDW point estimator, its overall bias distribution has minimum values similar to the ones of the kriging predictors at low sample fractions, $f = 0.05$ and $f = 0.10$; whereas the former has higher maximum. As the sampling fraction increases, the resulting overall bias distribution of the kriging predictor becomes more shrunk around its median.

In terms of RMSE, there is not any appreciable difference between the medians and means of the overall distributions of the IDW point estimator and f the kriging predictor; the minimums are almost the same too, while the former's maximum slightly exceeds the latter's. Both of them outperform the SRSWoR estimator in predictive form since its overall distribution presents higher values.

The overall coverage distribution of the kriging predictor has higher mean and median values. The IDW point estimator's distribution ones are slightly lower, whereas the SRSWoR has a poor behaviour in terms of coverage.

The coverage maps for the three techniques at different sampling fractions are collected in Figure 5.6. It can be clearly seen that both the IDW point estimator and the kriging predictor outperform the SRSWoR estimator in predictive form. Quite surprisingly, given the high range parameter used for generating population F, at the lowest sampling fraction (5%) the IDW point estimator performs not much worse than expected compared to the kriging predictor. As the sampling fraction increases both get to perform better; however, the coverage of the kriging predictor gets closer to the nominal level quicker than the IDW point estimator's one. The SRSWoR estimator's

Table 5.22: Population F: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-1.137	-1.026	-0.953	-0.891	0.127	0.083	0.061	0.051	0.045	0.100	0.146	0.213
	1st Qu.	-0.191	-0.166	-0.158	-0.152	0.227	0.179	0.152	0.134	0.554	0.498	0.506	0.543
Point	median	-0.001	0.005	0.009	0.008	0.324	0.275	0.246	0.226	0.768	0.718	0.694	0.694
Estimator	mean	0.015	0.013	0.010	0.009	0.383	0.324	0.293	0.272	0.705	0.677	0.670	0.682
	3rd Qu.	0.270	0.244	0.212	0.198	0.490	0.416	0.383	0.359	0.909	0.886	0.862	0.848
	max	1.161	1.007	0.916	0.846	1.242	1.104	1.051	1.008	0.976	0.971	0.971	0.961
	min	-1.175	-1.011	-0.882	-0.772	0.132	0.088	0.066	0.053	0.035	0.076	0.125	0.173
	1st Qu.	-0.188	-0.184	-0.161	-0.144	0.248	0.197	0.170	0.152	0.624	0.635	0.659	0.692
Kriging	median	0.001	0.009	0.021	0.010	0.324	0.268	0.231	0.213	0.843	0.828	0.824	0.826
Predictor	mean	0.009	0.004	0.002	0.002	0.382	0.313	0.276	0.250	0.749	0.755	0.770	0.787
	3rd Qu.	0.246	0.176	0.153	0.137	0.471	0.380	0.341	0.313	0.922	0.912	0.914	0.911
	max	1.070	0.811	0.689	0.605	1.235	1.099	0.991	0.893	0.967	0.969	0.973	0.950
	min	-1.575	-1.490	-1.417	-1.339	0.132	0.088	0.066	0.053	0.035	0.076	0.125	0.173
	1st Qu.	-0.404	-0.380	-0.363	-0.344	0.240	0.214	0.201	0.192	0.061	0.107	0.158	0.211
SRSWoR	median	-0.032	-0.029	-0.030	-0.029	0.445	0.422	0.408	0.395	0.230	0.183	0.213	0.280
Estimator	mean	0.001	0.002	0.000	0.000	0.515	0.487	0.468	0.452	0.372	0.338	0.352	0.390
	3rd Qu.	0.431	0.409	0.381	0.361	0.735	0.708	0.686	0.665	0.718	0.566	0.532	0.530
	max	1.561	1.468	1.387	1.292	1.621	1.573	1.534	1.491	0.949	0.950	0.945	0.947

coverages are far from the nominal level.

The results for inference on the population total are presented in Table 5.23. The expansion estimator is pretty much unbiased; however, the biases of the IDW point estimator and kriging predictor are sufficiently low.

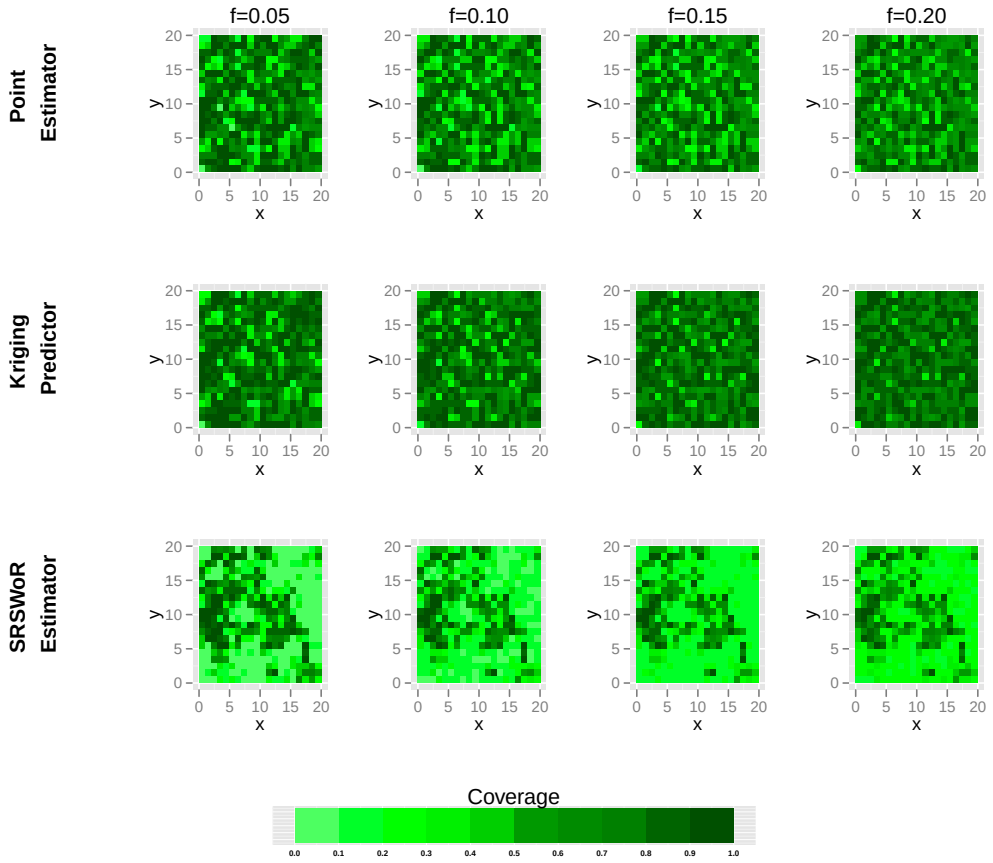

Figure 5.6: Population F: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

The RMSE of the expansion estimator is higher compared to the ones of the other techniques which have quite the same values.

The IDW point estimator fails in reaching the nominal level of 95% at the lowest sampling fraction and nearly reaches it at the higher fractions. The kriging predictor and the expansion estimator have coverages at least equal to the nominal level.

Table 5.23: Population F: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	6.194	5.131	3.986	3.574	44.186	28.098	21.021	17.201	0.944	0.949	0.949	0.949
Kriging Predictor	3.435	1.661	0.907	0.942	40.825	23.886	17.193	13.426	0.953	0.955	0.953	0.953
Expansion Estimator	0.319	0.906	0.150	-0.003	54.578	37.450	29.164	24.471	0.950	0.951	0.952	0.953

Table 5.24 collects the main indicators of the overall bias distributions of the estimators of the variance of the IDW point estimator for individual quantities. Once again, estimator $\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$ has the more conservative behaviour among those proposed in Subsection 4.2.6 even though it mostly assumes negative values as well as the other variance estimators.

Table 5.24: Population F: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-0.146	-0.163	-0.181	-0.207
	1st Qu.	-0.051	-0.036	-0.032	-0.029
	median	-0.034	-0.021	-0.016	-0.014
	mean	-0.040	-0.029	-0.025	-0.024
	3rd Qu.	-0.024	-0.013	-0.009	-0.007
	max	-0.011	-0.005	-0.002	-0.001
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-0.072	-0.116	-0.160	-0.197
	1st Qu.	-0.003	-0.011	-0.017	-0.023
	median	0.002	-0.002	-0.005	-0.007
	mean	-0.001	-0.008	-0.014	-0.018
	3rd Qu.	0.005	0.001	0.000	-0.001
	max	0.019	0.010	0.006	0.004
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.069	-0.116	-0.160	-0.197
	1st Qu.	-0.003	-0.011	-0.017	-0.023
	median	0.002	-0.002	-0.005	-0.007
	mean	0.000	-0.008	-0.014	-0.018
	3rd Qu.	0.006	0.001	0.000	-0.001
	max	0.020	0.011	0.006	0.004

For the estimator of the variance of the IDW point estimator of the population total, the conservative jackknife estimator involving the correction for not negligible sampling fractions is the most conservative among those studied in the Monte Carlo simulation (Table 5.25), nevertheless assuming negative values at low sampling fractions.

Table 5.25: Population F: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	-57.786	-12.770	3.516	0.683
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	-54.149	-12.199	3.709	0.773

Population G

Population G (Appendix E.7) is generated using a range parameter barely exceeding the mean distance observable in the domain and a rather large sill parameter. The results of the simulation study for inference on individual quantities are summarized in Table 5.26. Both the IDW point estimator and the SRSWoR estimator seem to be slightly biased in terms of the median of their overall distribution, whereas the corresponding mean are close enough to zero. The kriging predictor is unbiased both in terms of mean and median of the overall bias distribution. Regardless of the sampling fraction, the overall bias distributions of the IDW point estimator and of the kriging predictor are quite similar, while the SRSWoR has a poorer behaviour. At the lowest sampling fraction the IDW point estimator has an overall bias distribution slightly more concentrated around its central tendency than the one of the kriging predictor; as the sampling fraction increases the latter has a more desirable behaviour.

At a sampling fraction of $f = 0.05$, the IDW point estimator has lower RMSE since its overall distribution generally shows lower values. For higher sampling fractions, the kriging predictor improves and generally the overall RMSE distribution presents lower values, while medians and means remain quite similar. The SRSWoR estimator in predictive form has minimum values of the overall RMSE distribution close, or even equal, to those of the other techniques; however, its medians, means and maximums are substantially higher.

In terms of coverage of the 95% confidence interval, the kriging predictor seems the one to be preferred regardless of the sampling fraction: its overall distribution has higher median and mean. For $f = 0.05$, the main indicators

Table 5.26: Population G: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-4.352	-3.855	-3.535	-3.339	0.590	0.392	0.296	0.239	0.043	0.091	0.146	0.201
	1st Qu.	-1.073	-0.977	-0.903	-0.835	1.039	0.874	0.776	0.703	0.484	0.455	0.476	0.510
Point	median	0.140	0.127	0.099	0.088	1.440	1.257	1.155	1.087	0.724	0.673	0.656	0.662
Estimator	mean	0.050	0.043	0.036	0.033	1.660	1.458	1.340	1.252	0.663	0.637	0.640	0.653
	3rd Qu.	1.189	1.100	1.006	0.944	2.077	1.854	1.734	1.657	0.881	0.858	0.836	0.814
	max	4.468	4.041	3.781	3.534	4.698	4.323	4.140	3.970	0.962	0.953	0.954	0.950
	min	-4.663	-3.803	-3.371	-3.027	0.489	0.325	0.247	0.202	0.038	0.079	0.123	0.172
	1st Qu.	-1.127	-0.966	-0.886	-0.826	1.079	0.956	0.834	0.753	0.514	0.576	0.622	0.665
Kriging	median	0.053	0.062	0.028	-0.012	1.485	1.306	1.167	1.081	0.749	0.759	0.776	0.792
Predictor	mean	0.004	0.010	0.008	0.005	1.709	1.485	1.344	1.233	0.674	0.707	0.738	0.762
	3rd Qu.	1.171	1.016	0.859	0.772	2.155	1.869	1.681	1.557	0.892	0.891	0.892	0.889
	max	4.553	3.787	3.507	3.164	4.949	4.081	3.832	3.580	0.967	0.959	0.957	0.958
	min	-6.347	-6.027	-5.683	-5.336	0.489	0.325	0.247	0.202	0.038	0.079	0.123	0.172
	1st Qu.	-1.439	-1.351	-1.279	-1.191	0.871	0.770	0.721	0.689	0.064	0.106	0.156	0.212
SRSWoR	median	0.115	0.118	0.105	0.103	1.523	1.435	1.382	1.336	0.298	0.207	0.246	0.305
Estimator	mean	-0.009	-0.009	-0.011	-0.007	1.897	1.791	1.720	1.659	0.389	0.352	0.367	0.403
	3rd Qu.	1.397	1.316	1.241	1.170	2.634	2.543	2.461	2.387	0.733	0.590	0.544	0.557
	max	5.831	5.534	5.212	4.944	6.519	6.341	6.155	5.962	0.945	0.943	0.938	0.939

of the overall coverage distribution of the IDW point estimator are not much different from those of the kriging predictor; however, this difference gets bigger for increasing sampling fractions. The SRSWoR estimator has a poor coverage.

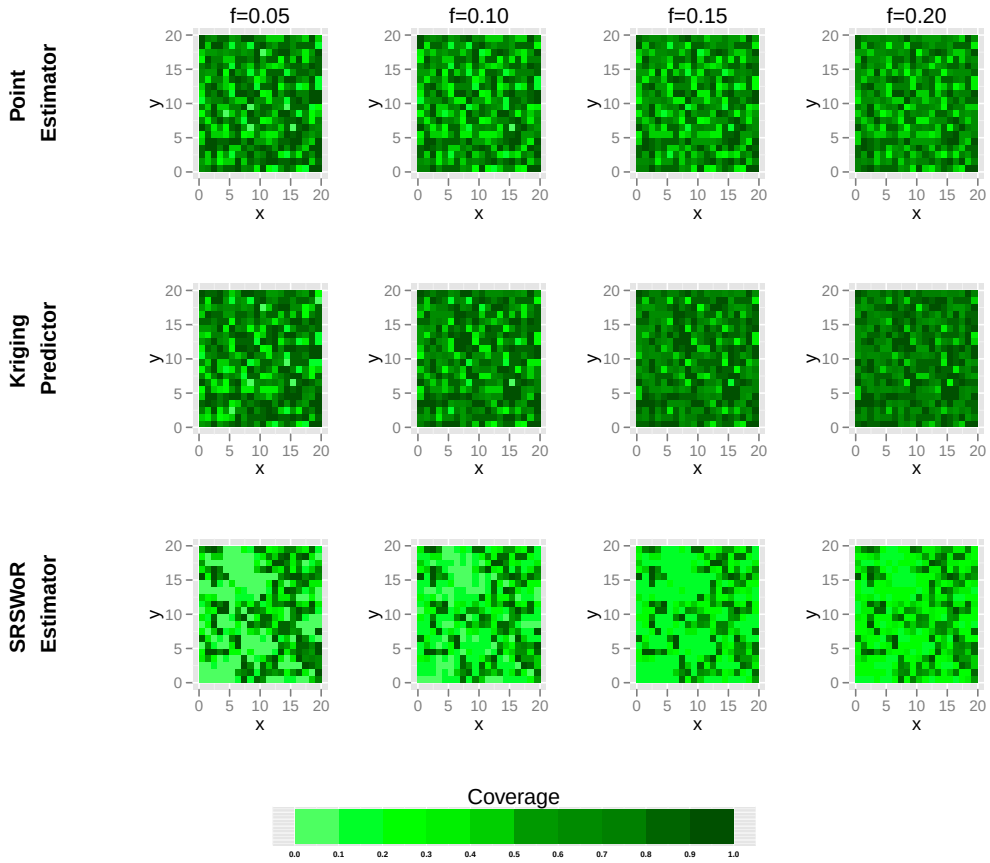

Figure 5.7: Population G: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

Figure 5.7 maps the coverage of the three methods in the domain at different sampling fractions. As pointed out for the overall distribution, the coverage of the IDW point estimator is generally lower than the one of the kriging predictor. More specifically, at a sampling fraction of 5%, the difference is quite small; as the sampling fraction increases the kriging predictor's coverage improves more quickly.

Table 5.27: Population G: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	20.016	17.358	14.349	13.044	174.764	113.402	86.760	71.669	0.947	0.941	0.947	0.942
Kriging Predictor	1.536	4.074	3.026	1.992	171.698	104.048	75.667	60.175	0.946	0.950	0.955	0.951
Expansion Estimator	-3.647	-3.514	-4.319	-2.737	202.622	137.918	108.477	91.375	0.950	0.944	0.945	0.947

The results of the Monte Carlo experiment for inference on the population total are collected in Table 5.27. The kriging predictor and the expansion estimator are practically unbiased and both outperform the IDW point estimator in terms of bias. The bias of the first two techniques has not a monotone behaviour for increasing sampling fractions.

Table 5.28: Population G: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-1.918	-2.213	-2.656	-3.076
	1st Qu.	-0.881	-0.680	-0.616	-0.608
	median	-0.633	-0.441	-0.366	-0.305
	mean	-0.697	-0.541	-0.497	-0.487
	3rd Qu.	-0.431	-0.288	-0.229	-0.172
	max	-0.220	-0.098	-0.056	-0.031
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-1.007	-1.921	-2.532	-3.051
	1st Qu.	-0.069	-0.247	-0.363	-0.477
	median	0.021	-0.068	-0.141	-0.192
	mean	-0.024	-0.170	-0.279	-0.360
	3rd Qu.	0.068	0.011	-0.019	-0.038
	max	0.445	0.248	0.122	0.109
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.996	-1.919	-2.531	-3.051
	1st Qu.	-0.059	-0.243	-0.363	-0.476
	median	0.026	-0.067	-0.140	-0.192
	mean	-0.013	-0.168	-0.278	-0.359
	3rd Qu.	0.077	0.012	-0.019	-0.038
	max	0.486	0.256	0.124	0.110

Despite having rather high bias, the IDW point estimator has a RMSE similar to the the kriging predictor's one which is the lowest. The values of the expansion estimator are substantially higher than those of the other two techniques.

The coverage of the IDW point estimator and expansion estimator are

generally lower than the nominal level. The kriging predictor's ones exceed the 95% except for the lowest sampling fraction.

Table 5.28 collects the main indicators of the overall bias distribution of the estimators of the variance of the IDW point estimator. Estimator $\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$ is the more conservative one despite presenting negative values.

Table 5.29: Population G: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	818.788	592.312	446.865	264.123
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	869.959	600.503	449.534	265.527

Analogously, the corresponding estimator $\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$ of the variance of the IDW point estimator of the population total is to be preferred to $\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$ as shown in Table 5.29.

Population H

Table 5.30 summarizes the results of the Monte Carlo experiment on population H (Appendix E.8) which has been generated with an exponential semivariogram having the same sill parameter as population G and a larger range parameter. The overall bias distributions of the three methods have almost null means; the same goes for the medians except for the SRSWoR estimator in predictive form. The kriging predictor shows a nicer behaviour, especially at large sampling fractions since the overall bias distribution shrinks more quickly around the central tendency values. For $f = 0.05$, the overall bias distribution of the IDW point estimator is almost equal to the kriging predictor's one; as the sampling fraction increases the performance of the latter improves more quickly than the other techniques do.

In terms of RMSE, the kriging predictor has lower median and mean values of the the overall distribution. At the lowest sampling fraction the difference between its overall distribution and the IDW point estimator's one is rather null. Both of them outperform the SRSWoR estimator.

Table 5.30: Population H: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-3.612	-3.224	-2.994	-2.788	0.395	0.292	0.232	0.192	0.050	0.087	0.132	0.193
	1st Qu.	-0.743	-0.651	-0.591	-0.542	0.722	0.565	0.485	0.428	0.551	0.518	0.528	0.553
Point	median	0.009	-0.001	-0.001	0.003	1.012	0.840	0.728	0.687	0.778	0.734	0.708	0.696
Estimator	mean	0.030	0.016	0.014	0.015	1.200	1.012	0.912	0.846	0.711	0.682	0.674	0.682
	3rd Qu.	0.763	0.633	0.586	0.553	1.569	1.343	1.216	1.139	0.910	0.879	0.861	0.844
	max	2.669	2.427	2.266	2.135	3.782	3.462	3.286	3.142	0.969	0.960	0.956	0.946
	min	-3.595	-2.979	-2.606	-2.323	0.430	0.291	0.221	0.181	0.028	0.079	0.115	0.168
	1st Qu.	-0.644	-0.481	-0.399	-0.360	0.734	0.601	0.534	0.472	0.659	0.692	0.712	0.728
Kriging	median	-0.002	0.014	0.013	0.025	1.006	0.777	0.686	0.629	0.846	0.848	0.840	0.839
Predictor	mean	0.014	-0.002	0.000	0.000	1.154	0.933	0.821	0.743	0.764	0.772	0.785	0.797
	3rd Qu.	0.674	0.549	0.484	0.434	1.467	1.180	1.007	0.910	0.919	0.914	0.918	0.914
	max	2.761	2.303	2.004	1.787	3.764	3.220	2.885	2.644	0.963	0.962	0.956	0.958
	min	-5.342	-5.118	-4.801	-4.494	0.430	0.291	0.221	0.181	0.028	0.079	0.115	0.168
	1st Qu.	-1.381	-1.300	-1.240	-1.141	0.794	0.711	0.673	0.634	0.060	0.106	0.156	0.210
SRSWoR	median	-0.090	-0.085	-0.082	-0.071	1.503	1.426	1.374	1.325	0.218	0.180	0.214	0.287
Estimator	mean	-0.015	-0.013	-0.014	-0.007	1.704	1.613	1.550	1.495	0.370	0.343	0.362	0.399
	3rd Qu.	1.446	1.367	1.289	1.217	2.421	2.332	2.258	2.189	0.689	0.576	0.536	0.535
	max	4.010	3.778	3.570	3.339	5.509	5.382	5.209	5.037	0.952	0.945	0.944	0.930

The overall coverage distribution of the kriging predictor have rather higher medians and means than those of the other techniques. As usual, the difference increases along with the sampling fraction. The SRSWoR has the poorest coverage regardless of the sampling fraction.

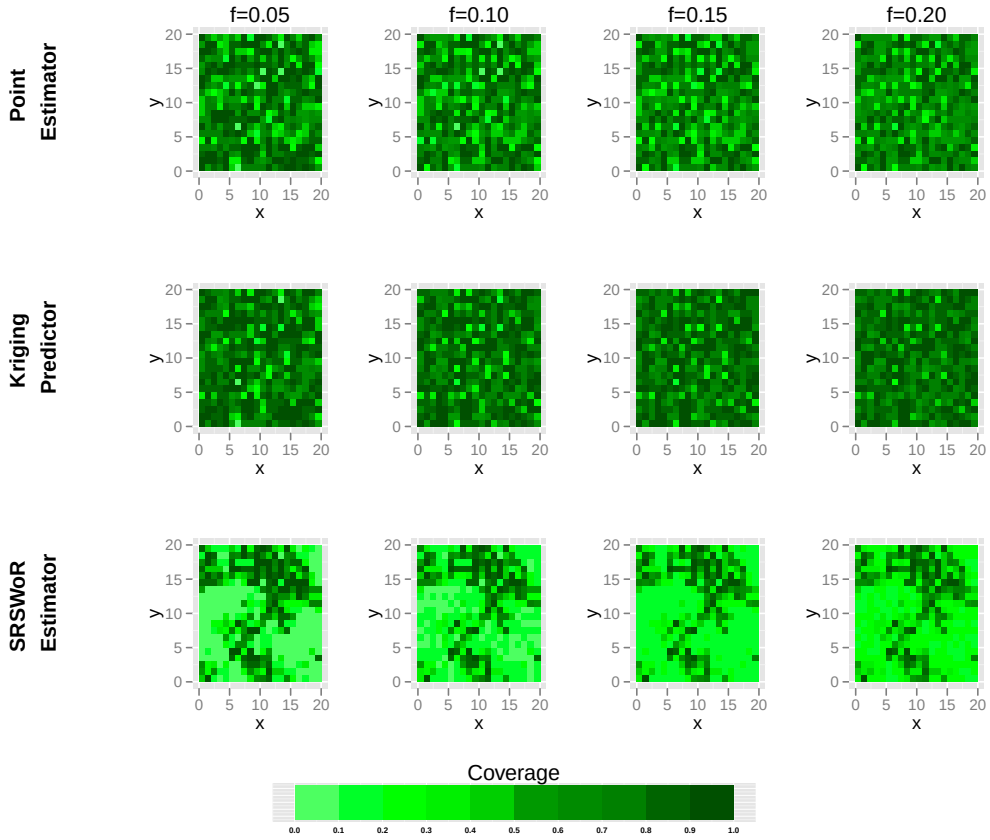

Figure 5.8: Population H: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

By looking at Figure 5.8, it is even clearer how the kriging predictor offers higher coverages. The results for the IDW point estimator are not far from those, however, presenting some problems in estimating values at some spatial locations. The results suggests not use the SRSWoR estimator in predictive form .

Table 5.31 summarizes the results of the Monte Carlo experiment for inference on the population total. The bias of the kriging predictor of the

Table 5.31: Population H: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	11.946	6.455	5.500	5.847	139.087	88.158	64.939	52.824	0.952	0.944	0.944	0.944
Kriging Predictor	5.667	-0.775	0.033	-0.065	118.968	69.948	49.999	39.447	0.958	0.950	0.934	0.950
Expansion Estimator	-6.010	-5.201	-5.601	-2.748	176.761	122.563	96.483	81.146	0.957	0.950	0.949	0.942

population total decreases quicker than the one of the other techniques and reaches the zero already at $f = 0.15$. At a sampling fraction of 5%, the expansion estimator's bias in absolute value has a similar value to the kriging's one; however, at higher sample sizes, it decreases less quickly. The IDW point estimator seems to not negligibly overestimate the population total regardless of the sampling fraction.

Table 5.32: Population H: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-1.382	-1.363	-1.630	-1.891
	1st Qu.	-0.479	-0.344	-0.310	-0.304
	median	-0.334	-0.205	-0.153	-0.130
	mean	-0.389	-0.270	-0.232	-0.224
	3rd Qu.	-0.234	-0.127	-0.088	-0.063
	max	-0.113	-0.056	-0.029	-0.012
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-0.629	-1.138	-1.526	-1.885
	1st Qu.	-0.022	-0.117	-0.185	-0.226
	median	0.022	-0.020	-0.046	-0.067
	mean	0.005	-0.074	-0.124	-0.165
	3rd Qu.	0.059	0.009	-0.001	-0.013
	max	0.208	0.142	0.094	0.053
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.617	-1.135	-1.525	-1.884
	1st Qu.	-0.016	-0.114	-0.185	-0.226
	median	0.026	-0.019	-0.046	-0.067
	mean	0.012	-0.073	-0.123	-0.165
	3rd Qu.	0.065	0.010	-0.001	-0.013
	max	0.225	0.146	0.095	0.053

The kriging predictor has the lowest RMSE, and the difference with the other techniques can not be neglected. Despite being not much biased, the expansion estimator has the poorest performances, while the RMSE of the IDW point estimator of the population total lies in between.

The coverage of the kriging predictor of the population total at least reaches the nominal level except for $f = 0.15$, whereas the other techniques exceeds it only at the lowest sampling fraction.

From Table 5.32 it can be seen that, despite assuming negative values, the conservative jackknife estimator of the variance of the IDW point estimator involving the modification for not negligible sampling fractions has the best performance in terms of bias.

Table 5.33: Population H: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	565.285	44.778	206.092	117.082
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	604.762	51.248	208.235	118.113

Analogously, according to Table 5.33, estimator $\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$ of the variance of the IDW point estimator of the population total is to be preferred.

Population I

Population I (Appendix E.9) has been generated from a superpopulation model having exponential semivariogram model with small nugget parameter and a range parameter slightly exceeding the domain mean distance. Table 5.34 collects the main indicators of the overall distributions of the bias, the RMSE and the coverage for the three methods at different sampling fractions. All the methods seems unbiased since the medians and means of the overall bias distributions are reasonably close to zero. The SRSWoR estimator in predictive form shows the worst behaviour among the three techniques analysed; the discrepancy with the other methods increases with the sampling fraction. At the lowest sampling fraction, the IDW point estimator seems to be slightly less biased than the kriging predictor; as the sampling fraction increases this relationship is reverted, and the latter shows better performances.

The overall RMSE distributions have a pattern similar to the overall bias distributions. At $f = 0.05$ the IDW point estimator has slightly lower

Table 5.34: Population I: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-4.900	-4.421	-4.136	-3.870	0.560	0.387	0.309	0.250	0.054	0.096	0.161	0.223
	1st Qu.	-0.902	-0.784	-0.709	-0.656	0.875	0.709	0.613	0.544	0.473	0.446	0.450	0.502
Point	median	0.062	0.049	0.032	0.037	1.264	1.073	0.955	0.880	0.758	0.708	0.690	0.685
Estimator	mean	0.018	0.017	0.018	0.022	1.432	1.244	1.136	1.061	0.676	0.651	0.649	0.665
	3rd Qu.	0.968	0.835	0.759	0.694	1.844	1.658	1.549	1.454	0.902	0.879	0.854	0.846
	max	4.419	4.086	3.776	3.517	5.130	4.744	4.513	4.316	0.968	0.973	0.972	0.974
	min	-5.046	-4.398	-3.909	-3.463	0.383	0.246	0.181	0.153	0.039	0.079	0.125	0.166
	1st Qu.	-1.014	-0.784	-0.711	-0.650	0.823	0.714	0.664	0.626	0.321	0.501	0.598	0.662
Kriging	median	0.036	0.044	0.066	0.085	1.244	1.045	0.932	0.833	0.662	0.743	0.774	0.798
Predictor	mean	-0.004	0.001	0.002	0.006	1.436	1.241	1.110	1.018	0.602	0.686	0.734	0.765
	3rd Qu.	0.999	0.776	0.636	0.570	1.886	1.608	1.444	1.313	0.879	0.897	0.900	0.892
	max	4.560	3.934	3.374	2.923	5.231	4.718	4.295	3.911	0.963	0.965	0.964	0.963
	min	-5.254	-4.979	-4.774	-4.537	0.383	0.246	0.181	0.153	0.039	0.079	0.125	0.166
	1st Qu.	-1.185	-1.114	-1.050	-0.977	0.733	0.654	0.615	0.585	0.067	0.106	0.156	0.210
SRSWoR	median	0.043	0.044	0.047	0.051	1.258	1.190	1.143	1.107	0.250	0.178	0.206	0.286
Estimator	mean	-0.015	-0.010	0.004	0.003	1.477	1.394	1.338	1.293	0.368	0.325	0.336	0.381
	3rd Qu.	1.141	1.086	1.030	0.977	2.003	1.934	1.870	1.815	0.684	0.517	0.468	0.507
	max	4.868	4.655	4.377	4.113	5.398	5.243	5.128	4.994	0.948	0.944	0.944	0.939

RMSE values, and for the other sampling fractions the performances of the kriging predictor become better. Moreover, despite assuming higher values, the overall RMSE distribution of the SRSWoR estimator in predictive form has median and mean values very close to the ones of the other methods at the lowest sampling fraction. For higher sampling fractions the SRSWoR estimator has worse performances.

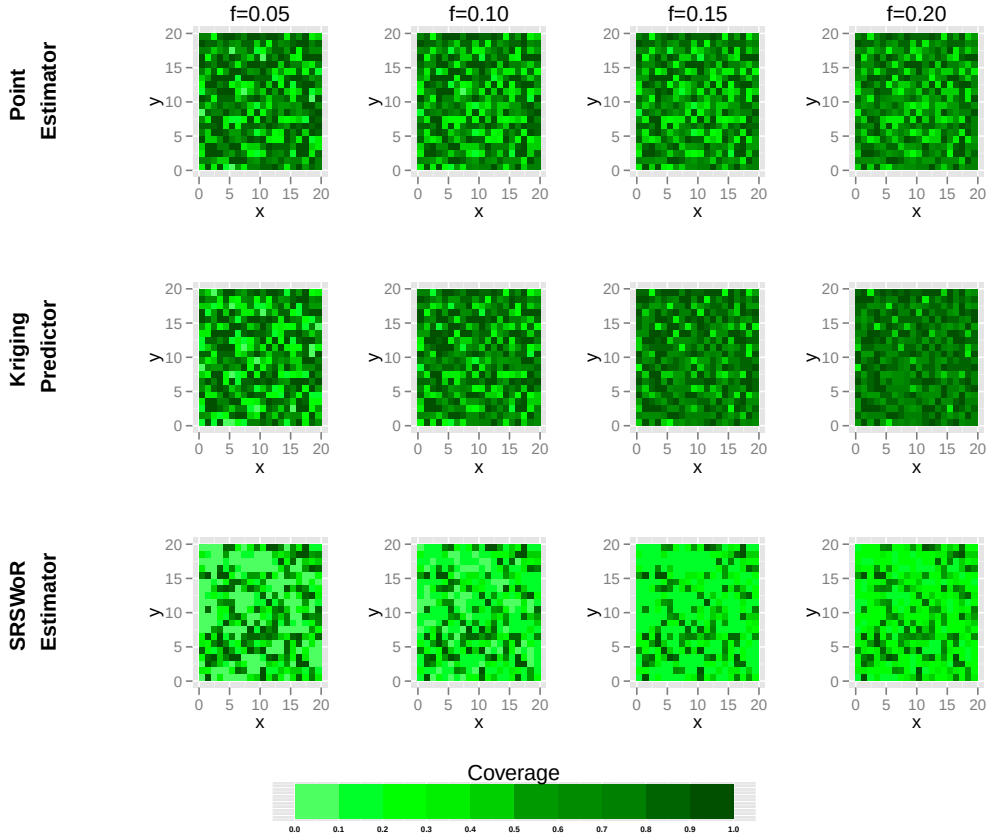

Figure 5.9: Population I: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

Quite surprisingly, the overall coverage distribution of the IDW point estimator presents decreasing medians and means for increasing sampling fractions. Hence, it has generally higher coverages than the kriging predictor at $f = 0.05$. As the sampling fractions increases, the coverages of the kriging predictor gets higher. The SRSWoR has very low coverages.

Table 5.35: Population I: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	7.255	6.702	7.025	8.828	152.902	98.244	72.490	61.155	0.949	0.954	0.957	0.948
Kriging Predictor	-1.648	0.497	0.818	2.252	152.539	94.514	67.055	53.533	0.955	0.954	0.952	0.947
Expansion Estimator	-5.930	-4.152	-1.776	1.025	158.117	104.913	79.359	68.415	0.949	0.947	0.953	0.951

Figure 5.9 maps the coverages over the domain. The poor behaviour of the SRSWoR estimator in predictive form can be easily spotted. From the maps corresponding to the IDW point estimator, the loss of effectiveness for increasing sample sizes can be easily seen; on the contrary, the kriging predictor's improves its coverages. The higher coverages of the IDW point estimator for the lowest sample size are clear.

Table 5.36: Population I: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-1.916	-2.580	-2.900	-3.279
	1st Qu.	-0.652	-0.498	-0.471	-0.485
	median	-0.478	-0.319	-0.253	-0.219
	mean	-0.547	-0.408	-0.368	-0.362
	3rd Qu.	-0.379	-0.226	-0.154	-0.115
	max	-0.205	-0.091	-0.054	-0.037
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-1.189	-1.976	-2.527	-3.130
	1st Qu.	-0.087	-0.193	-0.287	-0.378
	median	0.003	-0.036	-0.077	-0.111
	mean	-0.040	-0.127	-0.201	-0.267
	3rd Qu.	0.047	0.018	0.000	-0.017
	max	0.170	0.127	0.113	0.044
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-1.165	-1.968	-2.525	-3.129
	1st Qu.	-0.078	-0.191	-0.287	-0.378
	median	0.008	-0.035	-0.077	-0.111
	mean	-0.032	-0.126	-0.201	-0.266
	3rd Qu.	0.057	0.019	0.000	-0.016
	max	0.189	0.131	0.116	0.045

Let us now consider the results regarding inference on the population total collected in Table 5.35. The bias of the three techniques is reasonably low; however, while for the kriging predictor and the expansion estimator it has a decreasing trend, for the IDW point estimator it remains almost

constant.

At a sampling fraction of 5%, the RMSE values are pretty similar among them; as the sample size increases, the difference becomes relevant. The kriging predictor has the lowest RMSE, while the expansion estimator shows the worst performance despite having a nice bias behaviour.

The coverages of the three techniques are close to the nominal level of 95% and sometimes they even exceed it.

Table 5.36 shows that the jackknife estimator of the variance of the IDW point estimator involving the modification for not negligible sampling fractions is the most conservative among the proposed one despite assuming negative values.

Table 5.37: Population I: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	-903.368	99.933	554.123	200.454
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	-876.650	104.196	555.544	201.168

Estimator $\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$ of the variance of the IDW point estimator of the population total is the more conservative, as pointed out by the literature (Wolter, 2007).

Population J

Population J (Apendix E.10) is the realization of a random field having exponential semivariogram model with a small nugget parameter and a rather large range parameter. The results of the Monte Carlo experiment for inference on individual values are collected in Table 5.38. The three techniques are practically unbiased in terms of the median and mean of the overall bias distributions. At the lowest sampling fraction the overall bias distributions of the IDW point estimator and of the kriging predictor can be considered equal. As the sample size increase the kriging predictor has better performances. The values of the main indicators of the overall bias distribution

Table 5.38: Population J: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-2.364	-2.118	-1.948	-1.805	0.351	0.235	0.188	0.154	0.074	0.088	0.138	0.189
	1st Qu.	-0.568	-0.494	-0.445	-0.414	0.612	0.475	0.396	0.344	0.532	0.490	0.495	0.516
Point	median	-0.042	-0.037	-0.018	-0.015	0.826	0.688	0.608	0.549	0.797	0.754	0.722	0.716
Estimator	mean	-0.002	-0.002	0.000	0.001	0.958	0.815	0.736	0.681	0.699	0.671	0.666	0.676
	3rd Qu.	0.629	0.538	0.510	0.468	1.189	1.049	0.975	0.939	0.908	0.888	0.867	0.851
	max	2.495	2.299	2.124	1.983	2.646	2.470	2.352	2.258	0.961	0.964	0.956	0.951
	min	-2.388	-2.063	-1.841	-1.714	0.332	0.224	0.170	0.136	0.028	0.077	0.123	0.169
	1st Qu.	-0.567	-0.448	-0.407	-0.369	0.620	0.483	0.415	0.367	0.634	0.640	0.664	0.669
Kriging	median	-0.037	0.010	-0.009	-0.022	0.801	0.633	0.582	0.528	0.836	0.826	0.819	0.811
Predictor	mean	-0.008	-0.005	-0.003	-0.001	0.927	0.753	0.676	0.618	0.748	0.752	0.758	0.764
	3rd Qu.	0.535	0.419	0.382	0.349	1.130	0.950	0.864	0.809	0.918	0.917	0.910	0.901
	max	2.456	2.169	1.946	1.780	2.589	2.329	2.160	2.034	0.967	0.964	0.951	0.953
	min	-3.817	-3.651	-3.473	-3.249	0.332	0.224	0.170	0.136	0.028	0.077	0.123	0.169
	1st Qu.	-1.003	-0.950	-0.899	-0.856	0.608	0.541	0.510	0.486	0.060	0.105	0.157	0.214
SRSWoR	median	0.084	0.076	0.075	0.073	1.127	1.074	1.034	1.004	0.226	0.179	0.218	0.276
Estimator	mean	0.001	0.000	0.003	0.004	1.311	1.241	1.192	1.149	0.371	0.342	0.359	0.392
	3rd Qu.	1.088	1.019	0.962	0.912	1.831	1.771	1.717	1.654	0.714	0.558	0.509	0.516
	max	3.355	3.196	3.075	2.918	3.920	3.826	3.728	3.604	0.948	0.952	0.937	0.933

of the SRSWoR estimator in predictive form suggest that it is more biased than the other methods.

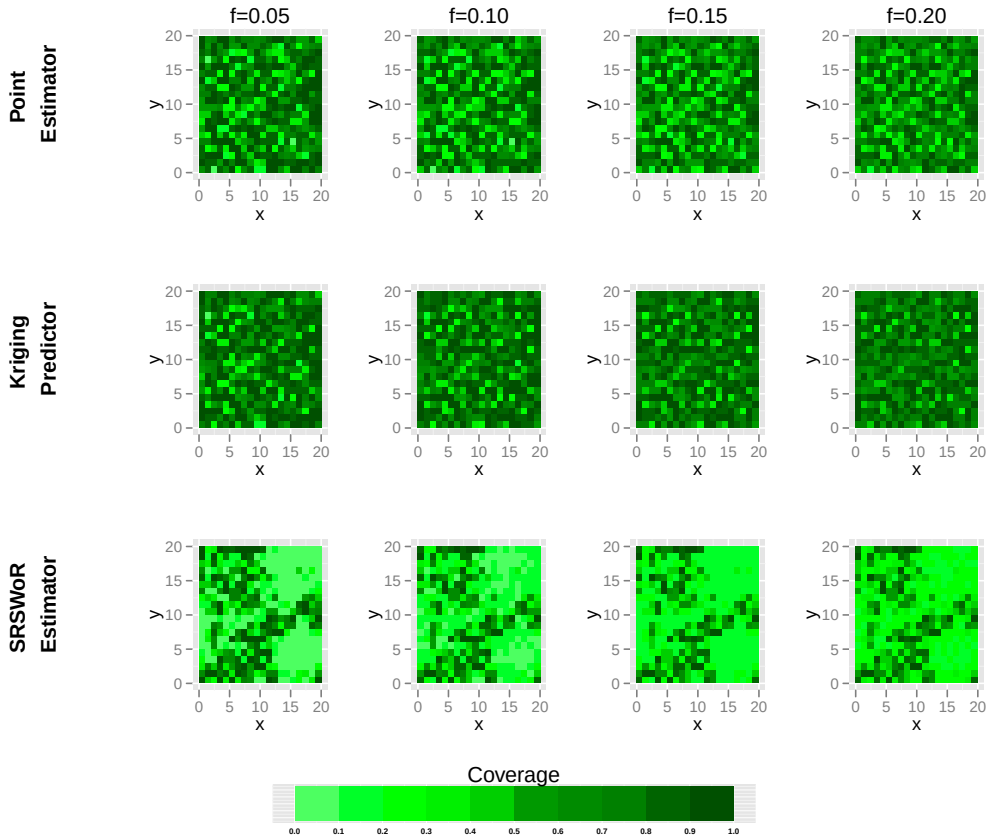

Figure 5.10: Population J: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

In terms of RMSE, the kriging predictor's overall distribution assumes slightly lower values than the IDW point estimator's one. The SRSWoR estimator's RMSE has higher overall values than the other techniques.

The SRSWoR estimator has substantially lower coverages than the ones of the other methods. According to the main position indicators of the overall bias distributions, the kriging predictor offers the best coverage of the 95% confidence interval. The medians and means of both the IDW point estimator and the kriging predictor have a decreasing trend for increasing sampling fractions.

Table 5.39: Population J: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	-0.682	-0.932	0.140	0.457	109.497	69.827	51.906	40.555	0.947	0.955	0.950	0.943
Kriging Predictor	-3.382	-1.936	-1.136	-0.415	96.000	55.294	41.057	31.415	0.946	0.951	0.956	0.946
Expansion Estimator	0.440	-0.159	1.151	1.413	136.845	94.439	74.487	61.304	0.952	0.953	0.945	0.941

Figure 5.10 collects the maps of the coverage at the different sampling fractions. From the maps regarding the performances of the kriging, it can be seen that it has slightly higher coverages than the other techniques. The maps on the bottom of the figure highlights how poor is the behaviour of the SRSWoR estimator in predictive form in terms of coverages.

Table 5.40: Population J: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-0.711	-0.713	-0.932	-1.084
	1st Qu.	-0.299	-0.206	-0.183	-0.183
	median	-0.218	-0.141	-0.101	-0.083
	mean	-0.245	-0.172	-0.149	-0.144
	3rd Qu.	-0.157	-0.096	-0.063	-0.043
	max	-0.097	-0.042	-0.018	-0.009
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-0.309	-0.562	-0.878	-1.073
	1st Qu.	-0.018	-0.074	-0.114	-0.149
	median	0.014	-0.014	-0.030	-0.040
	mean	0.006	-0.051	-0.083	-0.107
	3rd Qu.	0.040	0.006	-0.001	-0.005
	max	0.142	0.080	0.039	0.033
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.305	-0.561	-0.878	-1.073
	1st Qu.	-0.015	-0.073	-0.114	-0.148
	median	0.018	-0.014	-0.030	-0.040
	mean	0.011	-0.050	-0.083	-0.107
	3rd Qu.	0.043	0.007	-0.001	-0.005
	max	0.157	0.085	0.040	0.034

Table 5.40 collects the results of the Monte Carlo experiments for inference on the population total. Despite the biases of the different techniques are low and reasonably close to zero, the RMSE values are different. The kriging predictor has the lowest values; the IDW point estimator's ones are not far from those, and the expansion estimator has the poorest behaviour.

The coverages of the three methods are close to, or even exceeds, the nominal level, except for highest sampling fraction at which they fail to reach the 95%.

The results regarding the estimators of the variance of the IDW point estimator proposed in Subsection 4.2.6 are consistent to those of the previous subsections, as estimator $\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$ is the most conservative.

Table 5.41: Population J: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	511.993	-4.676	72.675	175.939
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	537.742	-1.083	73.809	176.458

Analogously, the jackknife estimator of the variance of the IDW point estimator of the population total involving the modification for not negligible sampling fractions is the one to be preferred.

Population K

Population K (Appendix E.11) is the realization of a superpopulation model having a range parameter barely exceeding the mean distance observable in the domain and a rather large nugget parameter. Table 5.42 collects the main position indicators of the overall bias, RMSE and coverage distributions of the three methods. Quite surprisingly and inconsistently to the results on the other populations, the IDW point estimator has an overall bias distribution very similar to, and sometimes even more appealing than, the kriging predictor's one, regardless of the sampling fraction. The SRSWoR estimator seems more biased than the other methods despite having almost identical median and mean of the overall bias distribution.

The overall RMSE distribution of the IDW point estimator shows that a reduction in the bias leads to slightly lower values; the medians and means are very similar to the kriging predictor's ones especially at the lower sampling fractions. As usual, the SRSWoR estimator in predictive form has the worst

Table 5.42: Population K: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-4.385	-3.963	-3.696	-3.416	0.573	0.449	0.346	0.282	0.045	0.079	0.125	0.189
	1st Qu.	-0.989	-0.897	-0.835	-0.773	0.892	0.754	0.660	0.603	0.467	0.434	0.440	0.493
Point	median	0.002	-0.016	-0.008	-0.008	1.267	1.125	1.033	0.981	0.704	0.659	0.649	0.661
Estimator	mean	-0.009	-0.020	-0.013	-0.009	1.435	1.290	1.200	1.130	0.651	0.626	0.626	0.643
	3rd Qu.	1.047	0.943	0.891	0.830	1.754	1.626	1.557	1.478	0.890	0.861	0.836	0.816
	max	3.430	3.220	3.034	2.846	4.580	4.267	4.081	3.879	0.959	0.976	0.976	0.974
	min	-4.515	-4.135	-3.861	-3.554	0.375	0.248	0.191	0.162	0.036	0.077	0.115	0.166
	1st Qu.	-1.026	-0.914	-0.861	-0.806	0.890	0.791	0.702	0.643	0.357	0.429	0.479	0.546
Kriging	median	0.030	-0.005	-0.006	-0.014	1.292	1.139	1.071	1.018	0.634	0.654	0.687	0.714
Predictor	mean	-0.003	-0.015	-0.007	-0.002	1.460	1.318	1.225	1.152	0.595	0.624	0.655	0.688
	3rd Qu.	1.174	0.992	0.897	0.814	1.850	1.654	1.551	1.483	0.863	0.860	0.859	0.855
	max	3.464	3.254	3.073	2.883	4.674	4.425	4.251	4.036	0.968	0.957	0.959	0.960
	min	-4.576	-4.297	-4.063	-3.819	0.375	0.248	0.191	0.162	0.036	0.077	0.115	0.166
	1st Qu.	-1.302	-1.235	-1.160	-1.091	0.736	0.673	0.632	0.605	0.058	0.105	0.157	0.213
SRSWoR	median	0.025	0.019	0.024	0.023	1.373	1.309	1.267	1.218	0.195	0.152	0.199	0.275
Estimator	mean	-0.004	-0.013	-0.007	-0.002	1.498	1.421	1.368	1.321	0.344	0.315	0.336	0.379
	3rd Qu.	1.252	1.191	1.144	1.074	2.058	1.995	1.920	1.858	0.641	0.496	0.487	0.511
	max	4.010	3.784	3.563	3.350	4.709	4.559	4.429	4.292	0.965	0.949	0.943	0.930

performances; however, the values of the overall RMSE distribution are not as far from the ones of the other techniques as for other populations.

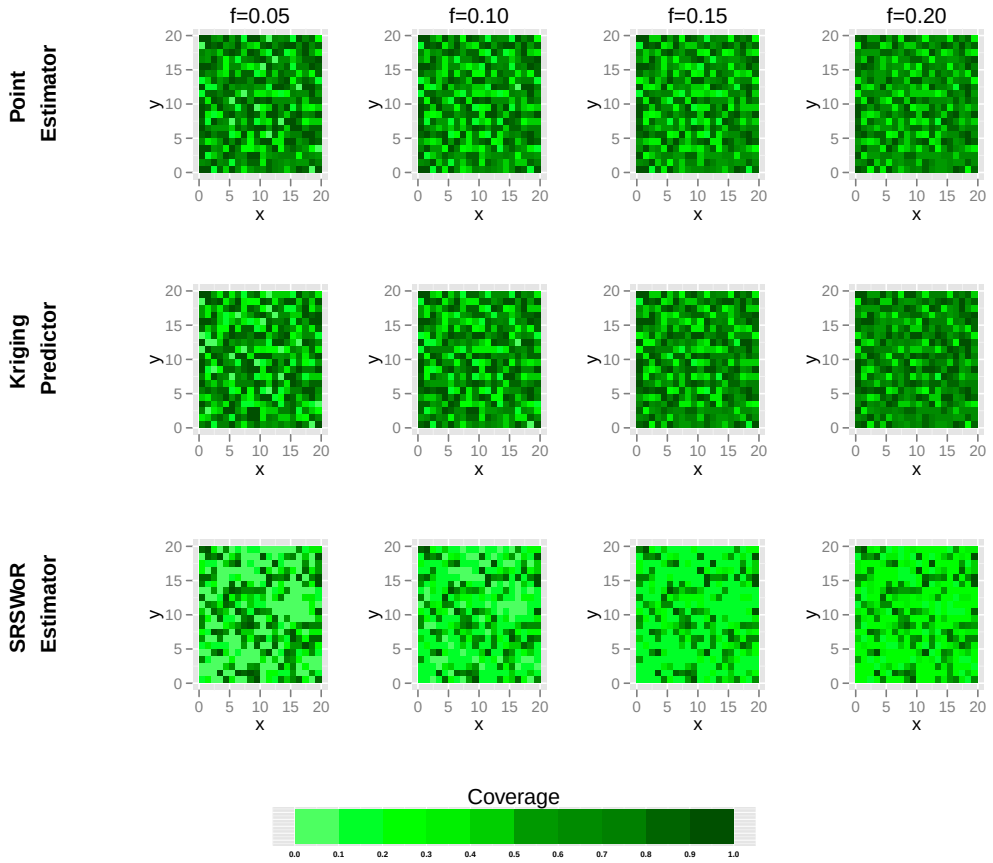

Figure 5.11: Population K: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

The coverages of the SRSWoR estimator are lower than those of the IDW point estimator and of the kriging predictor. The last one's overall coverage distribution has lower median and mean than the ones of estimator (4.2.5) at the lower sampling fraction, while the relationships is reverted at higher sample sizes.

From the comparison of the first and second row of Figure 5.11, it appears that IDW point estimator has slightly higher coverages than the kriging predictor at the lowest sampling fraction. At $f = 0.10$ and $f = 0.15$ it

Table 5.43: Population K: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	-3.726	-8.105	-5.015	-3.422	142.653	95.728	74.925	62.620	0.956	0.956	0.944	0.944
Kriging Predictor	-1.051	-5.978	-2.698	-0.792	145.868	95.516	74.132	61.535	0.961	0.951	0.947	0.944
Expansion Estimator	-1.590	-5.160	-2.601	-0.993	153.122	105.473	83.830	71.869	0.963	0.951	0.950	0.943

is hard to spot a substantial difference. Finally, at the highest sampling fraction the kriging predictor has higher coverages. As usual, the coverages of the SRSWoR estimator in predictive form are the lowest.

Table 5.44: Population K: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-1.472	-2.213	-2.706	-3.072
	1st Qu.	-0.595	-0.506	-0.491	-0.492
	median	-0.459	-0.340	-0.280	-0.255
	mean	-0.499	-0.414	-0.394	-0.398
	3rd Qu.	-0.366	-0.240	-0.178	-0.136
	max	-0.252	-0.123	-0.080	-0.047
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-0.986	-1.862	-2.523	-3.027
	1st Qu.	-0.072	-0.203	-0.308	-0.394
	median	0.003	-0.066	-0.115	-0.149
	mean	-0.040	-0.147	-0.233	-0.300
	3rd Qu.	0.045	0.000	-0.018	-0.032
	max	0.131	0.094	0.086	0.077
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.972	-1.858	-2.522	-3.026
	1st Qu.	-0.066	-0.199	-0.308	-0.394
	median	0.008	-0.065	-0.114	-0.149
	mean	-0.033	-0.146	-0.233	-0.300
	3rd Qu.	0.052	0.001	-0.017	-0.032
	max	0.140	0.097	0.088	0.078

Table 5.43 collects the results of the Monte Carlo experiments for inference on the population total. The bias of the three methods are close enough to zero; however, none of these has a monotonous decreasing trend. The IDW point estimator of the population total seems slightly more biased than the other techniques.

As for the case of inference on individual values, the IDW point estimator gain from being less biased: its RMSE is practically the same as the kriging

predictor one. The expansion estimator has a RMSE closer to the values of the other methods than it has been for other populations.

The coverages of the three techniques for inference on the population total exceed the nominal level at the lower sampling fractions, and then they even fail in reaching it.

Estimator $\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$ of the variance of the IDW point estimator is resulted the more conservative one among those presented in Subsection 4.2.6 (Table 5.44).

Table 5.45: Population K: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	253.211	25.760	-105.710	-132.359
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	277.899	29.507	-104.572	-131.847

Analogously, Table 5.45 shows that the corresponding estimator of the variance of the IDW point estimator of the population total has a conservative behaviour when compared to the other jackknife estimator.

Population L

Population L (Appendix E.12) is the result of a simulated random field having a rather large nugget parameter and a rather large range parameter. The results of the Monte Carlo simulations for the bias, RMSE and coverage are collected in Table 5.46. In terms of the median and mean of the overall bias distributions, the kriging predictor is slightly better than the IDW point estimator; both have more advisable performances than the SRWoR estimator. At the lowest sampling fraction the overall bias distribution of the IDW point estimator is very similar to the kriging predictor's one.

The median and mean of the overall RMSE distribution of the IDW point estimator are almost the same as the kriging predictor's ones; however, the former has lower maximum values except for the smallest sampling fraction. In terms of median and mean of the overall RMSE distribution, the SRSWoR estimator in predictive form has a behaviour similar to the one of the

Table 5.46: Population L: overall bias, RMSE and coverage distributions for the three methods

		bias				*RMSE*				*coverage*			
	f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
	min	-3.490	-3.147	-2.872	-2.673	0.485	0.366	0.298	0.252	0.035	0.087	0.139	0.185
	1st Qu.	-0.997	-0.887	-0.823	-0.781	0.757	0.635	0.566	0.531	0.409	0.402	0.418	0.472
Point	median	0.014	0.012	0.014	0.018	1.142	1.017	0.936	0.889	0.724	0.662	0.634	0.645
Estimator	mean	-0.083	-0.070	-0.051	-0.047	1.278	1.147	1.068	1.007	0.646	0.623	0.624	0.641
	3rd Qu.	0.850	0.766	0.712	0.676	1.634	1.505	1.427	1.346	0.881	0.854	0.831	0.810
	max	4.005	3.695	3.443	3.206	4.132	3.903	3.740	3.591	0.970	0.968	0.958	0.945
	min	-3.513	-2.924	-2.581	-2.374	0.350	0.237	0.182	0.150	0.035	0.076	0.125	0.169
	1st Qu.	-0.941	-0.817	-0.777	-0.740	0.798	0.666	0.614	0.566	0.437	0.464	0.486	0.542
Kriging	median	0.076	0.026	0.016	0.014	1.167	1.017	0.950	0.911	0.718	0.704	0.703	0.718
Predictor	mean	-0.025	-0.023	-0.011	-0.008	1.306	1.143	1.057	0.995	0.650	0.655	0.667	0.691
	3rd Qu.	0.902	0.779	0.766	0.729	1.671	1.453	1.362	1.274	0.891	0.876	0.868	0.855
	max	4.129	3.615	3.279	2.998	4.257	3.825	3.573	3.369	0.965	0.955	0.949	0.950
	min	-4.382	-4.142	-3.873	-3.647	0.350	0.237	0.182	0.150	0.035	0.076	0.125	0.169
	1st Qu.	-1.049	-1.002	-0.933	-0.888	0.621	0.554	0.522	0.497	0.057	0.107	0.157	0.211
SRSWoR	median	0.106	0.097	0.101	0.093	1.216	1.157	1.121	1.086	0.216	0.169	0.215	0.284
Estimator	mean	0.001	-0.003	0.007	0.003	1.411	1.336	1.285	1.238	0.378	0.355	0.376	0.411
	3rd Qu.	1.147	1.086	1.038	0.972	2.070	2.008	1.949	1.885	0.712	0.604	0.560	0.564
	max	4.700	4.447	4.187	3.921	4.797	4.660	4.525	4.375	0.952	0.943	0.938	0.944

other techniques; nevertheless, it is more disperse around its central tendency values.

The median and mean overall coverages of the kriging predictor are slightly higher than those of the IDW point estimator except for the lowest sampling fraction where the values are more or less the same. The SRSWoR has a poor behaviour in terms of coverage.

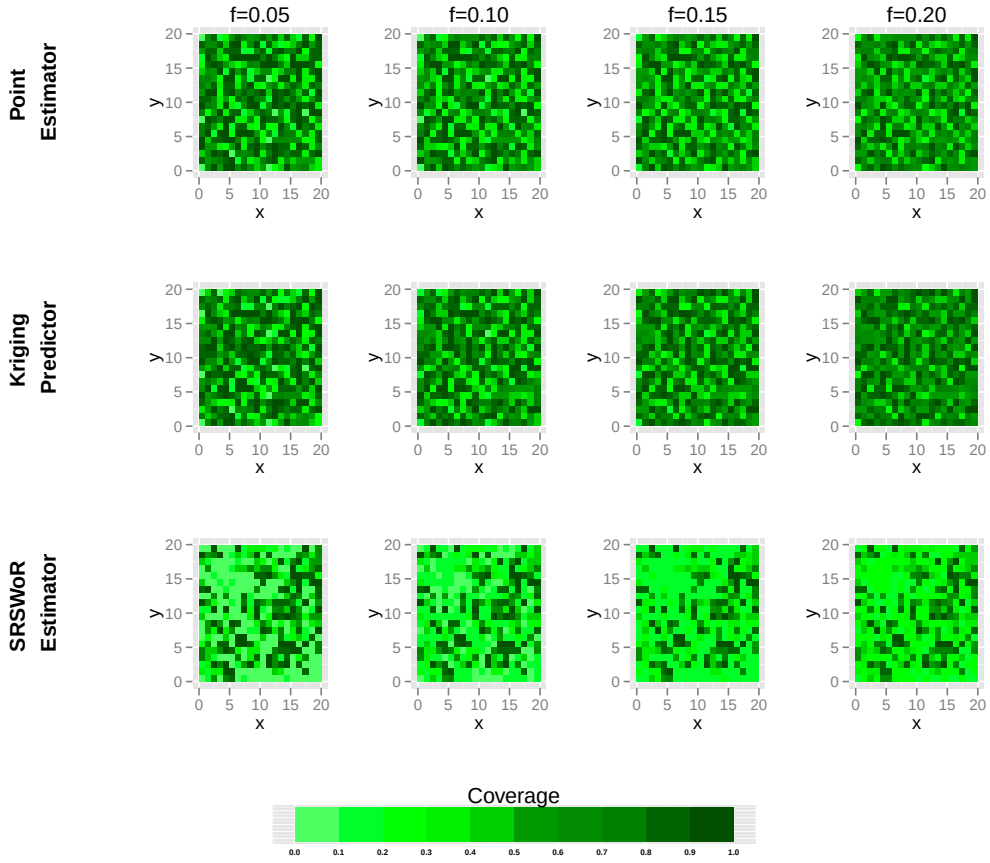

Figure 5.12: Population L: spatial distribution of the coverage of the 95% confidence interval centred in the estimated (predicted) value for the three methods

By comparing the top left coverage map and the one below it, it can be seen that the IDW point estimator and the kriging predictor give quite the same results in terms of coverage of the 95% confidence interval. The performances of the latter improve slightly more quickly as the sampling

Table 5.47: Population L: bias, RMSE and coverage of the estimators and predictor of the population total

	bias				*RMSE*				*coverage*			
f	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20	0.05	0.10	0.15	0.20
Point Estimator	-33.353	-28.043	-20.377	-18.656	128.959	88.039	68.913	57.206	0.940	0.939	0.937	0.930
Kriging Predictor	-10.113	-9.333	-4.389	-3.016	126.536	79.903	61.912	50.741	0.940	0.947	0.952	0.945
Expansion Estimator	0.413	-1.352	2.872	1.284	143.532	100.698	81.299	67.704	0.952	0.946	0.944	0.946

fraction increases. The bottom maps highlight the poor behaviour of the SRSWoR estimator in terms of coverage.

Table 5.48: Population L: distribution of the overall bias of the estimators of the variance of the IDW point estimator

	f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_n[\hat{z}(\mathbf{u}_i)]$	min	-0.915	-1.456	-2.012	-2.492
	1st Qu.	-0.479	-0.405	-0.388	-0.401
	median	-0.353	-0.273	-0.242	-0.218
	mean	-0.388	-0.323	-0.309	-0.314
	3rd Qu.	-0.256	-0.175	-0.127	-0.098
	max	-0.161	-0.076	-0.045	-0.031
$\hat{\mathrm{V}}_1[\hat{z}(\mathbf{u}_i)^*]$	min	-0.515	-1.274	-1.943	-2.492
	1st Qu.	-0.058	-0.166	-0.252	-0.328
	median	0.016	-0.054	-0.093	-0.129
	mean	-0.016	-0.112	-0.182	-0.239
	3rd Qu.	0.047	0.008	-0.011	-0.025
	max	0.130	0.078	0.046	0.039
$\hat{\mathrm{V}}_2[\hat{z}(\mathbf{u}_i)^*]$	min	-0.507	-1.273	-1.943	-2.492
	1st Qu.	-0.051	-0.166	-0.252	-0.328
	median	0.021	-0.053	-0.093	-0.129
	mean	-0.011	-0.111	-0.182	-0.239
	3rd Qu.	0.053	0.008	-0.011	-0.025
	max	0.138	0.080	0.048	0.040

From the analysis of Table 5.47 regarding the results of the Monte Carlo experiments for inference on the population total, it results that, despite being the more biased method, the IDW point estimator of the population total has a quite nice RMSE when compared to the other techniques. On the other hand, the expansion estimator has the highest RMSE even though it is the least biased.

The coverages of the 95% confidence interval of the three techniques almost ever fail in reaching the nominal level.

Again and consistently with the results on the previous populations, the conservative jackknife estimator of the variance of the IDW point estimator has the more appealing performances among those studied in the Monte Carlo simulations (see Table 5.48).

Table 5.49: Population L: bias of the jackknife estimators of the variance of the estimator of the population total based on the IDW point estimator

f	0.05	0.10	0.15	0.20
$\hat{\mathrm{V}}_1[\hat{\bar{t}}(z)^*]$	1560.817	437.918	94.067	94.808
$\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$	1588.035	443.565	96.340	96.119

Estimator $\hat{\mathrm{V}}_2[\hat{\bar{t}}(z)^*]$ of the variance of the IDW point estimator of the population total is the more conservative when compared to the other jackknife one.

5.2.2 The effect of the superpopulation parameters

In this subsection we focus the analysis on the effects that varying parameters of the superpopulation model have on inference. By comparing the results of the Monte Carlo experiments presented in the previous subsection, we would like to give an idea of when the IDW point estimator is more suitable than the kriging predictor or the SRSWoR estimator in predictive form. Firstly, we concentrate on the population generated by an exponential semivariogram model. According to Table 5.1, by comparing populations A, B, C and D we assess the effect of the range parameter, ϕ. Populations A, C, E, F, G and H give information on how different sill values, σ^2, affect inference. Finally, we compare populations A, B, I, J, K and L to assess the effects of the nugget parameter effect.

The range parameter

In order to understand how a different range parameter of the exponential semivariogram model superpopulation having exponential semivariogram affects the inference, we start by comparing Tables 5.6, 5.2, 5.10 and 5.14.

When we analysed populations A, B,C and D we already pointed out that, for inference on individual values, the SRSWoR in predictive form is outperformed by the IDW point estimator and the kriging predictor. Thus, the following analysis is focused on the comparison of only those last two.

The distribution of the overall bias of the IDW point estimator is similar to the kriging predictor's one when the range parameter does not exceed the maximum distance observable in the domain (in the study case 28.284), especially at low sampling fractions. On the contrary for $\phi = 45$ and $\phi = 90$ the kriging predictor seems to better capture the spatial nature of the phenomenon even at low sampling fractions. This aspect may depend on the choice of the inverse squared Euclidean distance in weighting system (4.2.1) which does not weighs the more distant observations properly; perhaps a lower exponents could lead to better results.

In terms of RMSE, the Monte Carlo experiments highlighted that the IDW point estimator has an overall distribution exhibiting lower values than the kriging predictor's one when the range parameter is smaller than the maximum distance observable in the domain. This peculiarity is more evident at small sampling fractions (e.g. $f = 0.05$ and $f = 0.10$). As the sampling dimension increases the overall RMSE distribution of the IDW point estimator still has lower minimum values, but its maximum exceeds the kriging predictor's distribution ones. The higher the sampling fraction, the more the RMSE distribution is affected by this phenomenon. Moreover, when the range parameter is greater than the maximum distance in the domain (e.g. $\phi = 45, 90$) the kriging predictor performs generally better, although the minimum of its RMSE distribution is greater than the one of the IDW point estimator's distribution. Again, the sensitivity of the RMSE to the range parameter of the semivariogram model adopted to generate the populations might suggest that the inverse squared Euclidean distance could not be the most suitable weighting system for highly spatial correlated populations. Perhaps, exponent values between $1 < \alpha < 2$ might produce better results.

Finally, by the comparison of Tables 5.6, 5.2, 5.10 and 5.14, not much can be said about the overall coverage distribution of the IDW point estimator and the kriging predictor. The latter seems to basically have a better

behaviour in terms of median and mean of the overall coverage distribution. However, attention must be paid to the smallest sampling fraction, $f = 0.05$: for the lower range parameter the IDW point estimator has higher overall median and mean coverage than the kriging predictor. These values are quite the same for $\phi = 15$ and are in opposite relationship when the range parameter exceeds the maximum distance observable in the domain. In order to better understand the spatial distribution of the coverage, we compare the plots of Figures 5.2, 5.1, 5.3 and 5.4, where each row concerns one of the analysed techniques. Looking at these maps it is clear that the IDW point estimator has a coverage comparable to the kriging predictor's one when the sampling fraction is small and the range parameter is not too great. Moreover, the technique proposed in this thesis has even a better coverage when the range parameter is $\phi = 6$ (i.e. population B) and the sampling fraction is small, $f = 0.05$.

In order to assess how inference on the population total is affected by the range parameter, we compare Tables 5.7, 5.3, 5.11 and 5.15. In terms of bias we note a substantial worsening of the performances of the estimator of the population total based on the IDW point estimator as the range parameter goes from $\phi = 6$ to $\phi = 15$. This is in contradiction with the behaviour of the expansion estimator and the kriging predictor. The latter presents few changes, whereas the former even reduces its bias. When the range parameter is greater than the maximum distance observable in the domain the kriging predictor of the population total shows a smaller bias than the estimator based on the IDW point estimator and the expansion estimator as it happens for inference on individual values. It seems that globally the expansion estimator has a more consistent behaviour in terms of bias at different range parameters.

Surprisingly, the RMSE decreases as the range parameter increases whatever the technique. The estimator of the population total based on the IDW point estimator has values not much higher than the kriging predictor's ones, despite being the most biased techniques.

In terms of coverage the three techniques present values close to the nominal level of 95% at different sampling fraction and for different range parameter values.

Regardless of the range parameter used for generating the populations and of the sampling fraction, estimator $\hat{\mathrm{V}}_2[\hat{\bar{z}}(\mathbf{u}_i)]$ along with modification (4.2.39) seems the best choice: it is the one with a behaviour closer to a conservative variance estimator. The same can be said for the corresponding estimator of the variance of the estimator of the population total based on the IDW point estimator.

The sill parameter

The effect of a varying sill parameter in the suprmodel generating the populations is taken into account by comparing populations A, E and G for a range parameter slightly exceeding the mean distance in the domain, $\phi = 15$, and populations C, F and H for a rather large range parameter, $\phi = 45$.

The performances of both the IDW point estimator and kriging predictor for inference on individual values are relatively high in comparison with those of the SRSWoR estimator in predictive form; thus, in the following analysis on the effects of the sill parameter for inference on individual values it will be omitted. Let us first start by analysing the results of the sill parameter for a rather low range parameter. By the comparison of Tables 5.2, 5.18 and 5.26 it can be seen that for the lowest sampling fraction ($f = 0.05$) the IDW point estimator's overall bias distribution is very similar to the kriging predictor's one. At increasing sampling fractions a small difference becomes more and more relevant as the distribution of the former technique becomes more shrunk around its central tendency values, regardless of the sill parameter value. Moreover, as the superpopulation parameter increases, the overall bias distributions of both methods become more disperse. For the lowest sill parameter, $\sigma^2 = 1$, the IDW point estimator's overall RMSE distribution has lower values than the kriging predictor's one except for the higher sampling fraction; at $f = 0.15$ the two distributions are rather similar. Regardless of the sampling fraction and of the sill parameter value, the kriging predictor has higher overall median and overall mean coverages than the IDW point estimator.

As pointed out in the previous subsection, by increasing the range param-

eter, the IDW point estimator becomes more biased. Thus, the comparison of Tables 5.10, 5.22 and 5.30 highlights that the IDW point estimator has an overall bias distribution similar to the kriging predictor's one only for a sampling fraction of $f = 0.05$, regardless of the sill parameter. For increasing sample sizes the kriging predictor is less biased. Analogously, the overall RMSE distributions are similar at $f = 0.05$ and $f = 0.10$ for the lowest sill parameter, $\sigma^2 = 1$, and at $f = 0.05$ for $\sigma^2 = 4$; in the other cases the kriging predictor generally presents lower values. In terms of median and mean of the overall coverage, the kriging predictor always performs better than the IDW point estimator.

From the comparison of Tables 5.3, 5.19 and 5.27, it can be seen that the expansion estimator is less biased than the other techniques for inference on the population total except for higher sill values in the case of a rather low range parameter. However, in terms of RMSE are far from those of the kriging predictor and the IDW point estimator, which have quite similar behaviours. Regardless of the technique used, the coverages are sufficiently close to the nominal level in the case of $\sigma^2 = 1$ and $\sigma^2 = 4$; for the highest sill value the kriging predictor is the only one to reach a coverage of 95% except for the lowest sampling fraction.

In contradiction to the results on the previous populations, a larger range parameter (Tables 5.11, 5.23 and 5.31) reduces the bias of the IDW point estimator of the population total despite remaining the most biased method. In this case the kriging predictor of the population total is less biased than the other techniques except for the lowest sill value. Again, the RMSE of the IDW point estimator and ofthe kriging predictor are lower than the expansion estimator's one; the differences become more relevant for higher sill values. Coverages are affected by the increase of the range parameter: the kriging predictor is the only one having close to the nominal level except for the lowest sampling fraction at which all the methods have good performances.

Regardless of the range and sill parameters the jackknife estimator of the variance of the IDW point estimator involving the modification for not negligible sampling fractions is the most conservative one despite presenting negative values. Analogously, the corresponding variance estimator of the IDW point estimator of the population total is the one to be preferred.

The nugget parameter

By comparing the results on populations A, C, I, J, K and L, we address the extent of the nugget parameter on inference for populations having a range parameter barely exceeding the mean distance observable in the domain, $\phi = 15$, and a rather large parameter, $\phi = 45$.

Let us first consider the populations generated with a smaller range parameter (A, I and K); the results of the Monte Carlo experiments for inference on individual values are collected in Tables 5.2, 5.34 and 5.42, respectively concerning populations with $\tau^2 = 0$, $\tau^2 = 0.25$ and $\tau^2 = 1.00$. The following analysis regards the comparison of the IDW point estimator with the kriging predictor only since the SRSWoR estimator in predictive form has always given worse results. At the lowest sampling fraction the IDW point estimator seems to be slightly less biased since its overall bias distribution is more concentrated around its central tendency values than the kriging predictor's one. As the sampling fraction increases, the kriging improves its performance more quickly: at $f = 0.10$ the two overall bias distributions are quite similar, while higher sample sizes favour the model-based technique. The same can be said for the overall RMSE distribution. Moreover, it seems that even for a 10% sampling fraction the IDW point estimator has at least the same performances as the kriging predictor; higher fractions lead to better results of the kriging except for population L which has been generated with the highest nugget parameter. As the nugget parameter increases, the IDW point estimator tends to reach the same coverages of the kriging predictor especially at lower sampling fraction.

By comparing Tables 5.6, 5.38 and 5.46, it can be noticed that a rather large range parameter favours the kriging predictor as it improves its performances consistently with the previous analyses. For a null nugget parameter the kriging predictor offers better performances than the IDW point estimator regardless of the sampling fraction. But as the nugget parameter increases the IDW point estimator improves especially at the lowest sampling fraction. For $\tau^2 = 1.00$ and $f = 0.05$, it seems to have slightly better performances in terms of the overall distributions of the bias, RMSE and coverage.

The results of the simulations for inference on the population total confirm

the idea that a high nugget effect improves the performances of the IDW point estimator. The comparison of Tables 5.3, 5.35 and 5.43 highlights that the IDW point estimator of the population total reduces its bias as the nugget parameter increases, regardless of the sampling fraction. The fact that the RMSE gets closer to the kriging predictor's one reflects this aspect. The coverages of the three techniques remain pretty close to the nominal level.

For higher values of the range parameter, the IDW point estimator of the population total still has RMSE values close to the kriging predictor's ones; however, the results on the bias are not as appealing as for the population having lower range parameter. Especially, it seems that in the case of the highest nugget parameter estimation is affected by a not negligible bias. The coverage of the 95% confidence interval for the three techniques barely reach the nominal level especially for $\tau^2 = 1.00$.

The Monte Carlo experiments highlights once again that the conservative jackknife estimator of the variance involving the modification for not negligible sampling fractions is to the one to be preferred among those proposed both for inference on individual or population values.

Chapter 6

Conclusions

This final chapter aims at summarizing the main ideas developed throughout the thesis. The characteristics of the IDW point estimator are highlighted along with some future lines of work. As a final thought, we believe that the IDW point estimator represents quite a novelty for the inference on individual values under a design-based framework since we have only been able to find a simpler estimator in the paper by Brus and de Gruijter (1997).

6.1 Overview

The major goal of this thesis were to develop a complete theory for design-based spatial estimation both for inference on individual and global values. The theoretical results were then tested through Monte Carlo experiments in order to asses

(1) whether the IDW point estimator is suitable for inference on individual values;

(2) whether it is able to infer on the global population value;

(3) the conditions favourable to their application.

The simulation study performed in Chapter 5 pointed out that the results of the IDW point estimator for inference on individual values are not far from the results of the kriging predictor which represents the benchmark for

spatial statistics. Moreover, it emerges that the employment of the spatial information represents a major boost to design-based inference on individual values when compared to the classic SRSWoR estimator in predictive form. Answer to question *(1)* is pretty straightforward: the technique we propose may not be as good as the kriging predictor, but definitely has appealing properties at least for the populations analysed in the simulation study.

Inference on population global values highlighted that the IDW point estimator is generally slightly more biased than the other techniques. However, its RMSE is much closer to the values of the kriging predictor of the population total than to those of the expansion estimator. Therefore, objective *(2)* has been proved to be fully satisfied in the sense that the use of spatial information at estimation level improves design-based inference on a population global value.

The answer to objective *(3)* is harder since it is not easy to manage all the different aspects of a spatial population. First of all, we can say that the IDW point estimator, both for individual and global values, has given the best results for low sampling fractions. In this case the results are at least comparable to the kriging ones. This is perhaps due to the difficulty to rightly estimate the semivariogram parameters when the the sample does not permit to observe enough spatial lags. Indeed, as the sampling fraction increases, the performances of the model-based technique improve faster than the IDW point estimator's ones which sometimes even remain constant. Moreover, in the simulation study we limited the analysis only to twelve populations generated in order to manage different aspects of a spatial superpopulation model. The simulation study has shown that populations generated by using a rather large range parameter of the exponential semivariogram model produce worse results of the IDW point estimator. This is perhaps due to the use of the inverse squared Euclidean distances; adopting a lower exponent, say $1 < \alpha < 2$, may lead to better results. The sill parameter of the exponential semivariogram corresponds to the variance of the random field at each point. It results that by introducing more variability the IDW point estimator's performances becomes poorer. Finally, the analysis pointed out that higher values of the nugget parameter improve the performances of the IDW point estimator. This is perhaps due to the difficulty of the semivariogram fitting

techniques to estimate the micro-scale process.

Finally, the simulation study confirmed that the jackknife estimator of the variance which is known to be conservative (Wolter, 2007), is indeed the more conservative among the two proposed in Subsections 4.2.6 and 4.4.3, despite assuming even negative values. The naïve estimator of the variance of the IDW point estimator of the individual values has always produced the worst results and, moreover, it is even computationally more requesting than the jackknife ones.

6.2 Open topics and future developments

The theoretical results presented in this thesis leave some questions open. First of all, the asymptotic properties of the IDW point estimator are not complete: the asymptotic p-unbiasedness and p-consistency (Särndal et al., 1992) have yet to be proved. Along with the asymptotic properties, it could help to obtain the asymptotic distribution of the proposed estimator in order to better understand its behaviour. The proposed jackknife variance estimators are not suitable for use in practical application since they may produce negative values. Therefore, the development of a conservative variance estimator is still an open topic.

The simulation study presented in the previous chapter does not fully cover the whole range of possibilities. The semivariogram model used for the prediction of the values is the same as the one used for generating the populations; moreover, in the computations, the values of its parameters have been initialized with the true population values. These aspects leave no room for the model misspecification problem typical of the model-based techniques. Therefore, the fact that the results of the Monte Carlo experiments are not too far from the best combination of the kriging somewhat defends the goodness of the proposed technique. However, a broader simulation study, which takes into account these aspects, may give interesting results in the case of model misspecification.

The simulation could be widened by the introduction of other superpopulation models: different semivariogram models and non-Gaussian random

fields may lead to different results. A white-noise process could be investigated since in this case the kriging predictor coincides with the SRSWoR estimator in predictive form. Finally, it may be interesting to analyse even fully deterministic surfaces. The use of auxiliary variable through the use of GWR may help the estimation and ad hoc populations may be analysed.

Appendix A

Approximating the properties of random variables through the delta method

Let us consider the Taylor series expansion of a function $f : \mathbb{R}^d \to \mathbb{R}$

$$f(\mathbf{x}) = \sum_{|\alpha|=0}^{k} \frac{D^\alpha f(\mathbf{x}_0)}{\alpha!}(\mathbf{x} - \mathbf{x}_0)^\alpha + o(\|\mathbf{x} - \mathbf{x}_0\|^k)$$
$$\sum_{|\alpha|=0}^{k} \frac{D^\alpha f(\mathbf{x}_0)}{\alpha!}(\mathbf{x} - \mathbf{x}_0)^\alpha + O(\|\mathbf{x} - \mathbf{x}_0\|^{(k+1)}),$$

where the multi-index notation is employed $|\alpha| = \alpha_1 + \cdots + \alpha_n$, $\alpha! = \alpha_1! \cdots \alpha_n!$ and $\mathbf{x}^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$.

Let the argument of $f(\cdot)$ be a d-dimensional random variable. Its kth-order Taylor series approximation around the expected value $\boldsymbol{\mu} = \mathrm{E}[\boldsymbol{\mathcal{X}}]$ is

$$f(\boldsymbol{\mathcal{X}}) = \sum_{|\alpha|=0}^{k} \frac{D^\alpha f(\boldsymbol{\mu})}{\alpha!}(\boldsymbol{\mathcal{X}} - \boldsymbol{\mu})^\alpha + o(\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mu}\|^k)$$
$$= \sum_{|\alpha|=0}^{k} \frac{D^\alpha f(\boldsymbol{\mu})}{\alpha!}(\boldsymbol{\mathcal{X}} - \boldsymbol{\mu})^\alpha + O(\|\boldsymbol{\mathcal{X}} - \boldsymbol{\mu}\|^{(k+1)}). \tag{A.1}$$

The kth-order approximation of the expected value of the function $f(\boldsymbol{\mathcal{X}})$ is obtained by taking the expectation of the kth-order approximated func-

tion (A.1) Analogously, we obtain the kth-order approximation of the variance of the same function of random variables as well as for the covariance between two different functions, $f(\cdot)$ and $g(\cdot)$, of two d-dimensional random variables, $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{Y}}$.

Let us consider the first-order Taylor approximation of the bivariate function $f(x, y) = x/y$ in the point (x_0, y_0)

$$\frac{x}{y} = \frac{x_0}{y_0} + \frac{1}{y_0}(x - x_0) - \frac{x_0}{y_0^2}(y - y_0) + O((x - x_0)^2 + (y - y_0)^2).$$

The approximation at the first-order of the ratio of two random variables $f(X, Y) = X/Y$ in the point $(\mathrm{E}[X], \mathrm{E}[Y])$ is

$$\begin{aligned} \mathrm{E}\left[\frac{X}{Y}\right] =& \frac{\mathrm{E}[X]}{\mathrm{E}[Y]} + o_p(n^{-1/2}) \\ =& \frac{\mathrm{E}[X]}{\mathrm{E}[Y]} + O_p(n^{-1}). \end{aligned} \tag{A.2}$$

Analogously, we obtain the variance of the same ratio

$$\begin{aligned} \mathrm{V}\left[\frac{X}{Y}\right] =& \frac{\mathrm{V}[X]}{\mathrm{E}[Y]^2} - 2\frac{\mathrm{Cov}(X, Y)\,\mathrm{E}[X]}{\mathrm{E}[Y]^3} + \frac{\mathrm{V}[Y]\,\mathrm{E}[X]^2}{\mathrm{E}[Y]^4} + o_p(n^{-3/2}) \\ =& \frac{\mathrm{V}[X]}{\mathrm{E}[Y]^2} - 2\frac{\mathrm{Cov}(X, Y)\,\mathrm{E}[X]}{\mathrm{E}[Y]^3} + \frac{\mathrm{V}[Y]\,\mathrm{E}[X]^2}{\mathrm{E}[Y]^4} + O_p(n^{-2}). \end{aligned} \tag{A.3}$$

Finally ,the covariance between two ratios of random variables, X/Y e W/Z, in the points $(\mathrm{E}[X], \mathrm{E}[Y])$ and $(\mathrm{E}[W], \mathrm{E}[Z])$ is

$$\begin{aligned} \mathrm{Cov}\left(\frac{X}{Y}, \frac{W}{Z}\right) =& \frac{\mathrm{Cov}(X, W)}{\mathrm{E}[Y]\,\mathrm{E}[Z]} - \frac{\mathrm{E}[W]\,\mathrm{Cov}(X, Z)}{\mathrm{E}[Y]\,\mathrm{E}[Z]^2} - \frac{\mathrm{E}[X]\,\mathrm{Cov}(Y, W)}{\mathrm{E}[Y]^2\,\mathrm{E}[Z]} \\ &+ \frac{\mathrm{E}[X]\,\mathrm{E}[W]\,\mathrm{Cov}(Y, Z)}{\mathrm{E}[Y]^2\,\mathrm{E}[Z]^2} + o_p(n^{-3/2}) \\ =& \frac{\mathrm{Cov}(X, W)}{\mathrm{E}[Y]\,\mathrm{E}[Z]} - \frac{\mathrm{E}[W]\,\mathrm{Cov}(X, Z)}{\mathrm{E}[Y]\,\mathrm{E}[Z]^2} - \frac{\mathrm{E}[X]\,\mathrm{Cov}(Y, W)}{\mathrm{E}[Y]^2\,\mathrm{E}[Z]} \\ &+ \frac{\mathrm{E}[X]\,\mathrm{E}[W]\,\mathrm{Cov}(Y, Z)}{\mathrm{E}[Y]^2\,\mathrm{E}[Z]^2} + O_p(n^{-2}). \end{aligned} \tag{A.4}$$

For a deeper look at the delta method, one can see Section 5.3 of the book by Bickel and Doksum (2001) or Chapter 4 of the book by Small (2010).

Appendix B

Relations between quantities involved in the calculus of the statistical properties of the IDW point estimator

The following Hadamard product is involved in the proof of the calculus of the approximated expectation in Theorem 4.2.1:

$$\mathbf{e}_i \circ \boldsymbol{\phi}_i = \mathbf{0}_{N\times 1}.$$

The following Hadamard products are involved in the proof of Theorem 4.2.2 regarding the calculus of the approximated variance of the IDW point estimator:

$$\begin{aligned}
\mathbf{e}_i\mathbf{e}_i^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top &= \mathbf{0},\\
diag(\mathbf{1}_N - \mathbf{e}_i) \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top &= diag(\boldsymbol{\phi}_i)^2,\\
(\mathbf{1}_N - \mathbf{e}_i)\mathbf{e}_i^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top &= \mathbf{0},\\
\mathbf{e}_i(\mathbf{1}_N - \mathbf{e}_i)^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top &= \mathbf{0},\\
\mathbf{1}_N\mathbf{1}_N^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top &= \boldsymbol{\phi}_i\boldsymbol{\phi}_i^\top.
\end{aligned}$$

The following equalities are used in the proof of Theorem 4.2.3 regarding the computation of the approximated covariance between two IDW point

estimators:

$$
\begin{aligned}
\mathbf{e}_i \circ \boldsymbol{\phi}_j &= \phi_{ij}\mathbf{e}_i \\
\mathbf{1}_N \circ \boldsymbol{\phi}_i &= \boldsymbol{\phi}_i \\
\mathbf{e}_i \circ \boldsymbol{\phi}_i &= \mathbf{0}_{N\times 1} \\
\mathbf{I}_N \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \operatorname{diag}(\boldsymbol{\phi}_i \circ \boldsymbol{\phi}_j) \\
\mathbf{1}_N\mathbf{1}_N^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top \\
\mathbf{e}_i\mathbf{e}_i^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \mathbf{0}_{N\times N} \\
\mathbf{e}_j\mathbf{e}_j^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \mathbf{0}_{N\times N} \\
\mathbf{e}_j(\mathbf{1}_N - \mathbf{e}_j)^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \phi_{ij}\mathbf{e}_j\boldsymbol{\phi}_j^\top \\
(\mathbf{1}_N - \mathbf{e}_j)\mathbf{e}_j^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \mathbf{0}_{N\times N} \\
\operatorname{diag}(\mathbf{1}_N - \mathbf{e}_j) \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \operatorname{diag}(\boldsymbol{\phi}_i \circ \boldsymbol{\phi}_j) \\
\mathbf{e}_i(\mathbf{1}_N - \mathbf{e}_i)^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \mathbf{0}_{N\times N} \\
(\mathbf{1}_N - \mathbf{e}_j)\mathbf{e}_j^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \phi_{ij}\boldsymbol{\phi}_i\mathbf{e}_i^\top \\
\mathbf{e}_i\mathbf{e}_j^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \mathbf{0}_{N\times N} \\
\mathbf{e}_j\mathbf{e}_i^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \phi_{ij}^2\mathbf{e}_j\mathbf{e}_i^\top \\
(\mathbf{1}_N - \mathbf{e}_i - \mathbf{e}_j)\mathbf{e}_i^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \phi_{ij}\left\{(\mathbf{1}_N - \mathbf{e}_j) \circ \boldsymbol{\phi}_i\right\}\mathbf{e}_i^\top \\
\mathbf{e}_i(\mathbf{1}_N - \mathbf{e}_i - \mathbf{e}_j)^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \mathbf{0}_{N\times N} \\
(\mathbf{1}_N - \mathbf{e}_i - \mathbf{e}_j)\mathbf{e}_j^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \mathbf{0}_{N\times N} \\
\mathbf{e}_j(\mathbf{1}_N - \mathbf{e}_i - \mathbf{e}_j)^\top \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \phi_{ij}\mathbf{e}_j\left\{(\mathbf{1}_N - \mathbf{e}_i) \circ \boldsymbol{\phi}_j\right\}^\top \\
\operatorname{diag}(\mathbf{1}_N - \mathbf{e}_i - \mathbf{e}_j) \circ \boldsymbol{\phi}_i\boldsymbol{\phi}_j^\top &= \operatorname{diag}(\boldsymbol{\phi}_i \circ \boldsymbol{\phi}_j).
\end{aligned}
$$

Appendix C

Jack-knife variance estimation

The jackknife provides a quick and reliable tool for estimating the variance of an estimator for the infinite case when using a SRSWoR. Jack-knife is basically a resampling technique consisting in randomly divide the original sample in k groups of size m, $n = mk$.

Let $\hat{\theta}$ denote an estimator and $\mathrm{V}_p[\hat{\Theta}]$ its variance calculated in respect of the estimator distribution over the σ-algebra $\mathcal{S}$ of all the possible samples of the sample space $\mathfrak{S}$. Then we define the jackknified version of a generic estimator as

$$\hat{\bar{\theta}} = \frac{1}{k} \sum_{\alpha=1}^{k} \hat{\theta}_\alpha,$$

where $\hat{\theta}_\alpha$ is the pseudo-value defined as

$$\hat{\theta}_\alpha = k\hat{\theta} - (k-1)\hat{\theta}_{(\alpha)} \tag{C.1}$$

and $\hat{\theta}_{(\alpha)}$ is the estimator of the same functional form as $\hat{\theta}$ calculated calculated for each subsample. In the case of a linear estimator, it results that the jackknified estimator is equal to the parent estimator, $\hat{\bar{\theta}} = \hat{\theta}$. However, for non-linear estimators it is generally not true.

Then, we define the variance of the jackknified estimator as

$$\hat{\mathrm{V}}_1[\hat{\bar{\Theta}}] = \frac{1}{k(k-1)} \sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \hat{\bar{\theta}})^2,$$

which can be used as an estimator of the variance of estimator $\hat{\theta}$. A conservative alternative to the previous variance is

$$\hat{\mathrm{V}}_2[\hat{\bar{\Theta}}] = \frac{1}{k(k-1)} \sum_{\alpha=1}^{k} (\hat{\theta}_\alpha - \hat{\theta})^2.$$

The previous variance estimators can likewise be used; however, the sampling fraction $f = n/N$ needs to be negligible. If not, the following modification of the estimator functional used in the subsamples provides better results:

$$\hat{\theta}^*_{(\alpha)} = \hat{\theta} - (1-f)^{1/2}(\hat{\theta}_{(\alpha)} - \hat{\theta}).$$

If $k = n$ each subsample has dimension $n - 1$ since only one unit of the original sample is omitted at a time.

For a deeper insight on the jackknife variance estimation techniques one can see Wolter (2007)

Appendix D

The expansion estimator and its predictive form

As Bolfarine and Zacks (1992) point out, design-based estimators are rarely presented in their predictive form. They refer to the paper by Rodrigues et al. (1985) as the main work on the subject.

In example 1.3.1 of their book, they rewrite the expansion estimator of the population total in predictive form as

$$\hat{t}_E(y) = \frac{N}{n} \sum_{i \in S} y_i = n\bar{y} + (N - n)\bar{y}.$$

Therefore, the unsampled values is estimated by the sample mean:

$$\hat{y}_j = \bar{y}, \quad \forall j \notin S.$$

Unfortunately, nothing is known about the statistical properties of such estimators.

Appendix E

The simulated populations

Each population from A to L is the realization of a Gaussian random field having superpopulation model and parameters as collected in Table E.1.

Table E.1: Random superpoluation models generating the populations analysed in the simulations

	Semivariogram	μ	σ^2	ϕ	τ^2
A	exponential	2	4	15	0.00
B	exponential	2	4	6	0.00
C	exponential	2	4	45	0.00
D	exponential	2	4	90	0.00
E	exponential	2	1	15	0.00
F	exponential	2	1	45	0.00
G	exponential	2	8	15	0.00
H	exponential	2	1	45	0.00
I	exponential	2	4	15	0.25
J	exponential	2	4	45	0.25
K	exponential	2	4	15	1.00
L	exponential	2	4	45	1.00

E.1 Population A

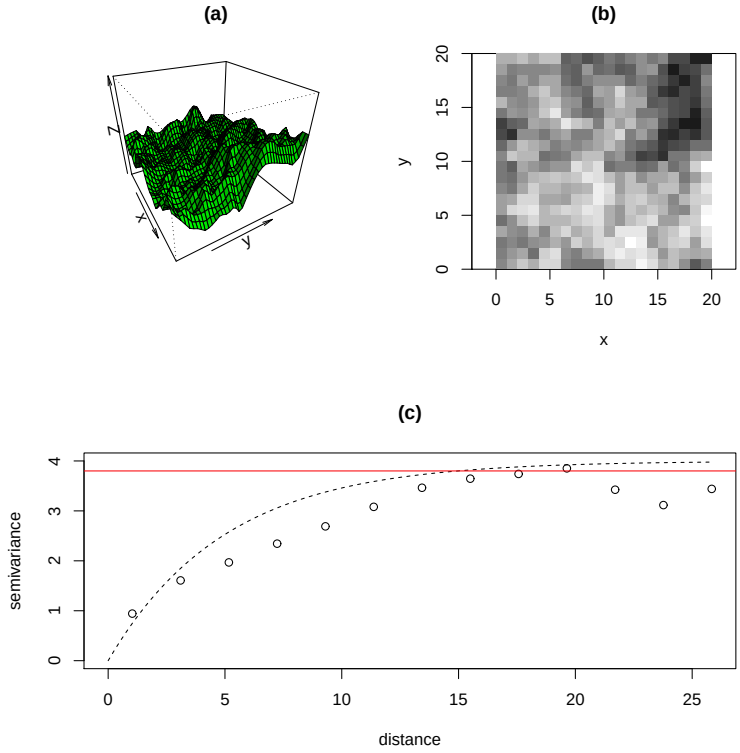

Figure E.1: Population A: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.2: Population A: descriptive statistics

mean	0.979	median	0.856
standard deviation	1.671	interquartile range	4.609
skewness	0.375	quartile skewness	0.040
kurtosis	-0.346	octile kurtosis	0.006

E.2 Population B

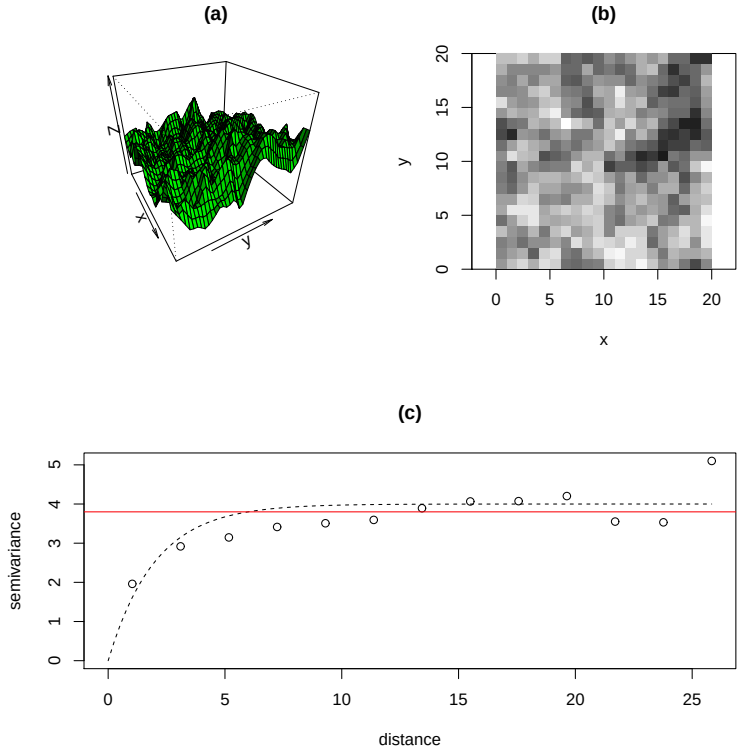

Figure E.2: Population B: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.3: Population B: descriptive statistics

mean	1.676	median	1.594
standard deviation	1.882	interquartile range	2.781
skewness	0.207	quartile skewness	0.079
kurtosis	-0.431	octile kurtosis	-0.107

E.3 Population C

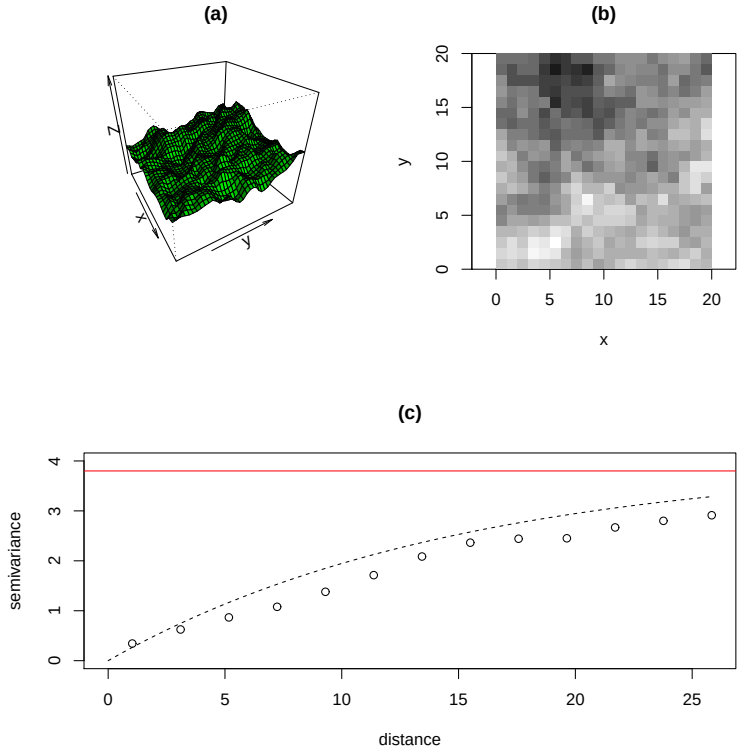

Figure E.3: Population C: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.4: Population C: descriptive statistics

mean	1.499	median	1.411
standard deviation	1.246	interquartile range	1.795
skewness	0.280	quartile skewness	0.084
kurtosis	-0.308	octile kurtosis	-0.059

E.4 Population D

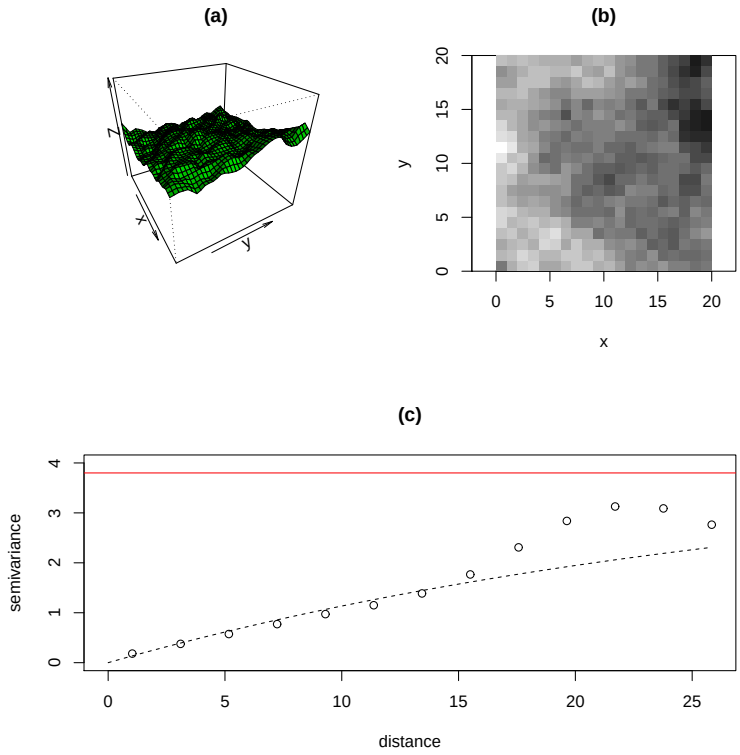

Figure E.4: Population D: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.5: Population D: descriptive statistics

mean	3.579	median	3.725
standard deviation	1.085	interquartile range	1.597
skewness	-0.011	quartile skewness	-0.296
kurtosis	-0.301	octile kurtosis	-0.155

E.5 Population E

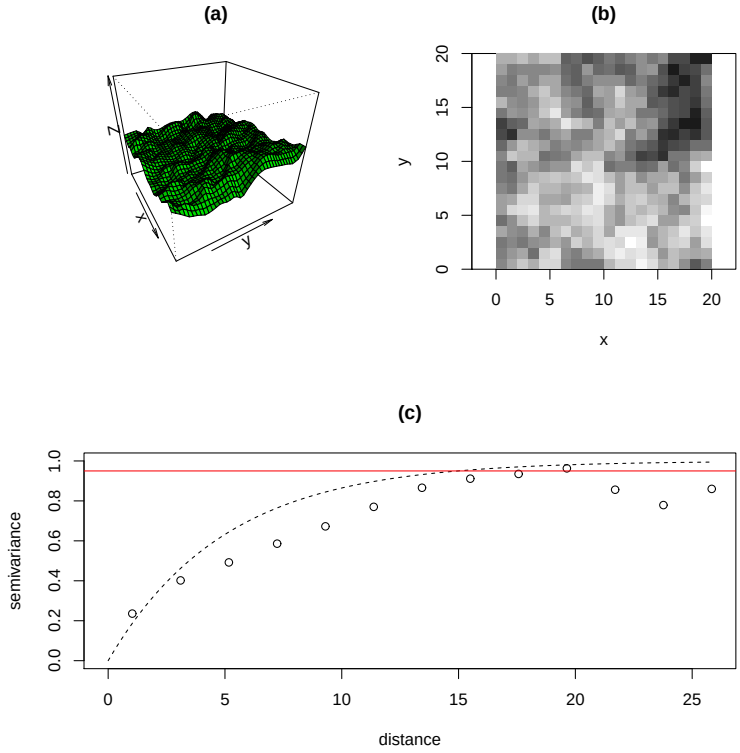

Figure E.5: Population E: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.6: Population E: descriptive statistics

mean	1.490	median	1.428
standard deviation	0.835	interquartile range	1.159
skewness	0.375	quartile skewness	0.040
kurtosis	-0.346	octile kurtosis	0.006

E.6 Population F

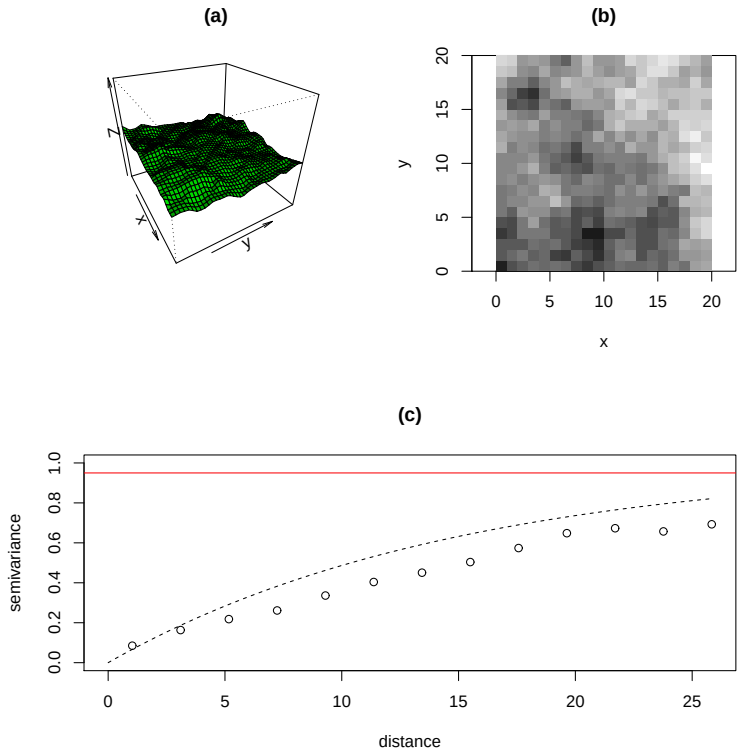

Figure E.6: Population F: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.7: Population F: descriptive statistics

mean	1.967	median	2.002
standard deviation	0.603	interquartile range	0.870
skewness	-0.048	quartile skewness	?-0.107
kurtosis	-0.530	octile kurtosis	0.016

E.7 Population G

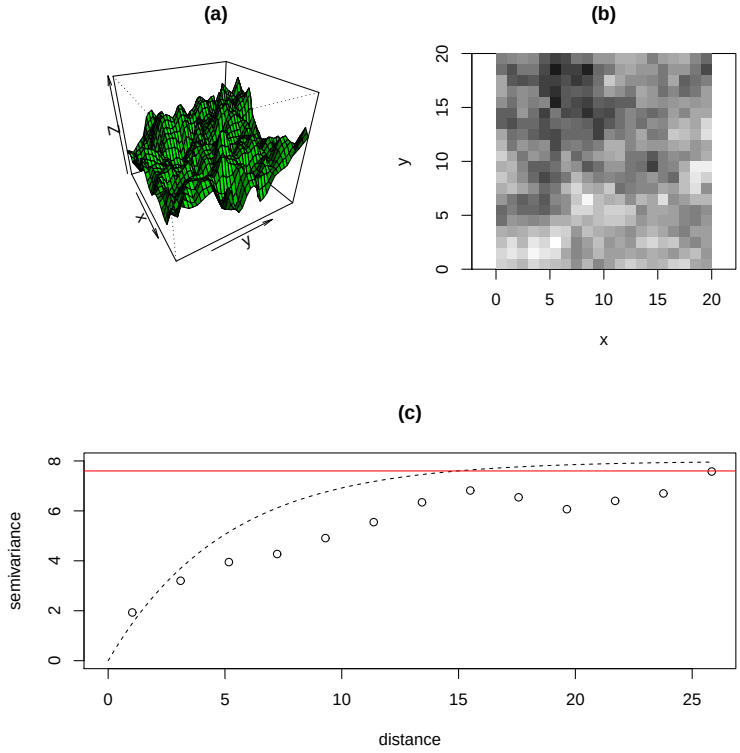

Figure E.7: Population G: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.8: Population G: descriptive statistics

mean	2.406	median	2.270
standard deviation	2.263	interquartile range	2.974
skewness	0.100	quartile skewness	0.102
kurtosis	-0.257	octile kurtosis	0.027

E.8 Population H

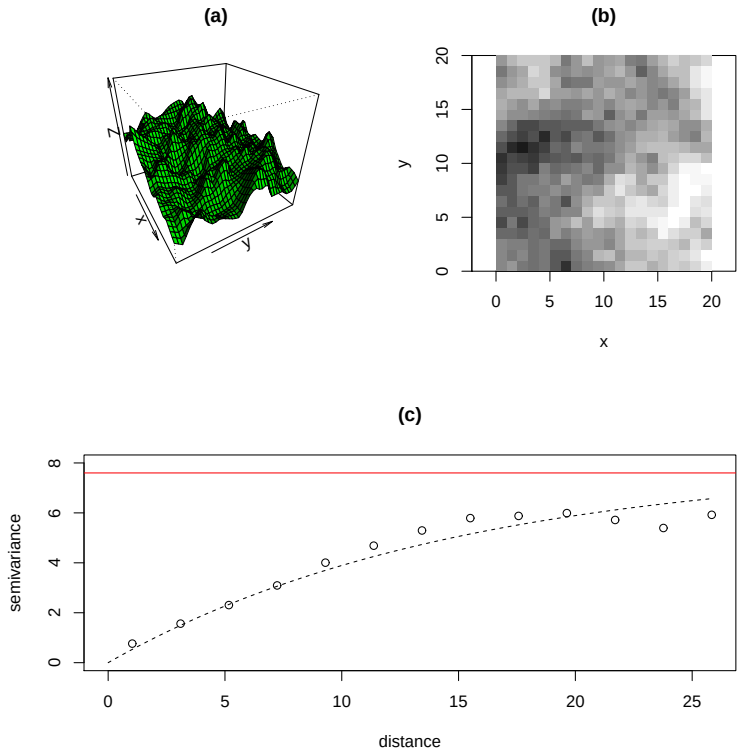

Figure E.8: Population H: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.9: Population H: descriptive statistics

mean	1.377	median	1.454
standard deviation	2.012	interquartile range	2.964
skewness	0.009	quartile skewness	-0.083
kurtosis	-0.534	octile kurtosis	-0.025

E.9 Population I

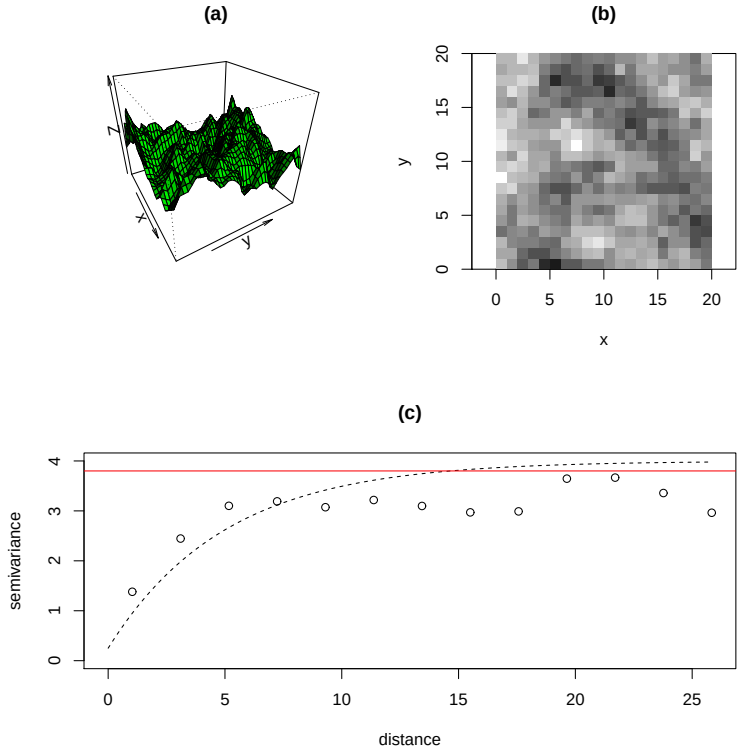

Figure E.9: Population I: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.10: Population I: descriptive statistics

mean	1.866	median	1.805
standard deviation	1.740	interquartile range	2.439
skewness	0.040	quartile skewness	0.055
kurtosis	-0.147	octile kurtosis	-0.100

E.10 Population J

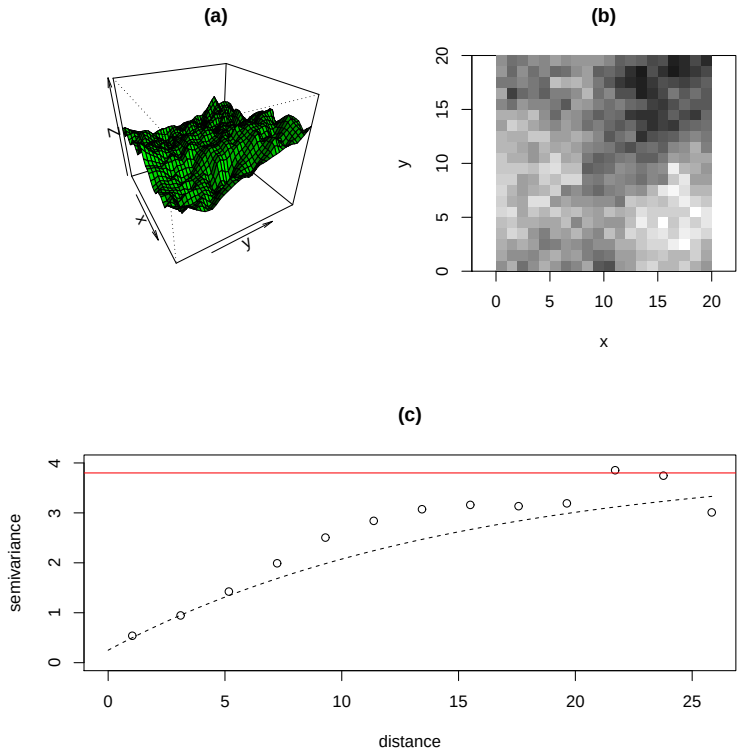

Figure E.10: Population J: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.11: Population J: descriptive statistics

mean	3.083	median	2.996
standard deviation	1.547	interquartile range	2.179
skewness	0.233	quartile skewness	0.039
kurtosis	-0.498	octile kurtosis	-0.004

E.11 Population K

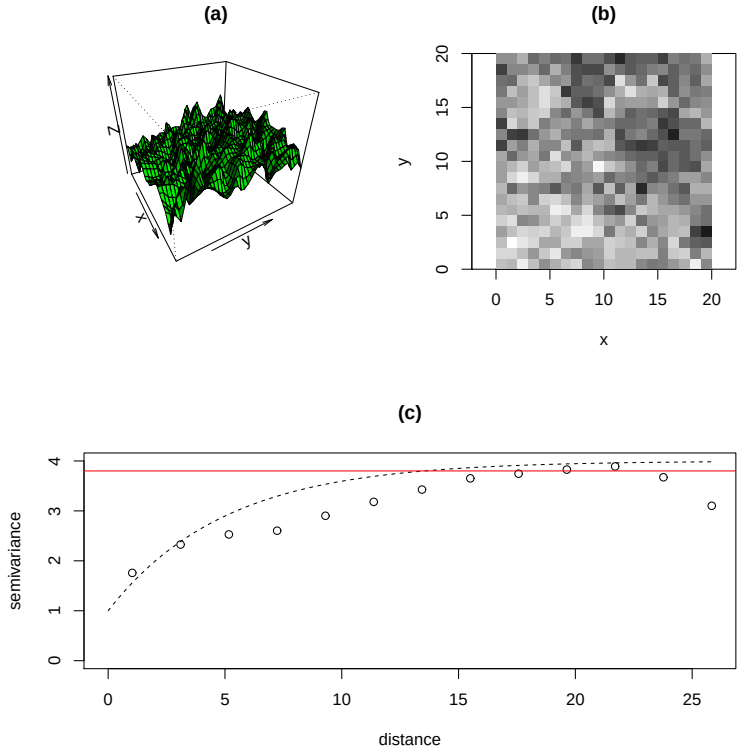

Figure E.11: Population K: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.12: Population K: descriptive statistics

mean	1.658	median	1.629
standard deviation	1.739	interquartile range	2.670
skewness	0.094	quartile skewness	0.037
kurtosis	-0.634	octile kurtosis	-0.177

E.12 Population L

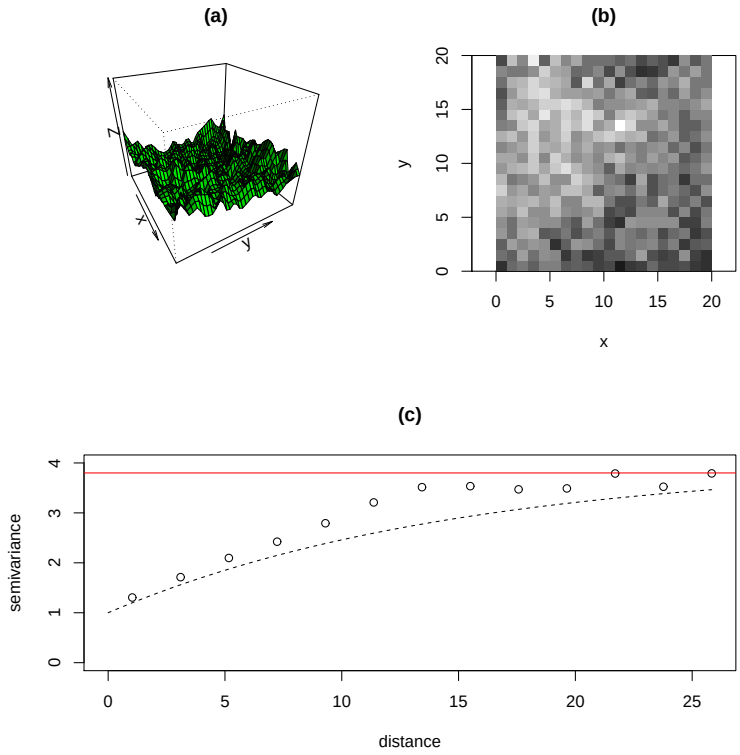

Figure E.12: Population L: (a) perspective and (b) tile plots, (c) true (dashed line) versus empirical (empty dots) semivariogram

Table E.13: Population L: descriptive statistics

mean	-0.811	median	-0.923
standard deviation	1.684	interquartile range	2.312
skewness	0.045	quartile skewness	0.053
kurtosis	-0.474	octile kurtosis	0.142

Bibliography

Adler, R. (1981). *The Geometry of Random Fields*. New York: Wiley.

Bickel, P. and K. Doksum (2001). *Mathematical Statistics: Basic Ideas and Selected Topics*. Upper Saddle River, New Jersey: Prentice-Hall.

Bolfarine, H. and S. Zacks (1992). *Prediction Theory for Finite Populations*. New York: Springer–Verlag.

Bruno, F., D. Cocchi, and A. Vagheggini (2012). Finite population properties of individual predictors based on spatial patterns. *Environmental and Ecological Society*, advance online publication.

Brunsdon, C., A. Fotheringham, and M. Charlton (1998). Geographically weighted regression-modelling spatial non-stationarity. *The Statistician 47*, 431–443.

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

Chilès, J.-P. and P. Delfiner (1999). *Geostatistics: Modeling Spatial Uncertainty*. New York: Wiley.

Cicchitelli, G. and G. Montanari (2012). Model-assisted estimation of a spatial population mean. *International Statistical Review 00*, 1–16.

Cochran, W. (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Annals of Mathematical Statistics 17*, 164–177.

Cordy, C. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics & Probability Letters 18*, 353–362.

Cox, D., L. Cox, and K. Ensor (1997). Spatial sampling and the environment: some issues and directions. *Environmental and Ecological Statistics 4*, 219–233.

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Mathematical Geology 17*, 563–586.

Cressie, N. (1990). Origins of kriging. *Mathematical Geology 22*, 239–252.

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

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

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

Diggle, P. and P. J. Ribeiro (2007). *Model Based Geostatistics*. New York: Springer.

Diggle, P., J. Tawn, and R. Moyeed (1998). Model-based geostatistics (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics) 47*, 299–350.

Dol, W., T. Steerneman, and T. Wansbeek (1996). Matrix algebra and sampling theory: the case of the horvitz–thompson estimator. *Linear Algebra and its Applications 237*, 225–238.

Fisher, R. (1966). *The Design of Experiments* (eighth ed.). London: Oliver and Boyd.

Fotheringham, A., C. Brunsdon, and M. Charlton (2000). *Quantitative Geography*. London: Sage.

Franke, R. (1982). Scattered data interpolation: tests of some methods. *Mathematics of Computation 38*, 181–200.

Gaetan, C. and X. Guyon (2010). *Spatial Statistics and Modeling. Modélisation et Statistique Spatiales*. New York: Springer.

Gandin, L. (1960). On optimal interpolation and extrapolation of meteorological fields. *Trudy Main Geophys. Obs. 114*, 75–89. (in Russian).

Gandin, L. (1963). *Objective Analysis of Meteorological Fields*. Leningrad: Gidrometeorotogicheskoe Izdatel'stvo (GIMIZ). (Translated by Israel Program for Scientific Translations, Jerusalem, 1965).

Gelfand, A., P. Diggle, M. Fuentes, and P. Guttorp (2010). *Handbook of Spatial Statistics*. Boca Raton, FL.: Chapman & Hall / CRC.

Ghosh, S., A. Gelfand, and T. Mølhave (2012). Attaching uncertainty to deterministic spatial interpolations. *Statistical Methodology 9*, 251–264.

Guttorp, P. and T. Gneiting (2006). Studies in the history of probability and statistics XLIX: On the Matérn correlation family. *Biometrika 93*, 989–995.

Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge University Press.

Handcock, M. and M. Stein (1993). A bayesian analysis of kriging. *Technometrics 35*, 403–410.

Journel, A. and C. Huijbregts (1978). *Mining Geostatistics*. London: Academic Press.

Kallenberg, O. (1997). *Foundations of Modern Probability*. New York: Springer.

Kitanidis, P. and E. Vomvoris (1983). A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations. *Water Resources Research 19*, 677–690.

Kolomogorov, A. (1941a). Interpolation and extrapolation of stationary random sequences. *Izvestya Akademiya Nauk SSSR, Seriia Matematicheskaia 5*, 85–95. (translation, Memo RM-3090-PR, Rand Corporation, Santa Monica, California, 1962).

Kolomogorov, A. (1941b). The local structure of turbulence in an incompressible fluid at very large Reynolds numbers. *Doklady Akademii Nauk SSSR 30*, 85–95. (reprinted in Turbulence: classic papers of statistical theory, S. K. Friedlander and L. Topping (Ed.): Interscience Publishers, New York, 1961, p..151-155).

Krige, D. (1951). A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa 52*, 119–139.

Li, J. and A. Heap (2008). *A Review of Spatial Interpolation Methods for Environmental Scientists*. Canberra: Geoscience Australia.

Lloyd, C. (2011). *Local Models for Spatial Analysis*. Boca Raton, Florida: CRC Press.

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

Matérn, B. (1960). Spatial variation. *Meddelanden fran Statens Skogsforskningsinstitut 49*.

Matérn, B. (1986). *Spatial Variation* (second ed.). Berlin: Springer-Verlag.

Matheron, G. (1962). *Traite de Geostatistique Applicque*, Volume I: Memoires du Bureau de Recherches Geologiques et Minieres, no. 14. Paris: Technip.

Matheron, G. (1963). *Traite de Geostatistique Applicque*, Volume II, Le Krigeage: Memoires du Bureau de Recherches Geologiques et Minieres no. 24. Paris: Bureau de Recherche Grologiques et Minidres.

Matheron, G. (1971). *The theory of regionalized variables and its application*. Fonatinebleu, France: Ecole des Mines.

Mercer, W. and A. Hall (1911). The experimental error of field trials. *Journal of Agricultural Sciences 4*, 107–132.

Mitas, L. and H. Mitasova (1999). *Spatial Interpolation*, Volume 1, Chapter 34. London: Wiley.

Parzen, E. (1960). *Modern Probability Theory and its Applications*. New York: Wiley.

Patterson, H. and R. Thompson (1971). Recovery of interblock information when block sizes are unequal. *Biometrika 58*, 545–554.

Pearson, E. (1990). *Student: A Statistical Biography of William Sealy Gosset*. Oxford: Oxford University Press.

Quenouille, M. (1949). Problems in plane sampling. *Annals of Mathematical Statistics 20*, 355–375.

Ribeiro Jr, P. and P. Diggle (2001). `geoR`: A package for geostatistical analysis. *R-NEWS 1*, 14–18.

Rodrigues, J., H. Bolfarine, and A. Rogatko (1985). A general theory of prediction in finite populations. *International Statistical Review 53*, 239–254.

Rubinstein, R. and D. Kroese (2007). *Simulation and the Monte Carlo Method*. New York: Wiley.

Särndal, C.-E., B. Swensson, and J. Wretman (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Schabenberger, O. and F. Pierce (2002). *Contemporary Statistical Models for the Plant and Soil Sciences*. Boca Raton, Florida: CRC Press.

Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. In *Proceedings of the 1968 23rd ACM national conference*.

Small, C. (2010). *Expansions and Asymptotics for Statistics*. Boca Raton, Florida: CRC Press.

Stehman, S. (2000). Practical implications of design-based sampling inference for thematic map accuracy. *Remote Sensing of Environment 75*, 35–45.

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

Stuart, A. and K. Ord (1987). *Kendall's Advanced Theory of Statistics: Distribution Theory*, Volume 1. London: Arnold.

Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography 46*, 234–240.

Webster, R. and M. Oliver (2007). *Geostatistics for Environmental Scientists*. Chichester: John Wiley and Sons.

Wolter, K. (2007). *Introduction to Variance Estimation*. New York: Springer.

Yates, F. (1948). Systematic sampling. *Philosophical Transactions of the Royal Society of London A 241*, 345–377.