Forest Ecosystems

Comparison of estimators of variance for forest inventories with systematic sampling - results from artificial populations

Steen Magnussen[1*], Ronald E. McRoberts[2], Johannes Breidenbach[3], Thomas Nord-Larsen[4], Göran Ståhl[5], Lutz Fehrmann[6] and Sebastian Schnell[7]

Abstract

Background: Large area forest inventories often use regular grids (with a single random start) of sample locations to ensure a uniform sampling intensity across the space of the surveyed populations. A design-unbiased estimator of variance does not exist for this design. Oftentimes, a quasi-default estimator applicable to simple random sampling (*SRS*) is used, even if it carries with it the likely risk of overestimating the variance by a practically important margin. To better exploit the precision of systematic sampling we assess the performance of five estimators of variance, including the quasi default. In this study, simulated systematic sampling was applied to artificial populations with contrasting covariance structures and with or without linear trends. We compared the results obtained with the *SRS*, Matérn's, successive difference replication, Ripley's, and D'Orazio's variance estimators.

Results: The variances obtained with the four alternatives to the *SRS* estimator of variance were strongly correlated, and in all study settings consistently closer to the target design variance than the estimator for *SRS*. The latter always produced the greatest overestimation. In populations with a near zero spatial autocorrelation, all estimators, performed equally, and delivered estimates close to the actual design variance.

Conclusion: Without a linear trend, the *SDR* and *DOR* estimators were best with variance estimates more narrowly distributed around the benchmark; yet in terms of the least average absolute deviation, Matérn's estimator held a narrow lead. With a strong or moderate linear trend, Matérn's estimator is choice. In large populations, and a low sampling intensity, the performance of the investigated estimators becomes more similar.

Keywords: Spatial autocorrelation, Linear trend, Model based, Design biased, Matérn variance, Successive difference replication variance, Geary contiguity coefficient, Random site effects

Introduction

Forest inventories have a long history of using systematic sampling (Spurr 1952, p 379) that continues to this date at both local, regional, and national levels (Brooks and Wiant Jr 2004; Kangas and Maltamo 2006; Nelson et al. 2008; Tomppo et al. 2010; Vidal et al. 2016). Since forests exhibit non-random spatial structures (Sherrill et al. 2008; Alves et al. 2010; von Gadow et al. 2012; Pagliarella et al. 2018), the main benefit of a uniform sampling intensity across a population under study (i.e. spatial balance) is an anticipated lower variance in an estimate of the population mean (total). However, the lack of a design-unbiased estimator of variance for the mean (total) remains a detractor (Gregoire and Valentine 2008, p 55). We do not have a design-unbiased estimator of variance for systematic sampling because the sampling locations are fixed by one independent random selection of a starting point and a sampling interval (*d*). With only one random draw, the systematic sample can be regarded as a random selection of one cluster with an undefined design-based variance (Wolter 2007, p 298).

* Correspondence: steen.magnussen@canada.ca
[1]506 West Burnside Road, Victoria, BC V8Z 1M5, Canada
Full list of author information is available at the end of the article

Without a design-unbiased estimator of variance, it becomes a challenge to quantify the advantage of systematic sampling, and to compute reliable confidence intervals for estimated population parameters. The wide-spread use of a variance estimator for *SRS* without replacement (Särndal et al. 1992, p 28) masks the advantage (efficiency) since this estimator tends to overestimate the actual variance (Wolter 1984; Fewster 2011). An overestimation that is, possibly, regarded as less problematic than an underestimation, and often referred to as a "conservative estimate".

The bias in the variance estimator for *SRS* when applied to data from a single systematic sample was recognized early on in Scandinavian countries by Lindeberg (1924), Langsæter (1926), and Näslund (1930), and in North America by Osborne (1942), and Hasel (1942). Lindeberg, Langsæter, and Näslund also proposed new estimators of variance that generated more realistic estimates of variance for line-transect surveys (Ibid.). Variations of these estimators were later credited to others (Wolter 2007, ch. 8.2).

To convince an inventory analyst – with sample data collected under a systematic design – to employ an alternative to the estimator for *SRS* requires assurance that the alternative is nearly design-unbiased. That is, the expected value of the alternative estimator, over all possible (K) systematic samples from a finite population, is equal to or close to the variance among the K sample means (Madow and Madow 1944). Assurances of this kind will have to come from simulated systematic sampling from actual or artificial populations.

The lack of a design-unbiased estimator of variance means that any applied estimator is biased for the actual design variance (Opsomer et al. 2012; Fattorini et al. 2018b). Variance estimators used in lieu of the design variance may carry the assumption that the sampling design is ignorable, or that any explicitly or implicitly stated model regarding the population is true (Gregoire 1998; Magnussen 2015). For example, when the estimator for *SRS* is applied to a systematic sample from a finite population, the design is ignored, and the variance is computed under the assumption that the sample values are independent.

In this study, we compare the performance of four alternatives to the estimator of variance for *SRS* in a suite of artificial populations with contrasting covariance structures and with or without a global linear trend. The performance in actual forest populations is deferred to a forthcoming study. The alternatives achieved – with respect to accuracy – a top ranking amongst 11 candidates in a preliminary study with 27 superpopulations described in Magnussen and Fehrmann (2019).

Although our primary focus is on systematic sampling designs with small populations (to expedite computations),

and higher than practiced sampling intensities, we demonstrate that a ranking of the relative performances of estimators will be preserved in larger populations and a lower sampling intensity. We extend the same expectations to non-aligned and quasi systematic designs (Särndal et al. 1992, 3.4.2; Grafström et al. 2014; Mostafa and Ahmad 2017; Wilhelm et al. 2017), and possibly the random tessellated stratified design (Stevens and Olsen 2004; Fattorini et al. 2009; Magnussen and Nord-Larsen 2019).

Materials and methods
Artificial populations
The four alternative estimators of variance are evaluated in realizations of two superpopulations: one ($\mathfrak{U}1$) with a stronger positive spatial autocorrelation between units in a single sample, and the other ($\mathfrak{U}0$) with a near zero spatial autocorrelation. Global linear trends ('strong', 'moderate', 'weak', or 'none') are present in both $\mathfrak{U}1$ and $\mathfrak{U}0$. Populations without a linear trend are weakly stationary (Cressie 1993, p 53). An attractive estimator of variance will generate estimates that are close to the actual variance regardless of the strength of a spatial autocorrelation or the presence of a global trend. In practice, the effects of a significant trend can be mitigated by formulating a model (parametric or non-parametric) for the trend (Valliant et al. 2000, p 57; Opsomer et al. 2012) or stratification (Dahlke et al. 2013).

The two superpopulations $\mathfrak{U}1$ and $\mathfrak{U}0$ are composed of $N = 57,600$ equal size (area) spatial units arranged in a regular array with 240 rows and 240 columns. Edge effects is therefore not an issue in our study (Gregoire and Scott 2003). In an attempt to generate unit level autocorrelation in values of y compatible with forest structures, we generated random realizations (populations) U_1, U_2, ... from $\mathfrak{U}1$ and $\mathfrak{U}0$ with three additive random 'site' effects ($s1$, $s2$, $s3$), operating at different spatial scales, plus unit-level random noise. The number of random spatial site effects is arbitrary. We know that forest attribute values depend on a multitude of factors operating at different spatial scales (Weiskittel et al. 2011). We consider three levels of site effects (e.g. soil, climate, and management) in our simulations of forest populations with a complex spatial structure.

To generate a site effect, the population under study was tessellated into a set of convex polygons (Møller 1994). Then a site effect was assigned to each polygon by a random draw from a distribution specified for the site effect in question. All units with at least half their area in a polygon inherit the site effect of the polygon. The number, size, and centroids of polygons for a site effect varies from one realization of a superpopulation to the next according to random draws from distributions for the number and placement of polygons. A complete population was then composed of three spatial layers of

polygon specific site effects (Fig. 1), and one complete (240×240) layer of unit-level random noise.

Accordingly, the unit-level value y_{ij} in the ith row and jth column ($i, j = 1, ..., 240$) in a realization from a superpopulation is the sum of three random site effects $s1$, $s2$, and $s3$, a global trend τ, and random noise (e). We have

$$y_{ij} = s1_{ij} + s2_{ij} + s3_{ij} + \tau_{ij} + e_{ij} \qquad (1)$$

where sT_{ij} ($T = 1, 2, 3$) is the random site effect associated with the polygon in which unit ij resides, τ_{ij} is a unit specific trend effect, and e_{ij} is an independent random Gaussian noise. All units within a polygon share the site effect assigned to the polygon, which gives rise to a positive covariance among unit site effects within the polygon (Searle et al. 1992, ch. 11.2). To control the total variance in a study variable, the sum of site effects and random noise was standardized to a mean of zero and a variance of one. Technical details are deferred to the Additional file 1.

In addition to the spatial autocorrelation, we simulated three levels of a non-null global linear trend (Table 1) in addition to the simulations without a trend ($\tau_{ij} = 0 \; \forall \; \{i, j\}$).

Six random realizations of population values of y_{ij} without a trend are shown in Fig. 2. They convey, as intended, a complex mosaic of the overlapping site effects. The visual resemblance of different realizations from a single superpopulation is low.

Sample-based maximum likelihood estimates of the autocorrelation function (*acf*, Anderson 1976, p 4) in the six populations in Fig. 2 are given in Fig. 3. One *acf* is shown for each of the possible samples under a given design. A considerable sample-to-sample variation is visible in some illustrations.

There is no variance heteroscedasticity in the simulated noise. To gauge its impact, we ran separate simulations with heteroscedasticity but only sketch the results in the discussion.

The population size in simulation studies are typically orders of magnitude smaller than actual finite populations.
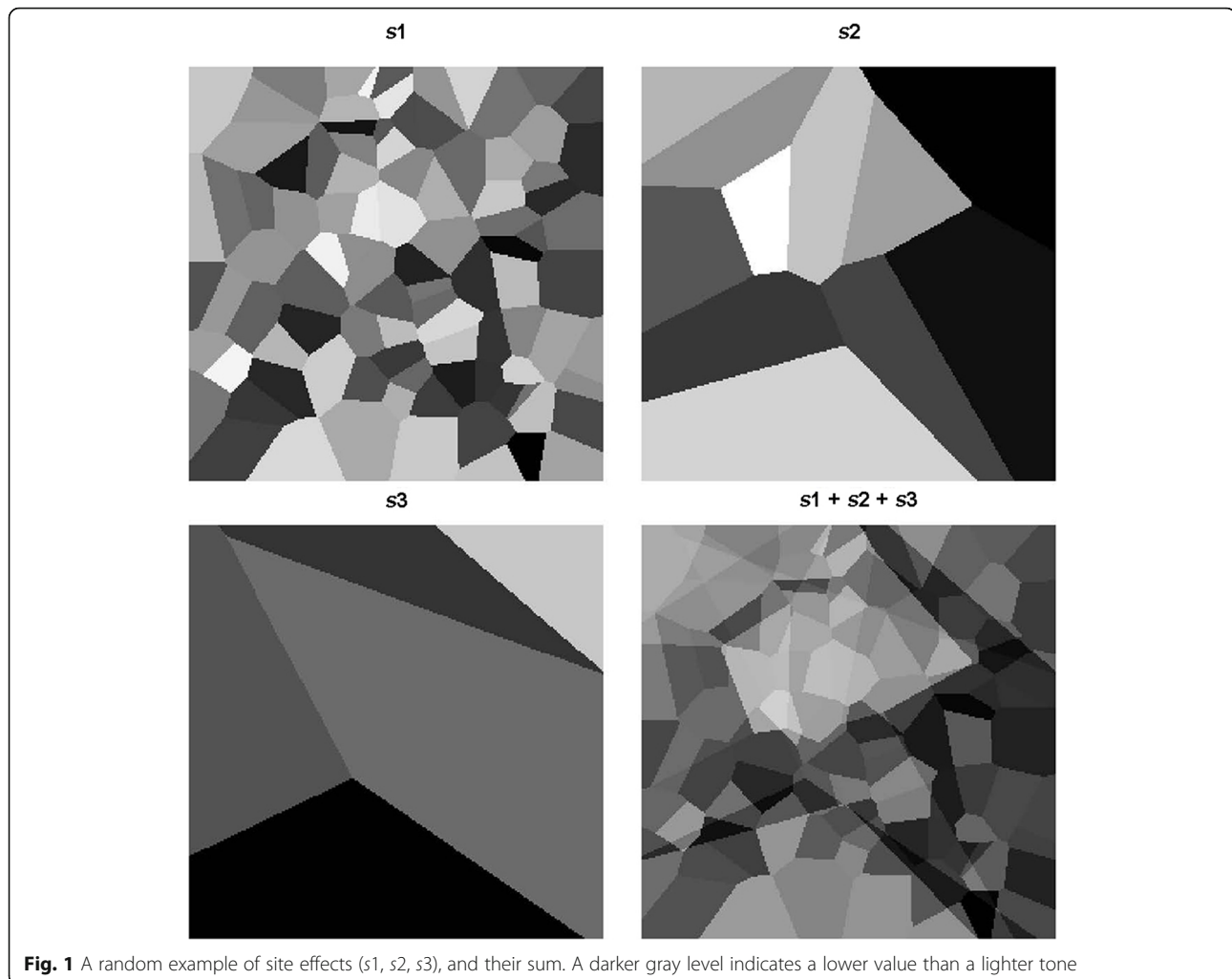

Fig. 1 A random example of site effects (*s1*, *s2*, *s3*), and their sum. A darker gray level indicates a lower value than a lighter tone

Table 1 Linear trend models and unit level trend components τ_{ij} ($i, j = 0.5, 1.5, \ldots, 239.5$; cf. (1))

Trend	τ_{ij}	$\min(\tau_{ij})$	$\text{mean}(\tau_{ij})$	$\max(\tau_{ij})$
Strong	$(i + j - 241)/120$	−2	0	2
Moderate	$(i + j - 241)/240$	−1	0	1
Weak	$(i + j - 241)/480$	−0.5	0	0.5

For the purpose of evaluating the relative performance of alternative variance estimators against a design variance, it is only important to stage: i) gradients of a spatial autocorrelation as done by choices of sample size; and ii) linear trends that will interact with sample size. A testing in a series of increasingly larger populations and across multiple spatial covariance structures is necessary if the relative performances of our estimators of variance are sensitive to sample size and/or trends. To assuage concerns about population size and sample size, we extended the simulations to include larger populations and a smaller sample size.

Sampling designs

Four systematic sampling designs are employed in the main study. Each design is defined by the sampling interval (d) in units in both of the two cardinal directions defining the population (here rows and columns) and a starting position (Cochran [1977], ch. 8.1; Särndal et al. [1992], ch. 3.4.1; Fuller [2009], ch. 1.2.4). We have $d = 6, 8, 10$, and 12. With a population matrix structure of 240 rows and 240 columns, the corresponding sample sizes were $n = 1600$ ($d = 6$), 900 ($d = 8$), 576 ($d = 10$),

and 400 ($d = 12$). We simulated all possible systematic samples (K) under a given design. The K starting positions by row and column were (d_i, d_j), (d_i, d_j) = 1, ..., d. Accordingly, $K = d^2$ or 36, 64, 100, and 144 for the designs with $d = 6, 8, 10$, and 12. All K samples for a fixed sample size were executed and replicated 30 times, each time with K samples from a new random realization of a superpopulation ($\mathfrak{U}1$ or $\mathfrak{U}0$). Hence our results come from 2 (*superpopulations*) × 4 (*linear trends*) × 4 (*sample sizes*) × 30 = 960 random realizations (480 from $\mathfrak{U}1$ and 480 from $\mathfrak{U}0$). With 30 realizations from a superpopulation, the relative standard error of the mean of a design variance was approximately 3% for sample sizes 400 and 576, and 5% for sample sizes 900 and 1600.

A sampling design was implemented by selecting all possible (K) different (or non-identical) systematic samples under the given sampling interval d. Specifically, we first divide a 240 × 240 population into $n = (240/d)^2$ square blocks each with d rows and d columns. To select a single systematic sample, one would pick a random integer (k) from the set {1, ..., $K = d^2$) and then select one unit at position k from each of the n blocks. An example with $d = 6$, and $k = 4$ and $k = 20$ is in Fig. [4].

Note, Thompson ([1992]) defines a systematic sampling by primary and secondary sampling units. For designs with one primary unit and n secondary units, as the case is here, and in most natural resource surveys, we can, without consequence, dispense with the notion of primary sampling units, consider the secondary units as sample units, and take n as sample size (Thompson [1992], p 113).
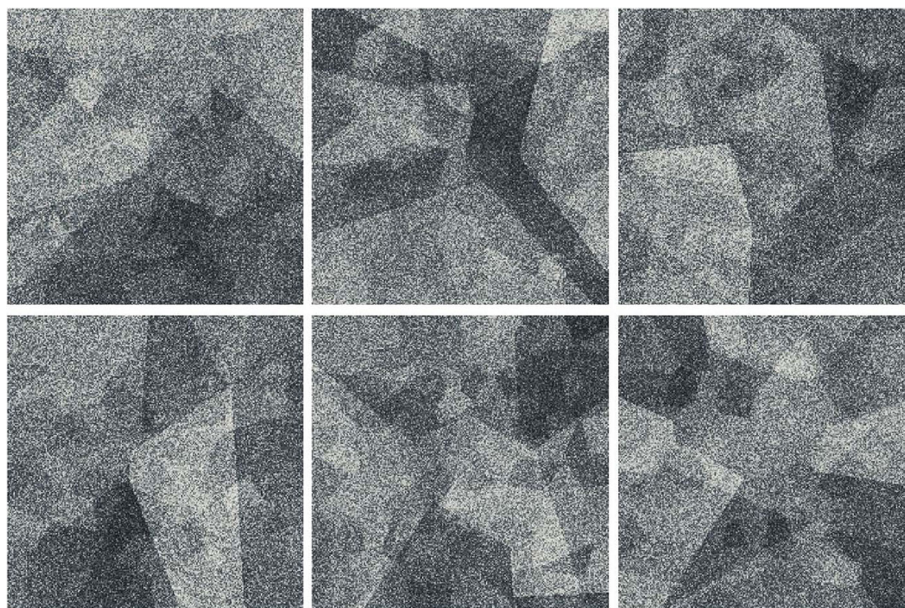

Fig. 2 Six random realizations of the superpopulation $\mathfrak{U}1$ (size 240 × 240 units) with an autocorrelation but no trend in unit level values (y_{ij}). The gray levels indicate scaled values of the study variable with darker tones for smaller values, and lighter tones for greater values
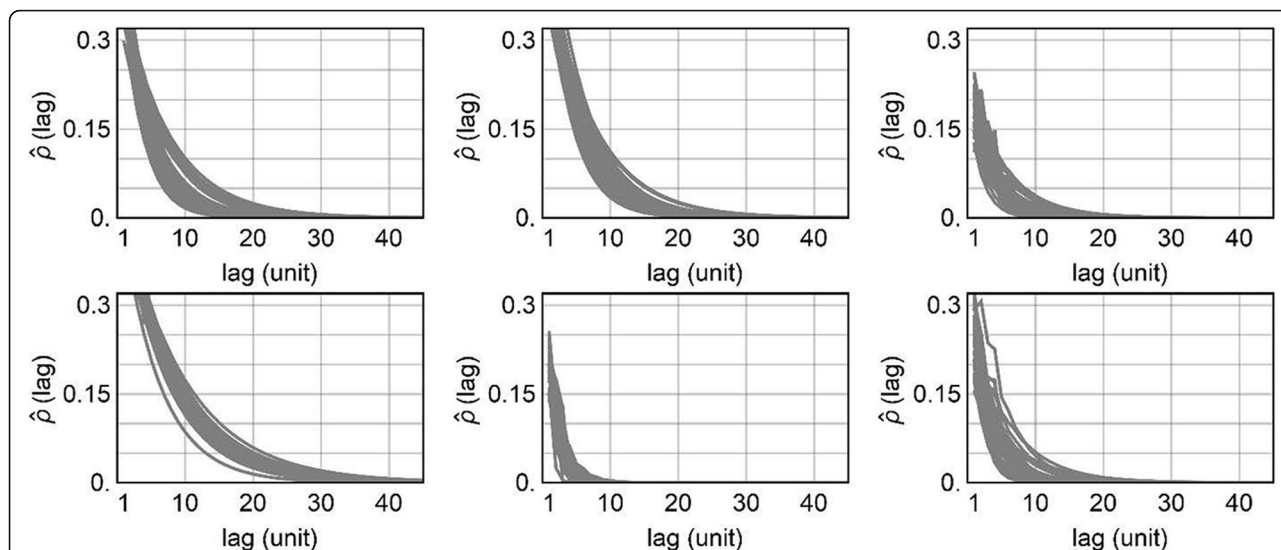
Fig. 3 Sample-based maximum likelihood estimates of the autocorrelation functions (*acf*) in a fractional Gaussian noise process (cf. (7)). The examples are from the populations in Fig. 2. The horizontal axis is the lag in units and the vertical axis is the autocorrelation. One *acf* is drawn for each possible sample under a given design

Supplementary populations and sample designs

A population size of $240 \times 240 = 57{,}600$ is orders of magnitude smaller than the size of actual finite regional or national forest populations. Conversely, even a sampling intensity of $n/N = 400/57{,}600$ or 0.7% is an order of magnitude greater than in practice. To augment the practical relevancy of our simulations, we gauged the impact of reducing the sample size to $n = 100$ in trendless populations with a spatial autocorrelation and sizes $N = 57{,}600$ unit (as in the main study), $N = 230{,}400$ units in a 480×480 array, and $N = 921{,}600$ units in a 960×960 array. The site effects were preserved at the levels detailed for the main study, but the number of polygons carrying a site specific effect was either defined as for the 240×240 unit populations in the main study, or doubled for $N = 230{,}400$ units, or quadrupled for $N = 921{,}600$ units. Thus the sample autocorrelation functions driving the variances

will depend exclusively on the sampling interval ($d = 24$, 48, or 96), the size (number) of the site polygons and their overlaps. Results with the *RIP* estimator of variance were dropped in consideration of the time required to compute the results with this estimator.

Variance estimators

In accordance with the populations under consideration, the variance estimators considered are cast for finite populations composed of N units. For these populations under a given systematic design there is a finite number (K) of distinct (non-overlapping) samples. With minor modifications the estimators also apply to infinite (continuous) populations of sample locations (points), but here $K = \infty$ {Mandallaz 2008 #10986} and there is no finite population correction in the variance estimators.
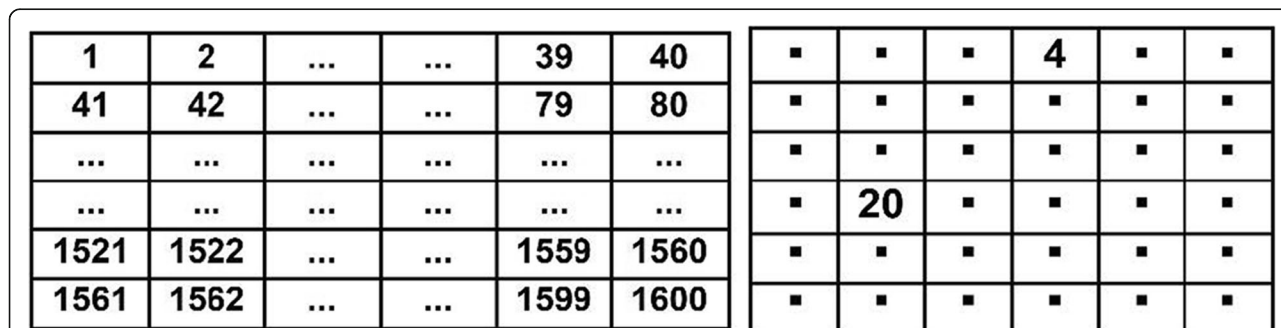

Fig. 4 Execution of a systematic sampling design with a sampling interval of $d = 6$ ($n = 1600$) from a population composed of 57,600 units arranged in an array with 240 rows and 240 columns. Left: the $n = 40 \times 40 = 1600$ sampling blocks. Right: A sampling block with indication of the position of the 4th and 12th of the $K = 36$ possible samples

Design variance

The design-based variance (*DES*) for systematic sampling in a finite population (Madow and Madow 1944) is

$$V_{DES}(\bar{y}_k) = K^{-1} \sum_{k=1}^{K} (\bar{y}_k - \bar{\bar{y}})^2, k = 1, ..., K \qquad (2)$$

where $\bar{y}_k$ is the mean of y in the *k*th systematic sample, $\bar{\bar{y}}$ is the population mean of y, and K is the number of possible samples under the design and population under study. To compute the design-based variance in Eq. (2), the sample mean from each of the K possible samples under a systematic sampling design must be known. Considering the finite populations in our simulation as described above, we have complete knowledge about the population and no uncertainty in the mean (total). Hence the design variance in Eq. (2) only serves as a benchmark in analytical developments, and in simulation studies like ours, where the value of y is known for every unit in a population under study.

Variance estimator for simple random sampling

The *SRS* estimator of variance – when applied to a sample selected under a systematic design – ignores the actual (spatial) ordering of the sampled units, and, by extension, any covariance between these units. Let y_i denote the *i*th unit in one of the K possible samples obtained under a systematic design. For a systematic sample of size *n*, taken from a population of N units, the estimator of variance is

$$\hat{V}_{SRS}(\bar{y}) = (n-1)^{-1} n^{-1} \left(1 - \frac{n}{N}\right) \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad (3)$$

where $\bar{y}$ is the sample mean of y_i. Subscripting to identify a specific sample out of the K possible is omitted here and forthwith. With a slight abuse of designation, we use the abbreviation *SRS* for the estimator in Eq. (3) as a synonym for the variance of an expansion estimator (Valliant et al. 2000, p 51).

Matérn's estimator of variance

Matérn (1947) proposed a per point (i.e. local) estimator of variance inspired, in part, by the pioneering work of Langsæter (1932), Langsæter (1926), and Lindeberg (1924). These authors suggested the use of first- and second-order differences as a mean to reduce the effect of local trends resulting in autocorrelation (Wolter 2007, ch. 8.2.1.). To our knowledge, the Swedish and Finnish national forest inventories (NFI) were the first to adopt a variant of his estimator (Ranneby et al. 1987; Heikkinen 2006).

In Matérn's estimator, the sample locations are split into Q non-overlapping groups of four nearest neighbours. An example is in Fig. 5. Two predictions of the local mean are constructed for each group, and the

squared difference of these predictions is taken as the per point variance.

With the notation in Fig. 5, the two local predictions are computed as $(y_{i, (j+1)} + y_{(i+1), j})/2$ and $(y_{i, j} + y_{(i+1), (j+1)})/2$. The final estimator of variance is the average per point variance. Modern parallels to this estimator can be found in texts on ordinary kriging (for example, Cressie 1989, ch. 3.2). Examples of practical applications with this estimator can be found in (Kangas 1993, 1994; Lappi 2001; Ekström and Sjöstedt-de Luna 2004; Tomppo 2006).

In populations where a sample location can be outside the domain of interest (here forest), at least one sample location in each group must be in the domain. Computation of Matérn's variance estimate is carried out with mean-centred values of y_{ij}. Within each group, the value of y_{ij} in locations outside the domain of interest is set to 0 (viz. the mean of all y_{ij} in the sample). We have (Matérn 1980, ch. 6.7, p 121; Ranneby et al. 1987)

$$\hat{V}_{MAT}(\bar{y}) = \frac{1}{Q} \sum_{q=1}^{Q} \frac{\left(\left(y_{q \ni \{i,j\}} + y_{q \ni \{i+1,j+1\}s}\right) - \left(y_{q \ni \{i+1,j\}} + y_{q \in \{i,j+1\}}\right)\right)^2}{n_q^2}$$

$$(4)$$

where n_q is the number of sample locations in a group in forest, and $q \ni \{i, j\}$ means that group q includes sample location $\{i, j\}$. Note, when all Q groups have four locations in the domain of interest, there is no need to mean-centre the observations. Conversely, the implicit imputation of the mean to location outside the domain of interest will, on average, inflate the variance in populations with autocorrelation.

Successive difference replication estimator of variance

The successive difference replication estimator of variance (*SDR*) was proposed by Fay and Train (1995). According to Fay and Train, *SDR* is an improvement over the first- and second-order difference estimators first proposed by (Lindeberg 1924) and later detailed in Wolter (2007). Like in a jackknife estimator of variance (Efron 1982), a number 2^r - with r an integer and $2^r - n - 2 \geq 0$ - of pseudo-values of the sample mean is produced, and then the variance among these pseudo-values is taken as an estimate of the design variance in Eq. (1). For a sample size of, for example 400, we take $r = 9$, and the number of pseudo-values becomes 512. Each pseudo-value is a weighted average of the n observations in a sample. To apply the *SDR* to a systematic sample from a spatial population, the sample units must be brought into an order compatible with a sample selected from a population with units arranged in a linear (one-dimensional) structure. *SDR* is applicable to a wide array of sampling designs (Opsomer et al. 2016).

The key feature of the *SDR* estimator of variance is that the r pseudo-values are independent. To achieve
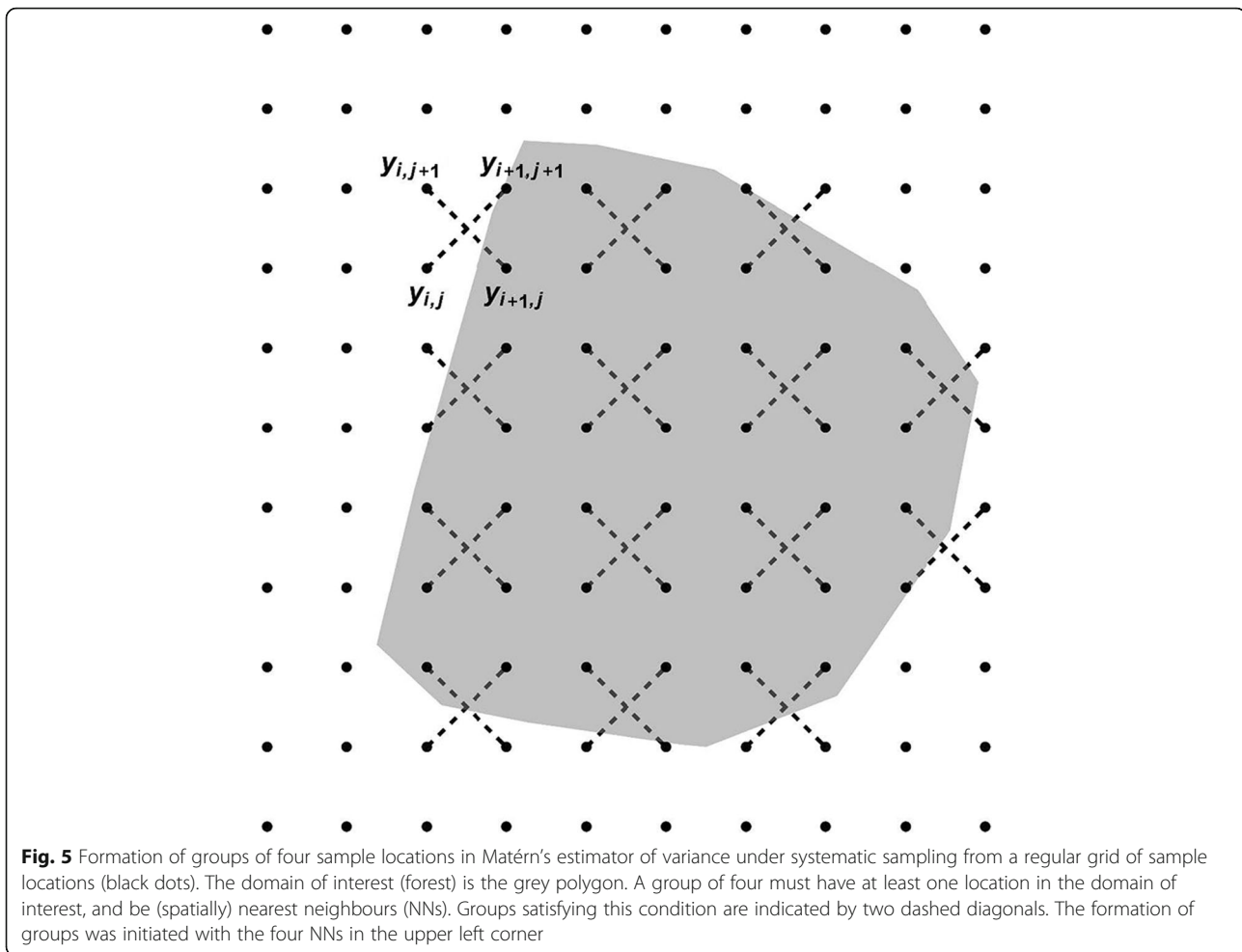
Fig. 5 Formation of groups of four sample locations in Matérn's estimator of variance under systematic sampling from a regular grid of sample locations (black dots). The domain of interest (forest) is the grey polygon. A group of four must have at least one location in the domain of interest, and be (spatially) nearest neighbours (NNs). Groups satisfying this condition are indicated by two dashed diagonals. The formation of groups was initiated with the four NNs in the upper left corner

this, a square Hadamard matrix ($\mathbf{H}$) with 2^r rows and 2^r columns is required with elements $h_{st} = 1$ or $h_{st} = -1$, and the first row is filled with 1 s. Also, $\mathbf{H}\mathbf{H}' = 2^r\mathbf{I}$ where $\mathbf{I}$ is an $2^r \times 2^r$ identity matrix. Each pseudo value is computed as a weighted average of the n sample observations, whereby the weight (w), in the sth *SDR* replication ($s = 1, ..., 2^r$) assigned to the ith sample unit, is $w_{st}^* = f_{st}w_t$ with $f_{st} = 1 + \frac{1}{2\sqrt{2}}(h_{s+1,t} - h_{s+2,t})$ and w_t as the original design weight (i.e. N/n). The distinct values of f_{st} are 1, $1 - \frac{1}{\sqrt{2}}$ and $1 + \frac{1}{\sqrt{2}}$. For $n = 400$, the frequencies of the three distinct values assigned to a unit are 256, 128, and 128, respectively.

With our population units, identified by their row and column position in a grid, we applied the *SDR* estimator of variance with the n sample units ordered row-wise, column-wise, and to a shortest path (with start in the first sampled unit) through the n sample locations (Fig. 6).

The simple average of the three *SDR* estimates of variance obtained with the row-wise, the column-wise, and

the shortest path ordering of the sample is our *SDR* estimate of variance for a single systematic sample. The *SDR* estimator applicable to an ordered sample with r pseudo-values of the population mean is:

$$
V(\bar{y}) = \frac{4}{512-1}\left(1 - \frac{n}{N}\right)\sum_{s=1}^{512}\left(\bar{y}_s - \bar{\bar{y}}_s\right)^2,
$$
$$
\text{with}\quad \bar{y}_s = \frac{1}{n}\sum_{t=1}^{n}f_{st}y_t \tag{5}
$$
$$
\text{and}\quad \bar{\bar{y}}_s = \frac{1}{512}\sum_{s=1}^{512}\bar{y}_s
$$

where $\bar{y}$ is the weighted sample mean (pseudo-value) in the sth replicate of successive differences.

Ripley's estimator of variance
Ripley's estimator Ripley (2004) is model based and applies to a continuous (in y) population with infinitely many possible sampling locations (Mandallaz 2008, pp. 60–62). Applied to a systematic sample of size n from a
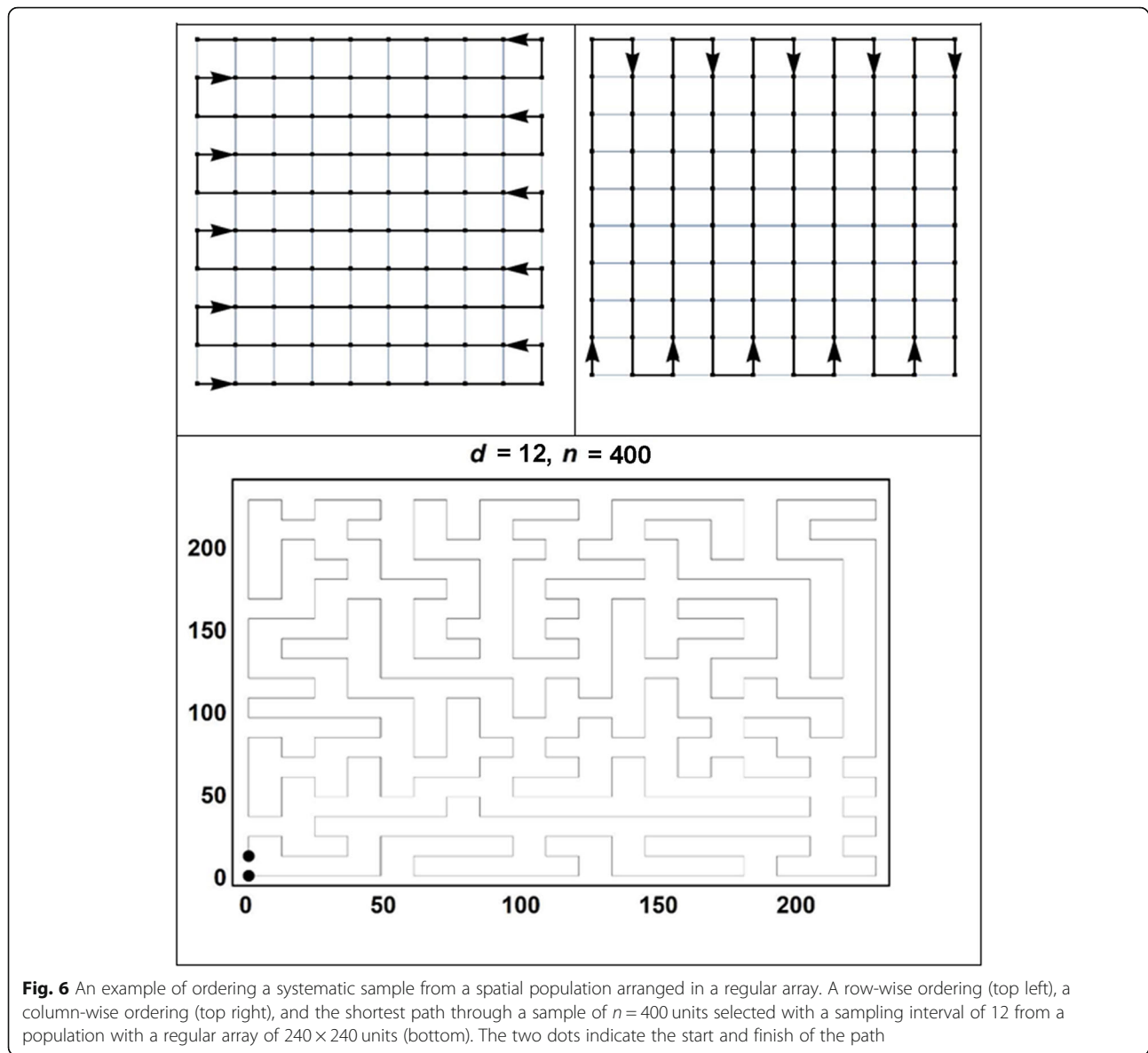
Fig. 6 An example of ordering a systematic sample from a spatial population arranged in a regular array. A row-wise ordering (top left), a column-wise ordering (top right), and the shortest path through a sample of $n = 400$ units selected with a sampling interval of 12 from a population with a regular array of 240×240 units (bottom). The two dots indicate the start and finish of the path

contiguous spatial area (A) equal to the extent of the finite populations under study, we have

$$\hat{V}_{RIP}(\bar{y}) = \frac{1}{n^2} \sum_{i,j} \hat{C}\left(y_i, y_j\right) - \frac{2}{n} \sum_i N^{-1} \int_A \hat{C}(y_i, y) dy$$
$$+ \int_A \int_A \hat{C}(y', y) dy' dy$$

$$(6)$$

where $\hat{C}(y_i, y_j)$ is an estimate of the covariance between sample observations of y in units i and j, $\int_A \hat{C}(y_i, y) dy$ is the integral of the covariance between the y-values in the sample and the y-values in the assumed continuous

surface of y-values in the area A defining the population under study. The last term (double integral) is the variance of the population mean. Stated differently, the first two terms on the r.h.s. of Eq. (6) is the expected variance of $\bar{y} - \bar{\bar{y}}$, while the last term is the variance of the expectation (i.e. the actual population mean $\bar{\bar{y}}$).

We chose the distance-dependent covariance function for an isotropic weakly stationary fractional Gaussian noise process (FNG, Baillie 1996). FNG's have been used to characterize 'long-term' memory processes (Johannesson et al. 2007; Nothdurft and Vospernik 2018). Accordingly, the covariance between observations from two units or two points separated by a distance h is

$$\hat{C}(h) = \hat{\sigma}^2 \left(|h-1|^{2\hat{t}} - 2h^{2\hat{t}} + (h+1)^{2\hat{t}} \right)/2 \qquad (7)$$

where $\hat{\sigma}^2$ and $\hat{t}$ are ordered sample-based maximum likelihood estimates (MLE) of the two parameters σ^2 (process variance), and $t \in (0, 1)$ controlling the rate of change in the covariance as a function of distance. Again, we used each of the three orderings outlined above, and took the average of the MLEs as our final estimates.

Computation of the last two terms in Eq. (6) can be demanding, in particular for large populations with an irregular spatial outline. In our computations we used Monte-Carlo integration (Robert and Casella 1999, ch. 5.3.2) over 2400 random points in A to obtain the second term on the r.h.s. of Eq. (6). To compute the third term on the r.h.s of Eq. (6) we exploited the fact that in a spatially continuous population with a simple geometric structure, we can integrate over all possible distances with a probability distribution function for the distance between two randomly selected points (Ripley 1977).

D'Orazio's estimator of variance

D'Orazio's estimator of variance (D'Orazio 2003) provides a correction (c) to the *SRS* estimator of variance intended to capture the effect of a spatial autocorrelation. The correction is through Geary's contiguity ratio c – a measure of the spatial association between a sample unit value of y and the y-values in its nearest (spatial) neighbours (Geary 1954). Geary's c takes a value of 1.0 when there is no association, while a $c < 1$ suggests a positive spatial association, and a $c > 1$ a negative association. The estimator showed promising results in a recent simulation study (Magnussen and Fehrmann 2019).

The idea behind D'Orazio's estimator, hereafter referred to as *DOR*, is simple. From Eq. (2) it is clear that the desired design variance is the variance among the K sample means whereas the *SRS* variance in Eq. (3) is the within sample variance of a sample mean. Consider a breakdown of the fixed total variance in a (finite) population into a within- and between sample variance. With a positive (negative) spatial covariance among units in a population the among-sample variance will decrease (increase) relative to a population without a spatial covariance. This follows because the sum of the within-sample variance is inflated (deflated) by the covariance. Since the *SRS* estimator does not account for the within sample covariance, it requires a correction. D'Orazio opted to use Geary's contiguity ratio as a correction factor since it represents an extension of the Durbin–Watson (*DW*) statistic (Durbin and Watson 1950) to a spatial context. The *DW* statistic was successful in explaining

the apparent efficiency of nearest-neighbour post-stratification in systematic sampling from populations arranged in a linear array (Ripley 2004, pp. 26 – 27). The *DOR* estimator of variance is

$$\hat{V}_{DOR}(\bar{y}) = \hat{c}\ \hat{V}_{SRS}(\bar{y}) \text{ with } \hat{c}$$
$$= \frac{\sum_{i=1}^{n} \sum_{j\sim i} w_{ij} \left(y_j - y_i\right)^2}{2\sum_{i=1}^{n} \sum_{j\sim i} w_{ij}}\ \hat{V}(y_i) \qquad (8)$$

where w_{ij} are distance dependent weights, and $\hat{V}(y_i)$ is the sample-based estimate of the population variance in y. The symbol $j\sim i$ indicates that sample unit j is a first-order neighbour of sample unit i. We assigned a weight $w_{ij} = 1.0$ if sample units i and j are separated by a distance d units equal to the sampling interval in the design under study (see next), and a weight of $1/\sqrt{2}$ to sample units separated by a distance $\sqrt{d^2 + d^2}$ units. Other weighting schemes are possible (Cliff and Ord 1981, ch. 1.4.2).

Monte-Carlo error in estimated variances

With 30 replications of K possible samples, the Monte-Carlo error (Koehler et al. 2009) on the average of an estimated variance was 4.6% ($n = 400$) to 1.6% ($n = 1600$) with the *SRS* estimator, 2.7% with the *MAT* estimator, 2.6% with the *SDR* estimator, and 1.2% ($n = 400$) to 13.8% ($n = 1600$) with the *RIP* estimator, and 2.4% with *DOR*.

Estimator performance

Two metrics are used to assess the expected performance of an estimator of variance. The first is the ratio $\frac{\text{mean}(\hat{V}_{EST})}{V_{DES}}$, $EST = \{SRS, MAT, SDR, RIP, DOR\}$ with $\text{mean}(\hat{V}_{EST})$ equal to the mean of the K estimates of variance. The second is the absolute difference $|1 - \text{mean}(\hat{V}_{EST})/V_{DES}|$ as a measure of bias. In practice, the anticipated performance (Isaki and Fuller 1982; Kish and Frankel 1974) in a single application is more relevant. Consequently we report on the distribution of the ratio $\frac{\hat{V}_{EST}}{V_{DES}}$, $EST = \{SRS, MAT, SDR, RIP, DOR\}$ and $|1 - \hat{V}_{EST}/V_{DES}|$ across all 10,320 combinations of sample sizes, samples, and realizations of a superpopulation.

Results

Populations with autocorrelation and no trend

The *SRS* variance estimator was consistently conservative (Fig. 7). In all but four out of 120 cases in the main study (4 sample sizes × 30 realizations of a superpopulation), the estimated variance was greater than the design based variance (V_{DES}). The average, over 30 realizations of a superpopulation, of the ratio
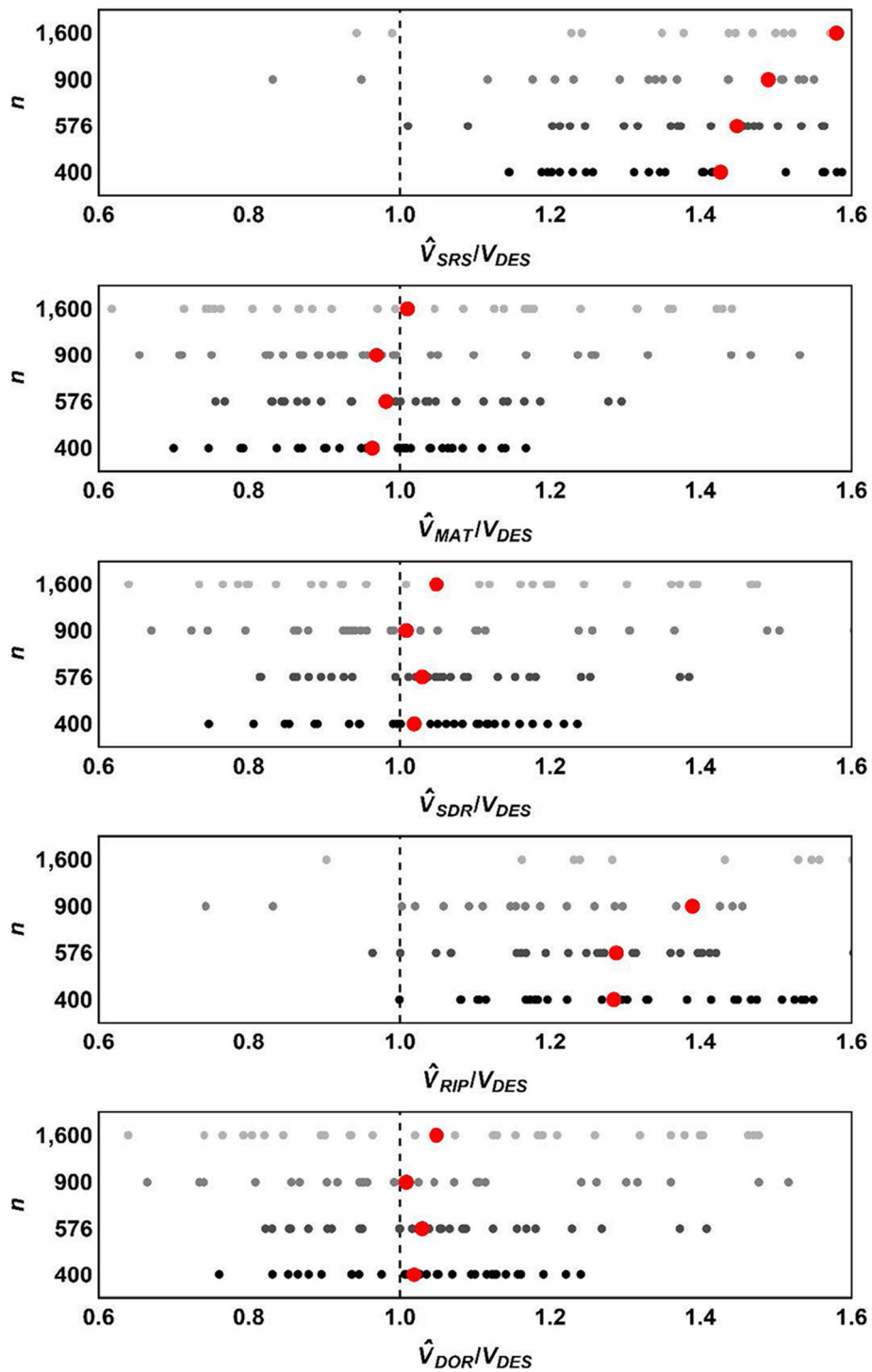
Fig. 7 Variance ratios ($\frac{\text{mean}(\hat{V}_{EST})}{V_{DES}}$, $EST = \{SRS, MAT, SDR, RIP, DOR\}$) in simulated systematic sampling with sample size $n = 400, 576, 900$, and 1600 (y-axis). A dot represents a ratio of the mean over K samples to the design-based variance in one realization of a superpopulation. The larger red dot is the mean ratio in 30 realizations. Dots above 1.6 have been clipped (34 values > 1.6 from *SRS* and 6 values > 1.6 from *RIP*)

mean($\hat{V}_{SRS}$)/V_{DES} - with the mean taken over the K samples - varied from 1.4 ± 0.04 ($n \le 900$) to 1.6 ± 0.08 ($n = 1600$). For all estimators, and visible in Fig. 7, the variation in this ratio increases with sample size because V_{DES} declines faster than the mean of the *SRS* estimator of variance.

The 30 averages of K SRS estimates of variance were perfectly and negatively correlated with V_{DES} $(\hat{\rho}(SRS, DES) = -1)$. With the total variance fixed at 1.0 in all cases - and recalling that the total variance in y is equal to the among-sample variance plus the within-sample variance (Särndal et al. 1992, p 78) - the result was expected inasmuch the SRS variance equals the within-sample variance (divided by n), and DES equals the among-sample variance. If one increases, the other has to decrease. Otherwise, the SRS estimator was negatively correlated (~ -0.6) with the remaining four estimators when sample size was 400. At larger sample sizes, the correlation between SRS and RIP estimates deteriorated to values around -0.2, but remained around -0.6 with MAT, SDR, and DOR for sample sizes ≤900. With $n = 1600$, the maximum correlation was -0.3.

Matérn estimates of variance were much closer to the design variance than the SRS estimates of variance (Fig. 7). The average ratio of mean$(\hat{V}_{MAT})/V_{DES}$, varied from 0.96 ± 0.02 ($n = 400$) to 1.01 ± 0.04 ($n = 1600$). The correlation between mean$(\hat{V}_{MAT})$ and V_{DES} (across the 30 realizations of a superpopulation) also decreased with an increase in n. From 0.64 ($n = 400$) to 0.29 ($n = 1600$). A confirmation that $\hat{V}_{MAT}$ decreases at a rate slightly slower than n^{-1}.

The performance of the SDR estimator was - by and large - similar to the performance of Matérn's estimator with a mean$(\hat{V}_{SDR})/V_{DES}$ varying from 1.01 ± 0.02 to 1.04 ± 0.04 across the four sample sizes (Fig. 7). The correlation between SDR and MAT variances was consistently strong (0.996 to 0.998). SDR estimates of variance from either a row-, a column-wise, or shortest path ordering of sample locations (cf. section on estimators) were always within 10% of each other.

Ripley's estimator of variance showed the strongest effect of sample size (Fig. 7). The ratio mean$(\hat{V}_{RIP})/V_{DES}$ increased from 1.28 ± 0.03 to 3.06 ± 0.52 as sample size increases from 400 to 1600. The increase was expected. By adding more sample units, the average covariance among unit observations in a sample increases; hence the numerical values of the first and second terms on the r.h.s. of Eq. (6) will increase, but the first term increases at a faster rate than the second term. Otherwise, the variability and correlation with V_{DES} was similar to what is reported for V_{MAT} and V_{SDR}. Again, $\hat{V}_{RIP}$ was strongly correlated (0.978–0.989) with both $\hat{V}_{MAT}$ and $\hat{V}_{SDR}$.

Results with D'Orazio's estimator of variance in Fig. 7 were nearly perfectly correlated with results from Matérn's (0.992–0.995) and the SDR estimator (0.999–1.000) and therefore not detailed separately.

In terms of absolute deviations from the design-based variance, Matérn's estimator was attractive when sample sizes were 400 and 576. In these settings, the average

MAT estimate of variance - over the K samples - was in 17 ($n = 400$) and 18 ($n = 576$) out of 30 realizations of a superpopulation the least biased (Fig. 8). With $n = 900$, the MAT estimator was in 13 realizations the least biased, and Ripley's estimator was 9 times the least biased. With the largest sample size ($n = 1600$) RIP and DOR are the least biased in 11 and 10 realizations, respectively.

The anticipated performance of an estimators in a single application is captured by the density distribution of $\frac{\hat{V}_{EST}}{V_{DES}}$, $EST = \{SRS, MAT, SDR, RIP, DOR\}$ across all settings of sample size, samples, and realizations of a superpopulation (Fig. 9). The almost perfectly correlated estimators SDR and DOR have distributions that are more concentrated around 1.0 than distributions for SRS, RIP, and MAT. The median squared difference of $1 - \frac{\hat{V}_{EST}}{V_{DES}}$ was 0.23 for SRS, 0.02 for MAT, 0.01 for SDR and DOR, and 0.09 for RIP.
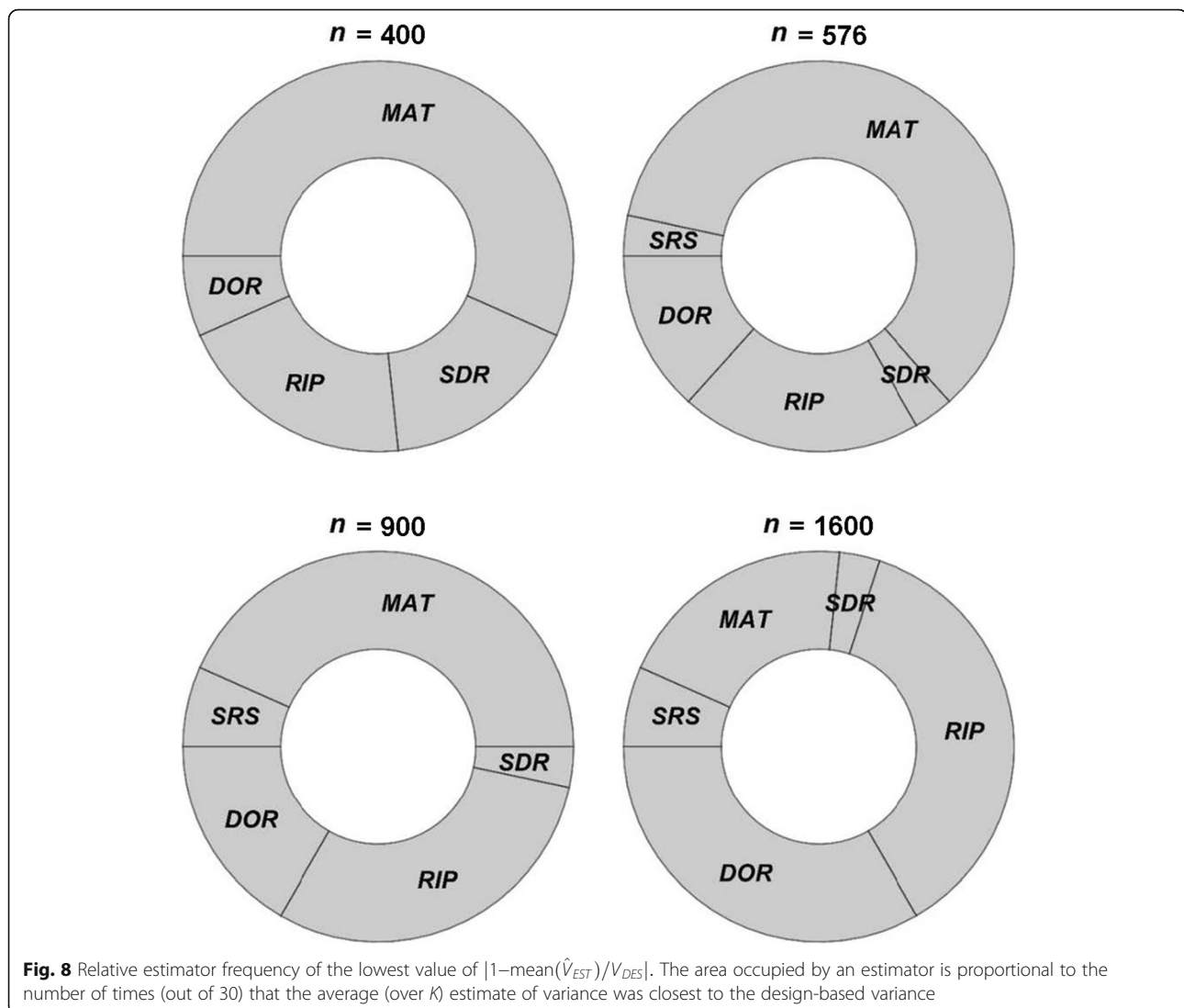
The anticipated performance in terms of $|1 - \hat{V}_{EST}/V_{DES}|$ is in Fig. 10 in the form of density distributions of $|1 - \hat{V}_{EST}/V_{DES}|$ across the settings of sample sizes, samples, and realizations of a superpopulation.

In terms of the distribution of absolute deviations, Fig. 10 indicates a better anticipated performance of DOR and SDR with MAT as the runner up. RIP is a distant fourth and closest to the distribution provided by SRS.

Should an analyst prefer an estimator that has a variance that is not only at least 20% below the variance with SRS, but also not underestimating the design variance, the choice would again be SDR and DOR with an estimated probability of 0.45 and 0.48 for satisfying this criterion in our simulations. Corresponding results for MAT and RIP were 0.34 and 0.12.

Populations with autocorrelation and a linear trend

A global linear trend, unless accounted for by modelling or stratification, will increase the variance in sampled values of y. In the scenarios with a linear trend the SRS estimator of variance was again, by a wide margin, the most conservative of the five tested estimators (Table 2). The overestimation of variance increased rapidly with the strength of the linear trend, from 50% with a weak trend to 188% with a strong trend. Sample size (from 400 to 1600) had, in comparison, only a minor effect. Results with the MAT estimator were better. Its poorest performance was an overestimation of 19% in populations with a strong linear trend and a sample size of 400 ($d = 12$), in all other settings the estimated variance was within 4 percentage points of the design variance. The performances of SDR and DOR were similar but consistently lagged that of MAT, especially in the populations with a strong linear trend. The RIP estimator performed

Fig. 8 Relative estimator frequency of the lowest value of $|1-\text{mean}(\hat{V}_{EST})/V_{DES}|$. The area occupied by an estimator is proportional to the number of times (out of 30) that the average (over K) estimate of variance was closest to the design-based variance

worse than *MAT*, *SDR*, and *DOR* in combinations of a strong or moderate trend and a sample size of 400. With a sample size of 1600 and a moderate or a weak trend, the performances of the four estimators *MAT*, *SDR*, *RI*P, and *DOR* were, from a practical perspective, similar.
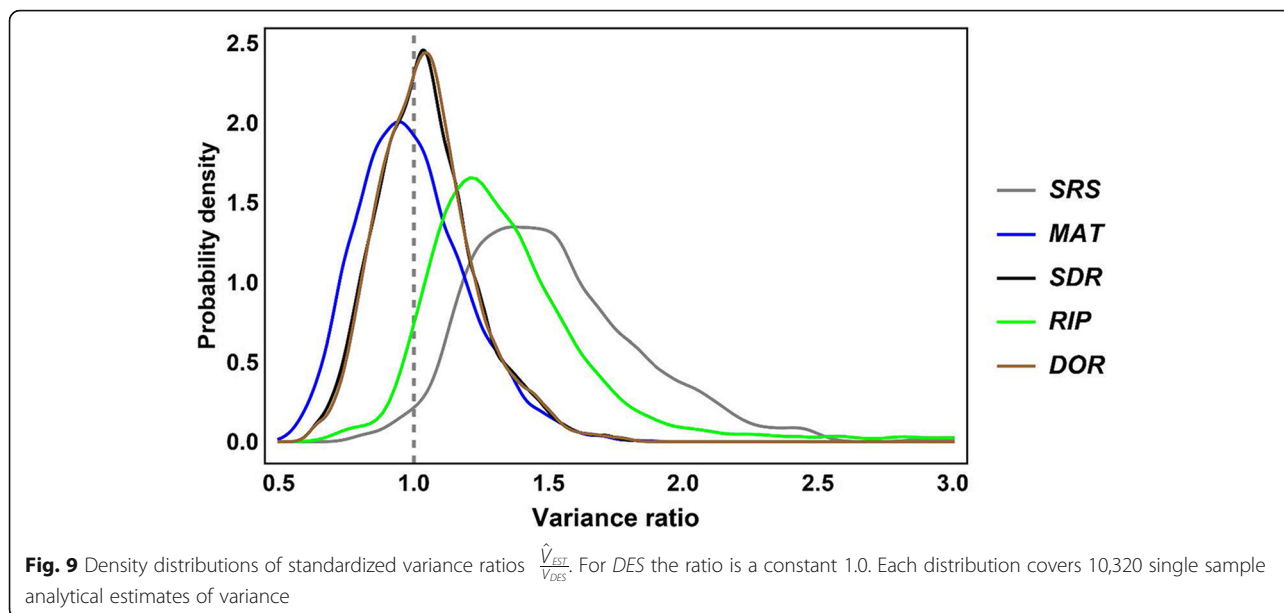
Populations with near zero autocorrelation and no trend

In a population with a near zero autocorrelation, the proposed alternative estimators (*MAT*, *SDR*, *RIP*, and *DOR*) and the *SRS* should, ideally, generate estimates of variance close to the actual design variance (*DES*). As can be taken from Table 3, this was the case across all sample sizes and realizations of a superpopulation (hint: paired equal variance *t*-test *p*-values for the null hypothesis of no difference were all greater than 0.05). The applicability of the *t*-distribution was ascertained with a KS-test (Kolmogorov-

Smirnov, $P > 0.34$, Barr and Davidson 1973). We failed in all cases to reject the null hypothesis of a *t*-distribution.

Populations with a near zero autocorrelation and a linear trend

In populations with a near zero autocorrelation and a linear trend, the *SRS* estimator of variance was consistently the most conservative (Table 4) with an overestimation that increased with sample size and strength of a linear trend. Estimates obtained with *MAT*, *SDR*, *RIP*, and *DOR* were all closer to the actual design variance than the *SRS* estimates of variance, but with a distinct sensitivity to the interaction between sample size (sampling interval) and strength of the linear trend. With $n = 1600$ the four alternative overestimated the actual variance by 22%–24%, but with $n = 400$ the *MAT*, *SDR*, and *DOR* estimator underestimated the

Fig. 9 Density distributions of standardized variance ratios $\frac{\hat{V}_{EST}}{V_{DES}}$. For *DES* the ratio is a constant 1.0. Each distribution covers 10,320 single sample analytical estimates of variance

variance by 1% to 5%, whereas *RIP* overestimated the design variance, most (22%) in presence of a strong trend, and least (2%) with a weak trend.

Scaling to larger populations and a smaller sample fraction

With the results from the supplementary populations and sample designs we gauge the scalability of the results from the main study in Table 5. *DES* and *SRS* variances were almost constant across the three population sizes (57,600; 230,400; and 921,600), as predicted by theory. Variances obtained with *MAT*, *SDR*,

and *DOR* were in all cases closer to the estimates of design variance than were the variances obtained with *SRS*. *MAT* was in each case closest to the *DES* variance but with a standard deviation across realizations almost twice as great as the standard deviation with *DES*. We also note that as the population size increased and the sample fraction decreased, the variances obtained with *MAT*, *SDR*, and *DOR* drifts – at a slow rate – towards the results of *SRS*. The *DOR* estimator has a much smaller standard deviation than *MAT* and *SDR*, suggesting that in even larger populations and smaller sample sizes, this estimator may, on
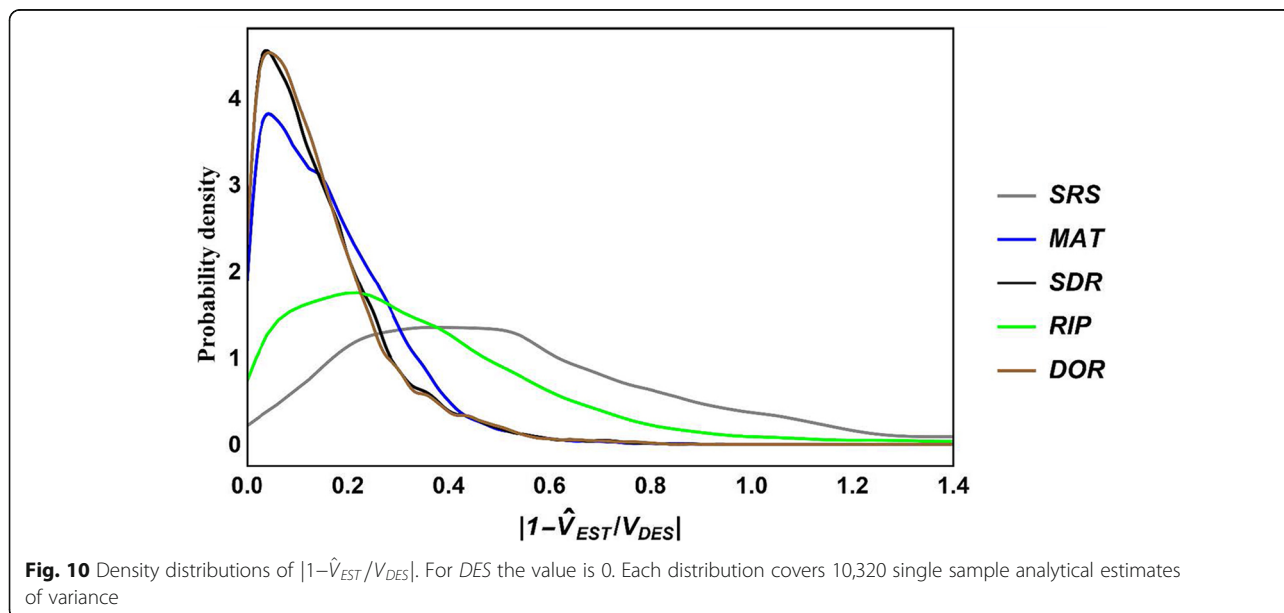

Fig. 10 Density distributions of $|1 - \hat{V}_{EST}/V_{DES}|$. For *DES* the value is 0. Each distribution covers 10,320 single sample analytical estimates of variance

Table 2 Estimated variances in populations with spatial autocorrelation and a strong, a moderate, and a weak global linear trend. All table entries are means across 30 realizations of a super-population with autocorrelation and a linear trend, and the K possible sample for a given sampling interval (d). Variances in parentheses are relative variances with the *DES* variance fixed at 100. $\tau_{i,j}$ is the trend component for unit in row i and column j ($i, j = 0.5, 1.5, \ldots, 239.5$)

Trend	Trend variance (%)	Site effect variance (%)	Sample size (units)	Sample interval (units)	Variance estimator $\times\ 10^3$					
					DES	*SRS*	*MAT*	*SDR*	*RIP*	*DOR*
Strong $\tau_{i,j} = (i+j-241)/120$	45	25	400	12	0.88 (100)	2.52 (288)	1.04 (119)	1.15 (131)	1.58 (180)	1.20 (137)
			1600	6	0.24 (100)	0.62 (262)	0.25 (104)	0.26 (107)	0.21 (88)	0.26 (110)
Moderate $\tau_{i,j} = (i+j-241)/240$	16	35	400	12	1.50 (100)	2.56 (171)	1.51 (101)	1.64 (109)	2.06 (138)	1.65 (110)
			1600	6	0.34 (100)	0.62 (182)	0.34 (101)	0.35 (104)	0.35 (105)	0.36 (106)
Weak $\tau_{i,j} = (i+j-241)/480$	4	40	400	12	1.71 (100)	2.56 (150)	1.67 (98)	1.81 (106)	2.20 (129)	1.81 (106)
			1600	6	0.38 (100)	0.62 (162)	0.38 (100)	0.39 (105)	0.40 (103)	0.39 (103)

a single-sample basis, frequently outperform *MAT* and *SDR*.

Discussion

Although we have a general methodology for constructing a model-based estimator of variance for systematic sampling that is model unbiased with a minimum root mean

Table 3 Paired *t*-tests under the hypothesis of equal variances. $\hat{\Delta}$ is the difference between the mean estimate derived with *SRS*, *MAT*, *SDR*, *RIP*, or *DOR* and the design based variance (*DES*). $|\hat{t}|$ is the absolute value of the *t*-statistics (effect size), and $P(|\hat{t}| \mid \Delta = 0)$ is the probability of a greater $|\hat{t}|$ under the null hypothesis of a zero difference

Estimator contrast	n	$\hat{\Delta} \times 10^4$	$|\hat{t}|$	$P(|\hat{t}||\Delta = 0)$
SRS-DES	400	0.8	1.31	0.20
	576	0.2	0.37	0.72
	900	0.0	0.16	0.87
	1600	0.0	0.11	0.91
MAT-DES	400	0.2	0.30	0.77
	576	0.3	0.52	0.61
	900	0.3	0.60	0.55
	1600	0.1	0.40	0.69
SDR-DES	400	0.3	0.54	0.60
	576	0.2	0.43	0.67
	900	0.2	0.54	0.59
	1600	0.1	0.40	0.69
RIP-DES	400	0.5	0.73	0.47
	576	0.2	0.34	0.74
	900	0.3	0.68	0.50
	1600	0.2	0.80	0.43
DOR-DES	400	−0.5	0.84	0.41
	576	−0.7	1.39	0.17
	900	−0.5	1.22	0.23
	1600	−0.2	0.73	0.47

squared error (Wolter 2007, ch. 8.2.2), and we appreciate Tobler's first law of geography ("units separated by a shorter distance are, on average, more similar than units separated by a longer distance", Tobler 2004), we are still evaluating model-based variance estimators for systematic sampling (McGarvey et al. 2016; Strand 2017; Magnussen and Fehrmann 2019) or proposing new ones (D'Orazio 2003; Clement 2017; Pal and Singh 2017; Fattorini et al. 2018a; Magnussen and Fehrmann 2019). A century long occupation that appears to have begun with the efforts by Lindeberg (1924) and Langsæter (1932).

There is a simple explanation as to why a single omnibus estimator for systematic sampling is unlikely to emerge, and that is the sensitivity of the design variance to a non-random ordering of the population units (Särndal et al. 1992, ch. 3.3.4). In forestry, we may have numerous non-random structures in any population of forest trees and associated vegetation that directly influence a study variable (Burslem et al. 2001; Scherer-Lorenzen and Schulze 2005). Spatial autocorrelation is just one of many manifestations of a non-random ordering, but it is pivotal for computation of variance in a spatial population.

The sensitivity of the design variance to a non-random ordering of population units calls for caution when we, from simulation studies, attempt to infer the performance of a variance estimator in a population with an unknown ordering. In particular when an estimator explicit or implicitly assumes a particular model for the study variable. Since "all models are wrong, but some are useful" (Box 1976), it is risky to assume that a model is true (Wolter 2007, p 305).

Yet, simulated systematic sampling from artificial or actual populations remains the most expedient method to screen variance estimators for systematic sampling. By necessity artificial populations will be simpler and smaller than actual ones. Ultimately, however, it is the spatial covariance structures in a population that drive the performance of a variance estimator (Fortin et al.

Table 4 Estimated variances in populations with a near zero autocorrelation and a strong, a moderate, and a weak global linear trend. All table entries are means across 30 realizations of a super-population and the K possible samples for a given sampling interval (d). Variances in parentheses are relative variances with the *DES* variance fixed at 100. $\tau_{i,j}$ is the trend component for unit in row i and column j ($i, j = 0.5, 1.5 \ldots, 239.5$)

Trend	Trend variance (%)	Site effect variance(%)	Sample size (units)	Sample interval (units)	Variance estimator $\times 10^3$					
					DES	*SRS*	*MAT*	*SDR*	*RIP*	*DOR*
Strong $\tau_{i,j} = (i+j-241)/120$	45	2	400	12	1.51 (100)	2.50 (166)	1.46 (97)	1.47 (98)	1.84 (122)	1.48 (98)
			1600	6	0.29 (100)	0.63 (214)	0.36 (124)	0.37 (125)	0.36 (124)	0.37 (126)
Moderate $\tau_{i,j} = (i+j-241)/240$	14	2	400	12	2.12 (100)	2.49 (117)	2.09 (98)	2.09 (98)	2.28 (107)	2.03 (96)
			1600	6	0.42 (100)	0.62 (149)	0.52 (123)	0.52 (124)	0.52 (124)	0.52 (123)
Weak $\tau_{i,j} = (i+j-241)/480$	4	2	400	12	2.36 (100)	2.49 (105)	2.34 (99)	2.33 (99)	2.41 (102)	2.25 (95)
			1600	6	0.47 (100)	0.62 (133)	0.58 (123)	0.58 (123)	0.58 (122)	0.57 (122)

2012). Casting the covariance process as the outcome of shared random effects is consistent with Tobler's first law of geography (Tobler 2004). With autocorrelation arising from three additive site effects – of different strength and operating at three spatial scales – our populations are one step closer to resemble actual forested landscapes than possible with a parametric spatial covariance model (Wolter 2007, ch. 8.3; Magnussen and Fehrmann 2019). A different approach was taken by Hou et al. (2015). They generated spatial covariance structures by manipulating the spatial distribution of live trees in an actual plantation. In terms of the sampling distribution in estimated means, their approach delivered results consistent with ours. A nearly constant relative performances of the five estimators of variance, in populations of different size and more realistic sample fractions, vouch for the scalability of our findings and main results.

Regional trends are commonplace in large-area forest inventories as they include sites with different climates, soils, species, forest structures, and associated forest management practices. If regional trends are not dealt with through modelling or a (post) stratification, they may drastically change the estimate of variance (Matérn 1980, ch. 4.6; Särndal et al. 1992, p 82; Breidt and Opsomer 2000; Wolter 2007, Table 8.3.1). Our results with a strong,

moderate, and weak global linear trend confirmed the sensitivity of the design variance to such trends. Each of the K sample means will differ by an amount determined by the average difference in y between adjoining population units. Regardless of the strength of a global trend, the four tested alternative variance estimators generated variances that were closer to the design variance than possible with the *SRS* estimator of variance. The *MAT* estimator is, in theory, robust against a unidirectional trend, but not against our bi-directional trends. In populations with a suspected trend, and no attempt to address the trend by modelling or a stratification, the *MAT* estimator emerged as most attractive followed by *DOR* and *SDR*. In larger populations, the less variable *DOR* estimator becomes more attractive. To be successful in populations with a trend, the *RIP* estimator requires a separation of trend and spatial covariance structures, otherwise the performance will be less predictable and more variable.

Heteroscedasticity is commonplace in data from actual forest inventories, but not included in our study settings. To gauge it importance, we ran simulations with a noise variance that increased linearly by a factor 3 across both rows and columns, and sample sizes 400 and 1600. The results (not shown) were similar to the results in Table 4 for a moderate trend and a near zero autocorrelation. That is, *SRS* was the most conservative with an

Table 5 Estimated variances with a systematic sample size of $n = 100$ in trend-less populations with autocorrelation, scaled sizes (N), and scaled expected number of site polygons ($n(sT), T = 1, 2, 3$). All results are based on 30 realizations of a superpopulation, and all or a maximum of 2000 random selections of all possible samples under a systematic sampling design. The standard deviation ($\hat{\sigma}$) across realizations is indicated ($\pm\hat{\sigma}$). Relative variances with *DES* fixed at 100 are in parentheses

N	Array	d	n (s1)	n (s2)	n (s3)	Variance $\times 10^2$				
						DES	*SRS*	*MAT*	*SDR*	*DOR*
57,600	240×240	24	96	12	6	0.73 ± 0.10 (100)	1.00 ± 0.13 (137)	0.73 ± 0.22 (100)	0.79 ± 0.22 (107)	0.76 ± 0.13 (104)
230,400	480×480	48	96	12	6	0.69 ± 0.08 (100)	1.00 ± 0.12 (145)	0.74 ± 0.22 (107)	0.79 ± 0.22 (114)	0.76 ± 0.12 (111)
–	–	–	192	24	12	0.70 ± 0.08 (100)	1.00 ± 0.13 (143)	0.76 ± 0.22 (108)	0.81 ± 0.22 (115)	0.78 ± 0.11 (112)
921,600	960×960	96	96	12	6	0.73 ± 0.08 (100)	1.00 ± 0.12 (136)	0.76 ± 0.22 (104)	0.80 ± 0.23 (110)	0.78 ± 0.12 (106)
–	–	–	384	48	24	0.71 ± 0.08 (100)	1.00 ± 0.13 (141)	0.81 ± 0.23 (114)	0.87 ± 0.23 (122)	0.83 ± 0.11 (118)

overestimation of variance of 40% ($n = 400$) and 22% ($n = 1600$), and *MAT* was the estimator with the performance closest to that of the target design variance (i.e. an overestimation of 10% for $n = 400$, and an underestimation of 6% for $n = 1600$). *SDR* was in this regard the runner up.

Despite marked differences in formulation of the *MAT*, *SDR*, *RIP*, and *DOR* estimators of variance, their expected performance was quite similar in populations without a global trend. The strong correlations among the variances obtained under these conditions, confirms the importance of a first-order autocorrelation since it alone was captured by all four. Higher order autocorrelations enters only in the *SDR* and *RIP* estimators.

The *MAT* estimator had the lowest expected absolute bias, i.e. the lowest expected risk of over- or underestimating the actual variance. Yet, in practical applications *MAT* will be sensitive to edge effects. In a fragmented forest landscape, its performance may suffer. The expected performance of *DOR* and *SDR* was close to that of *MAT*. From a practical perspective there is no strong rationale for preferring one over the other in populations with either no trends or with a trend dealt with through modelling or stratification. Otherwise the *MAT* estimator followed by *SDR* and *DOR* can be recommended.

The expected performance of *RIP* was sensitive to trends and varied with sample size which makes recommendations to practice more difficult. The sensitivity to sample size is largely a question of the number of integration points for computing the second term in Eq. (6). With 9600 integration points the performance was nearly constant across sample sizes (not shown), but with this number of integration points the computation time became impractical. Moreover, computational challenges in applications with large populations and a fragmented forest landscape, may further detract from its appeal in terms of supportive theory in spatial statistics (Thompson 1992, ch. 21; Ripley 2004).

In populations with a very weak autocorrelation, all estimators reproduced the design variance with a low margin of error. Thus, the risk of a counter-factual or a spurious result appears to be low with the four alternative estimators of variance.

In terms of the anticipated performance in a single application (without trends), the *DOR* and *SDR* estimators generated a higher frequency of estimates closer to the design variance than estimates from *MAT* and *RIP*. Moreover, *DOR* and *SDR* were also best in terms of the odds of generating a variance estimate that is at least 20% below the *SRS* without underestimating the actual variance.

DOR has two advantages over *SDR*, it is computationally simpler, and it provides a metric (Geary's c) of the first-order spatial correlation (association). The magnitude and sign of Geary's c provides a useful and interpretable statistic. It is fairly straightforward to implement a spatial randomization of the sample locations and repeat the estimation of Geary's c a large number of times to obtain the distribution of c under the null hypothesis of no spatial association amongst first-order neighbours. A rejection of the null hypothesis serves to argue against the *SRS* variance estimator.

As suggested from the supplementary yet limited simulations with larger populations (without a trend) and lower sample fractions, the estimates of variance obtained with the five estimators will gradually converge as N increases and n/N decreases. This was expected since d is inverse proportional to n/N and the average autocorrelation typically declines with an increase in d.

Several estimators of variance tailored to semi-systematic sampling (Stevens and Olsen 2003; Magnussen and Nord-Larsen 2019), quasi-systematic sampling (Wilhelm et al. 2017), or designs with a spatial balance in auxiliary space (Grafström et al. 2014) were beyond the scope of this study. Given the model-based nature of *MAT*, *SDR*, *RIP*, and *DOR*, we expect they will be of interest also for these variations on systematic sampling.

We made no use of auxiliaries from remote sensing although they are omnipresent. As pointed out by Opsomer et al. (2012) and Fattorini et al. (2018a) "… for a model that fit the data well, any variance estimation method that targets the residual variability will perform satisfactorily regardless of the autocorrelation in the sample data". In the forerunner to this study (Magnussen and Fehrmann 2019), we confirmed that the conservative nature of the *SRS* estimator of variance diminishes with the strength of the correlation between y and an auxiliary variable (x).

All our results apply to finite populations considered as realizations of a superpopulation (Bartolucci and Montanari 2006). We could have employed the infinite population paradigm on the finite-area populations under study with the constraint of equality in size (area) of a population unit and a sample plot. We would, in theory, have an infinite number of possible samples for a fixed sample size, but if we excluded samples with edge-effects and samples with overlapping plots – which violates the strong assumption of independent samples – we would generate results very similar to those presented.

We recognize that an analyst, accustomed to application of the *SRS* estimator of variance to data obtained under a systematic sampling design, may not be swayed by results from simulations or simulated sampling from actual populations. Yet, to continue with the *SRS* without an attempt to gauge the need for an alternative is not best practice. With today's powerful computers and

Magnussen *et al. Forest Ecosystems* (2020) 7:17

Page 17 of 19

readily available software for spatial analysis, it is not difficult to obtain statistics to guide an analyst towards a suitable estimator of variance. While issues of trend and heteroscedasticity may be addressed with a modelling, post-stratification (D'Orazio 2003; Westfall et al. 2011; Strand 2017; McConville and Toth 2017; Magnussen and Fehrmann 2019) or the one-per-stratum design proposed by Breidt et al. (2016), the issue of autocorrelation will persist across spatial scales.

Conclusions

The conservative nature of the *SRS* estimator of variance when applied to data collected under a systematic design was confirmed. The provision of conservative estimates of variance is counter-productive in an era where forest resource estimates are increasingly important in a number of policy issues where precise estimates are expected. Inflated estimates of variance may obscure opportunities for cost-savings from reduced sampling efforts that do not imperil targets set for precision. Additional computational complexities are encountered when switching from the *SRS* estimator to a better alternative, but they are not necessarily dissuasive.

In populations with spatial autocorrelation, the four alternative estimators of variance generated estimates of variance that were much closer to the actual design variance than possible with the *SRS* estimator. In the populations with near zero spatial autocorrelation the four alternatives closely tracked the actual design variance. No single alternative estimator emerged as uniformly best in terms of bias. In terms of expected performance in populations without a trend, *MAT* was slightly better than *SDR* and *DOR*. In terms of the anticipated (single sample) performance, *DOR* and *SDR* emerge as less variable than *MAT*. In populations with a strong or a moderate global linear trend, we would recommend *MAT*. Nevertheless, in a large population and a low sampling intensity, the performances of the investigated estimators will be less distinct.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s40663-020-00223-6.

Additional file 1. Appendix.

Author details
[1]506 West Burnside Road, Victoria, BC V8Z 1M5, Canada. [2]Raspberry Ridge Analytics, 15111 Elmcrest Avenue North, Hugo, MN 55038, USA. [3]Norwegian Institute of Bioeconomy Research, Høgskolevegen 8, 1431 Ås, Norway. [4]Department of Geosciences and Natural Resource Management, University of Copenhagen, Rolighedsvej 23, DK-1958 Frederiksberg C, Denmark. [5]Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden. [6]Forest Inventory and Remote Sensing, Faculty of Forest Sciences, University of Göttingen, Büsgenweg 5, D37077, Göttingen, Germany. [7]Alfred-Möller-Straße 1, Haus 41-42, 16225 Eberswalde, Germany.

References
Alves LF, Vieira SA, Scaranello MA, Camargo PB, Santos FAM, Joly CA, Martinelli LA (2010) Forest structure and live aboveground biomass variation along an elevational gradient of tropical Atlantic moist forest (Brazil). Forest Ecol Manag 260(5):679–691
Anderson OD (1976) Time series analysis and forecasting: the Box-Jenkins approach. Butterworths, London
Baillie RT (1996) Long memory processes and fractional integration in econometrics. J Econ 73(1):5–59
Barr DR, Davidson T (1973) A Kolmogorov-Smirnov test for censored samples. Technometrics 15:732–757
Bartolucci F, Montanari GE (2006) A new class of unbiased estimators of the variance of the systematic sample mean. J Stat Plan Infer 136(4):1512–1525
Box GEP (1976) Science and statistics. J Am Stat Assoc 71:791–799. https://doi.org/10.1080/01621459.1976.10480949
Breidt FJ, Opsomer JD (2000) Local polynomial regression estimators in survey sampling. Ann Stat 28(4):1026–1053
Breidt FJ, Opsomer JD, Sanchez-Borrego I (2016) Nonparametric variance estimation under fine stratification: an alternative to collapsed strata. J Am Stat Assoc 111(514):822–833. https://doi.org/10.1080/01621459.2015.1058264
Brooks JR, Wiant HV Jr (2004) Efficient sampling grids for timber cruises. North J Appl For 21(2):80–82
Burslem DF, Garwood NC, Thomas SC (2001) Tropical forest diversity--the plot thickens. Science 291(5504):606–607
Clement EP (2017) Estimation of population mean in calibration ratio-type estimator under systematic sampling. Elix Stat 106:46480–46486
Cliff AD, Ord JK (1981) Spatial processes. Pion, London
Cochran WG (1977) Sampling techniques. Wiley, New York
Cressie NAC (1989) Geostatistics. Am Stat 43:197–202
Cressie NAC (1993) Statistics for spatial data. Revised edition, 2nd edn. Wiley, New York
Dahlke M, Breidt FJ, Opsomer JD, Van Keilegom I (2013) Nonparametric endogenous post-stratification estimation. Stat Sin 23:189–211
D'Orazio M (2003) Estimating the variance of the sample mean in two-dimensional systematic sampling. J Agric Biol Envir S 8(3):280–295
Durbin J, Watson GS (1950) Testing for serial correlation in least squares regression: I. Biometrika 37(3/4):409–428
Efron B (1982) The jackknife, the bootstrap, and other resampling plans, vol 38. Regional Conference Series. Conference Board of Mathematical Science / National Science Foundation, Philadelphia

Ekström M, Sjöstedt-de Luna S (2004) Subsampling methods to estimate the variance of sample means based on non-stationary spatial data with varying expected values. J Am Stat Assoc 99(465):82–95

Fattorini L, Franceschi S, Pisani C (2009) A two-phase sampling strategy for large-scale forest carbon budgets. J Stat Plan Infer 139:1045–1055

Fattorini L, Gregoire TG, Trentini S (2018a) The use of calibration weighting for variance estimation under systematic sampling: applications to forest cover assessment. J Agric Biol Envir S 23(3):358–373. https://doi.org/10.1007/s13253-018-0325-x

Fattorini L, Marcheselli M, Pratelli L (2018b) Design-based maps for finite populations of spatial units. J Am Stat Assoc 113(522):686–697. https://doi.org/10.1080/01621459.2016.1278174

Fay RE, Train GF (1995) Aspects of survey and model-based postcensal estimation of income and poverty characteristics for states and counties. In: Proceedings of the Section on Government Statistics, vol 1995. American Statistical Association, Alexandria, VA, pp 154–159

Fewster RM (2011) Variance estimation for systematic designs in spatial surveys. Biometrics 67(4):1518–1531. https://doi.org/10.1111/j.1541-0420.2011.01604.x

Fortin M-J, James PMA, MacKenzie A, Melles SJ, Rayfield B (2012) Spatial statistics, spatial regression, and graph theory in ecology. Spat Stat 1(0):100–109. doi: https://doi.org/10.1016/j.spasta.2012.02.004

Fuller WA (2009) Sampling statistics. Wiley, New York

Geary RC (1954) The contiguity ratio and statistical mapping. The Incorp Statist 5(3):115–146

Grafström A, Saarela S, Ene LT (2014) Efficient sampling strategies for forest inventories by spreading the sample in auxiliary space. Can J For Res 44(10):1156–1164. https://doi.org/10.1139/cjfr-2014-0202

Gregoire TG (1998) Design-based and model-based inference in survey sampling: appreciating the difference. Can J For Res 28(10):1429–1447

Gregoire TG, Scott CT (2003) Altered selection probabilities caused by avoiding the edge in field surveys. J Agric Biol Envir S 8(1):36–47

Gregoire TG, Valentine HT (2008) Sampling strategies for natural resources and the environment. Chapman & Hall/CRC, Boca Raton, FL

Hasel A (1942) Sampling error of cruises in the California pine region. J For 40(3):211–217

Heikkinen J (2006) Assessment of uncertainty in spatially systematic sampling. In: Kangas A, Maltamo M (eds) Forest inventory - methodology and applications. Springer, NL, pp 155–176

Hou Z, Xu Q, Hartikainen S, Antilla P, Packalen T, Maltamo M, Tokola T (2015) Impact of plot size and spatial pattern of forest attributes on sampling efficacy. For Sci 61(5):847–860. https://doi.org/10.5849/forsci.14-197

Isaki CT, Fuller WA (1982) Survey design under the regression superpopulation model. J Am Stat Assoc 77(377):89–96

Johannesson G, Cressie N, Huang HC (2007) Dynamic multi-resolution spatial models. Environ Ecol Stat 14(1):5–25

Kangas A (1993) Estimating the parameters of systematic cluster sampling by model based inference. Scand J Forest Res 8:571–582

Kangas A (1994) Classical and model based estimators for forest inventory. Silva Fenn 28:3–14

Kangas A, Maltamo M (2006) Forest inventory: methodology and applications, vol 10. Springer, Dordrecht, NL

Kish L, Frankel MR (1974) Inference from complex samples. J Roy Stat Soc B 36(1):1–37

Koehler E, Brown E, Haneuse J-PA (2009) On the assessment of Monte Carlo error in simulation-based statistical analyses. Am Stat 63(2):155–162

Langsæter A (1926) Om beregning af middelfeilen ved regelmessige linjetakseringer. Medd. Norske Skogforsøksvesen, Norske Skogforsøgsvesen, Oslo

Langsæter A (1932) Nøiaktigheten ved linjetaksering av skog. Medd. Norske Skogforsøksvesen, vol 4

Lappi J (2001) Forest inventory of small areas combining the calibration estimator and a spatial model. Can J For Res 31:1551–1560

Lindeberg JW (1924) Über die Berechnung des Mittelfehlers des Resultates einer Linientaxierung. Acta Forestalis Fennica. Druckerei der Finnischen Literaturgesellschaft, Helsinki

Madow WG, Madow LH (1944) On the theory of systematic sampling, I. Ann Mat Stat 15(1):1–24

Magnussen S (2015) Arguments for a model-dependent inference? Forest Oxf 88(3):317–325. https://doi.org/10.1093/forestry/cpv002

Magnussen S, Fehrmann L (2019) In search of a variance estimator for systematic sampling. Scand J Forest Res 34(4):300–312. https://doi.org/10.1080/02827581.2019.1599063

Magnussen S, Nord-Larsen T (2019) A jackknife estimator of variance for a random tessellated stratified sampling design. For Sci 65(5):543–547. https://doi.org/10.1093/forsci/fxy070

Mandallaz D (2008) Sampling techniques for forest inventories. Chapman and Hall, Boca Raton, Florida

Matérn B (1947) Methods of estimating the accuracy of line and sample plot surveys [in Swedish]. Medd. från Statens Skogsforskningsinst., vol 36. Statens Skogsforskningsinst, Stockholm, Sweden

Matérn B (1980) Spatial variation: stochastic models and their applications to problems in forest surveys and other sampling investigations. Lecture notes in statistics, vol 36, 2 edn. Springer, New York

McConville KS, Toth D (2017) Automated selection of post-strata using a model-assisted regression tree estimator. arXiv preprint, arXiv:171205708

McGarvey R, Burch P, Matthews JM (2016) Precision of systematic and random sampling in clustered populations: habitat patches and aggregating organisms. Ecol Appl 26(1):233–248

Møller J (1994) Lectures on random Voronoi tesselations. Springer, New York

Mostafa SA, Ahmad IA (2017) Recent developments in systematic sampling: a review. J Stat Theory Pract 12(2):1–21

Näslund M (1930) Om medelfelets beräkning vid linjetaxering [on computing the standard error in line-surveys]. Svenska SkogsvFör Tidskr 28:309–342

Nelson R, Næsset E, Gobakken T, Ståhl G, Gregoire TG (2008) Regional forest inventory using an airborne profiling LiDAR. J For Plan 13(2):287–294

Nothdurft A, Vospernik S (2018) Climate-sensitive radial increment model of Norway spruce in Tyrol based on a distributed lag model with penalized splines for year-ring time series. Can J For Res 48(8):930–941

Opsomer JD, Francisco-Fernàndez M, Li X (2012) Model-based non-parametric variance estimation for systematic sampling. Scand J Stat 39(3):528–542. https://doi.org/10.1111/j.1467-9469.2011.00773.x

Opsomer JD, Jay Breidt F, White M, Li Y (2016) Successive difference replication variance estimation in two-phase sampling. J Surv Statist Meth 4(1):43–70. https://doi.org/10.1093/jssam/smv033

Osborne JG (1942) Sampling errors of systematic and random surveys of cover-type areas. J Am Stat Assoc 37(218):256–264

Pagliarella MC, Corona P, Fattorini L (2018) Spatially-balanced sampling versus unbalanced stratified sampling for assessing forest change: evidences in favour of spatial balance. Environ Ecol Stat 25(1):111–123

Pal SK, Singh HP (2017) A generalized efficient ratio-cum-product estimator in systematic sampling. Int J Agric Statist Sci 13(2):713–720

Ranneby B, Cruse T, Hägglund B, Jonasson H, Swärd J (1987) Designing a new national forest survey for Sweden. Studia forestalia Suecica, vol 177. Faculty of Forestry, Swedish University of Agricultural Sciences, Uppsala

Ripley BD (1977) Modelling spatial patterns. J Roy Stat Soc B Met 39(2):172–192. https://doi.org/10.1111/j.2517-6161.1977.tb01615.x

Ripley BD (2004) Spatial statistics, 2nd edn. John Wiley, Hoboken, NJ

Robert CP, Casella G (1999) Monte carlo statistical methods. Springer texts in statistics. Springer, New York

Särndal CE, Swensson B, Wretman J (1992) Model assisted survey sampling. Springer Series in Statistics. Springer, New York

Scherer-Lorenzen M, Schulze E-D (2005) Forest diversity and function: temperate and boreal systems, vol 176. Springer, Berlin, Heidelberg, New York

Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York

Sherrill KR, Lefsky MA, Bradford JB, Ryan MG (2008) Forest structure estimation and pattern exploration from discrete-return lidar in subalpine forests of the central Rockies. Can J For Res 38(8):2081–2096

Spurr SH (1952) Forest inventory. Ronald Press, New York

Stevens DL, Olsen AR (2003) Variance estimation for spatially balanced samples of environmental resources. Environmetrics 14(6):593–610

Stevens DL, Olsen AR (2004) Spatially balanced sampling of natural resources. J Am Stat Assoc 99(465):262–278

Strand G-H (2017) A study of variance estimation methods for systematic spatial sampling. Spat Stat 21:226–240

Thompson SK (1992) Sampling. Wiley, New York

Tobler W (2004) On the first law of geography: a reply. Ann Assoc Am Geogr 94(2):304–310

Tomppo E (2006) The Finnish multi-source national forest inventory - small area estimation and map production. In: Kangas A, Maltamo M (eds) Forest inventory - methodology and applications. Managing Forest ecosystems, vol 10. Springer, Dordrecht, NL, pp 195–224

Tomppo E, Gschwantner T, Lawrence M, McRoberts RE, Gabler K, Schadauer K, Vidal C, Lanz A, Ståhl G, Cienciala E (2010) National forest inventories.

Pathways for Common Reporting. European Science Foundation. Springer, Dordrecht

Valliant R, Dorfman AH, Royall RM (2000) Finite population sampling and inference. A prediction approach. Wiley series in probability and statistics. Survey methodology section. John Wiley & Sons, New York

Vidal C, Alberdi I, Hernández L, Redmond JJ (2016) National forest inventories: assessment of wood availability and use. Springer, Cham, CH

von Gadow K, Zhang CY, Wehenkel C, Pommerening A, Corral-Rivas J, Korol M, Myklush S, Hui GY, Kiviste A, Zhao XH (2012) Forest structure and diversity. In: Pukkala T, von Gadow K (eds) Continuous cover forestry. Managing Forest ecosystems, vol 23. Springer, Dordrecht, pp 29–83

Weiskittel AR, Hann DW, Kershaw JA, Vanclay JK (2011) Forest growth and yield modeling. Wiley, Chichester, UK

Westfall JA, Patterson PL, Coulston JW (2011) Post-stratified estimation: within-strata and total sample size recommendations. Can J For Res 41(5):1130–1139. https://doi.org/10.1139/x11-031

Wilhelm M, Tillé Y, Qualité L (2017) Quasi-systematic sampling from a continuous population. Comp Stat Data Anal 105:11–23. https://doi.org/10.1016/j.csda.2016.07.011

Wolter KM (1984) An investigation of some estimators of variance for systematic sampling. J Am Stat Assoc 79(388):781–790

Wolter KM (2007) Introduction to variance estimation. Statistics for social and behavioral sciences, 2nd edn. Springer, New York