ENVIRONMETRICS
Environmetrics 2000; 11: 13–41

RESPONSE DESIGNS AND SUPPORT REGIONS IN SAMPLING CONTINUOUS DOMAINS

DON L. STEVENS JR[1]* AND N. SCOTT URQUHART[2]

[1]*Dynamic International, Inc., 200 S.W. 35th Street, Corvallis, OR 97333, U.S.A.*
[2]*Department of Statistics, Oregon State University, Corvallis, OR 97331, U.S.A.*

SUMMARY

In many environmental samples, the target population is distributed over space in a more or less continuous manner, e.g., the waters of a lake or the trees in a forest. Attributes of such a population can be conceptualized as a continuous function defined on the spatial domain of the population. Some attributes (e.g., water temperature) can be observed at a point; others (e.g., species diversity) can only be determined over a finite extent or support region. A fixed-shape support with uniform weights leads to an unbiased estimator of the population total; however, it may be impossible to maintain a fixed shape near domain boundaries. From a purely formal standpoint, unbiasedness can be maintained by using differential weights or by changing the shape of the support region near the boundary. Both of these procedures raise some issues of interpretation that often are overlooked. We derive estimators that account for edge effects under several support strategies, and identify some interpretation issues, using examples from forestry and limnology. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS environmental sampling; plot size; response design; line intersect sampling

1. INTRODUCTION

Many environmental or biological populations can be viewed as spatial populations, by which we mean populations whose elements, be they points or objects, occupy fixed locations in a bounded spatial matrix. Examples include all lakes in Maine, wadeable streams in Oregon, and the Florida everglades. Attributes of spatial populations can sometimes be determined at a point by a measurement process that takes place at that point. For example, the concentration of a chemical in lake water and the depth of a stream are measured at a point. Many, if not most, attributes of spatial environmental populations require measurements taken over some neighborhood surrounding a point, e.g., any measure of a rate or density must be determined over some non-negligible support. Nevertheless, these attributes can be conceptualized as continuous surfaces defined on the domain of the population. For example, a measure of biomass per unit area or stand density in forestry, or a measure of fish abundance or diversity in surface waters, must be

* Correspondence to: D. L. Stevens, Jr., Dynamac International Inc., 200 S.W. 35th Street, Corvallis, OR 97333, U.S.A.

Contract grant sponsor: U.S. Environmental Protection Agency.
Contract grant number: 68-C6-0005.
Contract grant number: 68-C4-0019.
Contract grant number: CR 821738.
Contract grant number: CR 824682.

CCC 1180–4009/2000/010013–29$17·50

Received 10 October 1998
Revised 16 May 1999

determined over some support region, but these quantities can be, and often are, represented as surfaces or maps. The value of the surface at a point is the aggregate over a window surrounding the point. We generate the surface by moving the window over the extent of the population domain.

A common issue in sampling spatial populations is the selection of an appropriate sample area or plot on which measurements are actually carried out, that is, the determination of the support of the observation. A facet of this question is the appropriate support near an edge of the spatial domain of a population where a plot might cross the boundary. It is well-known (Finney 1948; Barret 1964; Fowler and Arvantis 1981; Gregoire 1982; Gregoire and Scott 1990) that inappropriate treatment of plots crossing the boundary can result in a discrepancy, sometime called edge-effect bias, between the mean of the plot-level values and the mean of the population. In this paper, we treat edge-effect bias as a special case of a more general problem, that of ensuring that the mean of the plot-level values is the same as the mean of the population. We term this requirement aggregation-unbiasedness, and derive a sufficient condition to ensure aggregation unbiasedness. We examine several popular plot layouts in light of this condition, and illustrate that it is sometimes easier to show the condition is satisfied than to establish unbiasedness directly. We also investigate several techniques for applying that condition to one- and two-dimensional resource domains.

2. AGGREGATION-UNBIASED RESPONSE DESIGNS

To introduce the notions, suppose R is some two-dimensional region, s is a point in R and $z(s)$ is an attribute expressed as a rate or density, e.g., a measure of fish abundance as number of fish per square meter of lake surface, or a measure of forest stand density as number of trees per hectare. A desirable property of any such attribute is that its integral over R be equal to the true total over R. If $z(s)$ were fish abundance, expressed as fish/m^2 in a lake, then

$$\int_{\text{lake area}} z(s)\, ds$$

should equal the number of fish in the lake. If $z(s)$ represents biomass per square meter in some region R, then

$$\int_R z(s)\, ds$$

should equal the total biomass in R.

If the field determination of the attribute value is an average over the plot, then on the surface there would seem to be little problem, since both averaging and integration are linear operations. Certainly, there is no problem if the size, shape, and orientation of the field plot is kept constant. However, it is not possible to do so for a bounded region, a feature of most real situations. A problem arises along a boundary when a plot, whose location is determined by a point, straddles the boundary (e.g., Finney 1948; Fowler and Arvantis 1981; Gregoire 1982). Either we adopt the Procrustean solution of redefining the boundary of R to fit the shape of the plot, we move the plot or modify its shape, or we introduce an analytic correction. All of these options have been used (Chhikara 1994; Gregoire and Scott 1990; Harrison and Dunn 1993; Scott and Bechtold 1995; Moisen *et al.* 1995). All raise the issue of potential bias, especially if edge conditions are

substantially different from interior conditions, or if R is of a shape with a large perimeter to area ratio.

Much of the statistical literature has treated this problem from the perspective of finite population sampling by treating the plot as a sample unit, with the population then consisting of a finite number of sample units. However, a development from the perspective of sampling points from a continuum has some advantages, and that is the viewpoint taken here. Sometimes called point-sampling, this approach is mentioned in the statistical literature (e.g., Yates 1960) and in the applied sampling literature (e.g., Grosenbaugh 1958). Cordy (1993) has published a continuous version of the Horvitz–Thompson theorem, and Campbell (1993) has investigated spatial support for samples from a continuum. Stevens (1997) has developed sampling designs that are directed explicitly to continuous spatially distributed populations. In this paper, we use that conceptual viewpoint to address some questions of plot shape, in particular, the impact of changing plot shape or size near a boundary.

Our conceptual viewpoint is that we are trying to infer properties of a continuous surface $z(s)$, which we will do by selecting points from the continuum that is the domain of the function z. However, the perspective of sampling from a continuous universe also provides the theoretical framework for sampling discrete populations distributed over space. One of the major problems in sampling such a population is obtaining an adequate frame. The populations are finite, but obtaining a list or catalog of all population elements is frequently prohibitively expensive or impractical, and probably not a good idea in any case. Forest sampling is a good example: sampling a stand of trees by labeling every tree in the stand, and then selecting from the list of labels is feasible only for relatively small stands. Furthermore, the population almost certainly has some spatial pattern, and random sampling from a list does not lead to a straightforward way to disperse the sample over space.

There are two conceptually separate and distinct design activities involved here. One design effort is determining what to measure, count, or observe given that we are at some point in the population domain, and how to combine or synthesize the measurements, counts, or observations collected. This effort is response design: the process of deciding what to measure and how to measure it; of defining and giving substance to the quantity or quantities we associate with the point. The other design activity is sampling design: the process of specifying how and where to select population units or points on the response surface. The response at these points will be used to estimate attributes of the response surface, which, if we have done the response design correctly, will bear some known relationship to the attributes of the population of real interest.

These two processes are conceptually distinct, but often are confused in practice. The task of finding suitable and efficient ways to sample a spatial environmental population by exploiting its spatial component is greatly simplified if we keep these two processes operationally distinct. The benefit of doing so is that we can develop efficient and practical response designs that exploit the local characteristics of the population, and develop efficient sampling designs that exploit the regional characteristics of the population. We then appeal to sampling theory to establish design-unbiasedness of proposed estimators of the response surface parameters, and appeal to the response design to establish the link between the response surface parameters and population parameters.

To formalize the concepts, we assume that the population we wish to describe exists within some domain R, a subset of one, two, or three dimensional Euclidean space, and that $\|R\|$, the measure (length, area, or volume) of R, is non-zero. Some populations of environmental interest are naturally viewed as continuous surfaces, e.g., a chemical concentration within a lake, or the

elevation of a terrestrial region. Others are more akin to point-like collections of discrete objects, e.g., the trees in a forest. Our approach covers both extremes: in the continuous case, we take the population to be an integrable function $z(t)$ defined on R. In the case of discrete objects located at the points $t_1, t_2, \ldots, t_N$ with values $z_1, z_2, \ldots, z_n$, we take the population as the generalized function

$$z(t) = \sum_i^N z_i \delta(t - t_i),$$

where $\delta(t)$ is the Dirac delta function. (The Dirac delta function is a generalized function, also called a distribution, with the properties that $\delta(t) = 0$, $t \neq 0$, and

$$\int_{-\infty}^{\infty} \delta(t)\, \mathrm{d}t = 1.$$

See Richards and Youn (1990), for a discussion of distributions.) For the discrete case, we will also need the generalized identity function

$$i(t) = \begin{cases} \sum_i^N \delta(t - t_i), & \text{discrete population} \\ 1, & \text{otherwise} \end{cases}.$$

In either case, we interpret

$$\int_R z(t)\, \mathrm{d}t = z_T$$

as the total of the population over R. Our objective is to estimate z_T and the mean value $\mu_z = z_T/\|R\|$. This is a quite general objective, as estimates of variances, proportions, and distribution functions can also be formulated as estimates of integrals over R. For example, the distribution function for $z(s)$ over R is defined as

$$F_z(x) = |\{s \in R : z(s) \leqslant x\}|/\|R\|.$$

Expressed as an integral,

$$F_z(x) = \int_R I_{\{s|z(s) \leqslant x\}}(s)\, \mathrm{d}s/\|R\|,$$

where $I_A(X)$ is the indicator function for A defined as

$$I_A(x) = \begin{cases} 1, & x \in A \\ 0, & \text{otherwise} \end{cases}.$$

Note that in even the discrete case, we are interested in the average spatial density of z, not the per element average,so that space is an intrinsic aspect of our objective. In the forest example, z_i might be the basal area of tree i, and we would be interested in the average basal area per acre of the forest, not in the average basal area of the trees in the forest. Space is also an intrinsic part of our strategy to estimate z_T and μ_z, in that we will do so via a sample $s_1, s_2, \ldots, s_n$ drawn from R.

Let $D(s) \subset R$ be some neighborhood of support around a point s such that all population elements in $D(s)$ are used in determining the value of the response surface at s. For example, if we were sampling trees, $D(s)$ might be the intersection of the area covered by the stand with a circle of fixed radius centered on s. If we were sampling benthic communities, $D(s)$ might be specified by the number, spatial arrangement, and depth of the bottom samples taken at a site. We call $D(s)$ the *plot configuration*: It is the particular arrangement of shape, orientation, and extent over which measurements are taken. Often, $D(s)$ is simply the plot associated with s. However, there are some techniques discussed in the environmental sampling literature where $D(s)$ is more general than a simple plot. An example is variable-radius point sampling in the forestry literature (Grosenbaugh 1958; Schreuder *et al.* 1993) in which $D(s)$ becomes a finite point set of tree locations.

We are explicitly allowing the plot configuration to change with the location s. In particular, the plot configuration near the boundary may be different from the plot configuration in the interior of R. For example, let $C(s)$ be a circle centered on s. $D(s)$ might be the entire circle in the interior of R but be some other shape, e.g., $C(s) \cap R$, near a boundary of R. In Figure 1, for example, the plot configuration might be the shaded areas at points s_1 and s_2.

We suppose that the measurement process on D can be carried out so that the resulting sample value is a faithful representation of the underlying surface. In practice, the sample value of the

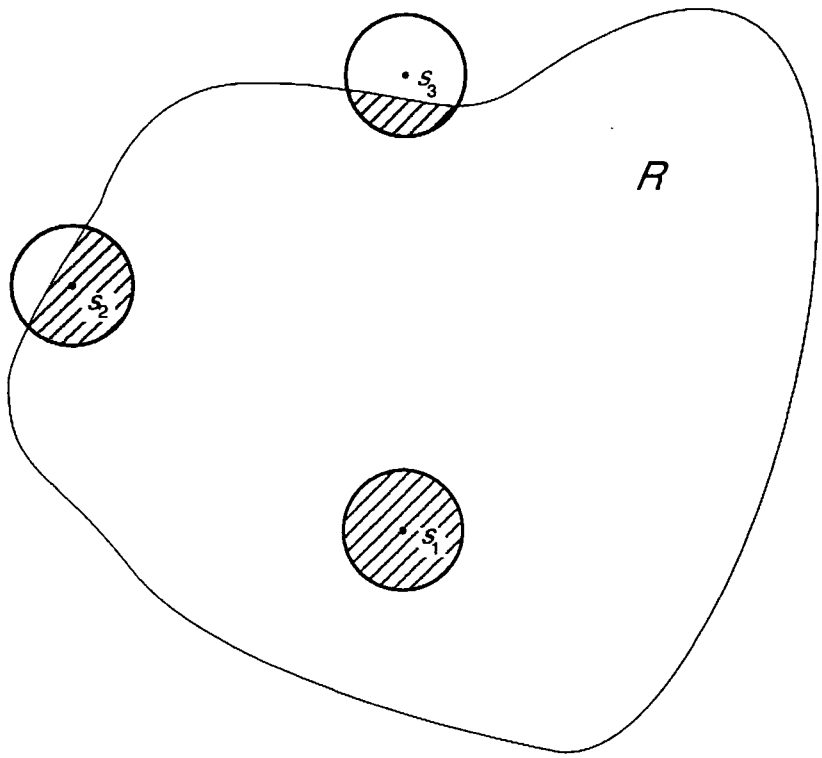

Figure 1. Plot configuration examples. The shaded regions around points s_1 and s_2 show possible plot configurations. We assume in this paper that inferences about R use only points within R, such as s_1 and s_2. No observation would be collected at s_3 even though its nominal plot intersects R

response at the point is usually calculated as an average or similar aggregation of observations taken along a transect, at some regular grid of points covering D, or as an average over some collection of discrete units within D (e.g., all trees larger than some minimum size). For generality of presentation, such finite sums will be replaced with an integral, so that the sample value at s will have the form

$$\bar{z}_D(s) = \int_R I_{D(s)}(t)z(t)g(s, t)\, dt. \tag{1}$$

where $g(s, t)$ is a weighting or aggregation function that provides the prescription for aggregation over $D(s)$, and $I_A(x)$ is the indicator function for the set A defined as

$$I_A(x) = \begin{cases} 1, & x \in A \\ 0, & \text{otherwise} \end{cases}.$$

We assume that $g(s, t) \geqslant 0$, and that $g(s, t)$ vanishes for t outside $D(s)$. The integral form allows us to treat both the continuous and discrete cases simultaneously, inasmuch as in the discrete case, (1) becomes

$$\bar{z}_D(s) = \sum_{t_i \in D(s)} z(t_i)g(s, t_i). \tag{2}$$

The aggregation function $g(s, t)$ is the weight given the population element at t when it is included in the response surface at s. We allow g to depend on both s and t because some response designs in use, e.g., the mirage method in forestry (Gregoire, 1982), require such generality.

The response surface defined by our response design is $\bar{z}_D$, and is the population we will actually sample. Inferences we make using a sample $\bar{z}_D(s_1), \ldots, \bar{z}_D(s_n)$ will be inferences about $\bar{z}_D$, not necessarily about z. Thus, the population we sample is defined and given substance by our choice of response design and aggregation function. We can carry through some properties of the underlying population to the response surface by making an appropriate choice of a response design and an aggregation function. We cannot hope to preserve all characteristics of z, however. For example, we would expect that the variance of $\bar{z}_D$, both locally and regionally, to be smaller than the variance of z. We need to place restrictions on the aggregation function $g(s, t)$ to insure that the response design preserves the population characteristic of interest. In particular, for our objective of estimating z_T, we want

$$\int_R \bar{z}_D(s)\, ds = \int_R z(s)\, ds.$$

a property which we label aggregation-unbiased. This property has been termed 'design-unbiased' (Gregoire and Monkevich 1994) or just 'unbiased' (Scott and Bechtold 1995) and analyzed as an aspect of sampling design. However, we view it as strictly a property of the response design, inasmuch as it has nothing to do with where sample points are located. It can be investigated and controlled independently of any sampling design that might be used to locate points at which to sample.

In formulating this requirement, we have implicitly eliminated some options for response designs. In particular, we have made an explicit choice to define the response only on the support of those points that are in R. Point s_2 in Figure 1 would be used in the estimator, perhaps with

modified support because of its proximity to the boundary. No observation would be collected at sample point s_3, even though $C(s_3)$ intersects R. An alternative would be to include points s outside of R but which have support circles that intersect R. We do not take this approach, but instead assume that no observation is collected for s outside of R.

If our response surface has the aggregation-unbiased property, then any sampling design can lead to a design-unbiased estimator via the Horvitz–Thompson theorem (Horvitz and Thompson 1952) or its continuous, infinite-population analog (Cordy 1993; Stevens 1997). In particular, if a sample design were to select points $s_1, \ldots, s_n$ with inclusion probability densities $\pi(s_1), \ldots, \pi(s_n)$, then

$$\hat{\bar{z}}_{D,T} = \sum \frac{\bar{z}_D(s_i)}{\pi(s_i)}$$

is an unbiased estimator of z_T. Moreover, a variance estimator can be obtained from formulas given in Cordy (1993), Stevens (1997), or Stevens and Kincaid (1998). For example, the continuous domain Horvitz–Thompson variance estimator is

$$\hat{V}_{HT}(\hat{\bar{z}}_{D,T}) = \sum_{s_i \in R} \frac{\bar{z}_D^2(s_i)}{\pi^2(s_i)} + \sum_{s_j \in R} \sum_{\substack{s_j \in R \\ j \neq i}} \left[\frac{\pi(s_i, s_j) - \pi(s_i)\pi(s_j)}{\pi(s_i, s_j)\pi(s_i)\pi(s_j)} \right] \bar{z}_D(s_i)\bar{z}_D(s_j). \tag{3}$$

The variance estimator depends only on the response surface observations, i.e., the aggregated observations, and properties of the sampling design through the inclusion and pair-wise inclusion probability functions evaluated at the sampling points. We stress that even in the finite, discrete population case, there is no need to calculate inclusion or pair-wise inclusion probabilities for the individual population units that get included in the sample. We need to evaluate inclusion information only at sampling design points. Thus, for example, if we were estimating total tree volume in a stand of trees, we do not need inclusion nor pair-wise inclusion probabilities for individual trees. We need only know the inclusion and pairwise inclusion density functions at the location of our sample points.

The concept dual to plot configuration is the *inclusion field*. The plot configuration for a point s is the set of all points t such that the value of $z(t)$ is used in calculating the aggregate value at s. The inclusion field for a point t is the set of all potential sample points s that include t in their plot configuration. Thus, the inclusion field for a point $t \in R$ is the region around t containing all s such that t is in the plot configuration of s, that is, a point s is in the inclusion field of t if and only if t is in the plot configuration of s. If we regard $D(s)$ as a set function mapping points in R to a collection of subsets of R, the inclusion field for t is the inverse image of $D(s)$, that is, the inclusion field is $D^{-1}(t) = \{s | t \in D(s)\}$. This relationship can also be expressed in terms of indicator functions as $I_{D(s)}(t) = I_{D^{-1}(t)}(s)$. We note that $D^{-1}(t)$ will either be an empty set, or else will have non-zero length, area, or volume in contrast to $D(s)$, which, in some circumstances, is a finite point set. As we show later, the inclusion field is critical to the determination of the properties of the response surface.

When we define a response design, we can choose to define the plot configuration $D(s)$ for every point $s \in R$ or the inclusion field $D^{-1}(t)$ for every point $t \in R$ (or at least at every t where there is a population element). The two approaches are two faces of the same coin. We can use whichever is most convenient since defining one implicitly defines the other. Both approaches have been used:

'point-sampling' as discussed by Grosenbaugh (1958) is based on specifying $D^{-1}(t)$; Kershaw (1964) and Grieg-Smith (1964) recommend quadrats positioned by randomly located points to sample vegetation.

To find conditions on $g(s, t)$ so that

$$\int_R \bar{z}_D(s)\,\mathrm{d}s = \int_R \int_R I_{D^{-1}(t)}(t) z(t) g(s, t)\,\mathrm{d}t\,\mathrm{d}s = z_T,$$

we use the identity $I_{D(s)}(t) = I_{D^{-1}(t)}(s)$, and interchange the order of integration to obtain

$$\int_R \int_R I_{D(s)}(t) z(t) g(s, t)\,\mathrm{d}t\,\mathrm{d}s = \int_R \int_R I_{D^{-1}(t)}(s) z(t) g(s, t)\,\mathrm{d}t\,\mathrm{d}s = \int_R z(t) \int_{D^{-1}(t)} g(s, t)\,\mathrm{d}s\,\mathrm{d}t.$$

It follows that for

$$\int_R \bar{z}_D(s)\,\mathrm{d}s = z_T$$

for arbitrary $z(\cdot)$, we must have that

$$\int_{D^{-1}(t)} g(s, t)\,\mathrm{d}s \equiv 1.$$

More generally, suppose that y and z are two variables defined on R, and let $\bar{y}_D(s)$ and $\bar{z}_D(s)$ be defined by (1). If $y(t) = h(z(t))$, and h is a non-linear function, then in general, $\bar{y}_D \neq h(\bar{z}_D)$, because integration does not preserve non-linear relationships. Thus, the choice of plot configuration has an impact on the observed relationship. Moreover, even in the linear case, where $y(t) = \beta_0 i(t) + \beta_1 z(t)$, the plot configuration can impact the relationship. Applying (1), it follows that

$$\bar{y}_D(s) = \beta_0 \int_R I_{D(s)}(t) i(t) g(s, t)\,\mathrm{d}t + \beta_1 \int_R I_{D(s)} z(t) g(s, t)\,\mathrm{d}t = \beta_0 \int_{D(s)} i(t) g(s, t)\,\mathrm{d}t + \beta_1 \bar{z}_D(s). \quad (4)$$

Thus, the linear relationship with $\beta_0 \neq 0$ is independent of plot configuration if and only if

$$\int_{D(s)} i(t) g(s, t)\,\mathrm{d}t = 1.$$

which becomes

$$\int_{D(s)} g(s, t)\,\mathrm{d}t = 1$$

in the continuous case or

$$\sum_{t_i \in D(s)} g(s, t_i) = 1$$

in the discrete case. If we set $g(s, t) = 0$ for s outside $D^{-1}(t)$, then the requirement to preserve mean values and to preserve linear relationships is that $g(s, t)$ be *doubly stochastic*, that is,

$$\int_R i(t)g(s, t)\, dt = 1$$

and

$$\int_R g(s, t)\, ds = 1.$$

If the relationship between y and z is subject to noise, or if observations have a noise component, than any analysis of the effect of aggregation on associations between variables should take that noise into account. See, for example, Robinson (1950), Yule and Kendall (1950), or Openshaw and Taylor (1979). Equation (4) shows that even in the perfect world of no noise, there can be an aggregation effect.

If g does not depend on s in the sense the weight assigned to the population element at $t \in D(s)$ is the same for any point s, the unbiasedness requirement becomes

$$g(s, t) = \frac{I_{D(s)}(t)}{\|D^{-1}(t)\|}. \tag{5}$$

Thus, in this case we can get aggregation-unbiasedness by taking the aggregation function as the reciprocal of the size of the inclusion field. This result is sufficiently counter-intuitive that it is worth stressing: We achieve aggregation-unbiasedness through scaling observations within the plot by the size of their inclusion fields, not by the plot size. For example, if we were sampling trees using a circular plot in the interior of R, and a plot determined by the intersection of the circle with R near the boundary, we do not achieve unbiasedness merely by giving those plots near the boundary less weight. The aggregation function would be tree-specific rather than plot-specific. Moreover, the only way we can avoid a location-specific weight is for $\|D^{-1}(t)\|$ to be constant. Thus, there can be a considerable advantage to selecting $D(s)$ so that $\|D^{-1}(t)\|$ is constant, or conversely, specifying the response design by specifying a constant-measure $D^{-1}(t)$. If, in addition, g is doubly stochastic, this implies that $\|D(s)\| = \|D^{-1}(t)\| = \|D\|$, i.e., the plot configuration and the inclusion field have equal and invariant area.

In some instances, we may be interested in more than just z_T. Environmental assessments, for example, may also require an estimate of the proportion of R that meets (or fails to meet) a requirement on $\bar{z}_D$ of the form $\bar{z}_D \geqslant z_0$. Because the variance of $\bar{z}_D$ depends on all of z, $D(s)$, and $g(s, t)$, such an assessment depends critically on our choice of response design and aggregation function. A response design with the property that $\|D(s)\| = \|D^{-1}(t)\| = \|D\|$ will ensure that the response design does not unduly influence the assessment by introducing local variability in $\bar{z}_D$ resulting from variation in $\|D(s)\|$.

A sufficient set of conditions for $\|D(s)\| = \|D^{-1}(t)\| = \|D\|$ is that $D(s)$ be both translation congruent ($t \in D(s) \Leftrightarrow t + h \in D(s + h)$, for any h) and radially symmetric ($t + h \in D(s) \Leftrightarrow t - h \in D(s)$, for any h). These conditions imply that $D^{-1}(t) = D(t)$, and are satisfied for plot configurations that are common geometrical shapes such as squares, rectangles, or hexagons. The radial symmetry condition is not satisfied by a triangular plot configuration, and the presence of edges inevitably guarantees that translation congruence fails to hold.

There are cases where either R has no boundaries, or we can construct a topology on R to eliminate edges. If R is the entire surface of the earth, then spherical-squares, -rectangles, or -hexagons satisfy the translation congruent and radially symmetric conditions. If R is a 1-dimensional domain, e.g., a stream segment, we can eliminate ends by forming R into a circle (cf., Fuller 1970). If R is a rectangle, we can eliminate its boundaries by deforming it into a torus, joining the top to the bottom edge and the left to the right edge. A plot that overlapped a boundary of R would cross one or more seams. The corresponding field plot would be located by cutting the torus along the seams, and mapping back to the plane. A plot crossing a seam would consist of disconnected fragments, e.g., a plot crossing the right boundary would have a piece appearing on the left boundary. Such constructions can be used to satisfy the theoretical requirements, but do not provide practical field protocols.

3. LINEAR DOMAIN EXAMPLES

The interplay between plot configuration, inclusion field, and aggregation function is most easily illustrated in the 1-dimensional case. The specifics of the examples are motivated by the sampling of streams where the plot configuration is often a length of stream around a sample point, say a fixed distance on either side of the point as at point s_1 in Figure 2. If the sample point happens to

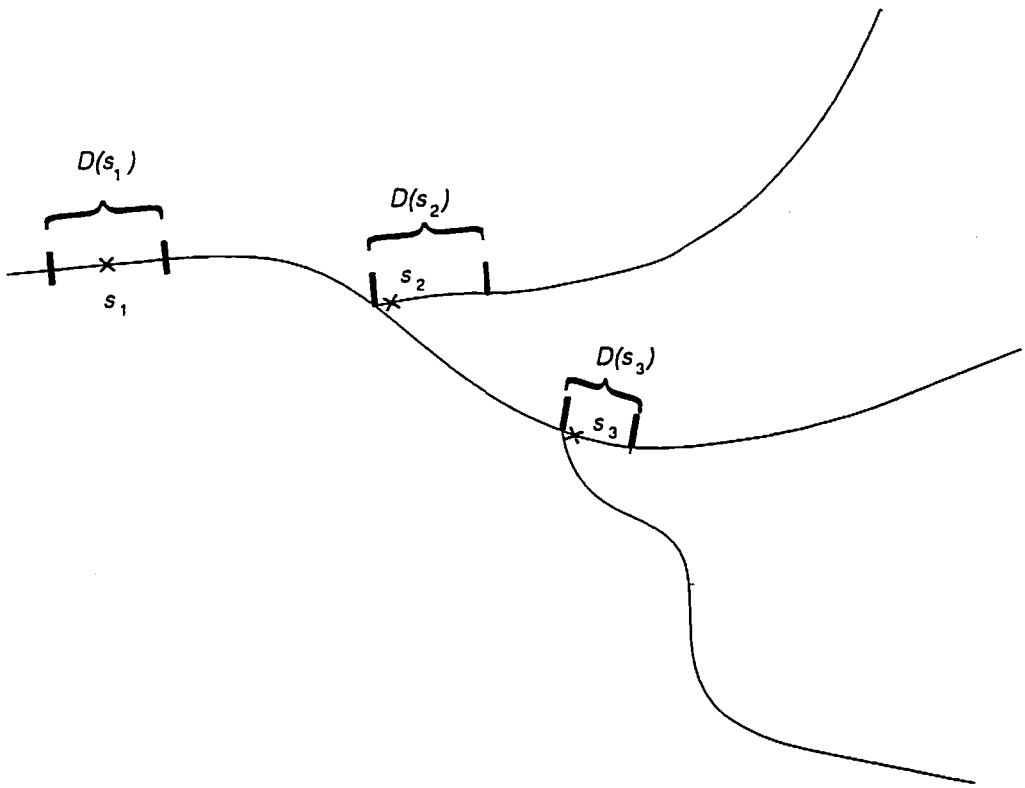

Figure 2. Sample locations and support regions on a hypothetical stream network. $D(s_1)$ is symmetric about the sample point; $D(s_2)$ is the same length as $D(s_1)$ but is asymmetric about the sample point, and $D(s_3)$ is truncated at a downstream confluence

be near a confluence with another stream, so that the plot would cross the confluence, some limnologists have kept the entire plot on one reach to avoid habitat changes that might occur at the confluence and that could obscure the interpretation of the observation. We will look at three strategies for doing this: (1) slide the plot back onto the reach so that a constant length of stream is sampled for all points (see point s_2 in Figure 2); (2) truncate the plot at the endpoints of the reach, resulting in shorter plot lengths for those points near the endpoints (see point s_3 in Figure 2), and (3) reflect the plot back onto itself when it overlaps the end of the stream reach. The reflect strategy is the 1-dimensional analog of the mirage method used in forestry (Schmid-Haas 1969; Gregoire 1982; Gregoire and Monkevich 1994) and discussed in Section 4.2. Physically, of course, the truncate strategy and the reflect strategy result in the same portion of the stream being sampled. The differences between the two lie in the aggregation function, and possibly in the on-site protocol, as discussed below.

Suppose the stream reach we wish to sample has length L, which we represent by an interval on the real line, so that $R = [0, L]$. For the 'slide' strategy, the plot configuration has constant length, and is given by

$$D_{SL}(s) = \begin{cases} [0, 2\delta], & 0 \leqslant s \leqslant \delta \\ [s - \delta, s + \delta], & \delta < s \leqslant L - \delta \\ [L - 2\delta, L], & L - \delta < s \leqslant L \end{cases}.$$

In contrast, for the 'truncate' and 'reflect' strategies, the plot configuration at s is the interval of width 2δ centered on s intersected with R. It is truncated at both the upper and lower ends of the reach and thus is given by

$$D_{TR}(s) = \begin{cases} [0, s + \delta], & 0 \leqslant s \leqslant \delta \\ [s - \delta, s + \delta], & \delta < s \leqslant L - \delta \\ [s - \delta, L], & L - \delta < s \leqslant L. \end{cases}$$

The upper half of Figure 3 shows plots of s versus the upper and lower endpoints of $D_{SL}(s)$ and $D_{TR}(s)$. For any s in R, the plot configuration is the projection onto the t-axis of the vertical line segment between the two endpoint curves. A schematic of this operation is shown in the upper left of Figure 3, where a vertical line is drawn upward from the s-axis, and the segment between the two endpoint curves is projected onto the t-axis. Conversely, the inclusion field is the projection onto the horizontal axis of the horizontal line segment between the two curves. For example, from the plot of the endpoints for the slide option, it can be seen that

$$D_{SL}^{-1}(t) = \begin{cases} [0, t + \delta], & 0 \leqslant t \leqslant 2\delta \\ [t - \delta, t + \delta], & 2\delta < t \leqslant L - 2\delta \\ [t - \delta, L], & L - 2\delta < t \leqslant L. \end{cases}$$

The lower half of Figure 3 shows the length of the inclusion field for the three strategies. For example, as can be read from the lower right plot in Figure 3, the length of the inclusion field for the truncate or reflect strategy is given by

$$\|D_{TR}^{-1}(t)\| = \begin{cases} t + \delta, & 0 \leqslant t \leqslant \delta \\ 2\delta, & \delta < s \leqslant L - \delta \\ \delta + L - t, & L - \delta < t \leqslant L. \end{cases}$$

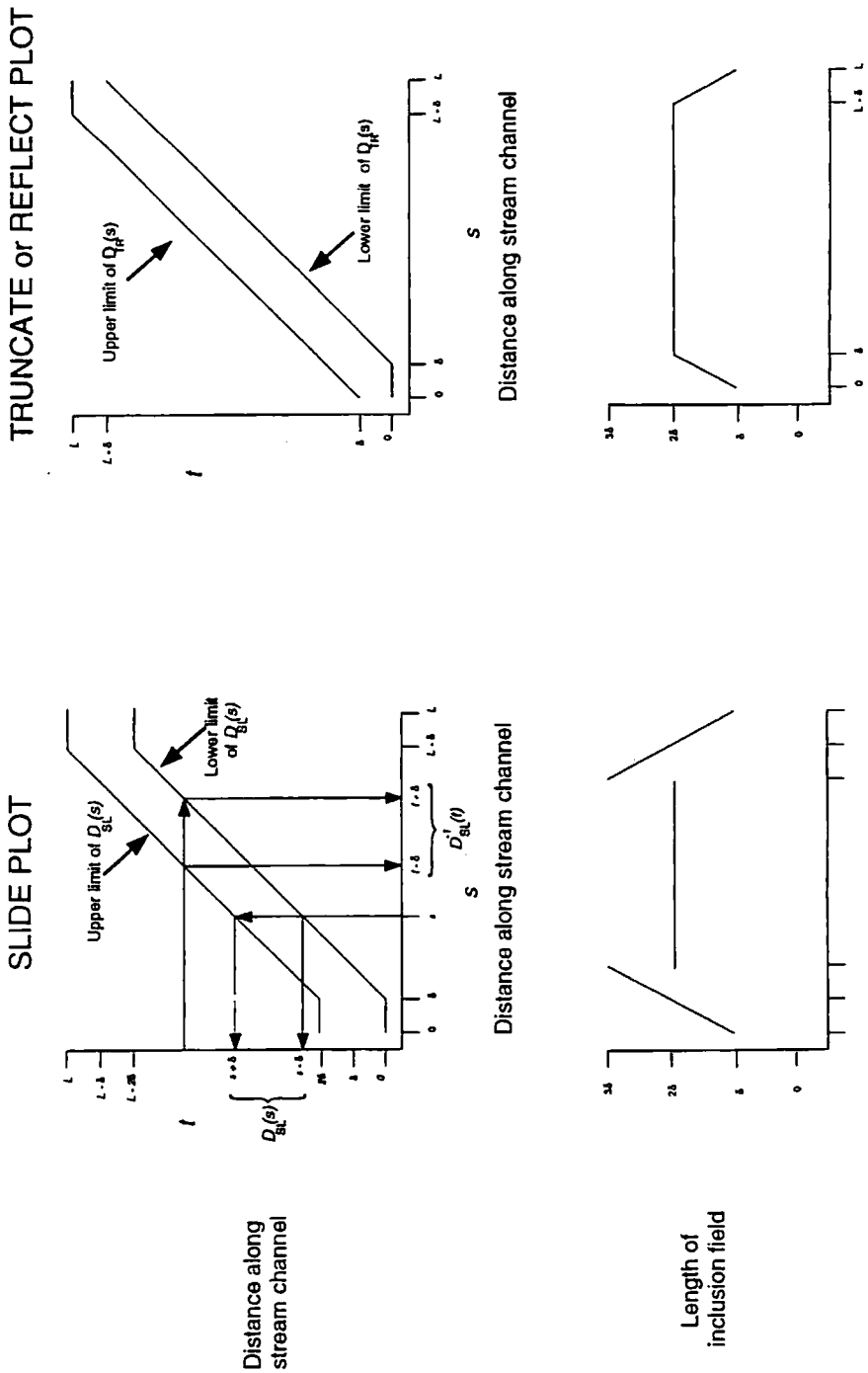

Figure 3. Schematics of the relationship between a point s and its support $D(s)$ and of the relationship between t and its inclusion field $D^{-1}(t)$ for strategies slide, truncate, and reflect for a linear resource. The length of the inclusion field as a function of distance along stream channel is in the lower half of the figure

An aggregation function that gives aggregation-unbiasedness for the slide and truncate strategies can be obtained by taking

$$g(s, t) = \frac{I_{D(s)}(t)}{\|D^{-1}(t)\|}.$$

so that

$$\int_{D^{-1}(t)} g(s, t)\, ds \equiv 1.$$

Schematics of these functions are given in Figure 4. In neither case is the aggregation function doubly stochastic, so linear relationships are not preserved by either the slide or truncate strategy.

The reflect strategy folds the portion of the plot that extends beyond the stream reach back onto the reach, with the intention that the portion of the stream that is twice-covered should receive twice the relative weight of the remainder of the plot. Figure 4 also has a schematic of this aggregation function: the function takes on the value $1/\delta$ in the two hatched regions; $1/2\delta$ in the central box; and vanishes elsewhere. This aggregation function is doubly stochastic, so both mean values and linear relationships are preserved. Application of the aggregation function could be accomplished analytically, during the data analysis stage, or, for some kinds of measurements, accomplished by a modification of the field procedure. For example, suppose the response is abundance of benthic organism, and the standard protocol calls for organisms to be collected at a rate of k benthic grabs per meter. If the collection intensity were changed to $2k$ grabs per meter in the reflected portion of the plot (Figure 5), and every grab received the same weight in the aggregation, then $\bar{z}_D$ would be aggregation-unbiased.

It is possible to obtain constant length inclusion field, but at the expense of constant length support. This alternative would be represented by interchanging the roles of $D(s)$ and $D^{-1}(t)$ in the slide strategy. The circular topology mentioned in Section 2, together with a symmetric support interval, gives constant plot length and constant inclusion field length, but is not an attractive field protocol. This has the disadvantage of giving a support that is the composite of two segments of R that may be widely separated, and would seem to be of little practical utility. In the stream sampling context, this option connects the upper end of a headwater reach to the confluence end, and potentially results in the disparate mixture of habitats we tried to avoid.

4. TWO-DIMENSIONAL EXAMPLES

In this section, we discuss several examples in which the domain of the resource is two-dimensional, even though the resource itself may be finite, e.g., a forest occupies area, but the trees within the forest constitute a finite population. First, we identify plot configurations, inclusion fields, and aggregation functions for two widely used techniques, line-intercept sampling and tree-concentric sampling in forestry. We then examine some techniques that ensure either constant area plot configuration or constant area inclusion field.

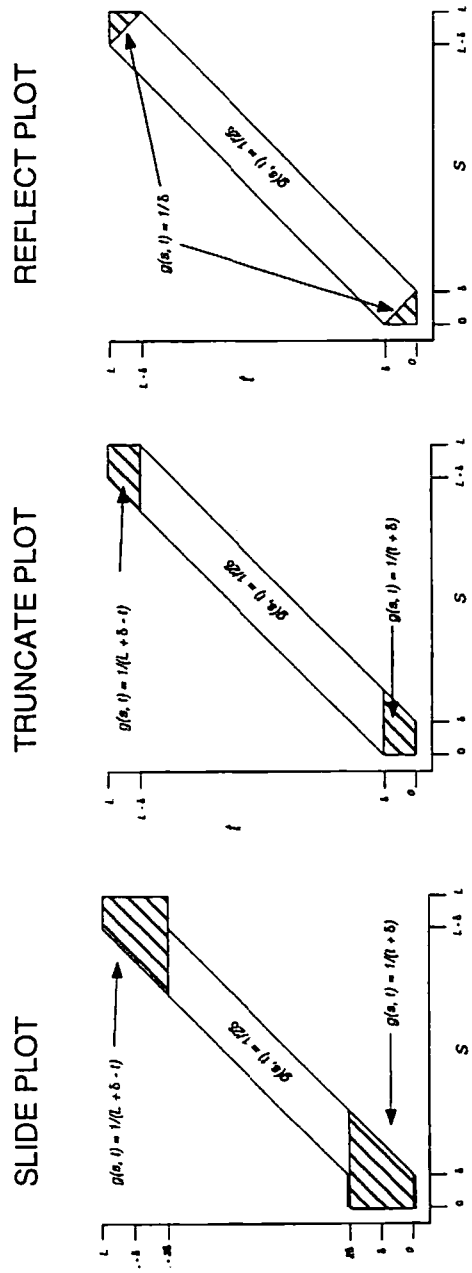

Figure 4. Sketches showing the definitions of aggregation functions that give aggregation-unbiased response surfaces for the slide, truncate, and reflect strategies. In each case, the functions vanish outside the indicated regions

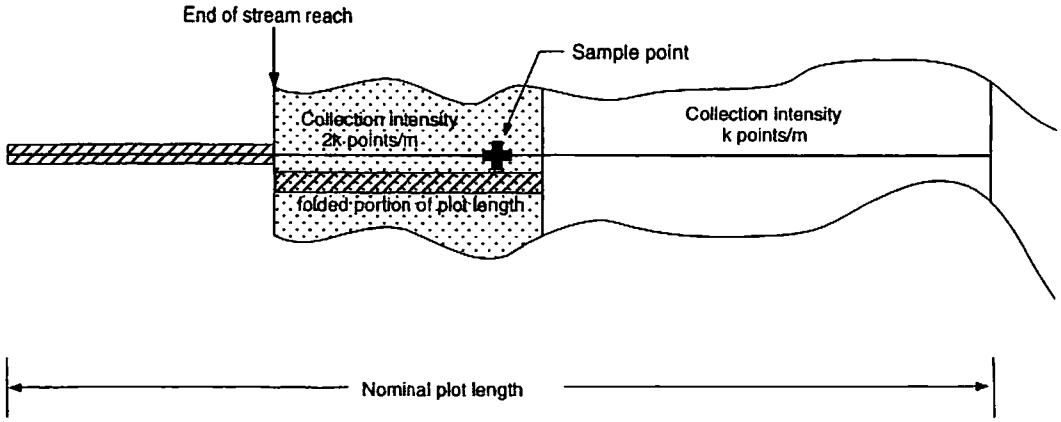

Figure 5. One way to apply the reflect strategy to a linear resource. The portion of the nominal plot extending beyond the end of the stream reach is folded back onto the reach. The collection intensity is doubled within the portion of the reach covered by the folded plot

4.1. Line-intercept sampling

Line-intercept sampling (LIS) (Kaiser 1983; Thompson 1992; Gregoire and Monkevich 1994) is a method of selecting a sample from a population of objects, usually called *particles*, which are scattered over some planar region R. The particles may be such things as trees, shrubs, animal dens, logging debris, roads, animal tracks, etc. In our view of LIS, each particle has an associated index location t_i, such as the particle centroid. We will consider the particle selected if the index location is selected. The method is applied by locating a line L at random in R, and including in the sample all those particles P_i intersected by the line. Some characteristic z_i is measured for each particle selected, and the object is to estimate Z_T or μ_z.

Several variations of the method use different line placements or different specific definitions of 'intersection'. One of the simplest methods (Thompson 1992, pp. 225–230) selects particles if they are intersected by a line passing through R perpendicular to some baseline that spans R (see Figure 6). The particle located at t_i is selected by any point that falls in the perpendicular projection of the particle onto baseline, so that $D^{-1}(t_i)$ is just that perpendicular projection. For a point s on the baseline, $D(s)$ consists of all t_i such that P_i is intersected by the perpendicular line through s, i.e., $D(s)$ is a finite point set of index locations. Referring to Figure 6, we have that $D(s_1) = \{t_2\}$, $D(s_2) = \{t_3, t_4\}$, and $D(s_3)$ is the empty set. It follows from (1) and (5) that

$$\bar{z}_D(s) = \sum_{t_i \in D(s)} \frac{z_i}{\|D^{-1}(t_i)\|}$$

is an aggregation unbiased response surface, in this case, a 1-dimensional 'surface'.

The inclusion field for this method, and for line-intercept sampling in general, depends on attributes of the population unit located at t_i, e.g., on the shape and orientation of P_i. If we want to achieve aggregation-unbiasedness, the aggregation function must be population dependent. In

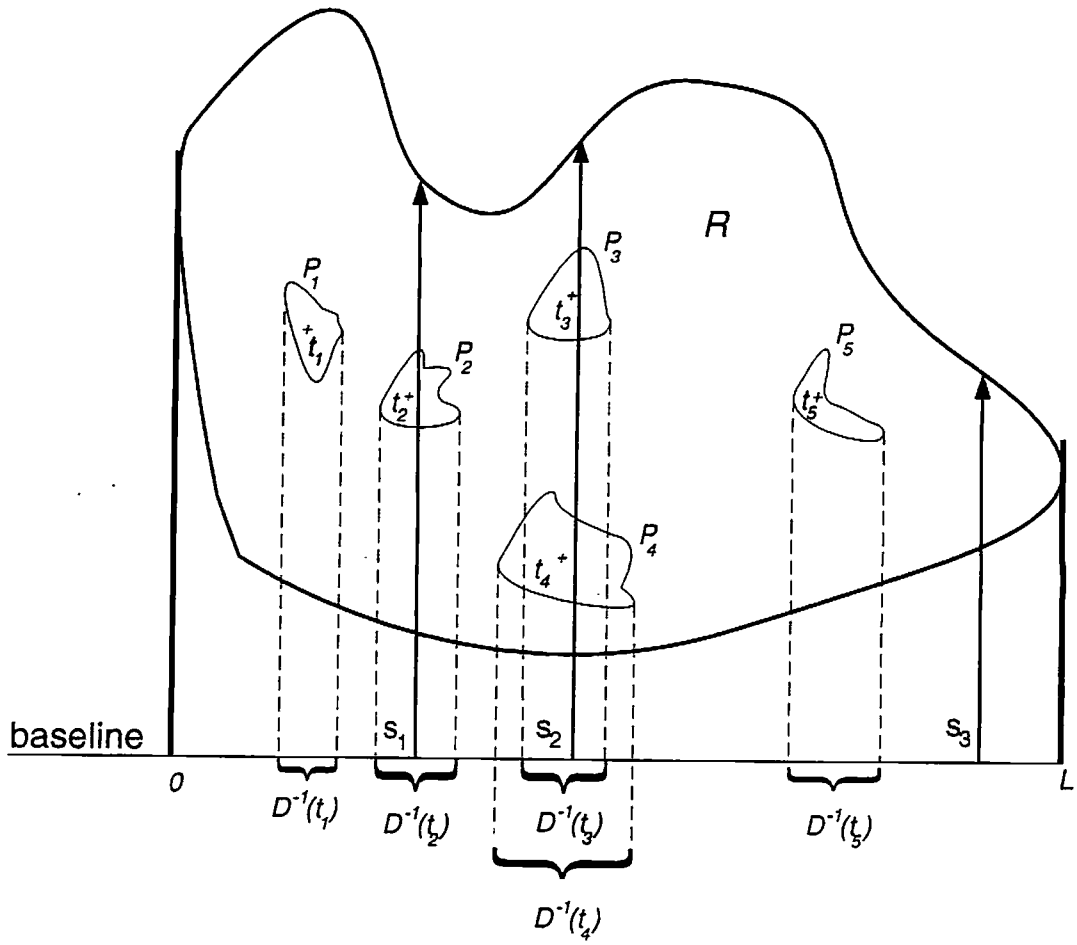

Figure 6. An application of line intercept sampling. R is projected onto the baseline of length L, as are the particles within R. Particles are selected if they are intercepted by lines perpendicular to the baseline at sample points, so that $D^{-1}(t_i)$ is the segment on the baseline in which a point must fall to select P_i. Thus, s_1 selects P_2, s_2 selects P_3 and P_4, and s_3 fails to select a particle

turn, this implies that we have no hope of obtaining a doubly stochastic aggregation function, so that relationships will inevitably be distorted by this method of sampling.

Suppose the projection of R onto the baseline has length L, and a sample is selected by picking n points independently and uniformly distributed over $[0, L]$. For such a sampling design, $\pi(s) = n/L$, and $\pi(s_i, s_j) = n(n-1)/L^2$, so that

$$\hat{\bar{z}}_D = \sum_{j=1}^{n} \frac{\bar{z}_D(s_j)}{\pi(s_j)} = \frac{L}{n} \sum_{j=1}^{n} \sum_{t_i \in D(s_j)} \frac{z_i}{\|D^{-1}(t_i)\|}$$

is an unbiased estimator of z_T, with variance estimator from Equation (3) given by

$$\frac{n}{n-1}\sum_{j=1}^{n}(\bar{z}_D(s_j)-\hat{\mu}_z)^2.$$

These are the same estimators given by Thompson (1992, p. 226), however, we have obtained them from a very different perspective. An advantage of our viewpoint is that we can easily change sampling designs without having to derive new estimators. For instance, we could use a linear analog to a randomized tessellation stratified design (Stevens 1997) to ensure that sample points were spread more or less evenly across the baseline, or use a variable probability design to concentrate sample points in regions where we knew or suspected a high particle concentration. Either of these designs would likely be more efficient than simple random sampling, and would entail no additional analytic effort from our perspective.

Kaiser (1983) and Gregoire and Monkevich (1994) discuss another version of line-intercept sampling, where L is a line of fixed length l located at random in R. Their discussions cover alternatives where the orientation θ of L is taken as fixed or random. We consider here only the case of fixed θ. Extension to the random case is straightforward. In this version of line intercept sampling, 'intersects' means either (1) L passes completely through P_i, or (2) an arbitrarily specified endpoint of L falls within P_i. In our terminology, the method specifies that the set $D^{-1}(t_i)$ for the particle P_i located at t_i consists of all points s such that a line L centered on s with fixed orientation θ intersects P_i. Figure 7 illustrates application of the intersection rule and determination of the inclusion field using the convention that a particle is intersected if L passes through the particle or the left endpoint of L is within the particle. Particles P_1 and P_2 are intersected by the lines at s_1 and s_2, respectively, while P_3 is not intersected by the line at s_3. The hatched area around P_4 is the region in which a point must fall in order to select P_4, i.e., it is $D^{-1}(t_4)$. Kaiser (1983) provides a derivation of the computations necessary to determine $\|D^{-1}(t)\|$.

Transects that cross the boundary, such as those near P_5, can be treated in two different ways. Let $C(t)$ be the area around the particle at t that covers the locations of all lines that intersect the particle. If the particle is in the interior of R, then $C(t)$ is the inclusion field, but if t is near a boundary, so that $C(t)$ overlaps the boundary, some other choice is necessary. Kaiser (1983) suggests translating the transect segment outside the boundary some distance to the left or right along the boundary. This has the effect of moving the part of $C(t)$ that falls outside the boundary back into R in such a way as to maintain the area of the inclusion field. In Figure 7, the inclusion field using this option would consist of the portion surrounding P_5 plus the cross-hatched portion above. The area of $D^{-1}(t)$ depends only on particle size and shape, not particle location, so $\|D^{-1}(t)\| = \|C(t)\|$. Setting

$$g(s,t)=\frac{I_{D(s)}(t)}{\|D^{-1}(t)\|}$$

in (1) gives an aggregation-unbiased response surface.

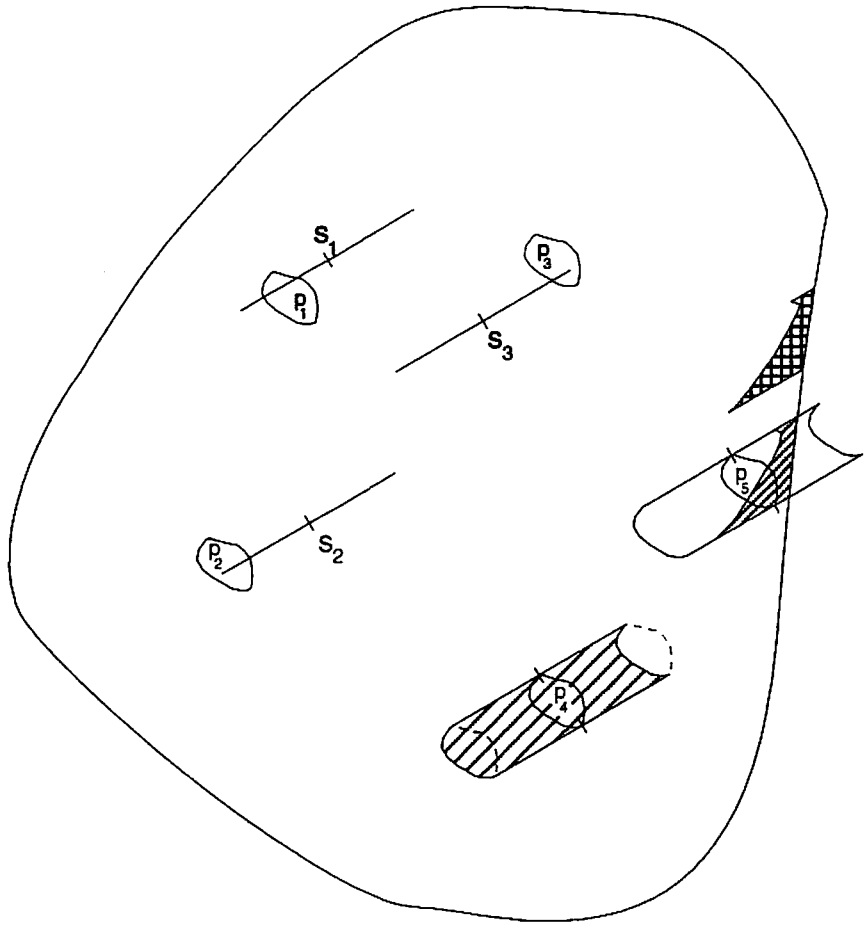

Figure 7. Line intercept sampling of particles by randomly placed lines. Particles are intercepted if the line passes completely through the particle, or the particle covers the left endpoint of the line. Particles P_1 and P_2 are intercepted by the lines at s_1 and s_2, respectively, while P_3 is not intercepted by the line at s_3. The hatched area around P_4 is $D^{-1}(t_4)$. The hatched area near P_5 illustrates the reflection strategy; the cross-hatched area illustrates the translation strategy

Gregoire and Monkevich (1994) showed how to achieve aggregation-unbiasedness by reflecting a transect crossing the boundary back along itself. Their approach sets $D^{-1}(t) = C(t) \cap R$, and sets the aggregation function equal to

$$2\frac{I_{D(s)}(t)}{\|C(t)\|}$$

if the particle is intersected by both the transect and its reflected portion, and equal to

$$\frac{I_{D(s)}(t)}{\|C(t)\|}$$

otherwise. In Figure 7, $C(t_5)$ overlaps the boundary, $D^{-1}(t_5)$ is the portion of $C(t_5)$ inside R, and

$$g(s, t) = \frac{2}{\|C(t)\|}$$

in the hatched portion of $D^{-1}(t_5)$. Clearly,

$$\int_{D^{-1}(t)} g(s, t)\, ds = 1,$$

so they obtain an aggregation-unbiased response surface.

4.2. Tree-concentric sampling

A sampling method sometimes used in forestry applications assigns each tree a circle, centered on the tree, with radius that may depend on some property of the tree. For example, a frequently used option makes the area of a tree circle proportional to the basal area of the tree. A point at s selects all trees such that s falls within the tree's circle. If $C(t_i)$ denotes the circle of radius $r_i(t_i)$ centered on the tree at t_i, then in our terminology, this method takes $D^{-1}(t_i) = C(t_i)$ for trees in the interior of R. Several variations have been used to handle trees whose circles intersect the boundary. See Schreuder *et al.* (1993, pp. 207–301) for a summary some of the methods. Gregoire and Scott (1990) compared the bias and mean square error for several of the methods. Here, we examine two of the methods, the tree-concentric method (Gregoire and Scott 1990) and the mirage method (Schmid-Haas 1969; Gregoire 1982). The tree-concentric approach takes $D^{-1}(t_i) = R \cap C(t_i)$ and achieves aggregation-unbiasedness by setting

$$g(s, t) = \frac{I_{D(s)}(t)}{\|D^{-1}(t)\|}.$$

The mirage method also uses $D^{-1}(t_i) = R \cap C(t_i)$, but uses a different aggregation function, derived by reflecting the portion of $C(t_i)$ that falls outside of R about the boundary. The reflected portion falls back into $C(t_i)$, and partitions the area of $D^{-1}(t_i)$ into two disjoint pieces, labelled $A_1(t_i)$ and $A_2(t_i)$ in Figure 8(a). Clearly, $\|C(t_i)\| = \|A_1\| + 2\|A_2\|$. Aggregation-unbiasedness results from setting

$$g(s, t) = \begin{cases} \frac{1}{\|C(t_i)\|}, & s \in A_1(t) \\ \frac{2}{\|C(t_i)\|}, & s \in A_2(t) \\ 0, & \text{otherwise} \end{cases},$$

so that

$$\int_{D^{-1}(t_i)} g(s, t)\, ds = \frac{\|A_1(t_i)\| + 2\|A_2(t_i)\|}{\|C(t_i)\|} = 1.$$

The procedure is applied in the field by reflecting the point s across the boundary to obtain a point s', and using the fact that s is in $A_2(t_i)$ if and only if both s and s' are in $C(t_i)$. Thus, in determining $\bar{z}_D(s)$, tree i is counted once if s is in $C(t_i)$, and counted again if s' is in $C(t_i)$. For

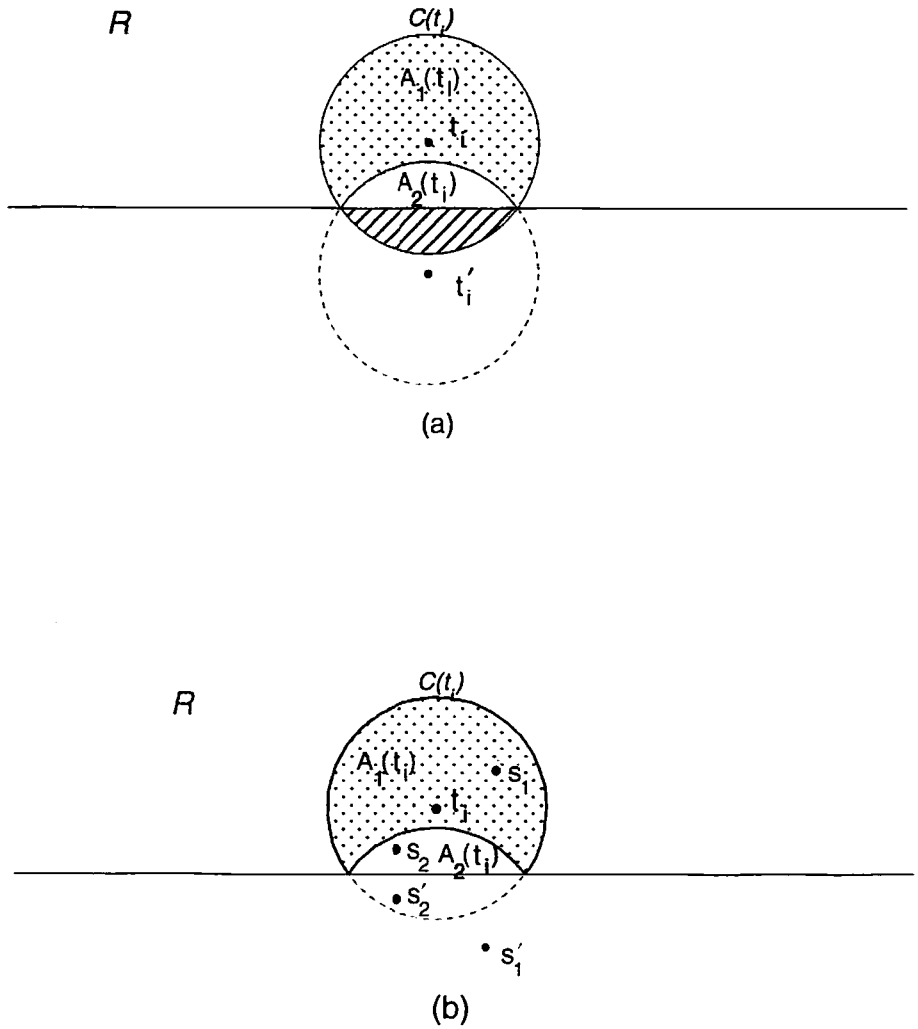

Figure 8. The mirage technique for determining $g(s, t)$. In (a), the portion of the tree circle at t_i that overlaps the boundary of R is reflected across the boundary. This partitions the tree circle within R into two regions $A_1(t_i)$ and $A_2(t_i)$. The reflected image of any point in $A_1(t_i)$, such as s_1 in (b), falls outside the tree circle at t_i, whereas the reflection of any point in $A_2(t_i)$, such as s_2, falls inside in the tree circle at t_i

example, the reflection s_1' of the point s_1 in Figure 8(b) falls in $C(t_i)$, so the tree at t_i is counted twice in forming $\bar{z}_D(s_1)$. Conversely, s_2 is outside of A_2, s_2' is outside of $C(t_i)$, and tree i is counted only once for $\bar{z}_D(s_2)$.

For the special case where the radius of a tree's circle is the same for every tree, then $D(s) = C(s) \cap R$. In this case, the mirage method yields a doubly stochastic aggregation function. This follows from the observation that $s \in A_2(t) \Leftrightarrow t \in A_2(s)$ so that

$$\int_{D(s)} g(s, t)\, dt = \frac{\|A_1(s)\| + 2\|A_2(s)\|}{\|C(s)\|} = 1.$$

4.3. Plot configurations or inclusion fields with invariant area

For the purpose of illustration, only the elementary case of a region R with a long straight boundary will be explored here. Real boundaries will be more complex, of course, but this case lets us investigate the consequences of several strategies, and suggests some guidelines that might be developed to handle more complex boundaries. First, consider the two-dimensional analog of the truncate strategy for linear resources: Use a fixed plot shape, and determine the support as the intersection of R with the plot. Further, let the plot be a circle with radius r centered on s, so that $D(s)$ consists of all points in R that are within a distance r of s. Then $D^{-1}(t) = D(t)$, that is, a point t is within a distance r of s if and only if s is within a distance r of t. In this case, $g(t)$ depends only on the distance between t and the boundary. If we pick a coordinate system so that the y-axis coincides with the edge of R, and R lies to the right of the y-axis (see Figure 9) then the distance from $t = (t_x, t_y)$ to the edge is just t_x, and

$$g(s, t) = \begin{cases} \dfrac{I_{D(s)}(t)}{\frac{\pi r^2}{2} + t_x\sqrt{r^2 - t_x^2} + r^2 \sin^{-1}\left(\frac{t_x}{r}\right)}, & t_x < r \\ \dfrac{I_{D(s)}(t)}{\pi r^2}, & t_x \geqslant r \end{cases}.$$

The solid line in Figure 10 shows $\|D^{-1}(t_x/r, 0)\|/\pi r^2$ as a function of t_x. This curve is the two-dimensional analog of the left-hand portion of the length of inclusion field curve for the truncate strategy (lower right plot in Figure 3), and has similar shape. (We will discuss the dashed curve shortly.)

A fixed-area plot configuration can be obtained by taking $D(s)$ as the locus of points in R closest to s with area πr^2. This is the converse of the enlarged-tree-circle strategy suggested by Barret (1964), which forces $D^{-1}(t)$ to have a constant area. The fixed-area plot is the intersection of R with a circle centered on s, with radius $\rho = \rho(s)$ satisfying

$$\rho = r, \quad \text{for} \quad s_x \geqslant r$$

$$\pi r^2 = \frac{\pi\rho^2}{2} + s_x\sqrt{\rho^2 - s_x^2} + \rho^2 \sin^{-1}\left(\frac{s_x}{\rho}\right), \quad \text{for} \quad s_x < r. \tag{6}$$

Unfortunately, this strategy only solves the problem of a variable plot area. The shape of the inclusion field for a point $t = (t_x, t_y)$ can be determined by using (6) to calculate $\rho(s_x)$ for a range of values for s_x, and then plotting the curve determined by

$$s_y = \sqrt{\rho(s_x)^2 - (s_x - t_x)^2}.$$

Both the shape and the area of the inclusion field vary as t moves away from the boundary. Figure 11 sketches plot configurations and inclusion fields for several distances from the boundary. The dotted regions are the inclusion fields at t, where t is the point at the large cross.

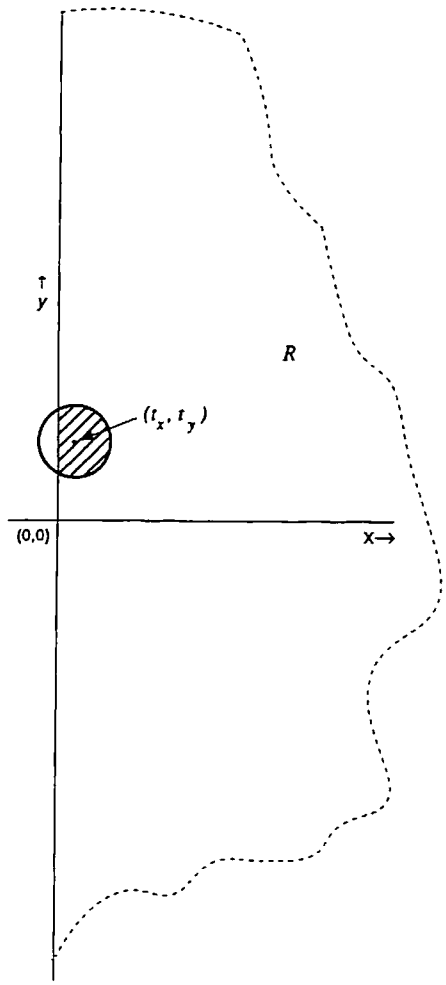

Figure 9. A portion of a region R with a long, straight boundary, showing the placement of a coordinate system. The hatched area is the inclusion field at the point $t = (t_x, t_y)$

Several examples of plot configurations (open figures bounded by solid line) are also sketched in the figure for various locations along the boundary of the inclusion field. The area of $D^{-1}(t)$ as a function of distance from the boundary is sketched as the dashed curve in Figure 10. This strategy is a 2-dimensional generalization of the slide strategy for linear resources. Note the similarity of the dashed curve in Figure 10, and the left-hand portion of the length-of-inclusion-field curve for the slide strategy (lower left in Figure 3).

There is a strategy such that both $D(s)$ and $D^{-1}(t)$ have constant area. If we examine the dashed curve in Figure 10, we note that regions close to the boundary have smaller than average inclusion fields, and regions from about $2r/3$ to $2r$ have larger than average inclusion fields. Heuristically, this occurs because the support strategy of using the closest points to s tends to 'replace' points in the circular region around s but outside of R with points that are inside R but further from the

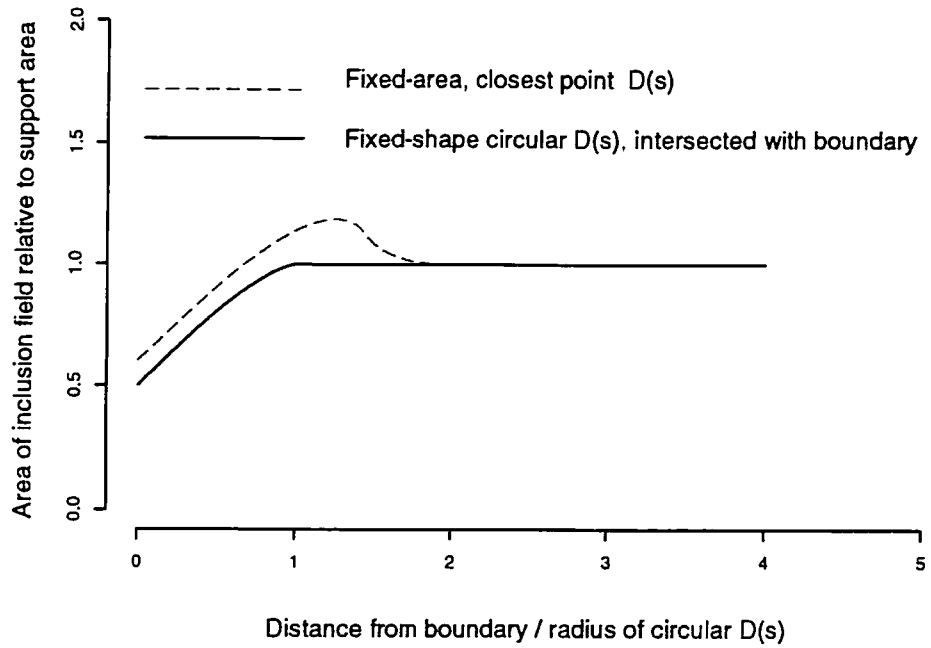

Figure 10. Variation in the area of the inclusion field as a function of distance from region boundary for two plot configuration strategies. Compare with the corresponding 1-dimensional cases in Figure 3

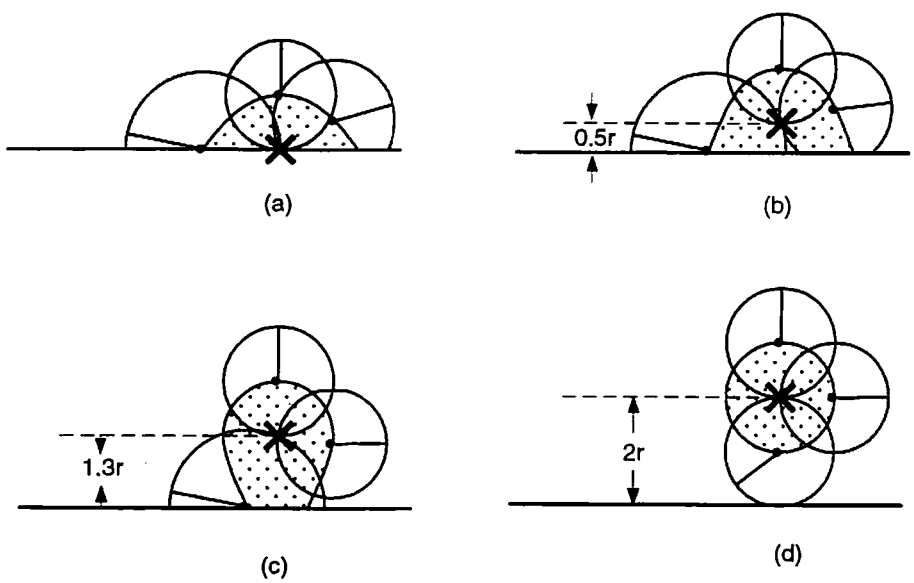

Figure 11. A plot configuration strategy that gives constant support area but variable inclusion field area. The dotted area is the inclusion field at the cross; the open figures are support regions for several points on the boundary of the inclusion field

boundary. Thus, the strategy tends to distort the support region in a direction away from the boundary. This suggests that a strategy that replaces points outside of R with points inside of R but the same distance away from the boundary might achieve both uniform plot area and uniform inclusion field area, and in fact, this is the case.

Figure 12 illustrates the process of clipping off the 'ears' of a circular support region that fall outside of R, and sticking them back onto the support inside of R. The piece outside of R is cut into thin strips. Each strip is then placed inside R so that it just touches the edge of the circle at the same distance from the boundary of R. The point t' outside of R is replaced by the point t inside of R. Clearly, the area of the $D(s)$ formed by the intersection of the circle with R plus the two 'ears' is constant, regardless of the location of s. Moreover, $D^{-1}(t) = D(t)$, again in the sense that the inclusion field at t has the same shape and area as the plot configuration at t. That shape is not invariant over all of R, however. In the interior of R, $D^{-1}(t) = D(t) = C(t)$, while for a point on

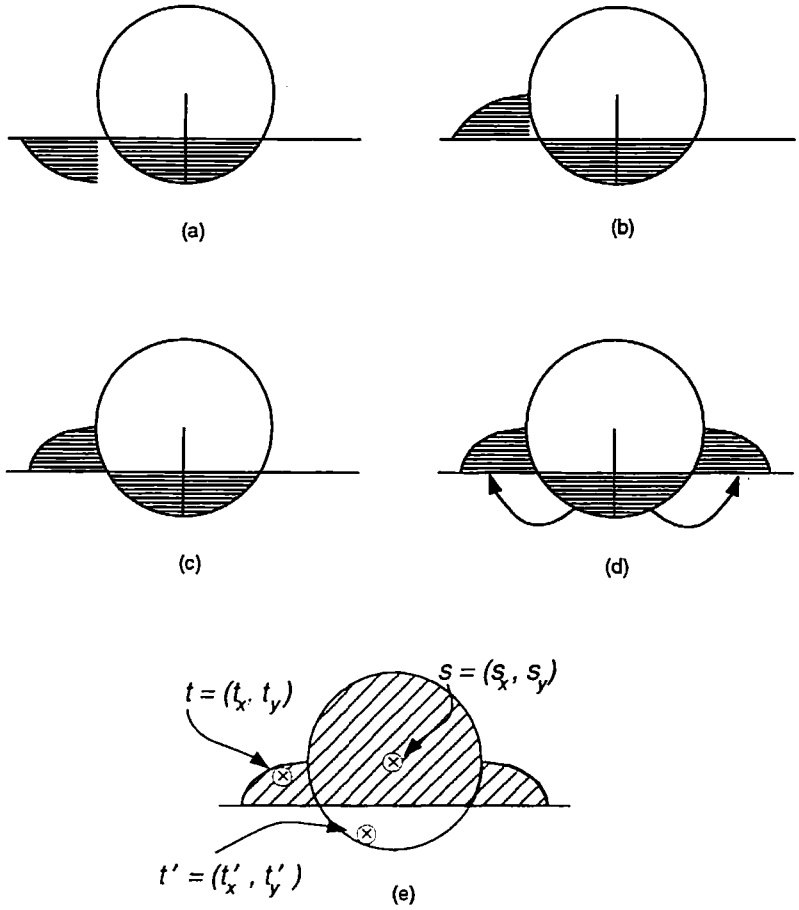

Figure 12. An example of the construction of plot configuration that leads to constant area $D(s)$ with the property that $D^{-1}(t) = D(t)$. The portions of a circular support region that overlap the boundary of R are clipped off and repositioned inside of R so as to preserve distance-to-boundary relationships

the boundary of R, $D(t)$ and $D^{-1}(t)$ are semi-ellipses. Thus, near the boundary, area is preserved, but the shape of the plot configuration changes.

To show that $D^{-1}(t) = D(t)$, we need to show that $t = D(s) \Leftrightarrow s \in D(t)$. Referring to Figure 12, $t \in D(s)$ if either $\|s - t\| \leqslant r$ or t is the image of a point t' given by

$$t' = (-t_x, t_y \pm \sqrt{r^2 - (t_x - s_x)^2})$$

with $\|s - t'\| \leqslant r$, where the sign of the radical is positive or negative as t_y is greater or less than s_y, respectively. If $\|s - t\| \leqslant r$, then $s \in D(t)$. If $\|s - t'\| \leqslant r$, then again $s \in D(t)$, since

$$r \geqslant \|s - t'\| = \sqrt{(t_x + s_x)^2 + (s_j - (t_y \pm \sqrt{r^2 - (s_x - t_x)^2})^2} = \|t - s'\|.$$

The inclusion fields resulting from applying this strategy are sketched in Figure 13 for several choices of distance from the boundary. The point $t = (t_x, t_y)$ is located at the large cross, and $D^{-1}(t)$ is the shaded region. We have sketched plot configurations (the unshaded figures) for several choices of points on the boundary of each $D^{-1}(t)$; in each case, note that the boundary of the plot configuration goes through t.

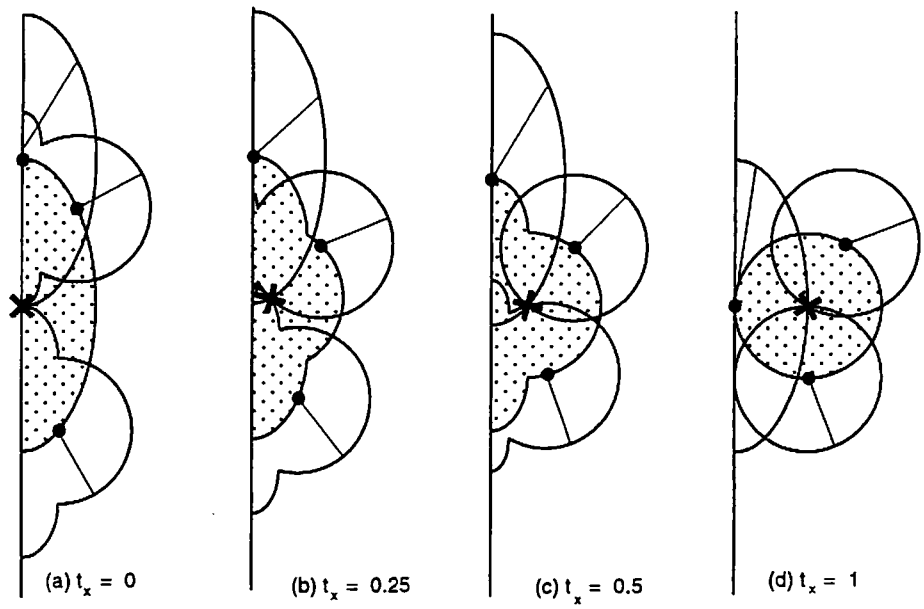

Figure 13. A plot configuration strategy that gives constant support area and constant inclusion field area. The dotted area is the inclusion field at the cross; the open figures are support regions for several points on the boundary of the inclusion field

5. SUMMARY AND DISCUSSION

A common objective in sampling a spatial environmental population is an inference about the mean or total of some attribute per unit area (Scott and Bechtold 1994, p. 55; Kaiser 1983). In sampling forest, we would be estimating forest properties, not tree properties. Our averages would be expressed in terms of a mean per unit area, not in a mean per tree. In sampling streams, we would be describing the population in terms of stream miles, not in terms of numbers of stream segments. Results would be expressed as an average per unit length, e.g., average number of pools per mile of stream, not as an average number of pools per stream segment. We noted that our perspective is that the population we are making inferences about (and hence sampling) is the continuous response surface, not the underlying unaggregated population. We believe that in most inferences about a spatial environmental population, this is the most logically consistent perspective.

If we were solely concerned with sampling a finite population, e.g., interested in tree properties rather than forest properties, then we need only compute inclusion probabilities for individual trees to apply a Horvitz–Thompson estimator to get unbiased estimates of totals. From this standpoint, there is no concern with aggregation bias, since we never aggregate to plot-level values. The entire analysis can be conducted at the level of tree-specific values. However, this perspective also commits one to a limited range of forest properties that depend on characteristics of individual trees. Descriptors that rely on characteristics of a neighborhood surrounding a tree are subject to an aggregation-effect bias. These descriptors include understory properties, vegetation classification, forest-type classification, and stocking. Similar comments apply to any spatial environmental population that can also be viewed as a collection of individuals. Attributes of the population that require a non-negligible support should be viewed and sampled as a continuous surface.

We have shown that some characteristics of finite populations of individuals can be preserved through aggregation by appropriate choice of a response design, and have given examples of application of two general techniques for doing so: one involving the translation of part of the plot and one involving reflection. Both can achieve doubly stochastic aggregation functions, so that mean values and linear relationships can be preserved.

We illustrated the interplay between plot configuration, inclusion field, and aggregation function, and showed how to obtain an aggregation function that gave an aggregation-unbiased response surface. For a given plot configuration, the aggregation function is not unique. For instance, the only difference between the reflect and truncate strategies for linear resources is the choice of aggregation functions. The distinction between the two is that the reflect aggregation function is doubly stochastic, and the truncate function is not.

There are other aspects to consider. Aggregation-unbiasedness is an adequate criterion only if one is interested solely in estimating means or totals. In many cases, interest extends beyond those simple statistics. For instance, if the response is a rate or density, such as stocking in forestry (trees/acre), or an indicator of ecological condition, such as an index of biological integrity (IBI), one may be more interested in estimating the proportion of R that meets some criterion, e.g., stocking is at least k trees per acre, or the IBI is greater than c. Some of the differences in aggregation-unbiased response designs will show up in the local smoothness of the response surface, which will in turn show up in the tails of the distribution function of the response surface. Local variability can be an artifice of the response design, influenced mostly by the size of the plot, but possibly also by the shape and orientation. In particular, the use of smaller plots near an edge may inflate variability in an area that is already more variable because it is near an interface. In these cases,

one should try to maintain equal area plots and inclusion fields. In those cases where the complexity of the boundary makes that difficult, one may opt to incur some (aggregation) bias in order to keep the plot configuration size constant.

Note that this is not a question of (sampling-design) bias. The response design determines the response surface, and proper sampling design and design-based analysis will ensure unbiased estimation of response surface characteristics. The question is the correspondence between the response surface and characteristics of the resource population.

The importance of maintaining equal area inclusion fields depends on the relationship between the surface $z(\cdot)$, the tangible, physical property that z represents, and the purpose or objective of the sampling. For instance, let $z(\cdot)$ be a measure of understory vegetation density, and suppose we wish to produce a map showing the spatial pattern of vegetation density. Before we can exhibit that map, we must settle on how to define the density at every point in R in both conceptual and operational terms. In particular, we must specify what 'vegetation density' means near a well-defined, sharp physical boundary. We must address such questions as the appropriate measurement protocols when some portion of a field plot is a paved parking lot. The correct approach depends to some extent on the intended use of the density surface. If we plan on using the density to construct an estimate of the total vegetation mass, then the density near the boundary should depend strictly on the area within R, and results on unbiased aggregation from this paper pertain. On the other hand, if the density is to be used to investigate the relationship between density and canopy cover, it may be appropriate for the density near the boundary to incorporate characteristics of areas outside of R. Such a density estimate will be 'biased' in the sense that the integral over R will not give the total biomass in R, but it may give a better picture of density as an attribute of the population on R.

We note that one of the advantages claimed for reflection-based methods is ease of field application (Gregoire 1982; Gregoire and Monkevich 1994). We recognize that a complex field protocol will be prone to errors, and do not discount the importance of a straightforward field procedure. However, in those cases where edge makes up a substantial portion of the population, and interest extends beyond totals, then plot a configuration that does not introduce an additional source of variation can be important. We have shown that the 'ears' plot configuration does satisfy the requirement of maintaining equal area plots and inclusion fields. Along a straight boundary, the ears plot can be approximated with straight lines fairly quickly, and even a crude approximation should serve to eliminate most of the edge effect. Along a complex boundary, the ears plot may be difficult to use; however, the principle of replacing area outside the boundary with an equal amount of area inside the boundary, and the same distance from the boundary, should serve to give a reasonable approximation.

We discussed some response designs where the plot configuration depends on local characteristics of the base population, e.g., line-intercept sampling and variable-radius tree-concentric sampling. In each of these cases, the plot configuration cannot be interpreted as a connected subset of R, but degenerates to a finite point set whose points index the location of population elements. These designs can lead to aggregation-unbiased response surfaces, but are not suitable for exploring relationships between multiple response variables. Because the plot configuration depends on population characteristics, no aggregation function can be doubly stochastic, so any relationship other than proportionality inferred from the aggregated observation will be influenced by the choice of plot configuration.

A final point to consider is that the notion of aggregation-unbiasedness, and the methods described herein to achieve it, depend very much on the boundary of R being known. Even if we

have an aggregation-unbiased response surface on R, that same surface will not in general be aggregation-unbiased for a region R' that is a proper subset of R.

ACKNOWLEDGEMENTS

The research described in this article has been funded by the U.S. Environmental Protection Agency. This document has been prepared at the EPA National Health and Environmental Effects Research Laboratory, Western Ecology Division, in Corvallis, Oregon, through Contract 68-C6-0005 to Dynamac International, Inc., Contract 68-C4-0019 to ManTech Environmental Research Services Corporation, and Cooperative Agreements CR 821738 and CR 824682 with Oregon State University. It has been subjected to the Agency's peer and administrative review and approved for publication. Mention of trade names or commercial products does not constitute endorsement for use.

REFERENCES

Barret, J. P. (1964). 'Correction for edge effect bias in point-sampling'. *Forest Science* 10(1), 52–55.

Campbell, K. (1993). 'Sampling the continuum: The spatial support of environmental data'. In *Proceedings of the Statistical Computing Section*. Alexandria, VA: American Statistical Association.

Chhikara, R. S. (1994). 'Efficient sampling designs for estimating ecological resources'. In *Statistics in Ecology and Environmental Monitoring*, eds. D. J. Fletcher and B. F. J. Manly. Dunedin, New Zealand: University of Otago Press.

Cochran, W. G. (1977). *Sampling Techniques*, 3rd Ed. New York: John Wiley & Sons.

Cordy, C. (1993). 'An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe'. *Probability and Statistics Letters* 18, 353–362.

Finney, D. J. (1948). 'The elimination of bias due to edge-effects in forest sampling'. *Forestry* 23, 31–47.

Fowler, G. W. and Arvantis, L. G. (1981). 'Aspects of statistical bias due to the forest edge: Horizontal forest sampling'. *Canadian Journal of Forestry Research* 11, 334–341.

Fuller, W. A. (1970). 'Sampling with random stratum boundaries'. *Journal of the Royal Statistical Society* B32, 209–226.

Gregoire, T. G. (1982). 'The unbiasedness of the Mirage correction procedure for boundary overlap'. *Forest Science* 28, 504–508.

Gregoire, T. G. and Monkevich, N. S. (1994). 'The reflection method of line intercept sampling to eliminate boundary bias'. *Environmental and Ecological Statistics* 1, 219–226.

Gregoire, T. G. and Scott, C. T. (1990). 'Sampling at the stand boundary: A comparison of the statistical performance among eight methods'. In *Research in Forest Inventory Monitoring, Growth, and Yield, Proceedings of the I.U.F.R.O. XIX World Congress, Montreal, Canada, 5–11 August 1990*, eds. H. E. Burkhart, G. M. Bonnor and J. J. Lowe. Publication FWS-3-90. Blacksburg, VA: School of Forestry and Wildlife Resources, Virginia Polytechnic Institute and State University, pp. 78–85.

Grieg-Smith, P. (1964). *Quantitative Plant Ecology*, 2nd ed. London: Butterworths.

Grosenbaugh, I. (1958). 'Point-sampling and line-sampling: Probability theory, geometric implications, synthesis'. Occasional Paper 160, Southeast Forest Experiment Station, Forest Service, U.S. Department of Agriculture.

Harrison, A. R. and Dunn, R. (1993). 'Problems of sampling the landscape'. In *Landscape Ecology and Geographic Information Systems*, eds. R. Haines-Young, D. R. Green and S. Cousins. London: Taylor & Francis.

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

Kaiser, L. (1983). 'Unbiased estimation in line-intercept sampling'. *Biometrics* 39, 965–976.

Kershaw, K. A. (1964). *Quantitative and Dynamic Ecology*. New York: American Elsevier Publishing Co.

Moisen, G. G., Stage, A. R. and Born, J. D. (1995). 'Sampling at the stand boundary'. *Forest Science Monographs* **31**, 62–81.

Openshaw, S. and Taylor, P. J. (1979). 'A million or so correlation coefficients: three experiments on the modifiable area unit problem'. In *Statistical Applications in the Spatial Sciences*, ed. N. Wrigley. London: Pion, pp. 127–144

Richards, J. I. and Youn, H. K. (1990). *Theory of Distributions: A Nontechnical Introduction*. Cambridge: Cambridge University Press.

Robinson, A. (1950). 'Ecological correlation and the behavior of individuals'. *American Sociological Review* **15**, 351–357.

Schmid-Haas, P. (1969). 'Stichen am waldrand'. *Mitt Schweiz Anst Forstl Versichswes* **45**(3), 234–303.

Schreuder, H. T., Gregoire, T. G. and Wood, G. B. (1993). *Sampling Methods for Multiresource Forest Inventories*. New York: Wiley.

Scott, C. T. and Bechtold, W. A. (1995). 'Techniques and computations for mapping plot clusters that straddle stand boundaries'. *Forest Science Monographs* **31**, 46–61.

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

Stevens Jr, D. L. and Kincaid, T. M. (1998). 'Variance estimation for subpopulations parameters from samples of spatial environmental populations'. In *1997 Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, pp. 86–95.

Thompson, S. K. (1992). *Sampling*. New York: John Wiley & Sons.

Yates, F. (1960). *Sampling Methods for Censuses and Surveys*. London: Griffin.

Yule, G. and Kendall, M. (1950). *An Introduction to the Theory of Statistics*. London: Griffin.

