Spatially balanced sampling: a review and a reappraisal

Roberto Benedetti

"G. d'Annunzio" University of Chieti-Pescara, Department of Economic Studies, Italy

Viale Pindaro, 42 – 65127 Pescara, Italy. E-mail: benedett@unich.it

Federica Piersimoni

Italian National Institute of Statistics (*ISTAT*), *Italy. E-mail:* piersimo@istat.it

Paolo Postiglione[1]

"G. d'Annunzio" University of Chieti-Pescara, Department of Economic Studies, Italy

Viale Pindaro, 42 – 65127 Pescara, Italy. E-mail: postigli@unich.it

Abstract

Spatially distributed data exhibit particular characteristics that should be considered when designing a survey of spatial units. Unfortunately, traditional sampling designs generally do not allow for spatial features, even though it is usually desirable to use information concerning spatial dependence in a sampling design. This paper reviews and compares some recently developed randomised spatial sampling procedures, using simple random sampling without replacement as a benchmark for comparison. The approach taken is design-based and serves to corroborate intuitive arguments about the need to explicitly integrate spatial dependence into sampling survey theory. Some guidance for choosing an appropriate spatial sampling design is provided, and some empirical evidence for the gains from using these designs with spatial populations is presented, using two datasets as illustrations.

Keywords: design-based inference, spatial units, finite populations, spatial dependence, spatial statistics.

[1] Corresponding author

1. Introduction

Geo-referenced data possess special characteristics that have strongly influenced the development of models used for spatial data analysis (Haining, 2003). Surprisingly, however, there is much less research effort aimed at integrating spatial information into sample designs and data collection methodology.

Estimates of totals, means, and/or proportions of some target variables are typically the main output of a sample survey. However, we could also be interested in analyses based on linear and/or generalised regression as is typical, for example, in spatial modelling studies. Unfortunately, practitioners usually apply infinite population methods and neglect to account for how the sample data were obtained. In many cases, the use of complex sampling schemes implies that weights should be used and that the variances of the survey estimators be computed in a way that reflects the complexity of the sample design. For a comprehensive discussion on this topic, see Chambers et al. (2012) and Benedetti et al. (2015).

When spatial units are randomly selected from a finite population, their main identifying characteristic is their geo-referencing information. Consequently, it is clear that this spatial distribution should be used as strategic information when designing the sample selection procedure (Vallée et al. 2015, Dickson & Tillé, 2016, Benedetti et al. 2016).

In this paper, we focus on probability samples that are well-spread over the population of interest. A sample is geographically well-spread if the number of selected units is close to what is expected on average in every part of the study region (Grafström & Lundström, 2013). These types of sampling designs avoid the selection of neighbouring geographical units. By spatially spreading the sample across the target population, such sample designs aim for a certain property, which is typically referred to as being *spatially balanced* (Steven & Olsen, 2004).

The rationale for a spatially balanced approach in sample selection is mainly intuitive. As a consequence, its potential impact on the estimation efficiency is not well understood. Because several sampling algorithms that aim to achieve spatial balance have been introduced (Christman, 2000; Wang et al. 2012), this gap in the sampling literature is worth addressing.

This paper reviews the current state of the art as far as spatially balanced sampling is concerned. In particular, we describe the main techniques for selecting spatially balanced samples that have been introduced in the recent literature, and we evaluate their advantages and drawbacks using comparisons based on two different datasets. To the best of our knowledge, this type of comparison is the first of its type because the methods have been introduced in very different frameworks and with reference to several fields of application. A unified presentation is therefore timely. In particular, we believe that the empirical comparisons set out in this paper should provide guidance for survey practitioners who wish to use spatial information in their sample designs. In this paper, we adopt the survey perspective of sampling from a finite population. Following this approach, it is also possible to tackle some problems in the monitoring of environmental and natural resources, although these problems are often explored using the infinite populations framework (see Thompson, 2013).

To formalise our approach to spatial sampling, let $U=\{1,2,\ldots,N\}$ be a finite population together with a set of q auxiliary variables $\mathbf{X}=\{\mathbf{x}_1, \mathbf{x}_2,\ldots, \mathbf{x}_j,\ldots, \mathbf{x}_q\}$ and a set of h coordinates obtained by the geo-coding of each population unit $\mathbf{C}=\{\mathbf{c}_1, \mathbf{c}_2,\ldots,\mathbf{c}_g,\ldots, \mathbf{c}_h\}$, where $\mathbf{x}_j=\{x_{1j}, x_{2j},\ldots, x_{ij},\ldots, x_{Nj}\}$ is the generic j-th auxiliary, and $\mathbf{c}_g=\{c_{1g}, c_{2g},\ldots, c_{ig},\ldots, c_{Ng}\}$ is the generic g-th coordinate. Given $\mathbf{C}$, we can always calculate, using any appropriate distance definition, a matrix $\mathbf{D}_U=\{d_{kl}; k=1,...,N, l=1,...,N\}$ that specifies the distances between all pairs of units in the population. We also note that to efficiently use the geographical position to design a sampling method, it is necessary to define methods that use the information in $\mathbf{D}_U$ and are therefore not simple extensions of sampling procedures that do not require this information (Müller, 2007).

On the basis of the intuitive notion that population units that are closer together provide less information about a population parameter than units that are farther apart, Benedetti & Palma (1995), Arbia & Lafratta (2002), Rogerson & Delmelle (2004), and Bohorquez et al. (2016) used spatial information about the population units to design well-spread samples. The selection of such samples corresponds to solving a combinatorial optimisation problem, which identifies the most spread-out sample that optimises an appropriate estimation criterion.

From a randomisation perspective, these approaches are unsatisfactory because they lead to a unique and fixed sample, and hence, they cannot be used to generate a randomisation-based sampling distribution for the population characteristic of interest. In particular, we are interested in an inference that assumes that the population values of the variable under investigation **y** are non-random (i.e., deterministic), with the uncertainty in inference characterised via the randomised sample design. However, this assumption of having a fixed **y** runs counter to assuming a model that explicitly characterises the spatial dependences in **y** (a standard feature of geo-referenced data) through a spatial autocorrelation parameter in the covariance matrix of a stochastic process. Such models are easily incorporated into either a model-based or model-assisted approach to defining the sampling schemes for spatial units. For a detailed distinction between these different approaches to inferences from survey data, see Pfeffermann & Rao (2009, Part 4).

The basic challenge that faces researchers who wish to account for spatial dependence in a design-based framework is the necessity to incorporate this dependence into the sampling plan. To achieve this goal, we model the unknown population values **y** of the survey variable in terms of a known matrix of auxiliaries, **X**, which is available from past surveys or administrative data. The basis of this approach therefore lies in the assumption that our prior knowledge suggests that the finite population can be viewed as if it were a sample from an infinite superpopulation and that there is a model ξ that defines the characteristics of the superpopulation (Isaki & Fuller, 1982). We then assume that this model is still valid for the current survey period (Baillargeon & Rivest, 2009; 2011). The sample design is then defined on the basis of the anticipated moments of **y** given **X**, i.e., the expected values under repeated sampling of the model-based moments of sample quantities.

Consider the estimator $\hat{t}$ of the total t of **y**. Under this approach, an optimal sample design will therefore seek to minimise the anticipated variance (AV) of the estimator $\hat{t}$ under an appropriate superpopulation model for **y**. This can be defined as the variance of the random variable ($\hat{t}-t$) under both the design s and the superpopulation model ξ

$$AV(\hat{t}-t) = E_\xi\{E_s[(\hat{t}-t)^2]\} - \left[E_\xi\{E_s(\hat{t}-t)\}\right]^2, \tag{1}$$

where E_ξ denotes the expectation with respect to the model, and E_s denotes the expectation with respect to the sample design.

In the case of a scalar survey variable **y**, it is quite common to assume a linear superpopulation model when designing a sample (Särndal et al. 1992, p. 449). This model has the form

$$\begin{bmatrix} y_k = x_k^t\beta + \varepsilon_k \\ E_\xi(\varepsilon_k) = 0 \\ Var_\xi(\varepsilon_k) = \sigma_k^2 \\ E_\xi(\varepsilon_k\varepsilon_l) = 0 \ k \neq l \end{bmatrix}, \tag{2}$$

where k represents the units, $\boldsymbol{\beta}$ is a vector of regression coefficients, and ε_k are random variables.

If we assume that $Var_{\xi}(\varepsilon_k) = \sigma_k^2$ and $Cov_{\xi}(\varepsilon_k, \varepsilon_l) = \sigma_k \sigma_l \rho_{kl}$, where ρ_{kl} is the dependence parameter, the AV of the Horvitz-Thompson (HT) estimator of the total of a variable y given **X** under (2) is (Grafström & Tillé, 2013)

$$AV(\hat{t}_{HT} - t) = E_s[(\sum_{k \in s} \frac{x_k}{\pi_k} - \sum_{k \in U} x_k)^t \beta]^2 + \sum_{k \in U} \sum_{l \in U} \sigma_k \sigma_l \rho_{kl} \frac{\pi_{kl} - \pi_k \pi_l}{\pi_k \pi_l}, \quad (3)$$

where U is a finite population. This superpopulation model can help us better understand the impact of spatial information when designing a sample for a population of geo-coded units. Our strategy is to use a design that minimises AV. We can see that Eq. (3) is minimised when the dependence parameter (ρ_{kl}) between each pair of units (k, l) is equal to 0. Assuming that ρ_{kl} decreases as the distance between k and l increases, we can therefore minimise AV by selecting the sampling units in such a way that we maximise the distance between them, thus ensuring that the realised sample is well-spread over the study region. This approach, which is based on the concept of anticipated variance, then represents one way of allowing for spatial dependence in a design-based inference. In fact, from (3), it can be seen that selecting a spatially well-spread sample implies a reduction in the anticipated variance if the underlying spatial dependence is positive, whereas if we assume that the observed phenomenon could exhibit a negative spatial dependence, a clustered sample should be preferred.

However, there are other possible reasons for selecting a spatially well-spread sample. For example, it appears to be reasonable that one should select such a well-spread sample when the observed phenomenon **y** is spatially stratified; i.e., its value varies significantly between the different sub-regions of the space that contains the target population. Alternatively, a well-spread sample is usually more efficient when the population exhibits a cluster structure because in such cases, a single sampled unit or a few such units within a cluster can characterise the whole cluster.

The layout of this paper is as follows. In Section 2, we briefly summarise a number of important spatial sampling methods that are described in the statistics literature. We focus on specific procedures that do not allow the selection of spatially contiguous units, on the Generalised Random Tessellation Stratified (GRTS) technique, which is routinely applied in environmental surveys and on various spatial extensions of *balanced sampling*. In Section 3, we then discuss the main features of the spatial sampling algorithms that make effective use of the matrix $\mathbf{D}_U$. Some empirical results are reported in Section 4. The designs are compared using two well-known datasets: the Mercer Hall wheat yield data and the Baltimore house sales price data. These allow one to compare the performances of the spatial sampling designs described in previous sections in terms of their design-based Mean Squared Errors (MSEs) using Simple Random Sampling without replacement (SRS) as a benchmark. Section 5 contains some concluding remarks that focus on outstanding research issues that are associated with spatial sampling designs.

2. Recent Developments in Spatial Sampling Selection Methods

The problem of the enhancement of estimates of population variables, using some additional knowledge on the sampling units, is a very debated topic in the surveyed literature. When addressing spatial data, this auxiliary information is represented by the geographic distribution of the units.

Following this idea, Hedayat et al. (1988b) suggested that more information on the reference population might be obtained, if the sample does not consider pairs of contiguous units. Hedayat et al. (1988b) proposed the use of a sampling design for which the second-order inclusion probabilities (i.e., π_{kl}s) are non-decreasing in the distance between units k and l.

In addition, Hedayat et al. (1988a) introduced a basic design denoted as Balanced Sampling design Excluding Contiguous units (BSEC). It is a fixed size *n* design with the constraint π_{kl}=0 if the units *k* and *l* are contiguous, while all of the other π_{kl}s are set to be equal to an appropriate constant.

This latter approach has been widely discussed in literature. Stufken (1993) defined the Balanced Sampling plans excluding Adjacent units (BSA). This design generalised the BSEC by excluding all of those pairs of units whose distance is less than or equal to a certain threshold *m*. Additionally, Stufken et al. (1999) introduced Polygonal Designs (PDs) and showed that PDs are equivalent to BSA plans. The reader can find other interesting contributions in works by Hedayat & Stufken (1998), Mandal et al. (2008), and Wright & Stufken (2008).

However, the most relevant drawback of the BSEC is the assumption of some naive linear or circular one-dimensional ordering of the units on which the method is based. This hypothesis appears to be too elementary if it is applied to geographically distributed data that are characterised by multidirectional relationships.

To obtain well-spread samples, a common choice is represented by the use of the units' stratification based on their geographical positions. Unfortunately, this strategy does not have a significant impact on the second-order inclusion probabilities. Therefore, it is not clear how to obtain a sample that has units that are not close each other. When addressing spatial phenomena that can be observed over a continuous space, a widely used simple option is the extension of the use of systematic sampling to two or more dimensions, overlapping a regular grid to the spatial domain (i.e., the study region).

These practices inspired the Generalised Random Tessellation Stratified (GRTS) design (Stevens & Olsen, 2004). The GRTS systematically selects the units and maps the two-dimensional spatial population into one dimension while attempting to preserve some of the multi-dimensional order.

The first contributions in the area of tessellation-based plans are due to Dalenius et al. (1961), Olea (1984), and Overton & Stehman (1993), who analysed the Random Tessellation Stratified (RTS) design. The RTS design selects randomly from a spatial point frame through a two-step procedure. In the first step, a tessellation that is coherent with a regular grid is randomly located over the domain. Note that the use of a regular grid does not at all limit the selection of the sample. In the second step, a random point is selected within each random tessellation cell.

Stevens (1997) introduced the Multiple-Density Nested Random Tessellation Stratified (MD-NRTS) design to provide for non-constant spatial sampling rates. The geometric assumption that underlies the MD-NRTS was the notion of coherent intensification of a grid. This methodology adds points to a regular grid in such a way that it produces a finer regular grid with similarly shaped but smaller tessellation cells. The MD-NRTS uses geographical data at multiple levels of spatial detail and can thus be considered a multiscale spatial method, similar to the quadtree method. For the application of the quadtree method in spatial sampling, see Minasny et al. (2007).

Stevens & Olsen (2004) applied this last approach to GRTS design, extending the methodology to potentially produce an infinite series of nested and coherent grids. This process generates a function $f(\cdot)$ that transforms a two-dimensional space into a one-dimensional space while preserving some spatial order and some proximity relationships.

The GRTS technique can be summarised in the following steps:

1. Assume that the sample frame consists of N geographically distributed units. Place a square grid over the frame of the geographic region.

2. Sub-divide the geographic region into four sub-regions and randomly assign numbers to the sub-regions. Sub-divide the sub-regions into another four sub-regions; randomly assign numbers independently to each new sub-region, creating hierarchical addresses. Continue sub-dividing until there is only one unit per cell. This process is denoted as the quadrant-recursive method. This process preserves the spatial relationships of the sample units (Steven & Olsen, 2004).
3. Identify each unit with a hierarchical address that was obtained according to the partitioning process. The sampling units are then arranged in line following the numerical hierarchical address order. The line has *N* lengths. This process maps a two-dimensional space into a one-dimensional space.
4. The line is divided into a number of equal-length segments depending on the requested sample size. A unit from each segment is randomly selected.

The GRTS provides an equal probability plan that is well-spread over the study area. For the definition of unequal probability samples, it is required to give a length that is proportional to its inclusion probability to each unit. In Figure 1, we present an example application of the GRTS technique (compared with the SRS method), selecting a sample from a finite list of points. Furthermore, we overlaid the Voronoi polygons to the selected sample (see Okabe et al. 2009). Finally, note that the circled points in the plot are selected in the sample.

The Voronoi polygons can be used to define a statistical measure that provides information on the spatial distribution of the selected samples. This index is very helpful in comparing the ability of an algorithm to produce a well-spread set of points over the study region.

<Fig. 1 about here>

Now define $v_k = \sum_{i \in VP(k)} \pi_i$ to be the sum of the first-order inclusion probabilities of the units of the population in the k-th Voronoi polygon. For any sample unit, we will have E(v_k)=1, and for a spatially balanced sample, all of the v_ks should be close to 1 (Steven & Olsen, 2004). The variance $Var(v_k)$ can be used as a measure of spatial balance for a sample. Obviously, a lower value of $Var(v_k)$ is an indicator of a *good* spatially balanced sample. Unfortunately, the index $Var(v_k)$ involves the first-order inclusion probabilities π_k that are fixed and cannot be modified. Hence, there are practical difficulties in adopting this index to directly design samples. For this reason, it may be more appropriate to use selection rules based on the distance between the sampled units (see the next paragraph) and the *ex-post* evaluation of the obtained $Var(v_k)$.

The GRTS design produces a sample that has a fixed π_k. In this case, the classical HT estimator can be applied to obtain estimates of population characteristics. Stevens (1997) provided exact expressions for the π_{kl}s in a specific case of the GRTS. Unfortunately, these expressions preclude the use of the variance estimator based on the HT or the Yates-Grundy-Sen estimators (Särndal et al. 1992) because they tend to be unstable due to the presence of several π_{kl} s that are very close to zero (Stevens & Olsen, 2004).

To overcome this problem, Stevens & Olsen (2003) proposed the contrast-based variance estimator for the GRTS design. This approach can be considered to be similar to the smoothed estimator (Overton & Stehman, 1993). The proposed variance estimator is defined as

$$\hat{V}_{NBH}(\hat{t}_y) = \sum_{k \in s} \sum_{l \in NB_k} wd_{kl} \left(\frac{y_k}{\pi_k} - \sum_{t \in NB_k} wd_{kt} \frac{y_t}{\pi_t} \right)^2 \qquad (4),$$

where NB_k is a local neighbourhood of unit k, and wd_{kl}'s are weights that decrease as the distance between unit k and l increases, with $\sum_k wd_{kl} = \sum_l wd_{kl} = 1$.

The efficiency of a spatially stratified sample, such as GRTS, improves as the number of strata increases and the per-stratum sample size decreases. The maximum efficiency is obtained for a one-unit per-stratum-design.

The GRTS is currently the most widely used method for designing spatially balanced samples. It has several advantages. In particular, the GRTS is a probability-based sampling technique that ensures a good degree of spatial balance. Also, it can be used to select samples that have unequal selection probabilities. The GRTS design can easily address problems that occur in sampling populations, such as poor frame information, inaccessibility, variable probability, irregular spatial patterns, missing data, and panel structures.

However, there are no theoretical results or sufficient empirical evidence on the gain in the efficiency that arises from the use of GRTS when addressing finite populations. Conversely, its application on continuous surface sampling is very desirable because it provides estimators that are very accurate and normally distributed for large samples, with a variance convergence rate of order $n^{-\gamma}$ with $1<\gamma\leq 3$ (Barabesi & Franceschi, 2011; Barabesi & Marcheselli, 2008).

The GRTS also has some disadvantages. In particular, the mapping that is used is not always very efficient because the units that are close in two-dimensional space could be mapped far apart in one-dimensional space.

However, the control of the quality of the selected sample is a very remarkable task for the researcher. This objective can be pursued by using some covariates **X** that are known for each unit of the population U. In this case, the interest is to evaluate the closeness of the HT estimator $\hat{t}_{HT,x_j}$ to the known population totals t_{x_j} for each of the q covariates. This argument is based on the reasonable assumption that a positive correlation between the auxiliary variable **x** and the survey variable **y** should hold and, consequently, that an error on an auxiliary variable **x** might be reproduced in the same way on the target variable **y**.

These considerations represent the rationale for the definition of the balanced sampling method. This scheme has been introduced in a non-spatial context, but if we use the coordinates of the units as auxiliary variables, the method can be straightforwardly considered a spatial technique. Furthermore, to corroborate its application for geographically distributed data, in the last part of this section, we describe a spatial extension of balanced sampling that makes effective use of spline regression.

It is important to highlight that the terms spatially balanced sampling and balanced sampling are also theoretically connected. As evidenced by Grafström & Lundström (2013), balance *means* global balance. Spatial balance is more restrictive and can be viewed as a form of local balance. Hence, spatial balance is more difficult to achieve than global balance, and thus, spatially balanced samples are a subset of balanced samples (Grafström & Schelin, 2014).

Balanced sampling is based on the simple idea that the estimated total of some of the auxiliary variables is equal to the known values of the population and is expressed as (Tillé, 2011)

$$\sum\nolimits_{k \in s} d_k x_{k,j} = \hat{t}_{HT,x_j} = t_{x_j} = \sum\nolimits_{k \in U} x_{k,j} \ \forall j = 1,\dots, \qquad (5)$$

where $d_k = 1/\pi_k$, for all $s \in \Omega$ (where Ω is the set of all possible samples) such that $p(s)>0$, and $x_{k,j}$ is the value of the *j*-th variable for the *k*-th unit. The main problem is that condition (5) can rarely be satisfied exactly (Tillè, 2011). The aim is thus to find a design that approximately satisfies balancing the equations in (5).

This approach has been mainly discussed within a model-based framework (Valliant et al. 2000). However, balanced sampling can be realistically acknowledged even in the design-based approach. This last idea led Deville and Tillé (2004) to introduce the CUBE method, which was later improved by Chauvet & Tillé (2006). The name of this sampling algorithm derives from the argument that each sample can be viewed as the coordinates of a vertex of the hypercube in the dimensional space $\mathbb{R}^N$.

Two main parts compose the algorithm: the flight and the landing phase. During the first phase, the constraints represented by Equation (5) are always exactly satisfied. The objective is to randomly round off to 0 or 1 almost all of the πs. The landing phase considers the possibility that the condition expressed in (5) cannot always be exactly realised. Unfortunately, the flight phase can be computationally quite expensive. For these reasons, Chauvet & Tillé (2006) developed a faster algorithm for implementing the three steps described above.

The CUBE method, particularly its faster version, was thoroughly discussed and analysed, with particular attention given to studying the properties of the selection method. Deville & Tillé (2005) investigated the use of balanced sampling in terms of its variance estimation, whereas Chauvet (2009) analysed the possibility of extending the constraints to the sub-populations. Finally, Chauvet et al. (2011) studied the optimal selection probabilities when addressing multivariate auxiliary variables.

As were well highlighted by Tillé (2011), the main advantages of balanced sampling are the following:

- the increase in the accuracy of the HT estimator because its variance depends on only the residuals of the regression of the variable of interest by the balancing variables;
- the protection against large sampling errors because the most unfavourable samples have a zero probability of being selected and go against a misspecification of the model within a model-based inference;

- the assurance that the sample sizes in the planned domains[2] are not too small or even equal to zero.

In a spatial context, the CUBE method can be successfully applied while imposing the condition that for any selected sample, the HT estimates of the first M moments of each coordinate should match the first M moments of the population. The underlying assumption is that the survey variable $\mathbf{y}$ follows a polynomial spatial trend of order M. Unfortunately, a sampling design that satisfies all of these conditions does not necessarily exist, even relaxing the conditions by adding an acceptable error. However, this situation in practical applications is quite rare unless the number of moments that we want to constrain is very high or the sample size is very low.

Recently, following this argument, Breidt & Chauvet (2012) extended the CUBE approach using linear mixed models at the design stage to include auxiliary information. They modified the variables included the balancing Equation (5), and suggested drawing penalised balanced samples. The method is based on the generation of an ordered new set of variables $\mathbf{B}$ through the use of penalised splines (Hastie & Tibshirani, 1990). These variables should be used in place of the covariates $\mathbf{X}$. The subjectivity of the method consists in the choice of the order of the splines and the number of knots K. This uncertainty may affect the final results, in particular the efficiency of the sampling design.

However, it is evident that distance is very useful in summarising the spatial distribution of the sample units. This consideration leads to the intuitive criterion that units that are close should rarely appear simultaneously in the sample. Therefore, the balanced sampling, even in the modified version with splines, is quite efficient in capturing linear and non-linear spatial

[2] A domain is denoted as planned if it is specified in advance of the sample allocation stage (Rao & Molina, 2015).

trends, but it is not able to catch the spatial dependence of the geographically distributed units of the population.

3. Sampling Designs Based on the Distance between Units

The motivation for the choice of selecting spatial well-spread samples is surely realistic if it is considered to be acceptable that increasing the distance between two units k and l increases the difference, observed at units k and l, namely, □$|y_k-y_l|$. In this situation, it is evident that the variance of the HT estimator will necessarily decrease if we set high joint inclusion probabilities to pairs that have very different **y** values.

Following this argument, Arbia (1993) introduced, in a model-based context, a draw-by-draw scheme, the Dependent areal Units Sequential Technique (DUST). However, the properties of this plan can also be analysed in a design-based framework because the method respects the randomisation principle. Arbia (1993) argued that "*it is intuitively clear that, when we have a clue of the spatial correlation structure underlying the spatial phenomenon to be sampled, it is desirable to exploit this information in the sampling design. In this way we could avoid duplicate information partly contained in areas already sampled and we can economize sampling costs without loosing reliability of the estimates*".

The DUST algorithm starts with the first unit selected at random, say k, at every step $t<n$; then, the algorithm updates the selection probabilities of any other unit l of the population according to the rule $\pi_l^{(t)} = \pi_l^{(t-1)}(1-\lambda^{d_{kl}})$, where λ is a tuning parameter that is useful for controlling the distribution of the sample over the study region, and d_{kl} is the distance between units k and l. The updating rule can be easily extended to follow the classical exponential decreasing autocorrelation function, $\pi_l^{(t)} = \pi_l^{(t-1)}(1-e^{-\lambda d_{kl}})$.

Note that Arbia (1993) suggested that several parameters λ should be estimated by using some auxiliary information to obtain a list of discretised distances. This procedure is very sensitive to the choice of λ; the results are very different according to the parameter used because it controls the distribution of the sample. Clearly, at each step, the updated inclusion probabilities should be normalised to have a sum of unity.

The main problem of this technique is represented by the difficulty of selecting random samples from a population and simultaneously producing a well-spread sample in more dimensions. The situation can be even more complicated when the objective of the sampling design is to use unequal inclusion probabilities. Bondesson & Grafström (2011) faced this problem, extending Sampford's method (Tillé, 2006) to select unequal probability samples over small strata when the inclusion probabilities have non-integer sums within strata.

Bondesson & Thorburn (2008) introduced the Correlated Poisson Sampling (CPS) to select units that have unequal inclusion probabilities. The CPS considers a sequential list of the visit of the population units. In practice, the algorithm randomly determines the inclusion of each unit k in the sample, according to probabilities that are modified at each step to produce correlations between the indicator variable of the unit visited, say I_k, and the indicator variables relative to all of the other units of the population, say I_l with $l \neq k$.

The logic of this sampling selection algorithm is the following. If unit 1 is included with probability $\pi_1^{(0)} = \pi_1$, we set $I_1 = 1$ and $I_1 = 0$ otherwise. At step t, we select the unit t with probability $\pi_t^{(t-1)}$. The inclusion probability for the unit $k \geq t+1$ is updated according to $\pi_k^{(t)} = \pi_k^{(t-1)} - (I_t - \pi_t^{(t-1)}) w_{k-t}^{(t)}$, where $w_{k-t}^{(t)}$ are weights that depend on I_1, $I_2, \ldots$, I_{t-1} but not on I_t. The weights depend only on previous and not on future units' random selection outcomes. Additionally, they are not restricted to be positive (Bondesson & Thorburn, 2008). Grafström (2012) observed that if the weights were positive, they produced a negative correlation between the indicator variables. However, if they were negative, the correlation would be positive (Grafström, 2012). Finally, note that in this sampling algorithm, the initial value of first-order inclusion probability is set to be equal to $\pi_k^{(0)} = \pi_k$.

It is worth noting that a negative correlation between the indicator variables of the units that are closer to those visited and selected is more appropriate in the case that the aim is to obtain a sample with close units that rarely appear simultaneously. The method is general: any design with fixed first-order inclusion probabilities can be implemented through CPS.

Grafström (2012) introduced a method called Spatially Correlated Poisson Sampling (SCPS) by adapting the CPS. The SCPS (Grafström, 2012) is a modification of the CPS method by introducing a measure of distance between the units. Note that for the implementation of the method, the distances between all of the pairs of units must be known. The SCPS design is based on two different strategies for choosing the weights: maximal weights and Gaussian preliminary weights. The maximal weights strategy provides samples of fixed size if the inclusion probabilities sum to an integer. After a decision on unit *t*, we will assign as much weight as possible to the closest unit (in terms of distance) among the remaining k=*t*+1,…,*N* units. Then, we proceed in the same way with the second closest unit. The procedure is similarly performed with the other units. Note that if the distance between two different units is equal, then the weight is equally distributed on these units. To ensure the fixed first-order inclusion probabilities, the weights should satisfy the following restriction:

$$-min(\frac{1-\pi_k^{(t-1)}}{1-\pi_t^{(t-1)}}, \frac{\pi_k^{(t-1)}}{\pi_t^{(t-1)}}) \leq w_{k-t}^{(k)} \leq min(\frac{\pi_k^{(t-1)}}{1-\pi_t^{(t-1)}}, \frac{1-\pi_k^{(t-1)}}{\pi_t^{(t-1)}}). \qquad (6)$$

The maximal weights strategy locally balances the sample size: in any local sub-region, the number of selected units is close to what is expected on average (Theorem 2, Grafström et al. 2012).

Using the second strategy, the weights are controlled by a Gaussian distribution that is centred on the position of unit *k*. It provides worse performance than the maximal weights strategy does (Grafström, 2012).

Unfortunately, the maximal weight strategy can provide many second-order inclusion probabilities that are equal to zero, making the definition of the design-based unbiased estimator of the variance infeasible. To solve this problem, it is possible to approximate the second-order inclusion probabilities to obtain an approximately unbiased estimator (Bondesson & Thorburn, 2008). In essence, the use of these weight strategies causes the sample to be less spatially balanced and the HT estimator to be less efficient.

Grafström et al. (2012) introduced two alternative procedures to draw samples with fixed π_k and correlated inclusion probabilities. These techniques are an extension of the Pivotal method that was introduced to select πps samples (Deville & Tillé, 1998).

The two methods presented by Grafström et al. (2012) are very similar. The only difference is in the criterion of the choice of the two nearby units *k* and *l*. These two techniques are denoted as the Local Pivotal Method 1 (LPM 1), which provides better *spatially balanced* samples, and the Local Pivotal Method 2 (LPM 2), which is simpler and faster.

The methods can be summarised in the following steps. A sample is obtained in *T* steps. At each step, the first-order inclusion probabilities are updated for two units, and in this way, the sampling outcome is decided for at least one of the two units. When the updated inclusion probability π_k^* is equal to 0 or 1, then a label, which is not selected or selected, respectively, is assigned to unit *k*. This unit is then removed from the population and cannot be chosen again. The updating procedure is repeated with updated inclusion probabilities until a label is assigned to all of the units of the population.

Deville & Tillé (1998) suggested randomly choosing a couple of units at each step, to maximise the entropy of the selected units. Conversely, Grafström et al. (2012) used the same updating rule of Deville & Tillé (1998), applied for two nearby units.

The LPM 1 randomly chooses the first unit k and then the closer unit l. Note that if two or more units have the same distance to k, the method randomly chooses between them. In this case, the inclusion probabilities are updated as follows.

If $\pi_k + \pi_l < 1$, then

$$(\pi_k^*, \pi_l^*) = \begin{cases} (0, \pi_k + \pi_l) & \text{with probability } \frac{\pi_l}{\pi_k + \pi_l} \\ (\pi_k + \pi_l, 0) & \text{with probability } \frac{\pi_k}{\pi_k + \pi_l} \end{cases}, \tag{7}$$

Otherwise, if $\pi_k + \pi_l \geq 1$, then

$$(\pi_k^*, \pi_l^*) = \begin{cases} (1, \pi_k + \pi_l - 1) & \text{with probability } \frac{1-\pi_l}{2-\pi_k-\pi_l} \\ (\pi_k + \pi_l - 1, 1) & \text{with probability } \frac{1-\pi_k}{2-\pi_k-\pi_l} \end{cases}. \tag{8}$$

This algorithm has an expected number of computations that, in the worst case, is O(N^3) and, in the best case, O(N^2).

The LPM 2 always updates the first-order inclusion probabilities with (7) and (8), without the constraint that the two units should be nearest neighbours. The expected number of computations that is needed to select a sample is, in this case, O(N^2).

Grafström et al. (2012) performed simulated comparisons among some different spatial sampling methods. They observed that the LPM and the SCPS produce samples that are much more spatially balanced than the GRTS design. Nevertheless, the LPM 1 appears to be slightly better than the LPM 2 for several sample sizes and for equal or unequal inclusion probabilities.

In analogy to all spatially balanced sampling algorithms, the LPM methods provide some second-order inclusion probabilities that are equal to zero. Hence, a design-based variance estimator of the HT estimator is infeasible. This problem can be overcome by, for example, using the local neighbourhood variance estimator (see Formula (4)). This estimator also produces encouraging results for SCPS and LPM.

Grafström & Tillé (2013) combined their techniques (i.e., the LPM and the CUBE) and proposed a new method that aims to achieve a double property of balancing. This technique ensures that the sample is well-spread and avoids the selection of neighbouring units (i.e., such as the LPM). In addition, the method also allows satisfying balancing equations on auxiliary variables that are available on all of the sampling spatial units (i.e., as in the CUBE). This method is denoted as Doubly Balanced Spatial Sampling (DBSS).

Unfortunately, the estimation (especially variance estimation) can be very difficult for many sampling schemes that have been described in this paper. In fact, the explicit derivations of the first- and second-order inclusion probabilities might be prohibitive for most of the summary indices of distance adopted. The computation of the unbiased HT variance estimator can be precluded because it is based on the knowledge of π_{kl}>0. For these reasons, to overcome this problem and to make operational use of these methods, it is necessary to estimate the inclusion probabilities. See Fattorini (2006) for an effective solution to this problem.

4. Empirical Evidence

To evaluate the relative strengths and weaknesses of the designs presented in the previous sections, simulation-based studies would be desirable. However, in this paper, for simplicity, we choose to verify the spatial designs on only two well-known datasets. These examples will be motivated purely as illustrations. The designs have been evaluated in terms of their statistical performance. It is worth noting that the sampling designs will be compared on the basis of the design-based properties. All of the codes of the examples were developed using the R software (R Core Team, 2016). In particular, we used the following R packages: sampling (Tillè & Matei, 2015), survey (Lumley, 2014), spsurvey (Kincaid & Olsen, 2016), and BalancedSampling (Grafström & Lisic, 2016). For further details about the code that can be employed to estimate spatially balanced samples, see Benedetti et al. (2015).

The data used in this design-based application derive from two different sources: the Mercer-Hall and the Baltimore datasets. These data are largely debated and analysed in the spatial statistics literature. The Mercer-Hall dataset concerns the uniformity trial of wheat in 1910. The data consist of 500 observations, and the main variables are the grain yield and the straw yield in pounds. The Baltimore dataset regards the house sales price and other characteristics for a spatial hedonic regression in Baltimore (Maryland, United States). The data entail 211 observations on 17 variables, which include the sales price of the house, the number of rooms, the number of bathrooms, the number of car spaces in the garage, and the coordinates. The following two datasets are also available in R: the first is the agridat package (Wright, 2015), and the second is the spdep package (Bivand & Piras, 2015).

For the purpose of this paper, we treated the two datasets as populations from which we select a sample. It is important to note that even if the spatial dependence in the sample does not necessarily reflect the spatial dependence in the population, the spatially balanced designs gain in efficiency with respect to spatially unbalanced designs that do not use the spatial dependence. On these samples, we estimated the total, using the HT estimator, of the variables of interest, the wheat and house prices.

The Mercer-Hall dataset is composed by areal geographical units. Conversely, the Baltimore dataset consists of point-referenced units. From looking at Figure 2, it is evident that both datasets show a significant geographical trend.

<Fig. 2 about here>

In Table 1, we summarise the main results in terms of the relative efficiency of the different plans compared with the SRS technique.

<Tab. 1 about here>

The GRTS captures the existence of any spatial trends and takes advantage of them to better select the units. Note that in this example, the GRTS provides a sensible gain in the relative efficiency with respect to the SRS. Lower values are usually obtained in correspondence to SCPS, LPM 1, LPM 2, DBSS 1 and DBSS 2, which show that the distance based methods select units that are well-spread across the study region.

The spatial distribution of the sample induced by SCPS, LPM 1, LPM 2, DBSS 1 and DBSS 2 leads to an increase in the efficiency of the different designs. Finally, it is worth noting that the worst results are usually obtained with DUST.

Considering these illustrations, it is possible to derive some recommendations for the best practices in spatial sampling. In that case, we expect that our data show an important spatial trend, and we suggest the use of the CUBE method, especially in its spatial version. Further, if we believe that the spatial dependence is remarkably present in the dataset, we should use a sampling method that is based on the distance between the spatial units (in particular, we suggest LPM 1). Finally, if both of the spatial effects are shown in the dataset, then a method that lets us simultaneously consider the spatial dependence and spatial trend, such as doubly balanced sampling, is highly recommended. Note that these two effects are intertwined and have an additive and separate impact on the anticipated variance, as is clear from Formula (3).

5. Conclusions

The definition of appropriate spatial sampling designs represents an enormous challenge for statisticians and researchers who work with geographically distributed data. The specific characteristics of the georeferenced populations should be considered when designing a sample (Grafström et al. 2014, Benedetti et al. 2015, Benedetti et al. 2016).

Many populations in environmental, agricultural, and forestry studies are distributed over space. It is clear that spatial units cannot be sampled as though they had been generated under the classical independent urn model. This argument is mainly due to some of the effects that spatial data show, such as clustering of the coordinates, homogeneity, spatial trends, and local homogeneity.

The main challenge for the researchers is how to include these effects in the sampling designs to reduce the variance of the estimators. The commonly used methods of spatial systematic sampling and spatial stratified sampling only partially use these spatial effects. For these reasons, in recent decades, many sampling designs that explicitly consider these spatial characteristics have been introduced in the literature.

This paper aims to review the main recent contributions in the thematic literature of spatially balanced sampling. In particular, the aim is to describe and compare several spatial sampling methods that have been only partially presented in the previous papers (see Wang et al. 2012). The main issue concerns the ability of a sample to be well-spread to take advantage of the presence of any spatial structure that is present in the analysis of geocoded populations. We compared the methods using two real populations. The main results are that if these spatial characteristics of the data exist and the method considers them, then there can be a remarkable reduction in the sampling error compared with SRS.

Several issues remain open for future research. In particular, it is necessary to theoretically derive the second-order inclusion probabilities π_{kl}s that are often unknown in the sampling plans described. This improvement might guarantee, for example, the use of the classical HT variance estimator that otherwise would not be possible to employ.

References

Arbia, G. (1993). The use of GIS in spatial statistical surveys. *Int. Stat. Rev.*, **61**, 339–359.

Arbia, G. & Lafratta, G. (2002). Anisotropic spatial sampling designs for urban pollution. *J. Roy. Stat. Soc. Ser. C ,* **51**, 223-234.

Baillargeon, S. & Rivest, L.P. (2009). A general algorithm for univariate stratification. *Int. Stat. Rev.*, **77**, 331-344.

Baillargeon, S. & Rivest, L.P. (2011). The construction of stratified designs in R with the package stratification. *Surv. Methodol.,* **37**, 53–65.

Barabesi, L. & Franceschi, S. (2011). Sampling properties of spatial total estimators under tessellation stratified designs. *Environmetrics*, **22**, 271–278.

Barabesi, L. & Marcheselli, M. (2008). Improved strategies for coverage estimation by using replicated line-intercept sampling. *Environ. Ecol. Stat.*, **15**, 215–239.

Benedetti, R. & Palma, D. (1995). Optimal sampling designs for dependent spatial units. *Environmetrics*, **6**, 101-114.

Benedetti, R., Piersimoni, F. & Postiglione, P. (2015). *Sampling spatial units for agricultural surveys*. Advances in Spatial Science Series. Berlin Heidelberg: Springer.

Benedetti, R., Piersimoni, F., & Postiglione P (2016). Advanced methods to design samples for land use/land cover surveys. In: *Topics on methodological and applied statistical inference*, T. Di Battista, E. Moreno & W. Racugno W (Eds), pp. 31-42. Springer, Heidelberg, Berlin.

Bivand, R. & Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *J. Stat. Softw.*, **63**, 1-36. URL http://www.jstatsoft.org/v63/i18/.

Bohorquez, M., Giraldo, R. & Mateu, J. (2016). Optimal sampling for spatial prediction of functional data. *Stat. Methods Appl.,* **25**, 39. doi:10.1007/s10260-015-0340-9.

Bondesson, L. & Grafström, A. (2011). An extension of Sampford's method for unequal probability sampling. *Scand. J. Stat.*, **38**, 377-392.

Bondesson, L. & Thorburn, D. (2008). A list sequential sampling method suitable for real-time sampling. *Scand. J. Stat.,* **35**, 466-483.

Breidt, F. J. & Chauvet, G. (2012). Penalized balanced sampling. *Biometrika*, **99**, 945–958.

Chambers, R.L., Steel, D.G., Wang, S. & Welsh, A. (2012). *Maximum likelihood estimation for sample surveys*. Boca Raton, USA: Chapman & Hall/CRC.

Chauvet, G. (2009). Stratified balanced sampling. *Surv. Methodol.,* **35**, 115–119.

Chauvet, G. & Tillé, Y. (2006). A fast algorithm of balanced sampling. *Comput. Statist.,* **21**, 53-62.

Chauvet, G., Bonnéry, D. & Deville, J.C. (2011). Optimal inclusion probabilities for balanced sampling. *J. Statist. Plann. Inference,* **141**, 984–994.

Christman, M.C. (2000). A review of quadrat-based sampling of rare, geographically clustered populations. *J. Agric. Biol. Environ. Stat.*, **5**, 168–201.

Dalenius, T., Hájek, J. & Zubrzycki, S. (1961). On plane sampling and related geometrical problems. *Proceedings of the 4th Berkeley Symposium on Probability and Mathematical Statistics*, pp. 125-150.

Deville, J.C. & Tillé, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika,* **85**, 89–101.

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

Deville, J.C. & Tillé, Y. (2005). Variance approximation under balanced sampling. *J. Statist. Plann. Inference,* **128**, 411–425.

Dickson, M.M. & Tillé, Y. (2016). Ordered spatial sampling by means of the traveling salesman problem. *Comput. Stat.,* doi:10.1007/s00180-015-0635-1.

Fattorini, L. (2006). Applying the Horvitz–Thompson criterion in complex designs: a computer-intensive perspective for estimating inclusion probabilities. *Biometrika,* **93**, 269–278.

Grafström, A. (2012). Spatially correlated Poisson sampling. *J. Statist. Plann. Inference,* **142**, 139–147.

Grafström, A. & Lundström, N.L.P. (2013). Why well spread probability samples are balanced. *Open J. Stat.*, **3**, 36–41.

Grafström, A., Lundström, N.L.P. & Schelin, L. (2012). Spatially balanced sampling through the pivotal method. *Biometrics,* **68**, 514-520.

Grafström, A. & Lisic, J (2016). BalancedSampling: Balanced and spatially balanced sampling. R package version 1.5.2. https://CRAN.R-project.org/package=BalancedSampling.

Grafström, A., Saarela, S., Ene, L.T. (2014). Efficient sampling strategies for forest inventories by spreading the sample in auxiliary space. *Can. J. For. Res.*, **44**, 1156-1164.

Grafström, A. & Schelin, L. (2014). How to select representative samples. *Scand. J. Stat.,* **41**, 277-290.

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

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

Hastie, T. & Tibshirani, R. (1990). *Generalized additive models*. London: Chapman and Hall.

Hedayat, A., Rao, C.R. & Stufken, J. (1988a). Sampling designs excluding contiguous units. *J. Statist. Plann. Inference,* **19**, 159–170.

Hedayat, A., Rao, C.R. & Stufken, J. (1988b). Designs for survey sampling avoiding contiguous units. *Handbook of statistics vol. 6: Sampling*, Eds. P.R. Krishnaiah, C.R. Rao, pp. 575–583. The Netherlands: Elsevier.

Hedayat, A. & Stufken, J. (1998). Sampling designs to control selection probabilities of contiguous units. *J. Statist. Plann. Inference,* **72**, 333–345.

Isaki, C.T. & Fuller, W.A. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.*, **77**, 89–96.

Kincaid, T. M. & Olsen, A. R. (2016). spsurvey: Spatial survey design and analysis. R package version 3.3.

Lumley, T. (2014). survey: analysis of complex survey samples. R package version 3.30.

Mandal, B., Parsad, R. & Gupta, V. (2008). IPPS Sampling plans excluding adjacent units. *Commun. Stat. – Theory,* **3**, 2532–2550.

Minasny, B., McBratney, A.B. & Walvoort D.J.J. (2007). The variance quadtree algorithm: Use for spatial sampling design. *Comput Geosci*, **33**, 383-392.

Müller, W.G. (2007). *Collecting spatial data. Optimum design of experiments for random fields*. Berlin Heidelberg: Springer-Verlag.

Okabe, A., Boots, B., Suguhara, K. & Chiu, S.N. (2000). *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams, Second Edition*. Chichester: Wiley.

Olea, R.A. (1984). Sampling design optimization for spatial functions. *Math. Geol.,* **16**, 369–392.

Overton, W.S. & Stehman, S.V. (1993). Properties of designs for sampling continuous spatial resources from a triangular grid. *Commun. Stat. Theory,* **22**, 2641–2660.

Pfeffermann, D. & Rao, C.R. (2009). *Handbook of Statistics, vol. 29B*. The Netherlands: Elsevier.

Rao, J.N.K & Molina, I. (2015). Small area estimation. New Jersey, USA: John Wiley & Sons.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rogerson, P. & Delmelle, E. (2004). Optimal sampling design for variables with varying spatial importance. *Geog. Anal.,* **36**, 177–194.

Särndal, C.E., B. Swensson, & Wretman, J. (1992). *Model assisted survey sampling*. New York: Springer.

Stevens, D.L. Jr (1997). Variable density grid-based sampling designs for continuous spatial population. *Environmetrics,* **8**, 167-195.

Stevens, D.L. Jr & Olsen, A.R. (2003). Variance estimation for spatially balanced samples of environmental resources. *Environmetrics,* **14**, 593–610.

Stevens, D.L. Jr & Olsen A.R. (2004). Spatially balanced sampling of natural resources. *J. Amer. Statist. Assoc.*, **99**, 262–278.

Stufken, J. (1993). Combinatorial and statistical aspects of sampling plans to avoid the selection of adjacent units. *J. Combin. Inform. System Sci.*, **18**, 81–92.

Stufken, J., Song, S.Y., See, K. & Driessel, K.R. (1999). Polygonal designs: some existence and non-existence results. *J. Statist. Plann. Inference,* **77**, 155–166.

Thompson, S.K. (2013). *Sampling. 3rd edition*. Hoboken, New Jersey: John Wiley and Sons, Inc.

Tillé, Y. (2006). *Sampling algorithms*. Springer series in statistics. New York: Springer.

Tillé, Y. (2011). Ten years of balanced sampling with the cube method: An appraisal. *Surv. Methodol.,* **37**, 215–226.

Tillé, Y. & Matei, A. (2015). sampling: Survey sampling. R package version 2.7. https://CRAN.R-project.org/package=sampling

Valliant, R., Dorfman, A.H., & Royall, R.M. (2000). *Finite population sampling and inference: a prediction approach*. New York: John Wiley & Sons, Inc.

Vallée, A.-A., Ferland-Raymond, B., Rivest, L.-P., & Tillé, Y. (2015). Incorporating spatial and operational constraints in the sampling designs for forest inventories. *Environmetrics,* **26**: 557–570. doi: 10.1002/env.2366.

Wang, J.F., Stein, A., Gao, B.B. & Ge, Y. (2012). A review of spatial sampling. *Spat. Stat.,* **2**, 1-14.

Wright, J. & Stufken, J. (2008). New balanced sampling plans excluding adjacent units. *J. Statist. Plann. Inference,* **138**, 3326–3335.

Wright, K. (2015). agridat: Agricultural Datasets. R package version 1.12. https://CRAN.R-project.org/package=agridat

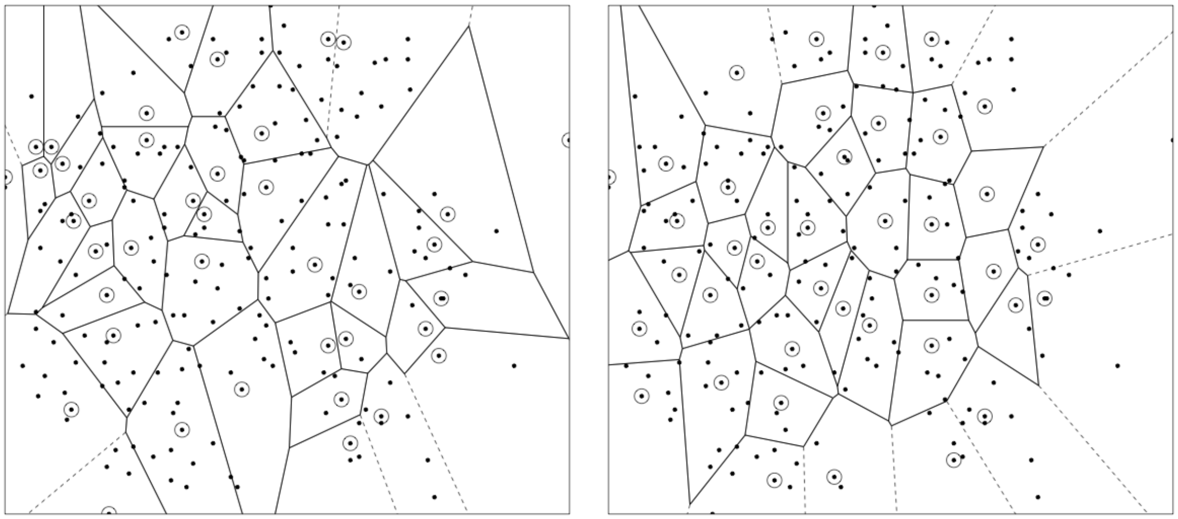

Fig. 1: Selected samples, with Voronoi polygons, from Baltimore population with n=40, SRS (*left*) and GRTS (*right*) where the axes labels are the geographical coordinates (x,y)

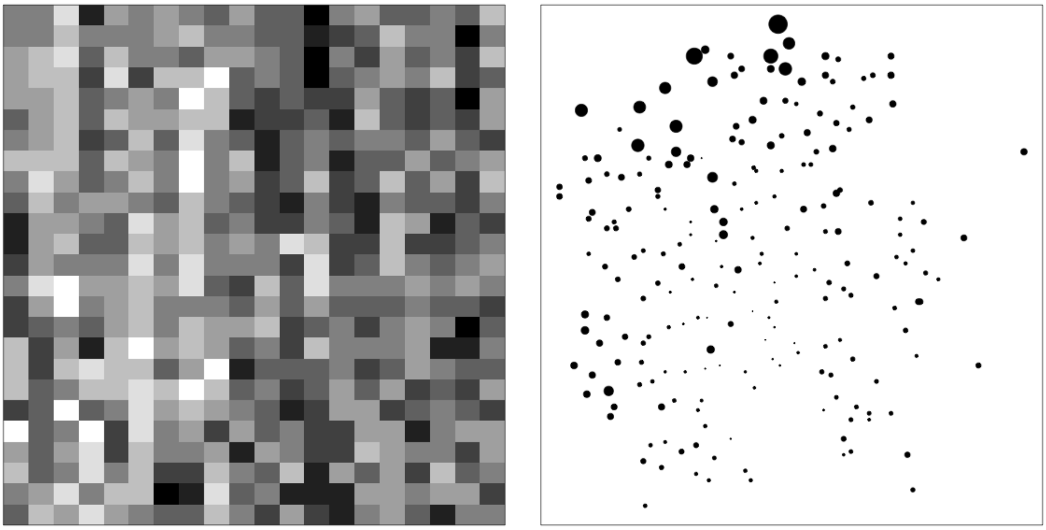

Fig. 2: Spatial distribution of the variable grain in Mercer-Hall dataset (*left*) and of the variable price in Baltimore dataset (*right*) where the axes labels are the geographical coordinates (*x*,*y*)

Design	Mercer-Hall			Baltimore		
	n	MSE/MSE_{SRS}	μ_{SBI}	*n*	MSE/MSE_{SRS}	μ_{SBI}
GRTS	10	0.54	0.14	10	0.58	0.16
CUBE 1	10	0.29	0.20	10	0.30	0.22
CUBE 2	10	0.32	0.16	10	0.35	0.17
DUST 1	10	0.57	0.22	10	0.80	0.30
DUST 2	10	0.59	0.21	10	0.76	0.28
SCPS	10	0.39	0.08	10	0.38	0.10
LPM 1	10	0.44	0.10	10	0.45	0.12
LPM 2	10	0.44	0.10	10	0.45	0.12
DBSS 1	10	0.34	0.09	10	0.36	0.11
DBSS 2	10	0.34	0.09	10	0.38	0.11
GRTS	50	0.28	0.10	25	0.41	0.15
CUBE 1	50	0.14	0.25	25	0.20	0.27
CUBE 2	50	0.15	0.23	25	0.24	0.24
DUST 1	50	0.32	0.92	25	1.70	0.79
DUST 2	50	0.32	0.78	25	1.55	0.72
SCPS	50	0.16	0.07	25	0.26	0.12
LPM 1	50	0.23	0.08	25	0.33	0.12
LPM 2	50	0.23	0.08	25	0.33	0.12
DBSS 1	50	0.16	0.07	25	0.26	0.12
DBSS 2	50	0.16	0.07	25	0.28	0.12
GRTS	100	0.23	0.12	50	0.35	0.17
CUBE 1	100	0.10	0.23	50	0.15	0.29
CUBE 2	100	0.11	0.21	50	0.18	0.26
DUST 1	100	0.34	0.87	50	2.86	0.89
DUST 2	100	0.34	0.90	50	2.8	0.97
SCPS	100	0.10	0.10	50	0.21	0.14
LPM 1	100	0.18	0.10	50	0.27	0.14
LPM 2	100	0.18	0.10	50	0.27	0.14
DBSS 1	100	0.12	0.10	50	0.21	0.14
DBSS 2	100	0.12	0.09	50	0.22	0.14

Tab. 1: Relative efficiency of the sample mean (MSE/MSE_{SRS}) and mean (μ_{SBI}) of the spatial balance indices for each design estimated in 10,000 replicated samples in the Mercer –Hall (variable wheat) and in the Baltimore dataset (variable house price) for different sample sizes.