Triangulation based inclusion probabilities: a design-unbiased sampling approach

Lutz Fehrmann · Timothy G. Gregoire · Christoph Kleinn

Abstract   A probabilistic sampling approach for design-unbiased estimation of area-related quantitative characteristics of spatially dispersed population units is proposed. The developed field protocol includes a fixed number of 3 units per sampling location and is based on partial triangulations over their natural neighbors to derive the individual inclusion probabilities. The performance of the proposed design is tested in comparison to fixed area sample plots in a simulation with two forest stands. Evaluation is based on a general approach for areal sampling in which all characteristics of the resulting population of possible samples is derived analytically by means of a complete tessellation of the areal sampling frame. The example simulation shows promising results. Expected errors under this design are comparable to sample plots including a much greater number of trees per plot.

Keywords   Design based inference · Inclusion probability · Delaunay Triangulation · Plot design · Continuous population

L. Fehrmann (✉) · C. Kleinn
Chair of Forest Inventory and Remote Sensing, Georg-August-Universität Göttingen, Büsgenweg 5, 37077 Göttingen, Germany
e-mail: lfehrma@gwdg.de

C. Kleinn
e-mail: ckleinn@gwdg.de

T. G. Gregoire
Yale School of Forestry and Environmental Studies, 195 Prospect Street, New Haven, CT 06511-2104, USA
e-mail: timothy.gregoire@yale.edu

1 Introduction

Probabilistic sampling for area related quantitative characteristics of spatially dispersed populations is based on knowledge about the probability of each population unit to become part of a sample. For each included unit, this inclusion probability is the reciprocal of the expansion factor that has to be applied to derive an unbiased estimator of the total for Horvitz-Thompson estimation. A typical example is the estimation of density per area or the estimation of some other area related population characteristics like e.g. the volume per area or biomass of a forest stand.

A common approach in aeral sampling is to distribute dimensionless sample points over the whole area of interest based on a prescribed design and to define a field protocol (or plot design) that (a) is a "decision rule" defining which population units are to be included at each sample location and (b) allows an unequivocal determination of the inclusion probability for each unit. In the case of fixed area sample plots, their predefined area size directly determines the inclusion probability of all elements. Under the assumption of randomly placed sample points, the inclusion probability of a population unit can be calculated as the area proportion of all sample point locations, leading to its inclusion in a sample under the stipulated design, on the total area of the sampling frame. This locus of points is also known as the element's inclusion zone (Gregoire and Valentine 2008).

As the number of units included per fixed area plot is a random variable, it might result in a great variation in the number of elements selected in each plot in populations characterized by large variability in local density. The last property implies that the number of units included will vary randomly from one plot to another, rather than being prescribed a priori by the sampling design. This might lead to impractically large numbers in areas characterized by high local density and vice versa.

An alternative to fixed plot designs are field protocols which include a fixed number, say k, of nearest units at each sampling location. The latter group of protocols is known as restricted k-tree sampling in forestry literature, fixed count distance sampling or point-to-plant sampling. In ecology sometimes the term 'plotless density estimator' (PDE) is used. In general those protocols are a form of unequal probability sampling for populations of discrete objects (Affleck et al. 2005). They present the analyst with difficult measurements if the determination of design-unbiased inclusion probabilities is targeted (Kleinn and Vilčko 2006b).

In brief, in order to determine actual design-based inclusion probabilities for each individual element when $k > 1$, a higher order tessellation of the sampling frame into mutually exclusive polygons is required. Each polygon is the locus of all points including a specific unordered set of k units as nearest neighbors, is required (Fig. 1). The corresponding tessellation of the areal sampling frame is known as order k Voronoi diagram (Okabe et al. 1999; Aurenhammer 1991; Sibson 1980; Delincé 1986). A single element's inclusion zone is the union of all those polygons encompassing points that would include it as element in their individual sets of k nearest neighbors (Kleinn and Vilčko 2006a).

The element's inclusion probability under this design depends on the value of k and the spatial distribution of neighboring elements in its surrounding. Because the
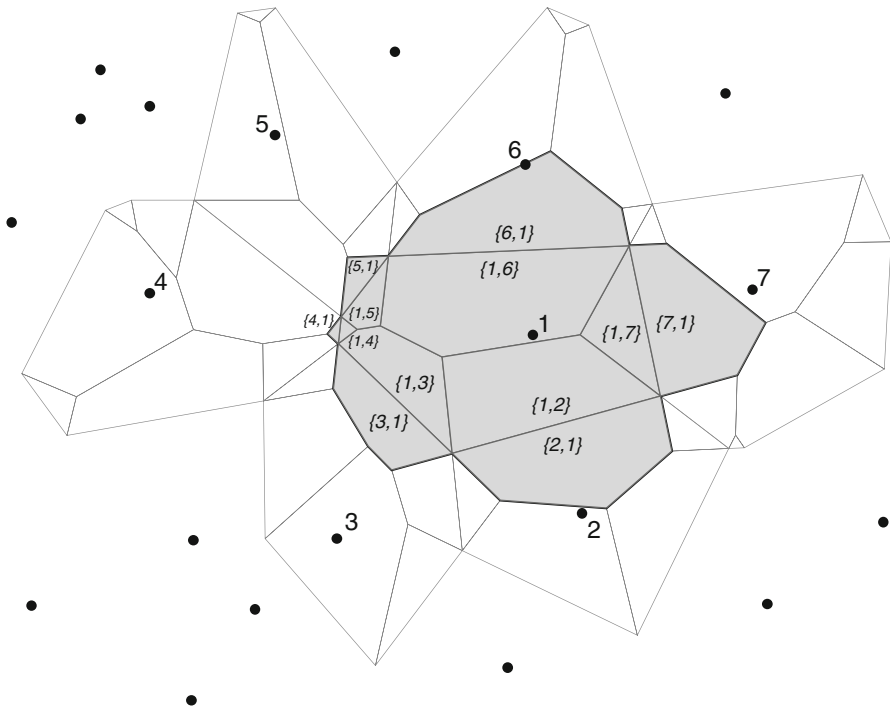
Fig. 1 Example for the design-based determination of an inclusion zone (*gray*) for the central element 1 and $k = 2$ following Kleinn and Vilčko (2006a). In a partial (here ordered) order k Voronoi tessellation all mutually exclusive polygons consisting of points that have element 1 in the vector $\{k_1, k_2\}$ of their k nearest neighbors are constituting its inclusion zone

complete construction of an order k Voronoi tessellation over the set of all included population units requires information about the location of a large number of neighboring elements, implementing it in the field is impractical. The conclusions of Kleinn and Vilčko (2006a) in regard to the determination of design-based inclusion probabilities are illustrative of the general difficulties of design-unbiased estimation following fixed-count distance sampling.

Therefore inference is often based on a presumed model of the population aiming towards an explanation of the stochastic process that determines the spatial pattern of elements. The assumption that circular sample plots with variable radius are able to approximate a suitable reference area for the observation can be seen as a model in this context, as such plots have no bearing on the virtual element's inclusion zones. Gregoire (1998) explains that both, model-based or design-based inference are valid, yet the notion of unbiasedness of an estimator appeals to different reference distributions.

As carefully articulated in Särndal et al. (1992, p. 8), a crucial precept of probability sampling is that each unit of the population must have a non-zero probability of being included in a sample. Lacking this condition, design-unbiased estimation of population parameters is not possible. In contrast, model-unbiased estimation is

possible, even in the absence of a probability sample, but only when the presumed model of the population distribution, and perhaps spatial pattern, is valid. When the unit-location pattern is one of complete spatial randomness (csr), then model-unbiased estimation of abundance is possible following either one of the sampling designs described above. In principle the same is true in the design-based framework when using the Horvitz-Thompson estimator (HTE).

Many different estimators of plant abundance or mean area per plant have been proposed for use following fixed-count distance sampling. Most involve a measure of mean distance from a point to the kth nearest element (Clark and Evans 1954; Thompson 1956; Eberhardt 1967; Prodan 1968; Batcheler 1971; Cox 1976; Barabesi 2001; Picard et al. 2005; Magnussen et al. 2008a,b; Staupendahl 2008; Law et al. 2009).

Moreover, there have been a number of empirical and simulation studies that have concentrated on comparing the performance of alternative estimators under different conditions or on comparing distance sampling to sampling with fixed area plots. An aim of these investigations has been to assess the magnitude of the model-based bias resulting from the departure from complete spatial randomness (csr) of population units (Cottam and Curtis 1956; Payandeh and Ek 1986; Engeman et al. 1994; Lynch and Rusydi 1999; Lessard et al. 2002; Lowell 1997; Nielson et al. 2004; Engeman et al. 2005; White et al. 2008). Recently Nothdurft et al. (2010) proposed an estimation approach for 6-tree sampling that is based on a simulated reconstruction of point patterns.

Fixed-count distance sampling has been widely applied also for the estimation of population density in plant communities that are well known to grow in more or less clustered or dispersed patterns (Lynch and Wittwer 2003; Steinke and Hennenberg 2006). In these cases a bias of unknown magnitude presumably is an acceptable trade-off for the relatively easy implementation of the sampling techniques in the field (Kleinn and Vilčko 2006b; Barabesi 2001; Magnussen et al. 2008b). The main argument for plotless approaches is that the fixed number of population units included per sample point result in a constant survey effort per sample point (Kronenfeld 2009; Nothdurft et al. 2010).

Inference under the design-based paradigm is free of the assumption about spatial pattern among elements. In particular, the assumption of csr is neither necessary nor relevant for purposes of design-based inference (Gregoire 1998; Affleck et al. 2005). Instead, the probabilistic inclusion of discrete stationary elements is based on the randomness introduced by the sampling design. A crucial point in model-based inference is that the model related error is constant, while the sampling error in design-based sampling usually decreases with increasing sample size (Mandallaz 1991).

The aim of this paper is to present a new field protocol that is based on a triangulation over the population elements. It includes a fixed number of $k = 3$ elements per sampling location and enables to determine actual inclusion probabilities as basis for design-unbiased inference under many conditions.

2 Materials and methods

2.1 Inclusion and selection probabilities

In accordance with standard sampling terminology, we use the term 'population parameter' to signify the actual aggregate value (or function thereof) of the quantitative characteristic of interest in the entire population, and define τ to signify this parametric value. Sampling provides a means to acquire data with which to estimate τ. The areal sampling frame from which a sampling location may be selected must have an extent that encompasses all elements of the population of interest, otherwise design-unbiased sampling is not possible. We let $\mathcal{A}$ denote the region encompassed by the areal sampling frame, and reserve A to indicate its area. Let further suppose that N suitably indexed population units are scattered over $\mathcal{A}$ at fixed locations. The continuum of the sampling frame can then be tessellated into distinct polygonal regions, each of which is the locus of all sample points that would include a specific unordered set of k population units in a sample. This tessellation arises from the intersection of all inclusion zones of distinct population units that are dependent on the field protocol to determine inclusion into the sample at any location within the sampling frame (Roesch et al. 1993; Williams 2001). It is obvious that the resulting number of polygons, or the entropy of the tessellation, is dependent on (a) the spatial distribution of population units within the area, and (b) the implication of the applied field protocol in regard to the joint inclusion of the spatially dispersed units. Single polygons represent support areas for univocal observations that can be drawn from the population. They are selected by randomly placed sample points with probability proportional to their area size (Mandallaz 1991; Gregoire and Valentine 2008).

In the following an inclusion probability π_i is the probability that the ith population unit is included in a sample, while the sample selection probability $p(s)$ denotes the probability that a distinct unordered set of k units is jointly selected as a sample according to the field protocol. Let A_i be the area spanned by the inclusion zone $\mathcal{A}_i$, of the ith population unit, where the unit of measure is identical to that of A. The inclusion probability of unit i is then given by $\pi_i = A_i/A$. Further let s index a sample comprised of a distinct set of k population units with the selection probability $p(s) = A_{\{s\}}/A$, where $A_{\{s\}}$ is the size of the support area $\mathcal{A}_{\{s\}}$ that is the intersection of the inclusion zones of all $i \in s$:

$$\mathcal{A}_{\{s\}} = \bigcap_{i \in s} \mathcal{A}_i \qquad (1)$$

Letting Ω signify the family of all possible, exclusive samples with $|\Omega|$ being the cardinality of this family, it is clear that $\sum_{s \in \Omega} p(s) = 1$, since each point in $\mathcal{A}$ can lead to the selection of a single set of k population units. As usual, an equivalent view of an element's inclusion probability is the sum of all the selection probabilities, $p(s)$, of samples in which it is a member (Gregoire and Valentine 2008):

$$\pi_i = \sum_{s \ni i} p(s) = \frac{\sum_{s \ni i} A_{\{s\}}}{A}, \qquad (2)$$

where $s \ni i$ indicates that this summation extends over all samples that include the ith element. As shown in Gregoire and Valentine (2008, p. 84), $\sum_{i=1}^{N} \pi_i$ is identical to the expected number of included elements per sample. In case that a fixed number of k population units is included at each sampling location, the above identity is equivalent to k in this case.

2.2 Triangulation based inclusion probabilities (TBIP)

The approach of TBIP sampling proposed here is based on a triangulation over the distinct population units and is designed to include $k = 3$ elements per sample location. The triangulation applied here is a Delaunay tessellation that is the dual graph of a first order Voronoi diagram. For detailed descriptions of the characteristics of this particular triangulation, see Aurenhammer (1991) or Okabe et al. (1999). A randomly placed sample point selects a single triangle of this tessellation with probability proportional to its size. In the proposed design, the included population units on which characteristics of the variable of interest are assessed are those corresponding to the vertices of the selected triangle. Following the above formulations, the inclusion probability of a population unit is proportional to the total area of the set of triangles of which it is a common vertex (Fig. 2).

As a triangulation does not exceed the extent of the population, let $\mathcal{H}$ be the convex hull originated by the unit-location pattern with area given by H. If $y_1, \ldots, y_N$ represent the values of the analyzed variable on the population units, the target parameter is given by

$$\tau = \sum_{j=1}^{N} y_j. \tag{3}$$

Considering the suggested design, it should be noticed that

$$\bigcup_{j=1}^{N} \mathcal{A}_i = \mathcal{H}. \tag{4}$$

Following Barabesi and Franceschi (2010) a useful definition of the estimate for τ at sample location u is

$$\hat{\tau}(u) = A \sum_{i=1}^{N} \frac{y_i}{A_i} I_{\mathcal{A}_i}(u), \tag{5}$$

where $I_{\mathcal{B}}(\cdot)$ represents the usual indicator function of a set $\mathcal{B}$. It should be remarked that the function $\hat{\tau}(u)$ with $u \in \mathcal{H}$ is actually depicted (up to scaling constants) in Fig. 6 for a dataset of tree locations. If u is the realization of the random variable U uniformly distributed on $\mathcal{A}$, then it is apparent that the estimator $\hat{\tau}(U)$ is unbiased for τ since
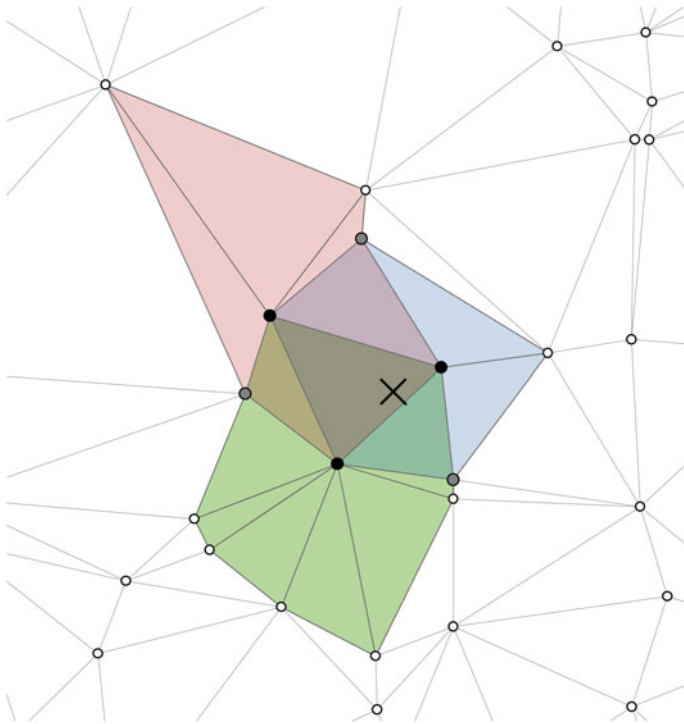
Fig. 2 Inclusion zones for 3 population units (*black dots*) included by a sample point (*cross*), their respective pairwise joint inclusion zones (overlapping individual inclusion zones) and the sample selection area (*central triangle*) for TBIP sampling

$$E[\hat{\tau}(U)] = A \sum_{i=1}^{N} \frac{y_i}{A_i} E[I_{\mathcal{A}_i}(U)] = \tau. \tag{6}$$

However, in this case the proposed design does not include exactly three units at each selected sampling location $u \in \mathcal{A}$, since the actual tessellation is restricted to $\mathcal{H}$ and $\hat{\tau}(u) = 0$ if $u \in \mathcal{A} - \mathcal{H}$. Hence the design might be less "efficient" when $\mathcal{A} - \mathcal{H}$ possesses a large area, a situation which is likely to occur for clustered population patterns. The generation of a uniform sampling location on $\mathcal{A}$ and discarding locations on $\mathcal{A} - \mathcal{H}$ would ensure that exactly three population units are included per sampling location. Nevertheless, if u is the realization of the random variable V uniformly distributed on $\mathcal{H}$, the estimator for $\hat{\tau}(V)$ is positively biased for $\hat{\tau}$ since it obviously holds that $\mathcal{H} \subseteq \mathcal{A}$:

$$E[\hat{\tau}(V)] = A \sum_{i=1}^{N} \frac{y_i}{A_i} E[I_{\mathcal{A}_i}(V)] = \frac{A}{H} \tau. \tag{7}$$

For those practical applications where the location patterns of population units implies that H is similar to A the bias might not be severe. For many population of

interest, like e.g. forest stands, it can be assumed that the the area encompassed by all population units can be approximated by a triangulation which would be similar to the forest area. Anyway, in case that H is unknown, an implementation considering the random variable U is useful, since unbiased estimation is achieved and null estimates should usually occur with relatively small probability in practice (see also Fig. 4). In other situations it might be possible to determine H by assessing the locations of those outer population units that define the convex hull and restrict sampling to this area.

Finally, as to the replicated sampling in a formal setting, we introduce n *iid* copies $U_1, \ldots, U_n$ of the random variable U, in such a way that the overall estimator be defined as

$$\hat{\tau}_n = \frac{1}{n} \sum_{i=1}^{n} \hat{\tau}(U_i). \tag{8}$$

Obviously, more improved schemes, based on stratification of $\mathcal{A}$, might be considered for the replicated placement of sampling locations (Gregoire and Valentine 2008; Barabesi and Franceschi 2010). Obviously, the estimator $\hat{\tau}_n$ is unbiased for τ and unbiased variance estimation is available since the $\hat{\tau}(U_i)$'s are in turn unbiased for τ. An unbiased estimator for $\sigma_n^2 = Var[\hat{\tau}_n]$ is given by

$$\hat{\sigma}_n^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} (\hat{\tau}(U_i) - \hat{\tau}_n)^2, \tag{9}$$

while the Central Limit Theorem and the Slutsky Theorem ensure that

$$\frac{\hat{\tau}_n - \tau}{\hat{\sigma}_n} \to \mathcal{N}(0, 1) \tag{10}$$

as $n \to \infty$.

2.3 Field implementation

In field application TBIP sampling is practically implemented by identifying three population elements that span a defined triangle around a randomly placed sample point. Regarding the Delauney tessellation that is used as example in this case, various criteria are suitable to identify the "correct" triangles. While most of them are applicable and implemented in computational geometry (Okabe et al. 1999; Halls et al. 2001), from a practical point of view they have drawbacks when trying to implement in field work. As there are multiple possible triangles that can be defined over neighboring population elements around a sample point, it is necessary to devise an efficient field protocol which will correctly identify triangles that are part of the same overall triangulation that is required for the application of TBIP.

Figure 3 shows the actual spatial constellation of population elements around a sample point. Obviously two triangles can be found that enclose the sample point here. In this case, a field protocol must be implemented which results in the correct

Fig. 3 Critical spatial constellation of population elements around the sample point. The *grey triangle* is identified as correct by comparing the distance $d_{1,3}$ and $d_{2,4}$
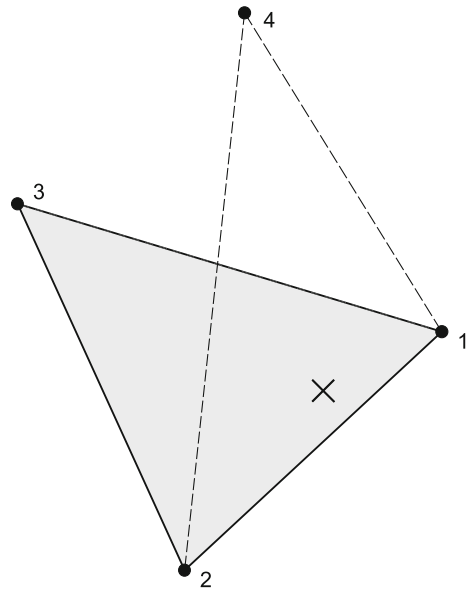
choice of either triangle 1, 2, 4 or triangle 1, 2, 3. This can be accomplished by measuring the Euclidean distances $d_{1,3}$ and $d_{2,4}$, that is, the lengths of the intersecting edges of the two candidate triangles. The triangle with shorter side length (in this example $d_{1,3} < d_{2,4}$) is the correct choice. With the proposed criterion correct triangles can be identified in many cases based on visual estimation; we suspect that distance measurements are only necessary in special cases.

Field work is restricted to the assessment of variables of interest on the three identified sample elements. Further, the position of the included elements and their natural neighbors have to be determined. Modern technologies like laser rangefinder devices significantly facilitate the necessary efforts if the design is implemented in the field. Even if a direct calculation of the single triangle areas is possible based on trigonometric principles, a more efficient implementation is to restrict field work to the measurement of distances and angles to population units. GIS applications include standard procedures that can be used to compute a triangulation over the calculated locations during post processing of field data.

2.4 Analytical case studies

We test the performance of TBIP in comparison to different configurations of fixed area circular sample plots in regard to the estimation of stem density with a small validation study based on two validation populations of real forest stands with completely mapped tree positions.

BCI is a 250×250 m subset subjectively selected from a 50 ha Smithsonian research plot on Barro Colorado Island (Condit 1998; Hubbel et al. 1999). Data of the 2,005
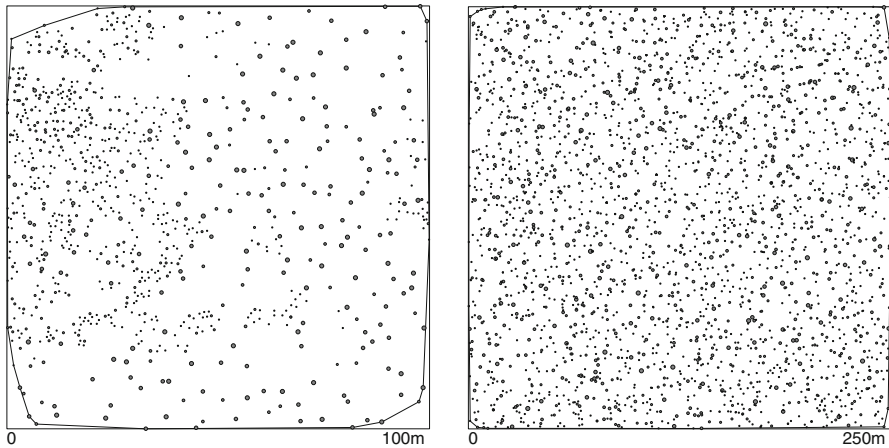
Fig. 4 Maps of tree locations for the two simulation stands LARIX (*left*) and BCI (*right*). The convex hull delineates the extent of the triangulation for TBIP sampling

census were reduced to trees with a dbh of more than 10 cm. The subset holds 2,587 trees. Density related to area of the convex hull over tree positions is 0.0417 trees m^{-2}.

LARIX is a 100×100 m long term observational plot in a mixed stand of beech and european larch with 897 trees from south Germany. Density related to the convex hull is 0.0927 trees m^{-2}.

Following Roesch et al. (1993) the complete family of possible sample outcomes (Ω) was derived with the aid of GIS software for both stands in Fig. 4. We restricted the analysis to the area of $\mathcal{H}$, so that $A = H$ in this example. All individual inclusion zones $\mathcal{A}_i$ were computed for each population unit for the different design alternatives. Possible outcomes were then calculated for each tree along formula 4. By intersecting all inclusion zones, the polygonal support areas $\mathcal{A}_{\{s\}}$ were constructed and the respective selection probability $p(s)$ was computed for each possible outcome. Finally each outcome was multiplied with its selection probability to simulate its selection under simple random sampling.

We were therefore able to portray graphically the function of $\tau(u)$ with $u \in \mathcal{H}$. For the LARIX data the randomization distribution under 6 m-radius plot sampling is displayed as Fig. 5, whereas Fig. 6 displays the randomization distribution under $k = 3$ TBIP sampling. To compare the performance of competing and alternative sampling strategies the parametric relative standard error ($SE\%$) and the coefficient of variation (CV) were calculated over all probability weighted outcomes.

3 Results

Table 1 shows the results for both stands and all different configurations of field protocols. As explained before the simulated configurations are design-unbiased by construction.
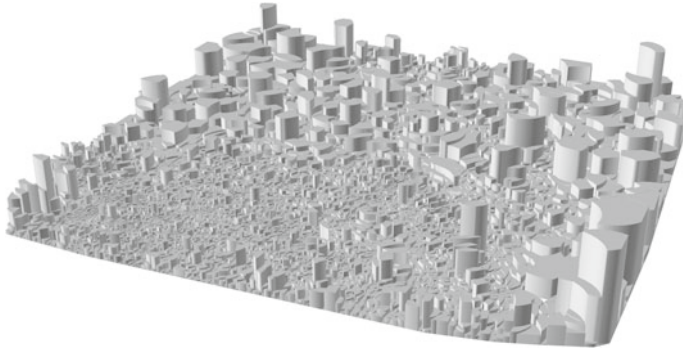
Fig. 5 Tessellation of the study area resulting from circular sample plots of 6 m radius for the LARIX data. Height of the sample selection areas $\mathcal{A}_{\{s\}}$ is proportional to $\tau(u)$ weighted with their selection probability
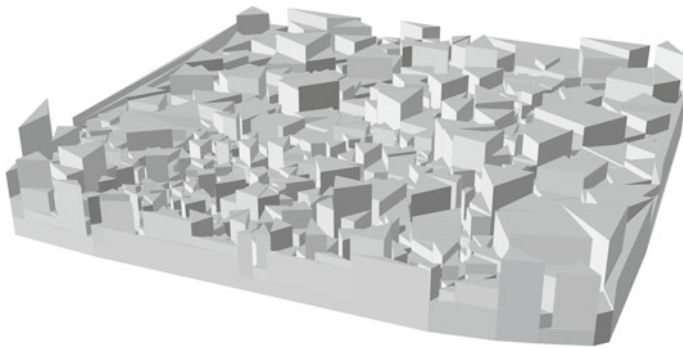
Fig. 6 Tessellation into possible samples for the LARIX data resulting from TBIP sampling. Height of the sample selection areas $\mathcal{A}_{\{s\}}$ is proportional to $\tau(u)$ weighted with their selection probability

Figures 5 and 6 display the "response surface" linked to fixed area plots with 6 m radius and the proposed strategy. Comparing the figures, it is obvious that TBIP results in a more flat surface than the fixed area plot design. The number of different observations, or the entropy, is increasing with larger radius of fixed area sample plots, as the number of intersections of single element inclusion zones is increasing. Even for relatively small sample plots of 4 m radius the number of possible sample outcomes $|\Omega|$ that describes the cardinality of Ω, is 17,729 for the BCI and 24,609 for the LARIX stand. In case of sample plots with 12 m radius 154,596 and 158,185 distinct outcomes can be drawn respectively.

TBIP sampling results in the smallest number of sample outcomes (see Fig. 6). Following Euler's formula, $|\Omega|$ under this design is always smaller than twice the number of elements in this case, depending on the number of "border" elements forming the convex hull.

The relative standard error ($SE\%$) of weighted estimates is decreasing over the plot radius as the variation of the number of trees per plot is decreasing. The $SE\%$ of TBIP sampling is in the magnitude of fixed area sample plots with 6–7 m radius in BCI and 5–6 m in LARIX respectively. Compared to the relatively high mean number of trees per plot with these configurations (>16 trees in BCI and >36 trees in LARIX), the

Table 1 Relative standard error ($SE\%$) and coefficient of variation (CV) calculated over all possible sample outcomes $|\Omega|$ from the simulation stands for circular fixed area plots of different radius and TBIP sampling. $\bar{k}_s$ is the actual mean number of trees included per sample (and maximum)

Stand	Plot design	$|\Omega|$	$SE\%$	CV	$\bar{k}_s$ (max.)
BCI	4 m	17, 729	1.13	1.51	3.0(15)
	5 m	27, 681	0.99	1.65	4.0(15)
	6 m	40, 189	0.87	1.75	5.4(16)
	7 m	54, 841	0.77	1.81	7.0(19)
	8 m	71, 157	0.69	1.83	8.9(24)
	9 m	88, 932	0.62	1.86	11.1(27)
	10 m	109, 257	0.59	1.96	13.5(30)
	11 m	130, 814	0.55	2.00	16.1(35)
	12 m	154, 596	0.51	2.00	19.1(37)
	TBIP	4, 876	0.81	0.56	3.0(3)
LARIX	4 m	24, 609	1.36	2.14	11.0(26)
	5 m	36, 554	1.18	2.25	16.1(36)
	6 m	50, 393	1.04	2.33	22.4(49)
	7 m	65, 655	1.02	2.62	29.4(63)
	8 m	82, 235	0.98	2.80	37.0(78)
	9 m	99, 839	0.95	3.00	45.5(93)
	10 m	118, 301	0.78	2.67	54.4(111)
	11 m	137, 719	0.73	2.72	63.8(126)
	12 m	158, 185	0.69	2.75	73.9(141)
	TBIP	1, 772	1.08	0.45	3.0(3)

number of included trees is constant with TBIP. Nevertheless it is important to note that on average 9 additional tree positions have to be mapped in order to delineate the inclusion zones in TBIP sampling. Figure 7 shows the distribution of probability-weighted sample outcomes resulting from TBIP sampling in both stands.

4 Discussion

To derive statistically sound estimations for area-related target variables of a population of interest, the inclusion probability of each population unit has to be known. Model-based approaches attempt to estimate a mean inclusion probability per element under the assumption that the spatial distribution of population units is a realization of a stochastic process. This process, if known, would allow to determine mean spacing between, as well as the resulting areas per elements (Ripley 1977; McIntire and Fajardo 2009).

In contrast, design-based approaches attempt to determine actual inclusion probabilities for each population unit based on the random selection of sample points from a continuum (Cochran 1977; Mandallaz 1991). As no model is applied to estimate mean spacing between the elements, it is necessary to determine the positions of a number of additional population elements around the included ones in order to
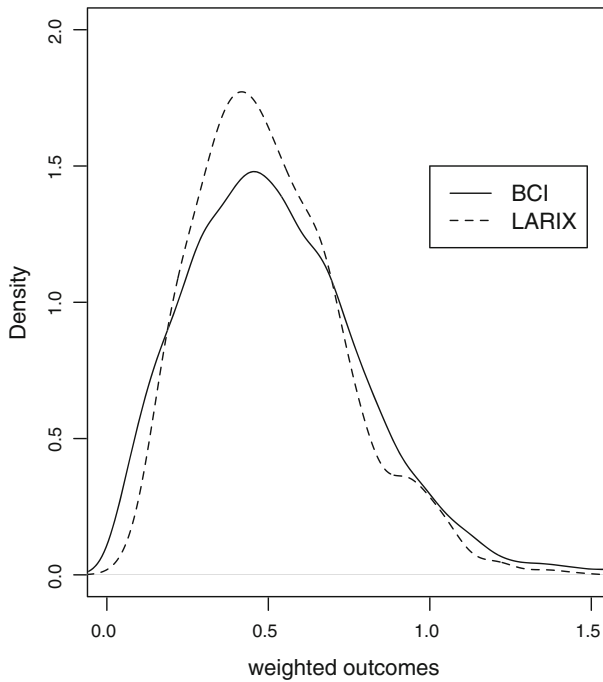
Fig. 7 Density plot of possible estimates, weighted with their selection probability resulting from TBIP sampling in BCI and LARIX

delineate their actual inclusion zones. Obviously the price of unbiasedness of design-based approaches have to be paid with this additional effort during field work.

Fraser (1977) proposed a field protocol referred to as triangle based probability polygons which is based on a triangulation over the set of population elements to determine weighted selection probabilities for single trees. Since then several studies concentrate on evaluating or modifying triangulation based approaches (Lowell 1997; Errico 1981).

Ward (1991) suggested selecting single triangles of a valid triangulation over a population of trees proportional to their size and to assign an uniform observation of 0.5 trees to each of them. This deterministic approach is based on the consideration that the internal angles of a triangle sum to π radians so that each triangle can be regarded as containing half an population unit. The mean area per element could then be estimated as twice the arithmetic mean of triangle areas and density follows. Contrary to this assumption, Euler's formula shows a different relation of the number of triangles and the set of elements over which this triangulation is derived. Considering the total number of elements as N and the number of 'border' elements forming the outer edges of the convex hull as n_e, the number of resulting triangles is given by $|\Omega| = 2N - 2 - n_e$ (Okabe et al. 1999). Lowell (1997) seizes suggestions for different triangulation-based approaches and compares them with fixed area response designs.

It is important to note that none of the mentioned approaches utilize the triangulation to assess the virtual design-based inclusion probabilities of all three elements

that are included by a triangle, but rather try to approximate the inclusion probability of single trees based on model assumptions.

One advantage of TBIP sampling is the relatively small number of additional element positions that have to be mapped in order to determine their inclusion probability. As result of the triangular constellation, the inclusion zones are overlapping in such a way that for each element 4 vertices are shared with one of its neighbors (see Fig. 2). Therefore the total number of positions that have to be mapped is reduced to $3 \cdot k = 9$ in mean. As the mean number of natural neighbors per element is assumed to be 6 (Aurenhammer 1991; Okabe et al. 1999) on average only 2 additional neighbors per element have to be mapped.

The necessary effort to determine single element inclusion probabilities is significantly reduced in comparison to the first design-based estimation approach for point to tree distance sampling proposed by Kleinn and Vilčko (2006a). They concluded that the required field work eventually disqualify their approach, as time efficiency decreases proportionally to the stipulated k. To determine actual inclusion probabilities per element in common distance sampling with $k = 3$, on average 24 neighboring elements have to be mapped, which is the set of all neighbors that unequivocal defines the order k Voronoi diagram around the included elements.

Kleinn and Vilčko (2006a) further argue that size and shape of the geometrically derived inclusion zones are not proportional to any characteristics of the respective elements and solely determined by the spatial arrangement of individuals in common distance sampling. For the response design proposed here it might be assumed that the single element inclusion probability tends to be proportional to the dimension of elements at least for some biological populations. As the individual inclusion zones are bounded by the natural neighbors of each element, size and shape are expressions of their individual demand of space which tends to be proportional to their size. On the example of forest stands it can be assumed that distances to the natural neighbors of bigger individuals are usually larger than in case of small trees, at least in even aged stands.

Potential weaknesses of TBIP sampling are related to practical field implementation, especially in the context of the determination of a valid triangulation over the set of elements at each sample point location. While the proposed field protocol can easily be applied in areas where the distribution of population elements tends to be more dense, correct triangles are less obvious and hard to determine if distances to the natural neighbors are increasing. This is especially evident near the boundary of a population, in gaps or in highly clustered populations. In this case, owing to the presence of Delaunay triangles with a very "stretched" shape, the field effort for computing the corresponding areas may be prohibitive (see Fig. 2). The proposed design might be satisfactorily applied in savannah like forest structures, where tree locations are relatively easy to determine. Further, beside possible applications in the field, the design might be suitable for sampling in aerial photos or high resolution satellite imagery, e. g. for tree resources outside forest. Many other applications in ecology, like the estimating the density of nesting sites in a bird colony are imaginable.

Further, the presented design is especially efficient for $k = 3$, a relatively low number of included elements, what might subsequently lead to relatively high variance among the derived estimations. Considering that the included elements are natural

neighbors in a compact constellation, it is expected that a relative high portion of covariance, typically occurring among nearby elements in natural populations, might be captured with this design.

Conclusions

Yet to our knowledge the proposed response design is new, even if other triangulation based techniques, mostly restricted to sampling of single sample elements, have been discussed previously (Fraser 1977; Sibson 1980; Ward 1991).

To date, practical experiences related to field implementation of TBIP are restricted to small example studies. A meaningful evaluation of the practical and statistical efficiency of TBIP sampling in comparison to alternative designs under various conditions of application is beyond the target of this study and will be addressed in future research.

From our perspective TBIP seems to be an alternative for target variables whose assessment is especially labor-intensive (because it is restricted to three elements) and the related additional effort that is necessary to determine the single element inclusion probabilities is warranted. Examples might be destructive sampling for tree biomass or other detailed measurements on single population elements. Populations in which the spatial positions of individuals are easy to determine are particular appropriate for an efficient implementation of this design. TBIP might be in particular useful if a density-adapted survey effort is welcome. Examples are the assessment of forest regeneration or other populations with high variability in local density within small distances, in which fixed area plot designs would lead to uneven and sometimes unreasonable efforts during field work.

There are many situations in which the implementation of TBIP will be impractical or less efficient than fixed area plot designs. This is for example the case in highly clustered populations or for large distances. Design-based approaches in distance sampling will therefore stay an alternative for special cases, while model-based estimators might be more efficient for other situations.

References

Affleck D, Gregoire T, Valentine H (2005) Design unbiased estimation in line intersect sampling using segmented transects. Environ Ecol Stat 12(2):139–154

Aurenhammer F (1991) Voronoi diagrams–a survey of a fundamental geometric data structure. ACM Comput Surv 23(3): 345–405. doi:10.1145/116873.116880

Barabesi L (2001) A design-based approach to the estimation of plant density using point-to-plant sampling. J Agric Biol Environ Stat 6(1):89–98

Barabesi L, Franceschi S (2010) Sampling properties of spatial total estimators under tessellation stratified designs. Environmetrics. doi:10.1002/env.1046

Batcheler CL (1971) Estimation of density from a sample of joint point and nearest-neighbor distances. Ecology 52(4):703–709

Clark PJ, Evans FC (1954) Distance to nearest neighbor as a measure of spatial relationships in populations. Ecology 35(4):445–453

Cochran WG (1977) Sampling Techniques. Whiley, New York

Condit R (1998) Tropical Forest Census Plots. Springer and R. G. Landes Company, Berlin, Germany, and Georgetown, Texas

Cottam G, Curtis JT (1956) The use of distance measures in phytosociological sampling. Ecology 37(3):451–460

Cox TF (1976) The robust estimation of the density of a forest stand using a new conditioned distance method. Biometrika 63(3):493–499

Delincé J (1986) Robust density-estimation through distance measurements. Ecology 67(6):1576–1581

Eberhardt LL (1967) Some developments in 'distance sampling'. Biometrics 23:207–216

Engeman RM, Sugihara RT, Pank LF, Dusenberry WE (1994) A comparison of plotless density estimators using monte carlo simulation. Ecology 75(6):1769–1779

Engeman RM, Nielson RM, Sugihara RT (2005) Evaluation of optimized variable area transect sampling using totally enumerated field data sets. Environmetrics 16(7):767–772

Errico D (1981) Some methods of sampling triangle based probability polygons for forestry applications. Master's thesis, Department of forestry, University of British Columbia

Fraser A (1977) Triangle-based probability polygons for forest sampling. For Sci 23:111–121

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

Gregoire TG, Valentine HT (2008) Sampling strategies for natural resources and the environment. Applied environmental statistics. Chapman Hall/CRC, London

Halls PJ, Bulling M, White PCL, Garland L, Harris S (2001) Dirichlet neighbours: revisiting dirichlet tessellation for neighbourhood analysis. Comput Environ Urban Syst 25(1): 105–117. doi:10.1016/S0198-9715(00)00035-1

Hubbel S, Foster R, O'Brian S, Harms K, Condit R, Wechsler B, Wright S, de Lao L (1999) Light gap disturbance, recruitment limitation, and tree diversity in a neotropical forest. Science 283:554–557

Kleinn C, Vilčko F (2006a) Design-unbiased estimation for point-to-tree distance sampling. Can J For Res 36:1407–1414

Kleinn C, Vilčko F (2006b) A new empirical approach for estimation in k-tree sampling. For Ecol Manag 237:522–533

Kronenfeld BJ (2009) A plotless density estimator based on the asymptotic limit of ordered distance estimation values. For Sci 55(4):283–292

Law R, Illian J, Burslem D, Gratzer G, Gunatilleke CVS, Gunatilleke IAUN (2009) Ecological information from spatial patterns of plants: insights from point process theory. J Ecol 97:616–628

Lessard VC, Drummer TD, Reed DD (2002) Precision of density estimates from fixed-radius plots compared to n-tree distance sampling. For Sci 48(1):1–6

Lowell KE (1997) An empirical evaluation of spatially based forest inventory samples. Can J For Res 27:352–360

Lynch T, Wittwer R (2003) n-Tree distance sampling for per-tree estimates with application to unequal-sized cluster sampling of increment core data. Can J For Res 33(7): 1189–1195. doi:10.1139/X03-036

Lynch TB, Rusydi R (1999) Distance sampling for forest inventory in indonesian teak plantations. For Ecol Manag 113:215–221

Magnussen S, Kleinn C, Picard N (2008a) Two new density estimators for distance sampling. Eur J For Res 127(3): 213–224. doi:10.1007/s10342-007-0197-z

Magnussen S, Picard N, Kleinn C (2008b) A Gamma-Poisson distribution of point to k nearest event distance. For Sci 54(4):429–441

Mandallaz D (1991) A unified approach to sampling theory for forest inventory based on infinite population models. Phd thesis, ETH Zürich

McIntire EJB, Fajardo A (2009) Beyond description: the active and effective way to infer processes from spatial patterns. Ecology 90(1): 46–56. doi:10.1890/07-2096.1

Nielson RM, Sugihara RT, Boardman TJ, Engeman RM (2004) Optimization of ordered distance sampling. Environmetrics 15(2):119–128

Nothdurft A, Saborowski J, Nuske R, Stoyan D (2010) Density estimation based on k-tree sampling and point pattern reconstruction. Can J For Res 40:953–967

Okabe A, Boots B, Sugihara K, Chiu SN (1999) Spatial tesselations. concepts and applications of voronoi diagrams. Wiley Series in Probability and Statistics. Wiley, New york

Payandeh B, Ek A (1986) Distance methods and density estimators. Can J For Res 16(5):918–924

Picard N, Kouyate A, Dessard H (2005) Tree density estimations using a distance method in mali savanna. For Sci 51(1):7–18

Prodan M (1968) Punktstichprobe für die Forsteinrichtung. Forst und Holzwirt 23(11):225–226

Ripley BD (1977) Modelling spatial patterns. J Royal Stat Soc Ser B (Methodological) 39(2):172–212

Roesch FA, Green EJ, Scott CT (1993) An alternative view of forest sampling. Surv Methodol 19(2):199–204

Särndal CE, Swensson B, Wretman J (1992) Model assisted survey sampling. Springer, New York 694

Sibson R (1980) The dirichlet tessellation as an aid in data analysis. Scand J Stat 7(1):14–20

Staupendahl K (2008) The modified six-tree-sample—a suitable method for forest stand assessment. Allgemeine Forst und Jagdzeitung 179(2–3):21–33

Steinke I, Hennenberg K (2006) On the power of plotless density estimators for statistical comparisons of plant populations. Can J Bot 84(3): 421–432. doi:10.1139/B05-135

Thompson HR (1956) Distribution of distance to nth neighbor in a population of randomly distributed individuals. Ecology 37(2):391–394

Ward D (1991) Triangular tesselation: a new approach to forest inventory. For Ecol Manag 44:285–290

White N, Engeman R, Sugihara R, Krupa H (2008) A comparison of plotless density estimators using monte carlo simulation on totally enumerated field data sets. BMC Ecol 8(1):6. doi:10.1186/1472-6785-8-6

Williams MS (2001) New approach to areal sampling in ecological surveys. For Ecol Manag 154:11–22

Author Biographies

Lutz Fehrmann is a postdoc at the Chair of Forest Inventory and Remote Sensing at Georg-August-Universität Göttingen, Germany. His current research covers sampling techniques in natural resources assessment and biomass modeling.

Timothy G. Gregoire is on the faculty of the School of Forestry and Environmental Studies at Yale University, where he is the J. P. Weyerhaeuser, Jr., Professor of Forest Management. His research is directed to the application and development of statistical methods for natural resources and environmental phenomena. One longstanding focus has been on probability sampling techniques and a second focus has been on statistical modeling of longitudinal and spatially correlated data.

Christoph Kleinn is professor at the Chair of Forest Inventory and Remote Sensing at Georg-August-Universität Göttingen, Germany. His research interests are the optimization of natural resources inventory, monitoring and assessment.