DEM Discussion Papers

Department of Economics and Management

2015/06

Ordered Spatial Sampling by Means of the Traveling Salesman Problem

Maria Michela Dickson and Yves Tillé

ISSN 2282-2801 DEM discussion papers [online]

UNIVERSITÀ DEGLI STUDI DI TRENTO

DEM DISCUSSION PAPER

ISSN 2282-2801 DEM Discussion Papers [online]
Università degli Studi di Trento

Department of Economics and Management, University of Trento, Italy.

Editors

Luciano ANDREOZZI luciano.andreozzi@unitn.it

Roberto GABRIELE roberto.gabriele@unitn.it

Technical officer

Marco TECILLA marco.tecilla@unitn.it

Guidelines for authors

Papers may be written in Italian or in English. Faculty members of the Department must submit to one of the editors in pdf format. Management papers should be submitted to R. Gabriele. Economics Papers should be submitted to L. Andreozzi. External members should indicate an internal faculty member that acts as a referee of the paper.

Typesetting rules:

1. papers must contain a first page with title, authors, abstract, keywords and codes. Page numbering starts from the following page;
2. a template is available upon request from the managing editors.

Ordered Spatial Sampling by Means of the Traveling Salesman Problem

Maria Michela Dickson[a] and Yves Tillé[b]

[a] *Department of Economics and Management, University of Trento, Via Inama 5, 38122,Trento, Italy*

[b] *Institute of Statistics, University of Neuchatel, Avenue de Bellevaux 51, 2000, Neuchâtel, Suisse*

Abstract: In recent years, spatial sampling has been the subject of a flourishing literature. Its use had become widespread due to the availability of topographical information about statistical units, especially in the environmental context. New algorithms enable us to take advantage of spatial locations directly. In this paper, we present a new way of using spatial information by using traditional sampling techniques as systematic sampling. By means of a famous optimization method, the Traveling Salesman Problem, it is possible to order the statistical units in a way that preserves the spatial correlation. Next ordered sampling methods are applied on the statistical units. Therefore we can render spatial some non-spatial methods. An economic application on real data is presented and different spatial and non-spatial methods are tested. Results are compared in terms of variance estimation and spatial balance, in order to establish the possibility of *spatializing* traditional sampling methods and of implementing them on data of different nature, among which economic ones.

***Keywords*:** sampling methods, TSP, variance estimation, spatial balance.

Correspondence to: Maria Michela Dickson, Department of Economics and Management, University of Trento, Via Inama 5, 38122, Trento, Italy. e-mail: mariamichela.dickson@unitn.it

1. Introduction

Frequently data-sets contain information about geographical position of the units in the space. Geo-referenced data are interesting in survey sampling theory in the sense that one can take advantage of the information about the localization of the units to improve the efficiency of the sampling design (Arbia 1992; Grafström and Tillé 2013; Dickson *et al*. 2014). Spatial sampling can be an efficient way of selecting samples in various contexts, such as, economic, social and environmental ones.

In the last years, a large set of new methods of spatial sampling have been proposed (Stevens and Olsen 2004; Grafström *et al*. 2011; Grafström 2012; Breidt and Chauvet 2012). Some of these methods are based on algorithms that compute distances between units in the selection process (Grafström 2012; Grafström *et al*. 2011). Another family of methods consists of selecting the units to include in a sample, according to a given order of units. The problem then consists of mapping the two-dimensional space into one-dimensional space while safeguarding the autocorrelation as in the Generalized Random Tessellation Stratified (GRTS) sampling method proposed by Stevens and Olsen (2004). The scope of this work consists of using the algorithm of the Traveling Salesman Problem (TSP) to order the units. We will show that it is an efficient way of mapping the two-dimensional space into one-dimensional space because it conserves a large part of the spatial information between the units.

Since the seminal work of Menger (1932), a large number of solutions for the TSP have been proposed. In the present work, the method is used in combination with classic sampling methods, such as systematic sampling (Madow 1949), ordered pivotal sampling (Deville 1998) and the sequential cube method (Deville and Tillé 2004; Chauvet and Tillé 2006), in a study on economic data. These three methods have been chosen because they utilize the order of units in the selection of the sample. The usual random order used for these methods is substituted with the order given by a TSP algorithm.

Moreover, this new family of ordered method is compared with two ordered spatial sampling methods, such as GRTS and the cube method. With the cube method, the sample is balanced on geographical coordinates, with random order of units. For all the designs, the Horvitz–Thompson (HT) estimator (Horvitz and Thompson 1952) of the population total is computed and the performances of the methods are evaluated in terms of variance of the HT estimator. The degree of spatial balance of each sampling method is also evaluated.

The structure of the paper is the following. In Section 2, the TSP problem is presented. Moreover several spatial and non-spatial sampling methods are defined. In Section 3, the use of TSP, in combination with sampling methods, is developed with a graphic comparison of the sample selections. Section 4 contains the results of the simulation study. Finally, some conclusions and directions for further studies in the field are given in Section 5.

2. The statistical methodological framework

As above-mentioned, the aim of this paper is to show that there exists a family of spatial sampling methods based on the Traveling Salesman Problem. To achieve this objective it is essential to use an ordered sampling method that is possible to combine with a TSP algorithm. Brewer and Hanif

(1983) proposed 20 (in a list of 50) different sampling methods which are 'exact', in sense that they are without replacement, they have a fixed sample size, they exactly satisfy the given inclusion probabilities, and they are applicable to any vector of inclusion probabilities. Here two classical sampling methods of this category are presented: ordered systematic sampling and ordered pivotal sampling. These methods can be combined with a TSP. Furthermore, we briefly present the Generalized Random Tessellation Stratified method (GRTS) and the cube method. The GRTS is a largely used spatial sampling method. The cube method enables us to select balanced samples and can also be adapted for spatial sampling.

2.1 The Traveling Salesman Problem

Since the seminal work in combinatorial optimization by Dantzig *et al*. (1954) and to arrive to the work by Lawler *et al*. (1985), which provides the state of the art description on the topic, the Traveling Salesman Problem (TSP) has known significant developments. The TSP consists of finding the best way of visiting a prescribed set of cities, starting from a given city (that can be selected randomly) and to return to it. The way must be the fastest in the sense that the total distance covered must be the smallest possible. In this sense, the TSP represents a typical 'hard' combinatorial optimization problem. This problem finds applications in several fields of research, e.g. vehicle routing, clustering of data arrays, machine scheduling (Lenstra and Kan 1975).

There is no univocal solution to the TSP without examining all the possible paths. So it is not possible to find an optimal path, but only groups of solutions which are close to the optimality.

The problem can be defined in terms of graph theory (see among others, Punnen 2002). Let $\mathcal{F}$ be the family of all Hamiltonian cycles (Kirkman 1856; Hamilton 1858) in a graph $G = (V, E)$, where V is a set of vertices (*nodes*) and E is a set of unordered *edges*. Each edge $e \in E$ has a weight c_e that is the distance between the nodes. The TSP problem consists of searching a Hamilton cycle, i.e. a tour in G, such that the sum of the weights of edges is as small as possible. If G is a complete graph and the vertices set $V = \{1, 2, \dots, N\}$, then matrix $D = \left(d_{ij}\right)_{n \times n}$ for the edge which connects nodes i and j in G, is called distance matrix. A distance $d(\dots)$ satisfies the properties:

i) $d\left(X_i, X_j\right) \geq 0$ for all X_i, X_j, (positivity)
ii) $d\left(X_i, X_j\right) = 0 \leftrightarrow X_i = X_i$ (identity);
iii) $d\left(X_i, X_j\right) = d\left(X_j, X_i\right)$ (symmetry);
iv) $d\left(X_i, X_j\right) \leq d(X_i, X_k) + d\left(X_k, X_j\right)$ for each triad $X_i, X_k, X_j \in A$ (triangular inequality).

A particular case is the Euclidean distance. When the vertices are points $P_i = (X_i, Y_i)$ in a plan, the Euclidian distance is defined by

$$d_{ij} = \sqrt{\left(X_i + X_j\right)^2 + \left(Y_i - Y_j\right)^2}$$

(Matai *et al*. 2010). The goal of TSP is to find a Hamilton cycle with the smallest sum of distances (Hoffman and Wolfe 1985).

Instead, in terms of permutation problems, with the distance matrix $D = [d(i,j)]$, where $i,j = 1,\dots,N$ and diagonal elements $d(i,j) = 0$, a tour is represented by a cyclic permutation π of $\{1,2,\dots,N\}$, where $\pi(i)$ represents the city following city i on the tour. In this case, the TSP consists of finding a permutation π that solves the problem about *'length of the tour'*, expressed as:

$$\min_{\pi} \sum_{i=1}^{N} d\left(i,\pi(i)\right).$$

Note that the TSP can also be applied to non-symmetric matrices (Volgenant 1990; Karp 1979) or to matrices of dissimilarities that do not satisfy the property of triangle inequality (Papadimitriou 1977; Bockenhauer *et al*. 2000). The present work is referred only to TSP with a matrix of distances.

To solve a TSP, it is possible to recur to different algorithms, which are divided in (the list does not want be exhaustive): exact solutions algorithms and heuristics algorithms. The first class includes algorithms of type branch-and-bound (Land and Doig 1960) and branch-and-cut (Grötschel and Holland 1991; Padberg and Rinaldi 1991). Concerning the heuristics algorithms, it is possible to include (Rosenkrantz *et al*. 1977): algorithms based on a tour construction (nearest neighbor algorithm and other insertion variants) and algorithms based on a tour improvement (2- opt, 3- opt). The nearest neighbor algorithm (Rosenkrantz *et al*. 1977) starts with a tour containing randomly chosen city. The algorithm then adds to the last city in the tour the nearest not visited city. It stops when all cities are on the tour. As concern insertion algorithms, they start with a tour consisting of an arbitrary city and choose at each step a city k that is not on the tour. This city is then inserted into the existing tour between two consecutive cities i and j. The algorithm stops when all cities are on the tour. Depending on the way wherein the city is inserted, there are algorithms *nearest insertion* (city k is chosen as the nearer to any other city on the tour), *farthest insertion* (city k is chosen as the farther to any other city on the tour), *cheapest insertion* (city k is chosen depending on the cost of insertion) and *arbitrary insertion* (city k is chosen randomly from all cities not on the tour).

The algorithms based on a tour improvement are particularly interesting. In optimization, a 2-opt algorithm (Croes 1958) is a simple local search algorithm. A complete 2-opt local search will compare every possible valid combination of the swapping mechanism. It solves the TSP taking a route that crosses over itself and reorder it, so that there is no more crossing. This procedure systematically exchanges two edges in the graph represented by the distance matrix, till no improvements are possible. The resulting tour is a cycle containing all nodes on the graph, minimizing the total length of the cycle, called 2-optimal. Similarly, a 3-opt algorithm (Lin 1965) involves three edges in a tour. Due to its diffusion, feasibility in computation terms and to availability of its use also to solve large problem sizes, a 2-optimal algorithm has been used in simulation presented in Section 3.

2.2 Sampling Methods

2.2.1 Population, sample and sampling design

Consider a population $U = \{1, \dots, k, \dots N\}$ of size N. A sample s is a subset of the population. A sampling design is a probability measure on the set of all the possible samples $p(s) \geq 0$ for all $s \subset U$ with

$$\sum_{s \subset U} p(s) = 1.$$

The random sample S is a random subset of U whose the distribution of probability is given by the sampling design $P(S = s) = p(s), s \subset U$. The inclusion probability π_k of unit k is the probability of selecting this unit in the random sample and can be derived from the sampling design

$$\pi_k = \sum_{s \ni k} p(s), k \in U.$$

Now consider a variable of interest y whose the value taken on unit k is denoted by y_k. The total

$$Y = \sum_{k \in U} y_k$$

can be unbiasedly estimated by the Horvitz-Thompson estimator

$$Y = \sum_{k \in S} \frac{y_k}{\pi_k},$$

provided that $\pi_k > 0$, for all $k \in U$.
There exists a large number of sampling designs with unequal inclusion probabilities. In the next two sections, we present the ordered systematic sampling and the ordered pivotal method.

2.2.2 Ordered Systematic Sampling

Systematic sampling (Madow 1949) is one of the most used sampling methods. Let $0 < \pi_k < 1$ the inclusion probabilities of units $k_i, i = 1, \dots, n$, with $\sum_{k \in U} \pi_k = n$, then $V_k = \sum_{l=1}^{k} \pi_l$, with $V_o = 0$ and $V_N = n$ (Tillé 2006). The method proceeds in this way. First, generate a uniform variable u in $[0,1]$. Next unit k_1 is selected, if $V_{k_1 - 1} \leq u < V_{k_1}$. The second unit k_2 is selected, if $V_{k_2 - 1} \leq u + 1 < V_{k_2}$. In fine, the jth unit is selected if

$$V_{k_{j-1}} \leq u + j - 1 < V_{k_j} \; . \quad (1.1)$$

In ordered systematic sampling, a lot of joint inclusion probabilities could be equal to zero. They can be computed by using:

$$V_{kl} = \begin{cases} \sum_{i=k}^{l-1} \pi_i , & if\ k < l \\ \sum_{i=k}^{N} \pi_i + \sum_{i=1}^{l-1} \pi_i = n - \sum_{i=l}^{k-1} \pi_i , & if\ k > l. \end{cases}$$

Next the joint inclusion probabilities are given by

$$\pi_{kl} = \min\{\max(0, \pi_k - \delta_{kl}), \pi_l\} + \min\{\pi_k, \max(0, \delta_{kl} + \pi_l - 1)\}, \quad k < l,$$

where $\delta_{kl} = V_{kl} - \lfloor V_{kl} \rfloor$ (see Tillé 2006).

2.2.3 *Ordered Pivotal Sampling*

Ordered pivotal method, also called Deville's systematic sampling (Deville 1998) is a particular systematic method. This technique selects only one unit in each interval $[i-1, i[$, with $i = 1, \ldots, n$ and the V_k is constructed according to (1.1). The method generates n non independent random variables u_i, $i = 1, \ldots, n$ with uniform distribution $u_i \sim U[0,1[$. For each variable i, the unit k is selected if

$$V_{k-1} \leq u_i + (i-1) < V_k.$$

The main idea of Deville is to introduce a dependency between the u_i in such a way that each unit can only be selected once in the sample.

Chauvet (2012) have shown that the Deville systematic sampling is the same as the ordered pivotal method. The pivotal method has been proposed by Deville and Tillé (1998). At each step of the pivotal method, only two units are considered. One of these two units is either definitively selected in the sample or definitively excluded from the sample. If these two units are selected sequentially the method is called the sequential pivotal method.

As systematic sampling, ordered pivotal method is an 'ordered' method of sampling in the sense that the selection of consecutive units is avoided. This method is particularly efficient when the neighboring units are similar.

2.3 Spatial Sampling Methods

2.3.1 Generalized Random-Tessellation Stratified Method

The Generalized Random-Tessellation Stratified Method (*GRTS*, Stevens and Olsen 2004) is one of the most popular methods in spatial sampling. Essentially, it assigns to the population units a defined order, according to a recursive hierarchical randomization process which preserves the spatial relationships of the sample units. Then, the units are arranged in order and mapped from two- or multi- dimensional space in one-dimensional space, through a quadrant recursive function

(Mark 1990).
The samples are then selected in one dimension, using systematic πps sampling and then mapped back in the original two- or multi- dimension, preserving the spatial order. GRTS design ensures that drawn samples are much more evenly distributed over space than ordinary probability designs. It uses an inclusion probability function, defined *intensity function* $\pi(s)$, which has the usual finite population sampling interpretation for discrete populations (Stevens and Olsen 2004).
The GRTS has been the most used probability method in spatial sampling, because it can manage linear, areal or not contiguous data. It is important to note that the mapping used in GRTS is not perfect, due to impossibility to exclude that units originally near in distance are mapped in the one-dimensional space far apart one from each other.

2.3.2 Balanced Sampling: the cube method

Balanced Sampling is a popular method to select units from a finite population, particularly in the case where the auxiliary variables are correlated with the variable of interest. Let $\boldsymbol{x}_k$ be a vector of values of p auxiliary variables. This vector is supposed to be known for each unit of the population. A sample is said to be balanced if the balancing equations given by

$$\sum_{k\in S}\frac{\boldsymbol{x}_k}{\pi_k}=\sum_{k\in U}\boldsymbol{x}_k$$

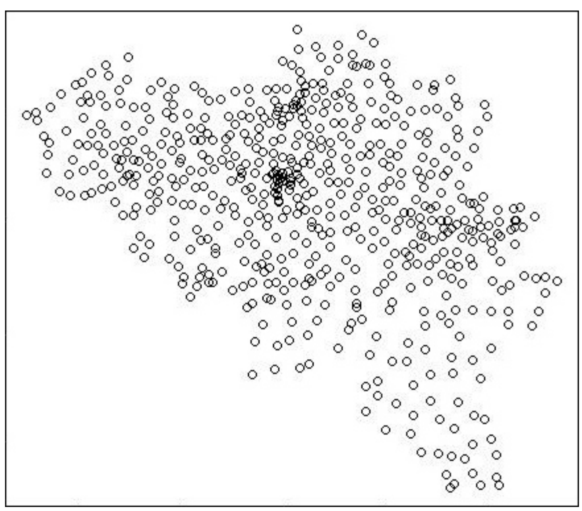
is satisfied at least approximately.
Some authors have proposed methods to select balanced samples (Yates 1949; Deville *et al*. 1988). The cube method (Deville and Tillé 2004) provides a general solution that can be applied to any set of equal or unequal inclusion probabilities with a large number of auxiliary variables.
The cube method thanks its name to a geometric representation of a sampling design. Indeed each sample can be represented as a vector of 0 and 1 and is thus a vertex of an hypercube in an N-dimensional space. This method enables the selection of balanced samples while satisfying predefined inclusion probabilities. The idea of the algorithm consists of transforming randomly at each step the vector of inclusion probabilities π while satisfying the balancing equations (Deville and Tillé 2004).
The cube method is structured in two phases: a *flight phase* and a *landing phase*. The flight phase is composed of a random walk that starts from the vector of inclusion probabilities and remains inside the intersection between the cube and the constraint subspace given by the balancing equations. This procedure stops when it arrives at a vertex of this intersection. If the vertex of this intersection is a sample, then this sample is selected. If this vertex is not a sample, it means that the balancing equations can only be approximately satisfied. In this case, the landing phase is applied and the sample is selected as near as possible to the constraint subspace (Deville and Tillé 2004).
In this paper the cube method has been applied in the case of ordered sampling, obtained by TSP, and as a spatial sampling method. In both cases, the geographical coordinates of units have been used as balancing variables. The cube method has been implemented on TSP path by using a sequential algorithm (called *sequential cube method*), which is computationally very fast (Chauvet and Tillé 2006).

3. Ordered Spatial Sampling using TSP: the case of Belgian municipalities

3.1 Data

The dataset used in the following example is composed by 589 observations, corresponding to the Belgian municipalities, and by 44 economic, social and demographic variables. For all the municipalities the geographical coordinates based on centroids of communes are available. The data are updated to December, 31, 2011. The map of Belgian municipalities is available in **Figure 1**. The country is divided in three regions: the Flemish Region (north), the Walloon Region (south) and the Brussels-Capital Region (center-north). The first Region is one of the most populated of Europe in terms of number of inhabitants. The Brussels-Capital Region is very small with 19 municipalities, visible as the agglomeration in the middle of the map. Instead, the Walloon Region is characterized by a larger municipalities in area.

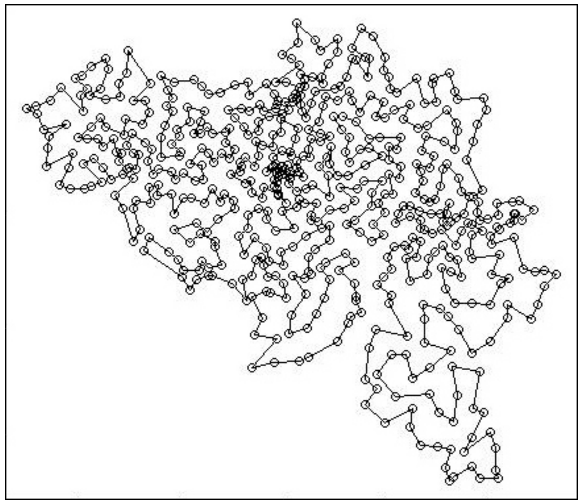
Figure 1. Map of the centroids of the Belgian municipalities.



3.2 A comparison between sampling with and without TSP

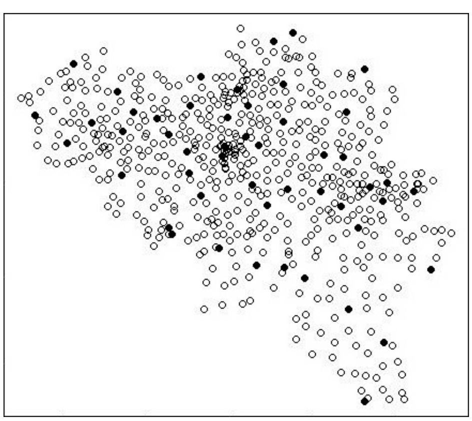
The sampling design consists of selecting samples of $n = 50$ among the 589 municipalities with equal inclusion probabilities based on the TSP. In what follows, all generated samples have been obtained by using the same random seed, so that all realizations are directly comparable and differences between them can be ascribed only to differences in the sampling methods used. By using the R *package* ‘TSP’ (Hahsler and Hornik 2007) the TSP can be solved for the 589 Belgian municipalities. The method used to solve the TSP is a 2-opt procedure (Croes 1958). **Figure 2** presents the best solution we have found for the Belgian municipalities.

Figure 2. Best path computed on Belgian municipalities using 2-opt method.

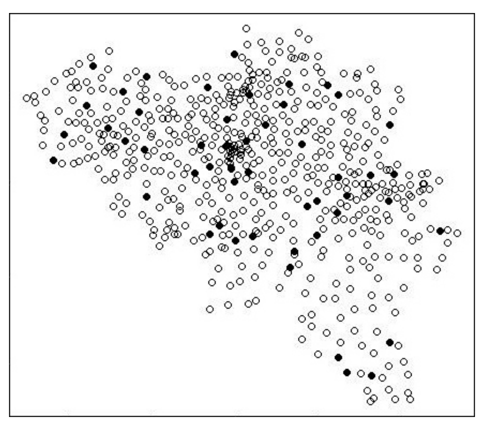


Once the unit of the population is ordered by the TSP, it is now possible draw samples of fixed size $n = 50$ with equal inclusion probabilities with an ordered sampling method. Samples are then selected by means of ordered systematic sampling, ordered pivotal sampling and ordered cube method, all implemented by using R *package* 'sampling' (Tillé and Matei 2005). The samples have been selected 10000 times. In **Figure 3**, different samples obtained are presented.

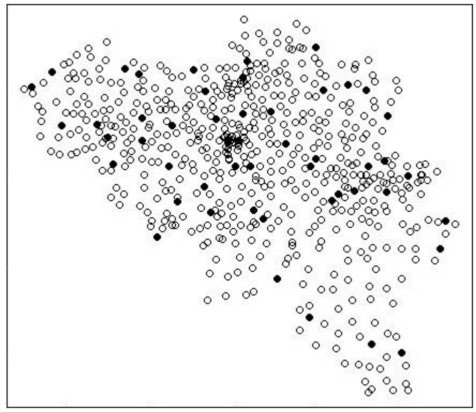
Figure 3. Examples of samples of $n = 50$ drawn with equal inclusion probabilities by ordered systematic sampling (a), ordered pivotal sampling (b) and ordered cube method (c) on a TSP path.



(a)



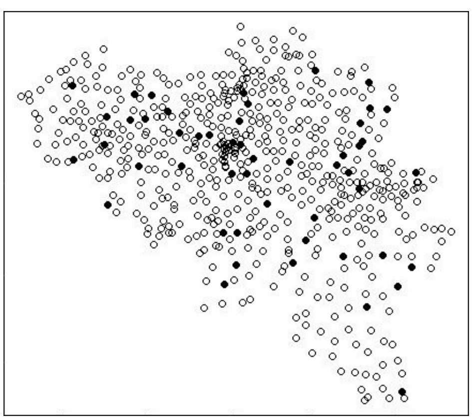
(b)



(c)

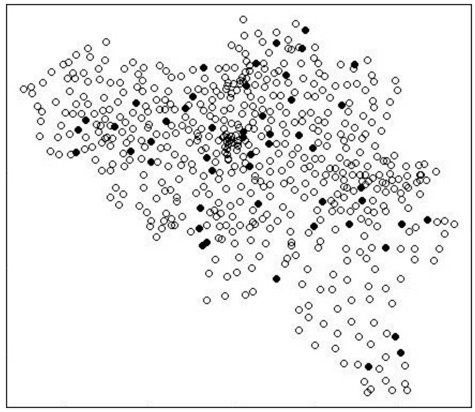
In order to compare the samples drawn with the order given by TSP, samples are also drawn with equal inclusion probabilities of size $n = 50$ by simple random sampling without replacement (SRSWOR) and by spatial sampling designs GRTS and cube method (**Figure 4**).

Figure 4. Samples of $n = 50$ drawn by SRSWOR (a), GRTS (b) and cube method (c).



(a)

(b)



(c)

As we can see, the samples are spread very differently in the space according to the sampling designs. In the case of methods constructed by using TSP, the algorithm orders the units in order to minimize the length of the tour. Then, the sampling methods have been applied on these ordered units. So that, all the methods used in combination with TSP select units according to the same order of units. For the other methods, the GRTS method and the cube method balanced on coordinates, the units are selected on a randomly sorted population. The selection is only based on the geographical positions of the units in the space.

4. An application on economic data

As said before, the dataset used in this paper is composed of many variables on the Belgian municipalities that have been evaluated in an economic perspective. Here, we take into account only four variables: the medium income of population ("*MediumIncome*"), the number of inhabitants ("*Residents*"), the number of migrants ("*Migrants*") and the total tax paid by population ("*Tax*").
To conduct an evaluation of different sampling methods used to select the samples, the variable *MediumIncome* has been estimated by drawing samples of size 50, 100 and 200, with the sampling methods presented before. The Horvitz-Thompson estimator has been used to estimate the total population and the variance estimation of the HT estimator has been calculated using the Deville's method (Deville 1993).
In **Table 1**, the values of the variance are approximated by simulations for the different sampling methods. Furthermore, in the same table, the values of the approximated variance are compared for the different methods:

- SRSWOR (simple random sampling without replacement),
- OSS TSP (ordered systematic sampling through TSP),
- OPS TSP (ordered pivotal sampling through TSP),
- OCM TSP (ordered cube method through TSP),
- GRTS (the GRTS method) and
- CM (cube method balanced on coordinates).

Simulations are only run for equal inclusion probabilities, because the variable of interest is strongly correlated with the other available variables. The samples have been all spread on the geographic coordinates, distinguishing between methods that use the order of units given by TSP, and spatial methods where the order is chosen randomly, according to the geographic position in the space.

Table 1. Results of 10000 simulations with equal inclusion probabilities for the variable *MediumIncome*.

Variance approximated by the simulations

Design	n=50	n=100	n=200
SRSWOR	1.91887e+16	1.42511e+16	8.96149e+15
OSS TSP	1.67055e+16	5.19828e+15	2.29457e+15
OPS TSP	1.30683e+16	5.14986e+15	2.62426e+15
OCM TSP	9.96893e+15	5.73177e+15	2.59631e+15
GRTS	1.66605e+16	6.32101e+15	2.44191e+15
CM	1.51785e+16	1.04383e+16	8.2236e+15

Variance approximated by the simulations in relation to SRSWOR (%)

Design	n=50	n=100	n=200
OSS TSP	87.06	36.48	25.60
OPS TSP	68.10	36.14	29.28
OCM TSP	51.95	40.22	28.97
GRTS	86.82	44.35	27.25
CM	79.11	73.25	91.77

Results of simulations show that all the sampling methods have smaller estimated variance with respect to simple random sampling. This result confirms the one expressed in Dickson *et al*. (2014) about the concept that spatial sampling methods are better in terms of estimation with respect to non-spatial designs. The use of TSP to *spatialize* the non-spatial designs gives a good improvement and makes these methods comparable or better than GRTS, which can be considered as a benchmark in spatial sampling.

When the sample size increases, the methods that use the order given by TSP are more and more efficient. In this case, the effect of spreading increases in function of the sample size, with respect to SRSWOR. In the case of cube method, the use of TSP causes an important improvement in the variance estimation with respect to the use of the same method only balanced on the coordinates.

The quality of spreading of the methods can also be evaluated by using the Voronoi polygons approach developed by Grafström *et al*. (2012). This index of spatial balanced can easily be computed by means of the R *package* 'BalancedSampling' (Grafström 2014). Results are shown in **Table 2.**

Table 2. Index of spatial balance for the different sampling designs. The designs with the smallest indices are better balanced.

Design	n=50	n=100	n=200
SRSWOR	0.3386	0.3177	0.2993
OSS TSP	0.0866	0.0939	0.1235
OPS TSP	0.2493	0.1273	0.1547
CM TSP	0.1642	0.1641	0.1883
GRTS	0.1355	0.1363	0.1640
CM	0.3003	0.2983	0.2902

All the methods under examination produce samples much more spatially balanced than non-spatial designs, such as SRSWOR. All the methods using TSP have a small index of spatial balance, with the best results obtained with ordered systematic sampling, especially with a small sample size. GRTS method has been confirmed as a good method also in terms of spatial balance, which slightly decreases when the sample size increases. The difference obtained in the case of the two applications of the cube method is also very interesting. Indeed, the ordered TSP version is much better than the version balanced on the geographic coordinates.

4. Discussion

The use of Traveling Salesman Problem in the context of spatial sampling gives an important advancement to this field of research. The simulations have shown that in most cases the efficiency is better than for GRTS. The use of TSP to render spatial some classical non-spatial sampling methods is an important innovation, because it gives the possibility to implement several sequential methods.

In the economic application presented, it has been showed that this new family of spatial methods is very functional to draw samples, both in terms of variance estimation, and consequently of efficiency in estimation, and of spatial balance. Samples drawn with *TSP family methods* showed an important reduction in the variance and enthusiastic results in terms of spatial balance, such to compete with GRTS, which is probably the most used spatial sampling method.

References

Arbia, G. (1993) The use of GIS in spatial statistical surveys, *International Statistical Review*, **61**, 339-359.

Bockenhauer, H.-J, Hromkovic, J., Klasing, R., Seibert, S., and Unger, W. (2000) An improved lower bound on the approximability of metric TSP and approximation algorithms for the TSP with sharpened triangle in3quality, in *Proceedings of 17th Symposium on Theoretical Aspects of Computer Science, Lecture Notes on Computer Science, 1770,* 382–394. Berlin: Springer-Verlag.

Brewer, K.R.W., Hanif, M. (1983) *Sampling with Unequal Probabilities*, New York: Spinger-Verlag.

Chauvet, G. and Tillé, Y. (2006) A fast algorithm for balanced sampling, *Computational Statistics*, **21**, 53-61.

Chauvet, G. (2012) On a characterization of ordered pivotal sampling, *Bernoulli*, **18**, 1320-1340.

Croes, G.A. (1958) A method for solving traveling salesman problems, Operations Research, **6**, 791-812.

Dantzig, G.B., Fulkerson, D.R., and Johnson, S.M. (1954) Solution of a large scale traveling salesman problem, *Operations Research,* **2**, 393–410.

Deville, J.-C., Grosbras, J.-M., and Roth, N. (1988) Efficient sampling algorithms and balanced sample, *COMPSTAT, Proceeding in Computational Statistics*, R. Payne, P. Green (eds.), Heidelberg: Physica Verlag, pp. 255-266.

Deville, J.-C. (1993) *Estimation de la variance pour les enquêtes en deux phases*, Manuscript, INSEE, Paris.

Deville, J.-C. (1998) Une nouvelle (encore une!) méthode de tirage à probabilités inégales. Tech. rept. 9804. Méthodologie Statistique, INSEE.

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

Deville, J.-C. and Tillé, Y. (2004) Efficient balanced sampling: The cube method, *Biometrika*, **91**, 893-912.

Dickson, M.M., Benedetti, R., Giuliani, D., and Espa, G. (2014) The use of spatial sampling designs in business surveys, *Open Journal of Statistics*, **4**, 345-354.

Grafström, A., Lundström, N.L.P., and Schelin, L. (2011) Spatially balanced sampling through the pivotal method, *Biometrics*, **68**, 514-520.

Grafström, A (2012) Spatially correlated Poisson sampling, *Journal of Statistical Planning and Inference*, **142**, 139-147.

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

Grafström, A. (2014) BalancedSampling: Balanced and spatially balanced sampling. R package version 1.4, http://www.antongrafstrom.se/balancedsampling.

Grötschel, M. and Holland, O. (1991) Solution of large-scale traveling salesman problems, *Mathematical Programming*, **5**, 141-202.

Hahsler, M. and Hornik, K. (2007) TSP – Infrastructure for the Traveling Salesperson Problem, *Journal of Statistical Software*, **23**,1–21.

Hamilton, W. R. (1858) Account of the Icosian Calculus, *Proceedings of the Royal Irish Academy*, **6**, 415–416.

Hoffman, A. and Wolfe, P. (1985) History. In E.L. Lawler, J.K. Lenstra, A.H.G. Rinnoy Kan, D.B. Shmoys (eds.) *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*. Chichester: Wiley, pp. 1-16.

Horvitz, D.G. and Thompson, D.J. (1952) A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association*, **47**, 663-685.

Karp, R.M. (1979) A patching algorithm for the nonsymmetric traveling-salesman problem, *SIAM Journal on Computing,* **8**, 561-573.

Kirkman, T.P. (1856) On the representation of polyhedra, *Philosophical Transactions of the Royal Society, Series A*, **146**, 413-418.

Land, A.H. and Doig, A.G. (1960) An Automatic Method of Solving Discrete Programming Problems, *Econometrica*, **28**, 497-520.

Lawler, E.L., Lenstra, J.K., Rinnooy Kan, A.H.G., and Shmoys, D.B. (eds) (1985) *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*. Chichester: Wiley.

Lenstra, J. and Kan, A.R. (1975) Some simple applications of the traveling salesman problem, *Operational Research Quarterly,* **26**, 717–733.

Lin, S. (1965) Computer solutions of the traveling salesman problem, *Bell System Technical Journal*, **44**, 2245-2269.

Madow, W.G. (1949), On the theory of systematic sampling, *Annals of Mathematical Statistics*, **20**, 333-354.

Mark, D.M. (1990) Neighbor-based properties of some orderings of two-dimensional space, *Geographical Analysis*, **2**, 145-157.

Matai, R., Singh, S.P., and Mittal, M.L. (2010) Traveling Salesman Problem: An Overview of Applications, Formulations, and Solution Approaches. In D. Davendra (eds.) *Traveling Salesman Problem, Theory and Applications*, InTech, pp. 1-24.

Menger, K. (1932) Das botenproblem, *Ergebnisse Eines Mathematischen Kolloquiums,* **2**, 11–12.

Padberg, M. and Rinaldi, G. (1991) A Branch-and-Cut Algorithm for the Resolution of Large-Scale Symmetric Traveling Salesman Problems, *Siam Review*, **33**, 60–100.

Papadimitriou, C. H. (1977) Euclidean TSP is NP-complete, *Theoretical Computer Science*, **4**, 237–244.

Punnen, A. P. (2002) The Traveling Salesman Problem: Applications, Formulations and Variations. In G. Gutin and A.P. Punnen (eds.) *The traveling salesman problem and its variations*, Boston: Kluwer, pp. 1-28.

Rosenkrantz, D.J., Stearns, R.E., and Lewis, P.M. II (1977) An Analysis of Several Heuristics for the Traveling Salesman Problem, *SIAM Journal on Computing*, **6,** 563–581.

Stevens, D. L. and Jr., Olsen, A. R. (2004) Spatially balanced sampling of natural resources, *Journal of the American Statistical Association*, **99**, 262-278.

Tillé, Y. and Matei, A. (2005) *The R package Sampling*, The Comprehensive R Archive Network, http://cran.r-project.org/, Manual of the Contributed Packages.

Tillé, Y. (2006) *Sampling algorithms*, New York: Spinger-Verlag.

Volgenant, A. (1990) Symmetric traveling salesman problems, *European Journal of Operational Research*, **49**, 153–154.

Yates, F. (1949) *Sampling methods for Census and Surveys*. London: Griffin.

List of recent papers published in the same series

2015/05 A spatial analysis of health and pharmaceutical firm survival, Giuseppe Arbia, Giuseppe Espa, Diego Giuliani, Rocco Micciolo

2015/04 Innovation, trade and the size of exporting firms, Letizia Montinari, Massimo Riccaboni, Stefano Schiavo

2015/03 La famiglia fa male all'internazionalizzazione dell'impresa?, Mariasole Bannò, Elisa Pozza, Sandro Trento

2015/02 Bilateral netting and contagion dynamics in Financial networks, Edoardo Gaffeo, Lucio Gobbi

2015/01 The Role of Firm R&D Effort and Collaboration as Mediating Drivers of Innovation Policy Effectiveness, Giovanni Cerulli, Roberto Gabriele, Bianca Potì

2014/10 Jointness in Sites: The Case of Migratory Beekeeping, Luciano Pilati, Vasco Boatto

2014/09 Sustainability vs. credibility of fiscal consolidation. A Principal Components test for the Euro Zone, Giuliana Passamani, Roberto Tamborini, Matteo Tomaselli

2014/08 Fitting Spatial Econometric Models through the Unilateral Approximation, Giuseppe Arbia, Marco Bee, Giuseppe Espa, Flavio Santi.

2014/07 Il modello di scoring del Fondo Centrale di Garanzia: un’analisi d’impatto, Davide Panizzolo

2014/06 Going abroad on regional shoulders: The role of spillovers on the composition of regional exports, Mariasole Bannò, Diego Giuliani, Enrico Zaninotto

2014/05 The role of covenants in bond issue and investment policy. The case of Russian companies, Flavio Bazzana, Anna Zadorozhnaya, Roberto Gabriele

2014/04 Efficiency or Bounded Rationality? Drivers of Firm Diversification Strategies in Vietnam, H. Thu Tran, E. Santarelli, E. Zaninotto

2014/03 Where Gibrat meets Zipf: Scale and Scope of French Firms, M. Bee, M. Riccaboni and S. Schiavo

2014/02 Competition in the banking sector and economic growth: panel-based international evidence, E. Gaffeo and R. Mazzocchi

2014/01 Macroprudential Consolidation Policy in Interbank Networks, E. Gaffeo and M. Molinari

2013/18 Bio-Economics of Allocatable Pollination Services: Sequential Choices and Jointness in Sites, L. Pilati and V. Boatto

2013/17 Resilience and specialization in volatile environments: evidence from the Italian Air Force Tornado crews learning practices, M. Laura Frigotto, Marco Zamarian

2013/16 Monetary Policy when the NAIRI is unknown: The Fed and the Great Deviation, Ronny Mazzocchi

2013/15 Intertemporal Coordination Failure and Monetary Policy, Ronny Mazzocchi

2013/14 Investment-Saving Imbalances with Endogenous Capital Stock, Ronny Mazzocchi

2013/13 Scope and Flaws of the New Neoclassical Synthesis, Ronny Mazzocchi

2013/12 Approximate Maximum Likelihood Estimation of the Autologistic Model, Marco Bee, Diego Giuliani, Giuseppe Espa

2013/11 Do middle managers matter?, Elena Feltrinelli, Roberto Gabriele, Sandro Trento

2013/10 Transatlantic contagion 2010-…, Roberto Tamborini

2013/09 Interbank contagion and resolution procedures: inspecting the mechanism, Edoardo Gaffeo e Massimo Molinari

2013/08 Using information markets in grantmaking. An assessment of the issues involved and an application to Italian banking foundations, Edoardo Gaffeo

2013/07 A spatial and sectoral analysis of firm demography in Italy, Giuseppe Espa, Danila Filipponi, Diego Giuliani, Davide Piacentino

2013/06 Reassessing the spatial determinants of the growth of Italian SMEs, Roberto Gabriele, Diego Giuliani, Marco Corsino, Giuseppe Espa

2013/05 At the core of the international financial system, Valentina Feroldi, Edoardo Gaffeo

2013/04 Uses And Motivations For Credit Derivatives: An Empirical Investigation Into Italian Banks, Eleonora Broccardo, Maria Mazzuca and Elmas Yaldiz

2013/03 Origins and prospects of the Euro existential crisis, Luigi Bonatti and Andrea Fracasso

2013/02 Employer moral hazard and wage rigidity. The case of worker-owned and investor-owned firms, Marina Albanese, Cecilia Navarra and Ermanno Tortia

2013/01 A regular multidistance among fuzzy numbers, Franco Molinari.