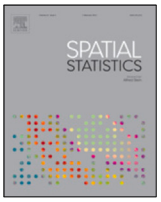
Spatial Statistics 47 (2022) 100613

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta



# Spatiotemporal sampling with spatial spreading and rotation of units in time

Esther Eustache *, Raphaël Jauslin, Yves Tillé

*Institute of statistics, University of Neuchâtel, Avenue de Bellevaux 51, 2000, Neuchâtel, Switzerland*

ARTICLE INFO

*Article history:*
Received 13 August 2021
Received in revised form 14 January 2022
Accepted 18 January 2022
Available online 4 February 2022

*Keywords:*
Balanced sampling
Cube method
Pivotal method
Spatial balance
Systematic sampling

ABSTRACT

When the sampled population belongs to a metric space, the selection of neighboring units will imply often similarities in the collected data due to their geographical proximity. In order to estimate parameters such as means or totals, it is therefore more efficient to select samples that are well distributed in space. Often, the interest lies not only in estimating a parameter at one point in time, but rather in estimating it at several points and studying its evolution. Because of the temporal autocorrelation of successive values from the same unit, a system of temporal rotation of the units in the samples must be provided. In other words, this type of problem forces us to consider two types of autocorrelation: spatial and temporal. In this article, we propose two new spatiotemporal sampling methods for equal or unequal inclusion probabilities. Systematic sampling is used to promote a rotation of the selection of the same unit over time, and thus address temporal spread. Both methods select samples that are well distributed in space at each sampling time. They differ by the fact that these samples are of random size for the first one, while for the second one, more complex, their sizes are controlled. Thus, the first method is called spatiotemporal sampling with random sample sizes (SPAR) and the second, spatiotemporal sampling with fixed sample sizes (SPAF). Simulations show that our methods outperform and generalize existing methods.

© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

## 1. Introduction

Sampling is almost always done to estimate unknown population parameters, for instance a total. When spatial data are considered, information from two neighboring units are generally very

* Corresponding author.
*E-mail address:* esther.eustache@unine.ch (E. Eustache).

https://doi.org/10.1016/j.spasta.2022.100613
2211-6753/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

similar. In this case, the selection of close units thus provides less information than the observation of spatially well-distributed units, and a less efficient estimator.

In addition, many applications require not only the selection of a spread sample at a given time, but also a rotation system of the selected units over time. For example, in some environmental monitoring such as that described in Tillé and Ecker (2013), a different part of the population is visited each year. Similarly, in the new census techniques applied in France and Italy, rotation groups of small municipalities are formed and one is selected each year. In these applications, if each annual sample is well spread, the gain in accuracy is likely to be significant. Therefore, the samples must remain spatially spread at each sampling time.

A large number of methods have been proposed to select spread samples. In one dimension, Quenouille (1949) has shown that systematic sampling is the optimal design for obtaining the most spread sample with equal inclusion probabilities. A first family of methods consists in transforming a multi-dimensional problem into a one-dimensional one in order to apply a systematic design. This is the case of the stratified sampling method by generalized random tessellation developed by Stevens and Olsen (1999, 2003, 2004), who use a quadrant-recursive partition of the unit square to map to a one-dimensional problem. The method has been implemented by Kincaid and Olsen (2015). Dickson and Tillé (2016) have used the traveling salesman problem to reduce the sampling problem to one dimension.

Another family of methods introduces a repulsion in the selection of neighboring units. Grafström (2011) has proposed the method called "Spatially correlated Poisson sampling" which generates a strong negative correlation between inclusion probabilities of close units using sampling weights. Grafström et al. (2012a) generalize the pivotal method proposed by Deville and Tillé (1998) to the selection of spread samples. This method has been modified by Grafström and Tillé (2013) to obtain samples that are both spread and balanced on totals of known auxiliary variables. Another generalization enables its application to continuous populations (Grafström and Matei, 2018). Grafström and Lundström (2013) recommend the use of spatially balanced samples on variables that are not geographic coordinates. Because of their similarities, groups of neighboring units can be seen as strata. Thus, the selection of spread samples can be compared to a multidimensional stratification, as in the method proposed by Jauslin and Tillé (2020).

Information from a unit collected at close sampling times will probably be similar. Zhao and Grafström (2020) propose a method of spatiotemporal sampling to improve estimators of change by selecting positively coordinated samples, i.e. by maximizing their overlap. If the goal is not to estimate the evolution of a parameter, it is preferable to do the opposite: select samples with negative coordination. Indeed, selecting samples that overlap as little as possible reduces redundant information. An appropriate rotation of the units must then be planned. This problem is complex when inclusion probabilities are unequal. Several solutions have been proposed in Deville and Tillé (2000) or Rivest and Ebouele (2020) but these methods do not take into account spatial autocorrelations. The selection of several spread samples from the same population over time becomes much more complex. Some solutions have already been proposed. Khavarzadeh et al. (2018) divide the space into primary units that are chosen with a balanced design and the units are selected to maximize the spread. Wang and Zhu (2019) propose a spatiotemporal sampling method based on a consecutive application of the local pivotal method. However, it cannot be applied when inclusion probabilities vary over time and does not allow to select a unit more than once over time.

In this paper, a set of new solutions for spatiotemporal sampling generalizing the method of Wang and Zhu (2019) is proposed. These methods can be applied to any temporal matrix of equal or unequal inclusion probabilities. If the inclusion probabilities allow the same unit to be selected several times, an appropriate rotation of the selected units over time is provided. The two proposed methods select spatiotemporal samples that are temporally and spatially spread. In other words, temporal spread means that a unit should not be selected at close sampling times, and spatial spread means that at each sampling time, the selected sample will be well distributed in space. The methods differ in that one produces random size samples at each sampling time while the other, much more complex, produces fixed size samples. Their effectiveness was demonstrated by a set of simulations on spatial biological data using the R package `SpotSampling` from Eustache et al. (2020). This package enables to apply the methods proposed in this paper.

## 2. Spreading in the context of spatial statistic trinity

In sampling theory, Hájek (1981) defines a pairwise strategy consisting of a design and an estimator. In survey sampling, we can also distinguish between a design-based approach and a model-based approach, depending on whether the inference is conducted according to the model or according to the design that generates the population (see among others Valliant et al., 2000, for the model-based approach, and Tillé, 2020; Lohr, 2021, for the design-based approach). Wang et al. (2020) define the trinity of spatial statistics as the triplet composed of a population, a sampling design and an estimator. In this section, we will justify the interest of our method in this context.

Consider a finite population $U$ of units denoted by $k \in \{1, \ldots, N\}$. A sample $s \subset U$ is selected by means of a sampling design $p(.)$ such that

$$p(s) \geq 0, \text{ for all } s \in U \text{ and } \sum_{s \subset U} p(s) = 1.$$

A variable $a_k$ has a Bernoulli distribution and takes the value 1 if unit $k$ is in the sample and 0 otherwise. The first and second order inclusion probability are respectively

$$\pi_k = \sum_{s \ni k} p(s) = \mathrm{E}_p(a_k) \text{ and } \pi_{k\ell} = \sum_{s \ni k,\ell} p(s) = \mathrm{E}_p(a_k a_\ell), \text{ for all } k, \ell \in U.$$

where $\mathrm{E}_p(.)$ is the expectation under the sampling design. Moreover, define $\Delta_{k\ell} = \mathrm{cov}_p(a_k, a_\ell) = \pi_{k\ell} - \pi_k \pi_\ell$ as the covariance under the sampling design between $a_k$ and $a_\ell$, with $\mathrm{cov}_p(., .)$ the covariance under the sampling design. In order to estimate a total

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

of a variable of interest $y_k$, $k \in U$, the Horvitz–Thompson estimator (Horvitz and Thompson, 1952)

$$\widehat{t}_y = \sum_{k \in s} \frac{y_k}{\pi_k}$$

gives the simplest unbiased estimator provided that all first order inclusion probabilities are not null.

Furthermore, suppose that the population is governed by a model $M$. As in Grafström and Tillé (2013), we can consider the general linear model, with possible heteroscedasticity and autocorrelation,

$$y_k = \boldsymbol{x}_k^\top \boldsymbol{\beta} + \varepsilon_k, \text{ for all } k \in U, \tag{1}$$

where $\boldsymbol{x}_k$ is a column vector of the values taken by $p$ auxiliary variables on unit $k$ and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of regression coefficients. Moreover, the $\varepsilon_k$ is a random variable such that $\mathrm{E}_M(\varepsilon_k) = 0$, $\mathrm{var}_M(\varepsilon_k) = \sigma_k^2$, for all $k \in U$, and

$$\mathrm{cov}_M(\varepsilon_k, \varepsilon_\ell) = \sigma_k \sigma_\ell \rho_{k\ell}, \text{ with } k \neq \ell \in U,$$

where $\mathrm{E}_M(.)$, $\mathrm{var}_M(.)$ and $\mathrm{cov}_M(., .)$ respectively denote the expectation, variance and covariance under model $M$. The spatial heterogeneity studied in Wang et al. (2016) is a particular case of model (1) when vector $\boldsymbol{x}_k$ contains the indicator variables of the strata and the $\sigma_k^2$ are equal within a stratum but can be unequal from one stratum to another. Usually the closer the units are, the more correlated they are. The $\rho_{k\ell}$ are thus supposed to be decreasing in function of a distance that can be computed between $k$ and $\ell$. For instance, the correlations could be written as $\rho_{k\ell} = \rho^{d(k,\ell)}$, where $d(k, \ell)$ is a distance between units $k$ and $\ell$.

Isaki and Fuller (1982) define the anticipated variance as

$$\mathrm{Avar}(\widehat{Y}) = \mathrm{E}_M E_p(\widehat{Y} - Y)^2.$$

The anticipated variance allows to evaluate the precision of an estimator under a given design and a model. It thus allows to conceive the best design according to a superpopulation model that would

have generated the population. Nedyalkova and Tillé (2008) compute the anticipated variance for a very general class of linear estimators. Grafström and Tillé (2013) prove that, under model (1), the anticipated variance of the Horvitz–Thompson estimator can be shown to be

$$\operatorname{Avar}(\widehat{Y}) = \mathrm{E}_p\left[\left(\sum_{k\in S}\frac{\mathbf{x}_k}{\pi_k} - \sum_{k\in U}\mathbf{x}_k\right)^{\top}\boldsymbol{\beta}\right]^2 + \sum_{k\in U}\sum_{\ell\in U}\sigma_k\sigma_\ell\rho_{k\ell}\frac{\Delta_{k\ell}}{\pi_k\pi_\ell}. \tag{2}$$

Grafström and Tillé (2013) also showed that the design that minimizes the anticipated variance should be balanced on $\boldsymbol{x}_k$ in the sense defined by Deville and Tillé (2004). Furthermore, it must have unequal inclusion probabilities proportional to $\sigma_k$ and must be as spread as possible in the space making the quantity $\rho_{k\ell}/\Delta_{k\ell}$ small. Neyman optimal allocation (Neyman, 1934) is also a particular case of this result. Variance estimators for spread samples are proposed in Grafström et al. (2012a).

Grafström and Lundström (2013) have also shown that spread samples are automatically well balanced on the variables used to compute the distance. The spread samples can then be considered approximately stratified over any compact set of units in space. Furthermore, Grafström et al. (2014) have shown, through a set of simulations, that a spread sample improves not only the accuracy of the Horvitz–Thompson estimator but also that of the nearest neighbor estimator. A rigorous proof of this result is given in Fattorini et al. (2021). There are thus multiple applications where it is interesting to select samples spread out with unequal probabilities. As we have seen, several methods exist to select such samples. However, the selection of several spread samples with unequal probabilities that are negatively coordinated from the same population is a problem that is not yet solved. We propose a solution in the following sections.

## 3. Spatiotemporal sampling notations and requirements

Suppose that each unit $k$ of the population $U$ belongs to a metric space of dimension $r \geq 2$ and the spatial coordinates of each unit are known. Consider also $T \in \mathbb{N}$ different moments spaced out in time. For example, these $T$ times may correspond to years or months. The selection of a spatiotemporal sample must satisfy given inclusion probabilities which can be equal or unequal. Let $\pi_k^t$ be the probability that unit $k \in U$ is selected at time $t \in \{1,\ldots,T\}$. Let $\boldsymbol{\Pi}$ denote the $N \times T$ matrix of temporal inclusion probabilities:

$$\boldsymbol{\Pi} = \begin{pmatrix} \pi_1^1 & \cdots & \pi_1^t & \cdots & \pi_1^T \\ \vdots & & \vdots & & \vdots \\ \pi_k^1 & \cdots & \pi_k^t & \cdots & \pi_k^T \\ \vdots & & \vdots & & \vdots \\ \pi_N^1 & \cdots & \pi_N^t & \cdots & \pi_N^T \end{pmatrix}.$$

The $t$th column of matrix $\boldsymbol{\Pi}$ is denoted by $\boldsymbol{\pi}^t = (\pi_1^t,\ldots,\pi_k^t,\ldots,\pi_N^t)^{\top}$ and contains the inclusion probabilities of all the units at sampling time $t$. The $k$th row of matrix $\boldsymbol{\Pi}$ is denoted by $\boldsymbol{\pi}_k = (\pi_k^1,\ldots,\pi_k^t,\ldots,\pi_k^T)$ and contains the inclusion probabilities of $k$ at each sampling time. The sum of the $t$th column of $\boldsymbol{\Pi}$ is denoted by $\psi^t = \sum_{k\in U}\pi_k^t$ and the sum of the $k$th row by $\psi_k = \sum_{t=1}^{T}\pi_k^t$. The sums $\psi^t$ and $\psi_k$ are not necessarily integer.

The aim is to generate a matrix of indicator random variables $a_k^t$ that are equal to 1 if plot $k$ is selected in the sample at sampling time $t$ and 0 otherwise. Matrix $\mathbf{A}$ is the $N \times T$ sampling indicator matrix:

$$\mathbf{A} = \begin{pmatrix} a_1^1 & \cdots & a_1^t & \cdots & a_1^T \\ \vdots & & \vdots & & \vdots \\ a_k^1 & \cdots & a_k^t & \cdots & a_k^T \\ \vdots & & \vdots & & \vdots \\ a_N^1 & \cdots & a_N^t & \cdots & a_N^T \end{pmatrix}.$$

The $t$th column of matrix $\mathbf{A}$ is denoted by $\mathbf{a}^t = (a_1^t, \ldots, a_k^t, \ldots, a_N^t)^\top$ and corresponds to the cross-sectional sample at time $t$. The $k$th row of matrix $\mathbf{A}$ is denoted by $\mathbf{a}_k = (a_k^1, \ldots, a_k^t, \ldots, a_k^T)$ and corresponds to the longitudinal sample of $k$. Let also $n^t = \sum_{k\in U} a_k^t$ be the number of units selected at the $t$th sampling time and $n_k = \sum_{t=1}^T a_k^t$ be the number of times that unit $k$ is selected during the $T$ times.

The objective is to select a spatiotemporal sample $\mathbf{A}$, which best meets the following three requirements:

(i) The sampling design satisfies the inclusion probabilities given in $\boldsymbol{\Pi}$, i.e. $\mathrm{E}_p(\mathbf{A}) = \boldsymbol{\Pi}$.
(ii) The longitudinal sample $\mathbf{a}_k = (a_k^1, \ldots, a_k^t, \ldots, a_k^T)$ is as spread over time as possible, for all $k \in U$, in the sense that once a unit has been selected, it should remain out of the following samples as long as possible.
(iii) The cross-sectional sample $\mathbf{a}^t = (a_1^t, \ldots, a_k^t, \ldots, a_N^t)^\top$ is as spread in space as possible, for all $t \in \{1, \ldots, T\}$, in the sense that we avoid selecting geographically neighboring units.

Requirement (i) is equivalent to have $\mathrm{E}_p(a_k^t) = \pi_k^t$, for each element $a_k^t$ of matrix $\mathbf{A}$ and implies $\mathrm{E}_p(n^t) = \psi^t$ and $\mathrm{E}_p(n_k) = \psi_k$.

A longitudinal sample corresponds to select or not the same unit at $T$ different times. The same variable measured at several different times on a unit $k$ is positively autocorrelated over time. For this reason, the objective is to obtain a sample $\mathbf{a}_k$ as spread as possible if the vector of inclusion probabilities $\boldsymbol{\pi}_k$ allows to select the unit $k$ more than once. By spreading each sample $\mathbf{a}_k$ (requirement (ii)), once a unit $k$ is selected, it remains out of the sample as long as possible, depending on the vector of inclusion probabilities $\boldsymbol{\pi}_k$. This generates an appropriate rotation of the units selected in the cross-sectional samples $\mathbf{a}^t$ and minimizes the overlap between successive samples.

If the units are geolocated, spatial autocorrelation must be taken into account. By selecting a spread sample based on the spatial coordinates at each sampling time $t$, the accuracy of the estimate should be better than with unspread samples. Requirement (iii) prevents the selection of similar units at the same time.

Finding a method to meet all of these requirements is not straightforward. In the following sections, we proceed step-by-step, first explaining how the rotation of units in time is optimized (ii), and then describing the two spatiotemporal methods.

## 4. Method of Wang and Zhu

The new methods proposed in this paper are based on an existing method, developed by Wang and Zhu (2019), for the problem of spatiotemporal sampling. It is based on consecutive applications of the local pivotal method developed by Grafström et al. (2012a). The local pivotal method generalizes the pivotal method, a sampling method without replacement described in Deville and Tillé (1998), to the selection of spread samples (see Appendices A and B).

The method of Wang and Zhu (2019) is described in Algorithm 1 and consists of two steps. First a spatially spread set of units, denoted by $G \subset U$, is selected using the local pivotal method. Then, samples $\mathbf{a}^1, \ldots, \mathbf{a}^T$ are selected from $G$ without replacement. The method satisfies the constraints (i)–(iii), but can only be applied under two conditions:

(*i*) The columns of matrix $\boldsymbol{\Pi}$ are proportional.
(*ii*) The sums of the rows of $\boldsymbol{\Pi}$ are equal to or smaller than one, i.e. $\psi_k \leq 1$ for all $k \in U$.

In the step 1 of Algorithm 1, probabilities of $\pi_k^\circ$ must remain smaller than one and $L$ must not be too large to have a good spatial balance. The authors recommend to take $L \leq \min\{2, \min_{k\in U}(\psi_k)^{-1}\}$.

In real sampling problems, the assumptions on $\boldsymbol{\Pi}$ are not always satisfied. Indeed, inclusion probabilities are not necessarily proportional as in condition (*i*) of Wang and Zhu, especially if

**Algorithm 1** Wang and Zhu method

1. Select an initial set from $U$, denoted by $G$, by the local pivotal method with probabilities $\pi_k^\circ = L\sum_{t=1}^T \pi_k^t$, where $L \geq 1$ is a predefined value.
2. For $t = 1, \ldots, T$, repeat the following steps.
   (a) Select a sample $\mathbf{a}^t$ from $G$ of size $n^t$ by the local pivotal method with equal inclusion probabilities.
   (b) Update $G$ by $G^* = G\backslash\{\mathbf{a}^1 \cup \ldots \cup \mathbf{a}^t\}$.

they are based on a variable that changes over time. Condition (*ii*) of Wang and Zhu is also very restrictive. In practical problems, the sums of the rows of $\boldsymbol{\Pi}$ could be larger that one. In this case, a unit would be selected several times during the period using a rotation scheme. With Wang and Zhu's method, each unit can only be selected once for the entire time period. The new methods proposed in this paper are not restricted to these two requirements.

## 5. Preliminary step to spatiotemporal sampling: a two-phase sampling approach

Spatiotemporal sampling methods can begin by the selection of an initial spread set of units, as in the method of Wang and Zhu. This allows to obtain a better spreading of the cross-sectional samples $\mathbf{a}^t$. This first sampling phase consists in selecting a first well-spread set $U'$ of $N'$ units, $N' \leq N$, which will then be considered for the spatiotemporal design. Thus, the inclusion probabilities used for this design become conditional on the first sampling phase.

All units $k \in \{U\backslash U'\}$ will therefore be permanently excluded from all the cross-sectional samples, which means that their inclusion probability will be $\pi_k^t = 0$ for each sampling time $t$. This first sampling phase is a generalization of the first step of the Wang and Zhu method without restricting $\boldsymbol{\Pi}$ to condition (*i*). Concerning the second condition (*ii*), this first phase should not be applied if $\boldsymbol{\Pi}$ does not satisfy it. Indeed, if we have $\psi_k \geq 1$ for a unit $k$, this unit must be selected in at least one cross-sectional sample $\mathbf{a}^t$, and then cannot be completely excluded.

This selection is made using the local cube method (Grafström and Tillé, 2013). The local cube method is based on two methods: the cube method, that allows to select balanced samples on totals of auxiliary variables (Deville and Tillé, 2004), and the local pivotal method. Similarly to the cube method, the local cube method is divided in two phases: the *flight phase* and the *landing phase* (see Appendix B). More precisely, this preliminary step uses only the flight phase of the local cube method (see Appendix C). Algorithm 2 describes the main steps of the selection of this initial spread set.

**Algorithm 2** Preliminary step: selection of an initial spread set

1. Compute inclusion probabilities $\boldsymbol{\pi}^\circ = (\pi_1^\circ, \ldots, \pi_k^\circ, \ldots \pi_N^\circ)^\top$ such that $\pi_k^\circ = \min(L\sum_{t=1}^T \pi_k^t, 1)$.
2. Run the flight phase of the local cube method with $\boldsymbol{\pi}^\circ$ as inclusion probabilities and column of $\boldsymbol{\Pi}$ as balancing constraints. A vector $\boldsymbol{\pi}^\bullet = (\pi_1^\bullet, \ldots, \pi_k^\bullet, \ldots \pi_N^\bullet)^\top$ of inclusion probabilities is obtained.
3. Update matrix $\boldsymbol{\Pi}$ with $\boldsymbol{\Pi}^\bullet$ such that $\boldsymbol{\Pi}^\bullet = \text{diag}(\boldsymbol{\pi}^\bullet)\text{diag}(\boldsymbol{\pi}^\circ)^{-1}\boldsymbol{\Pi}$.

The same recommendation as in the Wang and Zhu method is applied to the choice of $L$, i.e. $L \leq \min\{2, \min_{k\in U}(\psi_k)^{-1}\}$.

**Proposition 1.** *During the process of Algorithm 2, the sums of the columns of matrix $\boldsymbol{\Pi}$ are equal to that of matrix $\boldsymbol{\Pi}^\bullet$.*

**Proof.** After step 2 of Algorithm 2, because of the flight phase of the local cube method, many lines of $\boldsymbol{\Pi}^\bullet$ may contain only zeros and we have $\boldsymbol{\Pi}^\top\boldsymbol{\pi}^\bullet = \boldsymbol{\Pi}^\top\boldsymbol{\pi}^\circ$. This allows to deduce that $\boldsymbol{\Pi}^{\bullet\top}\mathbf{1}_N = \boldsymbol{\Pi}^\top \mathrm{diag}(\boldsymbol{\pi}^\circ)^{-1}\boldsymbol{\pi}^\bullet\mathbf{1}_N = \boldsymbol{\Pi}^\top \mathrm{diag}(\boldsymbol{\pi}^\circ)^{-1}\boldsymbol{\pi}^\circ\mathbf{1}_N = \boldsymbol{\Pi}^\top\mathbf{1}_N$, so the sums of the columns of $\boldsymbol{\Pi}^\bullet$ are exactly the same as the ones of $\boldsymbol{\Pi}$, with $\mathbf{1}_N$ a column vector composed of $N$ ones.

Proposition 1 will allow us to obtain the fixed sample size based on the original matrix $\boldsymbol{\Pi}$ during the second sampling phase, i.e. the spatiotemporal sampling methods presented below. The sums of the rows of $\boldsymbol{\Pi}^\bullet$ are either equal to zero or not greater than $1/L$.

Matrix $\boldsymbol{\Pi}$ can be singular. In this case, it is more efficient to remove the columns of $\boldsymbol{\Pi}$ that are linearly dependent on the others. Indeed, this reduces the number of balancing constraints. If all the columns of matrix $\boldsymbol{\Pi}$ are proportional (i.e. linearly dependent), as in condition $(i)$ of the Wang and Zhu method, only one column can be used and the local cube method is reduced to the local pivotal method. This shows that this first phase generalizes Wang and Zhu's.

## 6. Temporal spreading

As explained in Section 3, samples $\mathbf{a}^t$ must be as spread over time as possible while $\mathbf{a}_k$ must be as spread as possible in the space of dimension $r \geq 2$. These two spreads are difficult to manage simultaneously. The method used to address the first problem of temporal spreading is presented in this section.

The first constraint to be managed is the temporal spreading. In other words, each sample $\mathbf{a}_k$ must be spread in a one-dimensional space corresponding to the $T$ different times. To that end, the longitudinal samples $\mathbf{a}_k$ are generated using systematic sampling, $k \in U$. Systematic sampling for unequal probabilities was proposed by Madow (1949) (see also Iachan, 1982, 1983; Bellhouse, 1988; Bellhouse and Sutradhar, 1988). This sampling method selects a sample according to a random starting point but with a fixed, periodic interval based on the inclusion probabilities (see Appendix D). In one dimension, Quenouille (1949) and Bellhouse (1977) proved that systematic sampling is the best design to obtain the most spread sample when the inclusion probabilities are equal. Therefore, systematic sampling is a good way to select each longitudinal sample $\mathbf{a}_k$ meeting requirement (ii).

For each vector of inclusion probabilities over time $\boldsymbol{\pi}_k$, all possible systematic samples are computed. Let $h(k)$ denote the number of possible systematic samples with non-zero probabilities for $k$. This number $h(k)$ is not greater than $T$, if $\psi_k$ is integer, and is not greater than $(T+1)$ otherwise (Pea et al., 2007). Let also $H = \sum_{k\in U} h(k)$ be the total number of longitudinal samples. In addition to taking into account autocorrelation as explained above, the advantage of systematic sampling is that the total number $H$ of possible samples is relatively small compared to other designs for which $H = N \times 2^T$. This makes it possible to describe all possible systematic samples in a simple way.

Consider $\mathbf{S}_k$ the matrix containing in rows the $h(k)$ possible longitudinal samples of a unit $k$ such that $\mathbf{S}_k = (\mathbf{s}_{k,1}^\top, \ldots, \mathbf{s}_{k,i}^\top, \ldots, \mathbf{s}_{k,h(k)}^\top)^\top$, where $\mathbf{s}_{k,i} = (s_{k,i}^1, \ldots, s_{k,i}^t, \ldots, s_{k,i}^T)$ is the $i$th possible systematic sample. Consider also $\mathbf{p}_k = (p_{k,1}, \ldots, p_{k,i}, \ldots, p_{k,h(k)})^\top$ the probabilities of selecting the samples of $\mathbf{S}_k$. We have $\mathbf{S}_k^\top\mathbf{p}_k = \boldsymbol{\pi}_k$. All matrices $\mathbf{S}_k$ and vectors $\mathbf{p}_k$ are respectively concatenated in a matrix $\mathbf{S}$ and a vector $\mathbf{p}$, such that $\mathbf{S} = (\mathbf{S}_1^\top, \ldots, \mathbf{S}_k^\top, \ldots, \mathbf{S}_N^\top)^\top$ and $\mathbf{p} = (\mathbf{p}_1^\top, \ldots, \mathbf{p}_k^\top, \ldots, \mathbf{p}_N^\top)^\top$. The rows of matrix $\mathbf{S}$ thus contain $H$ longitudinal samples of size $T$. Vector $\mathbf{p}$ contains the selection probabilities of the $H$ longitudinal samples in $\mathbf{S}$.

**Example 1.** Consider the matrix of inclusion probabilities $\boldsymbol{\Pi}$ with $N = 3$ and $T = 4$

$$\boldsymbol{\Pi} = \begin{pmatrix}\boldsymbol{\pi}_1\\ \boldsymbol{\pi}_2\\ \boldsymbol{\pi}_3\end{pmatrix} = \begin{pmatrix}0.4 & 0.6 & 0.2 & 0.8\\ 0.5 & 0.5 & 0.5 & 0.5\\ 0.1 & 0.9 & 0.3 & 0.7\end{pmatrix}.$$

The longitudinal sampling designs using systematic sampling are computed:

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_1 \\ \hline \mathbf{S}_2 \\ \hline \mathbf{S}_3 \end{pmatrix} = \left(\begin{array}{cccc} \mathbf{s}_{1,1} \\ \mathbf{s}_{1,2} \\ \mathbf{s}_{1,3} \\ \hline \mathbf{s}_{2,1} \\ \mathbf{s}_{2,2} \\ \hline \mathbf{s}_{3,1} \\ \mathbf{s}_{3,2} \\ \mathbf{s}_{3,3} \end{array}\right) = \left(\begin{array}{cccc} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ \hline 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{array}\right) \text{ and } \mathbf{p} = \begin{pmatrix} \mathbf{p}_1 \\ \hline \mathbf{p}_2 \\ \hline \mathbf{p}_3 \end{pmatrix} = \left(\begin{array}{c} p_{1,1} \\ p_{1,2} \\ p_{1,3} \\ \hline p_{2,1} \\ p_{2,2} \\ \hline p_{3,1} \\ p_{3,2} \\ p_{3,3} \end{array}\right) = \left(\begin{array}{c} 0.2 \\ 0.2 \\ 0.6 \\ \hline 0.5 \\ 0.5 \\ \hline 0.1 \\ 0.2 \\ 0.7 \end{array}\right),$$

and the number of samples in $\mathbf{S}$ are $H = h(1) + h(2) + h(3) = 3 + 2 + 3 = 8$.

Once all longitudinal sampling designs have been computed, a systematic sample must be chosen for each unit $k \in U$. Let $\mathbf{q} \in \{0, 1\}^H$ denote the vector of indicators, with the same dimension as $\mathbf{p}$, that indicates which systematic sample is definitively selected. Each element $q_{k,i}$ is a Bernoulli random variable that is equal to 1 if the $i$th longitudinal sample of the unit $k$ is selected and 0 otherwise. Each realization of $\mathbf{q}$ corresponds to the selection of the longitudinal samples and we have $\mathbf{S}_k^\top \mathbf{q}_k = \mathbf{a}_k$. The aim is to define a method to select exactly one longitudinal sample for each unit $k$, this implies

$$\sum_{i=1}^{h(k)} q_{k,i} = \sum_{i=1}^{h(k)} p_{k,i} = 1, k \in U. \tag{3}$$

Vector $\mathbf{a}_k$ will be a systematic sample, the temporal spreading is therefore guaranteed. Some constraints must also be applied on $\mathbf{q}$ to ensure the spatial spreading of cross-sectional samples $\mathbf{a}^t$. If unit $k$ is in the neighborhood of $\ell$, the idea is to choose systematic samples that do not select $k$ and $\ell$ at the same time. In other words, it could be better to have as less as possible $a_k^t = a_\ell^t$. In the next section, two spatiotemporal sampling methods are explained. They use different methods to randomly select $\mathbf{a}_k$ in each submatrix $\mathbf{S}_k$ while satisfying the spatial spreading at each sampling time.

## 7. Spatiotemporal sampling

The constraint of knowing the size of the sample selected before sampling is often required. In the context of spatiotemporal sampling, this constraint must be taken into account for $T$ samples. This makes it much more complicated to satisfy. To obtain fixed sample sizes, vector $\mathbf{q}$ must satisfy the balancing equation

$$\mathbf{S}^\top \mathbf{q} = \mathbf{S}^\top \mathbf{p} = (n^1, \ldots, n^t, \ldots, n^T)^\top. \tag{4}$$

In this section, we propose two different spatiotemporal sampling methods that select spatially well-distributed samples at the sampling times. It is important to note that the methods differ in that one generates random sample size at each sampling time while the other generates fixed sample size. The second method is much more complex because of the complexity of satisfying Eq. (4) while still considering the requirements (i)–(iii) presented above. Both can be considered as generalizations of Wang and Zhu's method.

### 7.1. Spatiotemporal sampling with random sample sizes (SPAR)

The *spatiotemporal sampling with random sample sizes* (SPAR) method is a spatiotemporal sampling method with random sample size at each sampling time. The sizes of the samples $\mathbf{a}^t$ are fixed only if the inclusion probabilities in $\boldsymbol{\pi}^t$ are all equal. In this restricted case, this method is the same as that of Wang and Zhu. In the other cases, this method has only fixed sample size at the sampling time $t = 1$.

The SPAR method is described in Algorithm 3. The spread of the selected units is managed by recursively applying a spread sampling method at each sampling time. This spread sampling method

**Algorithm 3** SPAR Sampling

For $t = 1$ to $T$, apply the following instructions:

1. Apply a spread sampling method on vector of inclusion probabilities $\boldsymbol{\pi}^t$ of matrix $\boldsymbol{\Pi}$ and obtain vector $\mathbf{a}^t = (a_1^t, \ldots, a_k^t, \ldots, a_N^t)^\top$.
2. For each unit $k$, update the probabilities $\mathbf{p}_k = (p_{k,1}, \ldots, p_{k,j}, \ldots, p_{k,h(k)})^\top$ of the systematic sampling design as follows:

   - If $a_k^t = 1$,

$$p_{k,j}^* = \begin{cases} p_{k,j} \left( \sum_{j=1}^{h(k)} p_{k,j} s_{k,j}^t \right)^{-1} & \text{if } s_{k,j}^t = 1, \\ 0 & \text{if } s_{k,j}^t = 0. \end{cases}$$

   - If $a_k^t = 0$,

$$p_{k,j}^* = \begin{cases} p_{k,j} \left( \sum_{j=1}^{h(k)} p_{k,j} (1 - s_{k,j}^t) \right)^{-1} & \text{if } s_{k,j}^t = 0, \\ 0 & \text{if } s_{k,j}^t = 1. \end{cases}$$

3. Update matrix $\boldsymbol{\Pi}$ by $\boldsymbol{\Pi}^*$ such that the $k$th row of $\boldsymbol{\Pi}^*$ is $\boldsymbol{\pi}_k^* = \mathbf{S}_k^\top \mathbf{p}_k^*$, with $k = 1, \ldots, N$.

used could be for instance the local pivotal method, the generalized random tessellation stratified method (Stevens and Olsen, 1999, 2003, 2004), the wave sampling method (Jauslin and Tillé, 2020) or even the traveling salesman problem-systematic method (Dickson and Tillé, 2016). By updating $\mathbf{p}$ values to 0, some systematic samples are excluded at each iteration depending on the results of the spread sampling. For example, at time $t = 1$, if unit $k = 1$ is selected in sample $\mathbf{a}^1$, any systematic samples of unit $k$ for which the first element is 0 are excluded. The latest update of vector $\mathbf{p}$ in Algorithm 3 is only composed of 0 s and 1 s and corresponds to the vector $\mathbf{q}$. This sampling satisfies inclusion probabilities, requirement (i) is met. This vector $\mathbf{q}$ also satisfies requirements (ii) due to the systematic sampling and (iii) due to the first step, but does not have a fixed sample size.

### 7.2. Spatiotemporal sampling with fixed sample sizes (SPAF)

The *spatiotemporal sampling with fixed sample sizes* (SPAF) method allows to select spatiotemporal samples that are spatially, temporally spread, while controlling the size of cross-sectional samples. The procedure is based on the local pivotal method. Let $(\mathbf{S}_k, \mathbf{p}_k)$ denote the couple containing the longitudinal sampling design of a unit $k$ and $\mathrm{Im}(\mathbf{M})$ be the image of a matrix $\mathbf{M}$. The steps of the SPAF method are described in Algorithm 4.

Each of the $T$ sampling times is processed one by one. Two units $k$ and $\ell$ in $U$ are considered at main stages of Algorithm 4. The main idea is to recursively update the vectors $\mathbf{p}_k$ and $\mathbf{p}_\ell$ such that there is a repulsion in the selection of $k$ and $\ell$ at the same sampling time. All the couples of units $(k, \ell)$ are treated in a particular order. Indeed, they are sorted in increasing order according to the Euclidean distance between their coordinates in the metric space to which they belong. Nearby units are thus favored because they are treated first, and they will have less chances to be selected at the same sampling time. The update of $\mathbf{p}_k$ and $\mathbf{p}_\ell$ is similar to that of the local pivot method. The particularity here is that the update focuses on the decision making for the sampling time considered at this iteration. During the procedure, the $\mathbf{p}_k$ vectors are updated and some of their values are potentially set to 0. The probabilities updated to 0 imply the exclusion of the corresponding samples in the $\mathbf{S}_k$ matrices. These samples should therefore not be taken into account for the rest of the procedure, so sub-couple $(\widetilde{\mathbf{S}}_k, \widetilde{\mathbf{p}}_k)$, containing only systematic sampling with non-null probabilities of being selected in $\mathbf{p}_k$, must be defined for each $k$. These samples must not be taken into account for the rest of the procedure. It is thus necessary to define for each $k$ a

**Algorithm 4** SPAF sampling

(i) Define the vector $\mathbf{d}_t \in \mathbb{R}^T$ such that its $t$th element is equal to 1 and the others to 0, with $t \in \{1, \ldots, T\}$.

(ii) Consider all pairs of units $(k, \ell) \in (U \times U)$.

(iii) For each sampling time $t = 1$ to $T$, apply the following steps to each pair of units $(k, \ell)$ from the closest to the farthest in terms of Euclidean distance between their spatial coordinates.

1. Define the subdesign $(\widetilde{\mathbf{S}}_k, \widetilde{\mathbf{p}}_k)$ of $(\mathbf{S}_k, \mathbf{p}_k)$ such that $(\widetilde{\mathbf{S}}_k, \widetilde{\mathbf{p}}_k)$ contains only the $\widetilde{h}(k)$ samples of $(\mathbf{S}_k, \mathbf{p}_k)$ with non-null probabilities. Compute subdesigns $(\widetilde{\mathbf{S}}_k, \widetilde{\mathbf{p}}_k)$ and $(\widetilde{\mathbf{S}}_\ell, \widetilde{\mathbf{p}}_\ell)$.
2. Compute matrices $\mathbf{U}_k = \widetilde{\mathbf{S}}_k^\top - \boldsymbol{\pi}_k \mathbf{1}_{\tilde{h}(k)}^\top$ and $\mathbf{U}_\ell = \widetilde{\mathbf{S}}_\ell^\top - \boldsymbol{\pi}_\ell \mathbf{1}_{\tilde{h}(\ell)}^\top$.
3. While $\pi_k^t \notin \{0, 1\}$ and $\{\mathrm{Im}(\mathbf{U}_k) \cap \mathrm{Im}(\mathbf{U}_\ell)\} \neq \{\ b\}$, repeat the following instructions:
   (a) Let $\mathbf{u}_p \in \mathbb{R}^T$ be the orthogonal projection of $\mathbf{d}_t$ on the set $\{\mathrm{Im}(\mathbf{U}_k) \cap \mathrm{Im}(\mathbf{U}_\ell)\}$. If $\mathbf{u}_p$ is null, move on to another couple of units.
   (b) Find two vectors $\mathbf{b}_k \in \mathbb{R}^{\tilde{h}(k)}$ and $\mathbf{b}_\ell \in \mathbb{R}^{\tilde{h}(\ell)}$ such that $\widetilde{\mathbf{S}}_k^\top \mathbf{b}_k = \mathbf{u}_p$ and $\widetilde{\mathbf{S}}_\ell^\top \mathbf{b}_\ell = \mathbf{u}_p$.
   (c) Compute the largest values for $\gamma_{k1}$, $\gamma_{k2}$, $\gamma_{\ell 1}$ and $\gamma_{\ell 2}$ that satisfy $b \leq \widetilde{\mathbf{p}}_k + \gamma_{k1}\mathbf{b}_k \leq \mathbf{1}$, $b \leq \widetilde{\mathbf{p}}_k - \gamma_{k2}\mathbf{b}_k \leq \mathbf{1}$, $b \leq \widetilde{\mathbf{p}}_\ell + \gamma_{\ell 1}\mathbf{b}_\ell \leq \mathbf{1}$ and $b \leq \widetilde{\mathbf{p}}_\ell - \gamma_{\ell 2}\mathbf{b}_\ell \leq \mathbf{1}$.
   (d) Compute $\lambda_1 = \min(\gamma_{k1}, \gamma_{\ell 2})$ and $\lambda_2 = \min(\gamma_{k2}, \gamma_{\ell 1})$.
   (e) Update randomly vectors $\widetilde{\mathbf{p}}_k$ and $\widetilde{\mathbf{p}}_\ell$ such that

$$\begin{cases} \widetilde{\mathbf{p}}_k \leftarrow \widetilde{\mathbf{p}}_k + \lambda_1 \mathbf{b}_k \\ \widetilde{\mathbf{p}}_\ell \leftarrow \widetilde{\mathbf{p}}_\ell - \lambda_1 \mathbf{b}_\ell \end{cases} \text{ with probability } \lambda_2/(\lambda_1 + \lambda_2),$$

   or

$$\begin{cases} \widetilde{\mathbf{p}}_k \leftarrow \widetilde{\mathbf{p}}_k - \lambda_2 \mathbf{b}_k \\ \widetilde{\mathbf{p}}_\ell \leftarrow \widetilde{\mathbf{p}}_\ell + \lambda_2 \mathbf{b}_\ell \end{cases} \text{ with probability } \lambda_1/(\lambda_1 + \lambda_2).$$

   (f) Update non-null probabilities in $\mathbf{p}_k$ and $\mathbf{p}_\ell$ with the updated values of $\widetilde{\mathbf{p}}_k$ and $\widetilde{\mathbf{p}}_\ell$.
   (g) If there is at least one null value in vector $\widetilde{\mathbf{p}}_k$, remove them from $\widetilde{\mathbf{p}}_k$ and remove also their corresponding longitudinal samples in matrix $\widetilde{\mathbf{S}}_k$. Do the same for unit $\ell$ to considered only samples with positive selection probabilities in $\widetilde{\mathbf{S}}_\ell$ and $\widetilde{\mathbf{p}}_\ell$.
   (h) Update the inclusion probabilities $\boldsymbol{\pi}_k$ and $\boldsymbol{\pi}_\ell$ because $\widetilde{\mathbf{S}}_k$, $\widetilde{\mathbf{S}}_\ell$, $\widetilde{\mathbf{p}}_k$ and $\widetilde{\mathbf{p}}_\ell$ could be modified in the previous step.
   (i) Recompute also matrices $\mathbf{U}_k$ and $\mathbf{U}_\ell$.
4. Update $\mathbf{p}$ with $\mathbf{p}^*$ such that $\mathbf{p}_k^*$ and $\mathbf{p}_\ell^*$ are the updated $\mathbf{p}_k$ and $\mathbf{p}_\ell$.
5. Update $\boldsymbol{\Pi}$ with $\boldsymbol{\Pi}^*$ where the $k$th and $\ell$th rows of $\boldsymbol{\Pi}^*$ are respectively $\boldsymbol{\pi}_k^* = \mathbf{S}_k^\top \mathbf{p}_k^*$ and $\boldsymbol{\pi}_\ell^* = \mathbf{S}_\ell^\top \mathbf{p}_\ell^*$.

(iv) If $\boldsymbol{\Pi}$ still contains values not equal to zero or one, apply the following steps:

1. Consider the population of size $H$ of the systematic samples contains in $\mathbf{S}$.
2. Stratify the population in $N$ strata so that each stratum contains the subgroup of systematic samples of a unit $k \in U$, i.e. $\mathbf{S}_k$.
3. Apply a stratified balanced sampling on the population of systematic samples using vector $\mathbf{p}$ as inclusion probability, columns of $\mathbf{S}$ as balancing variables and strata defined in the previous step.

sub-couple $(\widetilde{\mathbf{S}}_k, \widetilde{\mathbf{p}}_k)$, containing only systematic samples having non-zero probabilities in $\mathbf{p}_k$ to be selected.

**Proposition 2.** *Consider the procedure described in Algorithm 4. Throughout this process, the following propositions are satisfied:*

*(i) The vectors* $\mathbf{b}_k$ *and* $\mathbf{b}_\ell$ *exist.*

*(ii) If the sums of the rows of* $\mathbf{S}_k$ *are all integer, the vectors* $\mathbf{q} \in \{0,1\}^H$ *and* $\mathbf{p} \in [0,1]^H$ *satisfy*

$$\sum_{i=1}^{h(k)} q_{k,i} = \sum_{i=1}^{h(k)} p_{k,i} = 1, k \in U,$$

*(iii)* $\mathrm{E}(\mathbf{A}) = \boldsymbol{\Pi}$,

*(iv) The units selected in each cross-sectional sample* $\mathbf{a}^t$, $t \in \{1,\dots,T\}$, *are well spread in the metric space in which they belong,*

*(v) The sum of the vector* $\boldsymbol{\pi}$ *remains unchanged after each of its updates:*

$$\sum_{k=1}^{n} \boldsymbol{\pi}_k^* = \sum_{k=1}^{n} \boldsymbol{\pi}_k.$$

**Proof.**

(i) The existence of $\mathbf{b}_k$ and $\mathbf{b}_\ell$ is guaranteed because $\mathbf{u}_p \in \{\mathrm{Im}(\mathbf{U}_k) \cap \mathrm{Im}(\mathbf{U}_\ell)\}$ can always be written as a linear combination of the samples contained in rows of $\mathbf{S}_k$ and $\mathbf{S}_\ell$

*(ii)* Throughout algorithm 4, $\mathbf{p}$ is iteratively modified in order to finally obtain $\mathbf{q}$. Only the non-zeros values of $\mathbf{p}$ are considered at each step, so a subvector $\widetilde{\mathbf{p}}_k$ of $\mathbf{p}_k$ containing only these relevant values is defined for each unit $k$. For a unit $k$, $\widetilde{\mathbf{p}}_k$ is updated by $(\widetilde{\mathbf{p}}_k + \lambda \mathbf{b}_k)$, with $\lambda \in \mathbb{R}$. If $\mathbf{b}_k$ is centered, the proposition *(i)* is immediately proven. Since $\widetilde{\mathbf{S}}_k^\top \mathbf{b}_k = \mathbf{u}_p$ and $\mathbf{1}_T^\top \mathbf{u}_p = 0$, we have $\mathbf{1}_T^\top \widetilde{\mathbf{S}}_k^\top \mathbf{b}_k = 0$. In the case where the sums of the rows of $\mathbf{S}_k$ are all equal because $\psi_k \in \mathbb{N}$, this implies $\mathbf{1}_T^\top \widetilde{\mathbf{S}}_k^\top = n_k \mathbf{1}_{h(k)}^\top$ and then $n_k \mathbf{1}_{h(k)}^\top = 0$. The vector $\mathbf{b}_k$ is thus centered and the proposition *(ii)* is proven.

*(iii)* For a unit $k$, the expectation of $\boldsymbol{\pi}_k$ under the random of the update, i.e. the expectation of $\boldsymbol{\pi}_k^*$, is:

$$\mathrm{E}_p(\boldsymbol{\pi}_k^*) = \frac{\lambda_2}{\lambda_1+\lambda_2} \widetilde{\mathbf{S}}_k^\top (\widetilde{\mathbf{p}}_k + \lambda_1 \mathbf{b}_k) + \frac{\lambda_1}{\lambda_1+\lambda_2} \widetilde{\mathbf{S}}_k^\top (\widetilde{\mathbf{p}}_k - \lambda_2 \mathbf{b}_k) = \boldsymbol{\pi}_k.$$

*(iv)* The update is made such that if a probability $\pi_k^t$ is increased by the update, the corresponding $\pi_\ell^t$ will be decreased and reciprocally. There is a repulsion in the selection of the neighboring units $k$ and $\ell$ at the same sampling time $t$, as in the local pivotal method. This allows to obtain a spread sample.

(v) At each iteration, only the inclusion probability vectors $\boldsymbol{\pi}_k$ and $\boldsymbol{\pi}_\ell$ of two units $k$ and $\ell$ are updated by $\boldsymbol{\pi}_k^*$ and $\boldsymbol{\pi}_\ell^*$. So the proposition *(v)* is proven if $(\boldsymbol{\pi}_k + \boldsymbol{\pi}_\ell) = (\boldsymbol{\pi}_k^* + \boldsymbol{\pi}_\ell^*)$. The sum $(\boldsymbol{\pi}_k^* + \boldsymbol{\pi}_\ell^*)$ can be computed:

$$\begin{aligned} \boldsymbol{\pi}_k^* + \boldsymbol{\pi}_\ell^* &= \widetilde{\mathbf{S}}_k^\top (\widetilde{\mathbf{p}}_k + \lambda \mathbf{b}_k) + \widetilde{\mathbf{S}}_\ell^\top (\widetilde{\mathbf{p}}_\ell - \lambda \mathbf{b}_\ell) \\ &= \widetilde{\mathbf{S}}_k^\top \widetilde{\mathbf{p}}_k + \lambda \widetilde{\mathbf{S}}_k^\top \mathbf{b}_k + \widetilde{\mathbf{S}}_\ell^\top \widetilde{\mathbf{p}}_\ell - \lambda \widetilde{\mathbf{S}}_\ell^\top \mathbf{b}_\ell \\ &= \boldsymbol{\pi}_k + \lambda \mathbf{u}_p + \boldsymbol{\pi}_\ell - \lambda \mathbf{u}_p = \boldsymbol{\pi}_k + \boldsymbol{\pi}_\ell, \end{aligned}$$

with $\lambda = \lambda_1$ or $\lambda = -\lambda_2$.

Only one longitudinal sample $\mathbf{a}_k$ must be selected among $\mathbf{S}_k$ for each unit $k$ (Eq. (3)). This is satisfied by keeping the sum of the components of $\mathbf{p}_k$ and $\mathbf{p}_\ell$ equal to one at each step, as in Proposition 2*(ii)*. This is satisfied only under the condition that the sums of the rows of $\boldsymbol{\Pi}$ are integer, i.e. $\psi_k \in \mathbb{N}$ for all $k \in U$. During each stage, if at least one of the two units $k$ and $\ell$ does not satisfy this condition, one can simply solve the problem by adding phantom sampling times to vectors of inclusion probabilities $\boldsymbol{\pi}_k$ and $\boldsymbol{\pi}_\ell$, as proposed by Grafström et al. (2012b), to sum to an integer. This trick allows to apply the algorithm, without restriction on the $\boldsymbol{\Pi}$ matrix. Proposition 2*(iii)* and *(iv)* correspond to requirements (i) and (iii) respectively. Then, the size of the cross-sectional samples is fixed at each sampling time $t$, this can be deduced from Proposition 2*(iv)*.

## 8. Simulations

First of all, the interest of this section is to compare our method to other existing methods in different situations. However, when the probability matrix $\boldsymbol{\Pi}$ is totally unequal, no method exists to select spatiotemporal samples. In this case, the general dispersion of each sample selected using the methods SPAR and SPAF must be evaluated and criticized, without any means of comparison.

### 8.1. Spreading measures

A commonly used spatial balanced index has been developed by Stevens and Olsen (2004). This index is based on the partition of space into Voronoï polygons and is particularly effective for comparing the spatial spread of different samples of the same population. Let $s^t = \{i \in U \mid a_i^t = 1\}$ be the set of selected units index at time $t$ in a cross-sectional sample $\mathbf{a}^t$, $t \in \{1, \ldots, T\}$. The Voronoï polygon of the $i$th selected unit in $\mathbf{a}_t$ is defined as the polygon which includes all non-selected units that are closest to $i$ than all other selected units, with $i \in s^t$. Let $z_i$ be the sum of the inclusion probabilities of units included into the $i$th Voronoï polygon. Grafström et al. (2012a) assert that a best spatially balanced sample is one with $z_i = 1$ for all selected unit $i$ and suggest that the variance $B(s^t) = 1/n_t \sum_{i \in s^t}(z_i - 1)^2$ can represent a measure of spatial balance for a sample of size $n_t = \sum_{k \in U} a_k^t$. The smaller its value, the better $\mathbf{a}^t$ is spatially balanced. Since this measure depends on the spatial pattern of the population, this allows to compare the spreading of different samples from the same population. It is very useful in determining which sampling design best selects well-spread samples at each sampling time.

Moran's $I$ index is a measure of spatial autocorrelation proposed by Moran (1950). It is based on the fact that the level of spatial autocorrelation of an indicator random variable as $\mathbf{a}^t$ shows its level of spatial spreading. Because of the limitations of this index, Tillé et al. (2018) have developed a normalized version of Moran's $I$ index: the $I_B$ index. This new index version can take values from $-1$ (perfect spatial balance) to 1 (maximum concentration) and a neutral value 0. It allows to evaluate the spatial spreading and the spatial balance of a sample.

### 8.2. Biological data

To evaluate our methods, the *Centre Suisse de Cartographie de la Faune* (CSCF) provided us with a spatial biological data set. These data list *odonata* (i.e. dragonflies and damselflies) species observed in land squares in Switzerland between 1840 and 2020. Data includes 1400 land squares with an area of 1 km$^2$ and 83 different *odonata* species located in the Swiss cantons of Fribourg, Neuchâtel and Vaud. Fig. 1 represents a map with the 1400 land squares.

We focused on a sampling design to study the rare species. The importance of a square is related to the number of rare species observed there. Let $\mathbf{M}$ denote the matrix that contains in rows the 1400 land squares and in columns the 83 species. Matrix $\mathbf{M}$ is composed of 0 s and 1 s that specify if a species has already been observed within a square. Consider column vector $\mathbf{g} \in \mathbb{N}^{83}$ that contains the inverse of the species occurrence rate. In other words, $g_i$ is equal to the inverse of the sum of the $i$th column of $\mathbf{M}$, $i = 1, \ldots, 83$. Consider also vector $\mathbf{c} \in \mathbb{R}^{1400}$, such that $\mathbf{c} = \mathbf{g}^\top \mathbf{M}$, containing a square importance measure based on rare species.

### 8.3. Results

To evaluate our sampling methods, we consider the problem of selecting a spatiotemporal sample of land squares that are both spread and with fixed size at each sampling time. We considered $T = 3$ sampling times. The simulations were run using the 'SpotSampling' R package (Eustache et al., 2020).

For the first scenario of simulations, we considered equal inclusion probabilities at each sampling time, with $n^1 = 200$, $n^2 = 250$ and $n^3 = 300$. The columns of $\boldsymbol{\Pi}$ are then proportional and the sum of its rows are all equal to 15/28. By taking this structure of inclusion probabilities, the method of Wang and Zhu can be applied and then compared to our methods. For the second scenario, inclusion

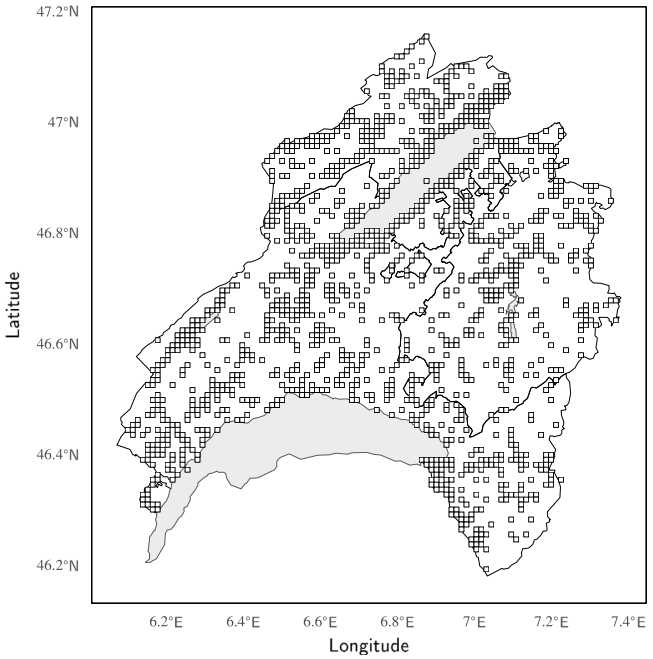



**Fig. 1.** Representation of land squares where *odonata* species have been observed within the Swiss cantons of Fribourg, Neuchâtel and Vaud. Lakes are filled in light gray.

probabilities are chosen totally unequal, with the idea of increasing emphasis on rare species over time. For the first sampling time, inclusion probabilities in $\boldsymbol{\pi}^1$ are all equal. For the second one, $\boldsymbol{\pi}^2$ is proportional to the number of species in land squares. Next, vector $\boldsymbol{\pi}^3$ of inclusion probabilities of the last sampling time is proportional to vector $\mathbf{c}$ to give more importance to squares potentially containing rare species. Sample sizes are $n^1 = n^2 = n^3 = 250$.

Fig. 2 shows a spatiotemporal sample selected with the SPAF method using equal inclusion probabilities from the first scenario of the simulations. Land squares filled in light gray represent the initial spread set selected using the preliminary step described in Section 5. This first step can be applied because the sums of the rows of $\boldsymbol{\Pi}$ are not greater than 1. This initial set is represented in Fig. 2 at the top left. Land squares definitively selected with the SPAF sampling are filled in black. Fig. 2 at top right, bottom left and bottom right respectively represent the selected samples at sampling times 1, 2 and 3.

We performed 10'000 simulations. For each compared method evaluated, the average values of the spread measures $I_B$ and $B$ during the simulations are calculated for each sampling time $t$. The results are summarized in Table 1. In the first scenario, SPAR, SPAF and Wang and Zhu methods are comparable in terms of $I_B$ and $B$ measures. For the second scenario, spreading measures show that samples selected with the SPAF sampling method are well spread. However, spreading of samples selected with SPAR is better than those selected with SPAF. This is easily explained by the fact that SPAR generates samples with random size.

## 9. Conclusion

The selection of spatiotemporal samples is a complex problem particularly when one wants to impose simultaneously constraints of temporal and spatial spreading. However, these constraints are important to optimize the collection of information.

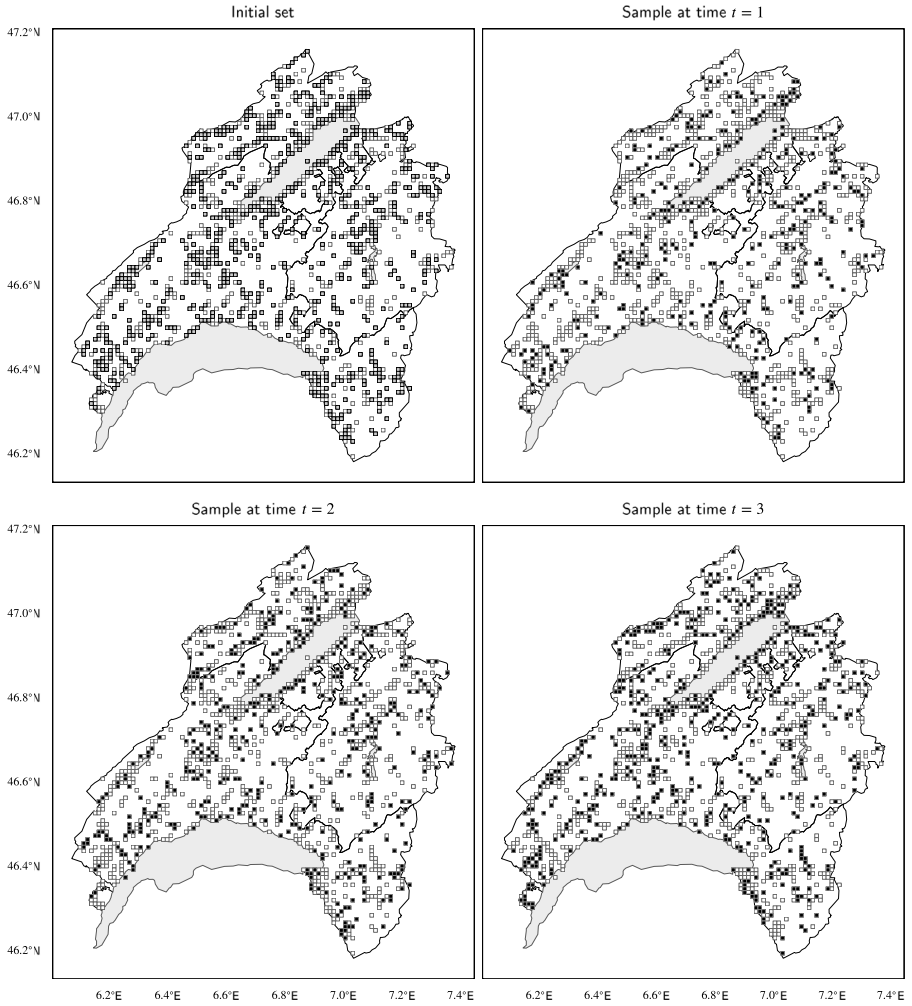



**Fig. 2.** Figures showing a spatiotemporal sample selected with the spot sampling method. The sampling population are composed of 1400 land squares located in the Swiss cantons of Fribourg, Neuchatel and Vaud. Three sampling times are considered with sample size respectively equal to 200, 250 and 300. Inclusion probabilities are equal at each sampling time. An initial spread set is selected before applying the SPAF method. Land squares that are not excluded from the population during this preselection are filled in light gray in figure at top left. Land squares definitively selected with the SPAF sampling are filled in black. Figures at top right, bottom left and bottom right respectively represent the selected samples at sampling times 1, 2 and 3.

In this paper, we have solved the problem in the most general case, i.e. when inclusion probabilities are unequal and variable over time. Two spatiotemporal sampling methods are described. They provide a random spatiotemporal sample that is well spread in time and in space. The first one, called SPAR, selects a sample at each sampling time that is of random size while the second one, the SPAF method, also allows to control their sizes. The SPAF method gives a random spatiotemporal sample containing longitudinal samples of fixed size and well spatially spread.

The proposed sampling methods are evaluated on spatial biological data given by the *Centre Suisse de Cartographie de la Faune*. Simulations show that samples selected with SPAR and SPAF are well spread in space. The SPAF method is the first spatiotemporal sampling method that allows to

**Table 1**
Spreading measures of spatiotemporal samples based on 10,000 simulations on the spatial biological data from CSCF. The WZ method cannot be applied on totally unequal inclusion probabilities.

| | Sampling design | | | | |
|---|---|---|---|---|---|
| | Equal probabilities | | | Unequal probabilities | |
| | SPAR | SPAF | WZ | SPAR | SPAF |
| $I_B$ | | | | | |
| $t = 1$ | −0.388 | −0.374 | −0.389 | −0.386 | −0.360 |
| $t = 2$ | −0.393 | −0.378 | −0.395 | −0.264 | −0.233 |
| $t = 3$ | −0.387 | −0.357 | −0.386 | −0.309 | −0.273 |
| $B$ | | | | | |
| $t = 1$ | 0.116 | 0.123 | 0.115 | 0.130 | 0.140 |
| $t = 2$ | 0.128 | 0.135 | 0.127 | 0.151 | 0.165 |
| $t = 3$ | 0.138 | 0.147 | 0.138 | 0.145 | 0.158 |

SPAR, spatiotemporal sampling with random sample sizes; SPAF, spatiotemporal sampling with fixed sample sizes; WZ, Wang and Zhu method.

consider unequal and time-varying probabilities. All of these results indicate that the spot method is very efficient to select a well spread sample. These methods can be very easily used by means of the 'SpotSampling' R package (Eustache et al., 2020).

## Acknowledgments

The authors warmly thank members of the *Centre Suisse de Cartographie de la Faune* for the trust they have given them by making available a set of data to evaluate our method. They are grateful for the time spent by Lionel Qualité and the reviewers to make relevant corrections and remarks on this paper. They also thank Zhonglei Wang and Zhengyuan Zhu who shared their R code, which made it possible to compare their methods.

## Appendix A. The local pivotal method

Consider a vector of inclusion probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)^\top$. The local pivotal method is described in Algorithm 5.

**Algorithm 5** Local pivotal method

Repeat the following steps until all elements in $\boldsymbol{\pi}$ are equal to zero or one.

1. Select two neighboring units $k \in U$ and $\ell \in U$ that still have non-integer inclusion probabilities $\pi_k$ and $\pi_\ell$.
2. Compute $\lambda_1 = \min(\pi_k + \pi_\ell, 1)$ and $\lambda_2 = \max(0, \pi_k + \pi_\ell - 1)$.
3. Update the probabilities $\pi_k$ and $\pi_\ell$ such that

$$\begin{cases} \pi_k \leftarrow \lambda_1, \pi_\ell \leftarrow \lambda_2 & \text{with probability } (\pi_k - \lambda_2)/(\lambda_1 - \lambda_2), \\ \pi_k \leftarrow \lambda_2, \pi_\ell \leftarrow \lambda_1 & \text{with probability } (\lambda_1 - \pi_k)/(\lambda_1 - \lambda_2). \end{cases}$$

In Algorithm 5, the definition of the neighborhood of unit $k$ may be based on spatial coordinates. Several variants of the method exist but they only differ by the way of selecting the two neighboring units from spatial coordinates. At each step, if $\pi_k$ is increased, $\pi_\ell$ is decreased and reciprocally. This repulsion in the selection of neighboring units allows to obtain a spread sample at the end of the algorithm.

## Appendix B. The cube method

Consider a vector of inclusion probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)^\top$. The cube method allows to generate a random vector $\mathbf{a} = (a_1, \ldots, a_N)^\top$ of Bernoulli variables such that $\mathrm{E}_p(\mathbf{a}) = \boldsymbol{\pi}$ and which is balanced on the totals of auxiliary variables. A sample $\mathbf{a}$ is said to be balanced on the $J$ auxiliary variables if it satisfies equation

$$\mathbf{M}^\top \mathbf{a} = \mathbf{M}^\top \boldsymbol{\pi}, \tag{5}$$

where $\mathbf{M} = (\mathbf{x}_1/\pi_1, \ldots, \mathbf{x}_N/\pi_N)^\top$. Expression (5) is called the system of balancing equations. Sometimes, this equality cannot be exactly satisfied.

The cube method is divided into two phases: the *flight phase* and the *landing phase*. The flight phase of the cube method is a random walk from $\boldsymbol{\pi}$ to $\boldsymbol{\pi}^*$ such that $\mathrm{E}_p(\boldsymbol{\pi}^*) = \boldsymbol{\pi}$, $\mathbf{M}^\top \boldsymbol{\pi}^* = \mathbf{M}^\top \boldsymbol{\pi}$ and $\#\{k \mid \boldsymbol{\pi}^* \notin \{0, 1\}\} \leq J$. So, there remain at most $J$ (i.e. the number of columns of $\mathbf{M}$) non-integer values in $\boldsymbol{\pi}^*$ at the end of the flight phase. The landing phase consists of rounding to 0 or 1 probabilities of the remaining units. If there are not a lot of remaining units, a solution satisfying exactly the balancing constraints can be found using linear programming. Otherwise, the constraints must be relaxed. One possibility consists of removing balancing variables one by one until a sample satisfying remaining balancing constraints can be selected. This landing phase by suppression of variables requires a priority order on the variables.

## Appendix C. The flight phase of the local cube method

Consider a vector of inclusion probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)^\top$. The flight phase of the local cube method is described in Algorithm 6.

**Algorithm 6** Flight phase of the local cube method

Repeat the following steps until there remains less than $(J + 1)$ non-integer values in $\boldsymbol{\pi}$.

1. Select $(J + 1)$ neighboring units with a non-integer inclusion probability in $\boldsymbol{\pi}$.
2. Apply the flight phase of the cube method only on these units and update $\boldsymbol{\pi}$.

At each step, only $(J + 1)$ neighboring units of $k$ are considered. At the end of the flight phase of the cube method, the updated vector of $\boldsymbol{\pi}$ contains mainly 0 s and 1 s, except for at most $J$ components.

## Appendix D. Systematic sampling

Consider a vector of inclusion probabilities $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_N)^\top$ and suppose that $\pi$ sums to an integer number $n$, i.e. $\psi = n$, with $n \in \mathbb{N}$. The usual systematic method is described in Algorithm 7.

**Algorithm 7** Systematic sampling

1. Compute the cumulative inclusion probabilities $V_k = \sum_{j \leq k} \pi_j$ with $k = 1, \ldots, N$ and $V^0 = 0$.
2. Generate a uniform continuous random variable $u$ on interval [0,1].
3. Next, for $i = 1, \ldots, n$, select the units $k(i)$ such that $V_{k(i)-1} \leq u + i - 1 < V_{k(i)}$.

Let $V_k = \sum_{j \leq k} \pi_j$ denote the cumulative inclusion probability, with $k = 1, \ldots, N$. Define $v_j \in [0, 1]$ such that $V_j \bmod 1 = v_j$, for $j = 0, \ldots, N - 1$. Let also $v_{(j)}$ be the $v_j$s sorted by increasing order with $v_{(N)} = 1$. Each interval $[v_{(k-1)}, v_{(k)}[$ corresponds to the selection of a unique sample and the length of this interval is the probability of selecting this sample, for $t = 1, \ldots, T$. The probability associated to each sample is thus $(v_{(k)} - v_{(k-1)})$.

## References


Bellhouse, D., 1977. Some optimal designs for sampling in two dimensions. Biometrika 64 (3), 605–611.
Bellhouse, D.R., 1988. Systematic sampling. In: Krishnaiah, P.R., Rao, C.R. (Eds.), Handbook of Statistics Volume 6: Sampling. Elsevier/North-Holland, New York, Amsterdam, pp. 125–145.
Bellhouse, D.R., Sutradhar, B.C., 1988. Variance estimation for systematic sampling when autocorrelation is present. Statist.: J. Inst. Statist. 37, 327–332.
Deville, J.-C., Tillé, Y., 1998. Unequal probability sampling without replacement through a splitting method. Biometrika 85, 89–101.
Deville, J.-C., Tillé, Y., 2000. Selection of several unequal probability samples from the same population. J. Statist. Plann. Inference 86, 215–227.
Deville, J.-C., Tillé, Y., 2004. Efficient balanced sampling: The cube method. Biometrika 91, 893–912.
Dickson, M.M., Tillé, Y., 2016. Ordered spatial sampling by means of the traveling salesman problem. Comput. Statist. 31 (4), 1359–1372.
Eustache, E., Jauslin, R., Tillé, Y., 2020. The R Package 'SpotSampling'. CRAN project, Vienna.
Fattorini, L., Marcheselli, M., Pisani, C., Pratelli, L., 2021. Design-based properties of the nearest neighbour spatial interpolator and its bootstrap mean squared error estimator. Biometrics 113, 463–475.
Grafström, A., 2011. 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. Statist. 3 (1), 36–41.
Grafström, A., Lundström, N.L.P., Schelin, L., 2012a. Spatially balanced sampling through the pivotal method. Biometrics 68 (2), 514–520.
Grafström, A., Matei, A., 2018. Spatially balanced sampling of continuous populations. Scand. J. Stat. 45 (3), 792–805.
Grafström, A., Matei, A., Qualité, L., Tillé, Y., 2012b. Size constrained unequal probability sampling with a non-integer sum of inclusion probabilities. Electron. J. Stat. 6, 1477–1489.
Grafström, A., Saarela, S., Ene, L.T., 2014. Efficient sampling strategies for forest inventories by spreading the sample in auxiliary space. Can. J. Forest Res. 44 (10), 1156–1164.
Grafström, A., Tillé, Y., 2013. Doubly balanced spatial sampling with spreading and restitution of auxiliary totals. Environmetrics 14 (2), 120–131.
Hájek, J., 1981. Sampling from a Finite Population. Marcel Dekker, New York.
Horvitz, D.G., Thompson, D.J., 1952. A generalization of sampling without replacement from a finite universe. J. Amer. Statist. Assoc. 47, 663–685.
Iachan, R., 1982. Systematic sampling: A critical review. Internat. Statist. Rev. 50, 293–303.
Iachan, R., 1983. Asymptotic theory of systematic sampling. Ann. Statist. 11, 959–969.
Isaki, C.T., Fuller, W.A., 1982. Survey design under a regression population model. J. Amer. Statist. Assoc. 77, 89–96.
Jauslin, R., Tillé, Y., 2020. Spatial spread sampling using weakly associated vectors. J. Agric. Biol. Environ. Stat. 25 (3), 431–451.
Khavarzadeh, R., Mohammadzadeh, M., Mateu, J., 2018. A simple two-step method for spatio-temporal design-based balanced sampling. Stoch. Environ. Res. Risk Assess. 32 (2), 457–468.
Kincaid, T.M., Olsen, A.R., 2015. spsurvey: Spatial Survey Design and Analysis. R package version 3.1.
Lohr, S.L., 2021. Sampling: Design and Analysis. Chapman and Hall/CRC.
Madow, W.G., 1949. On the theory of systematic sampling, II. Ann. Math. Stat. 20, 333–354.
Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. Biometrika 37 (1/2), 17–23.
Nedyalkova, D., Tillé, Y., 2008. Optimal sampling and estimation strategies under linear model. Biometrika 95, 521–537.
Neyman, J., 1934. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. J. R. Statist. Soc. 97, 558–606.
Pea, J., Qualité, L., Tillé, Y., 2007. Systematic sampling is a minimal support design. Comput. Statist. Data Anal. 51, 5591–5602.
Quenouille, M.H., 1949. Problems in plane sampling. Ann. Math. Stat. 20, 355–375.
Rivest, L.-P., Ebouele, S.E., 2020. Sampling a two dimensional matrix. Comput. Statist. Data Anal. 149, 106971.
Stevens, D.L., Olsen, A.R., 1999. Spatially restricted surveys over time for aquatic resources. J. Agric. Biol. Environ. Stat. 4, 415–428.
Stevens, D.L., Olsen, A.R., 2003. Variance estimation for spatially balanced samples of environmental resources. Environmetrics 14 (6), 593–610.
Stevens, D.L., Olsen, A.R., 2004. Spatially balanced sampling of natural resources. J. Amer. Statist. Assoc. 99 (465), 262–278.
Tillé, Y., 2020. Sampling and Estimation from Finite Populations. Wiley, Hoboken.
Tillé, Y., Dickson, M.M., Espa, G., Giuliani, D., 2018. Measuring the spatial balance of a sample: A new measure based on the moran's $I$ index. Spat. Statist. 23, 182–192.
Tillé, Y., Ecker, K., 2013. Complex national sampling design for long-term monitoring of protected dry grasslands in Switzerland. Environ. Ecol. Stat. 21, 1–24. http://dx.doi.org/10.1007/s10651-013-0263-2.
Valliant, R., Dorfman, A.H., Royall, R.M., 2000. Finite Population Sampling and Inference: A Prediction Approach. Wiley, New York.
Wang, J., Gao, B., Stein, A., 2020. The spatial statistic trinity: A generic framework for spatial sampling and inference. Environ. Model. Softw. 134, 104835.
Wang, J.-F., Zhang, T.-L., Fu, B.-J., 2016. A measure of spatial stratified heterogeneity. Ecol. Indic. 67, 250–256.
Wang, Z., Zhu, Z., 2019. Spatiotemporal balanced sampling design for longitudinal area surveys. J. Agric. Biol. Environ. Stat. 24 (2), 245–263.
Zhao, X., Grafström, A., 2020. A sample coordination method to monitor totals of environmental variables. Environmetrics 31 (6).