Spatially Balanced Sampling of Natural Resources

Don L. STEVENS Jr. and Anthony R. OLSEN

The spatial distribution of a natural resource is an important consideration in designing an efficient survey or monitoring program for the resource. Generally, sample sites that are spatially balanced, that is, more or less evenly dispersed over the extent of the resource, are more efficient than simple random sampling. We review a unified strategy for selecting spatially balanced probability samples of natural resources. The technique is based on creating a function that maps two-dimensional space into one-dimensional space, thereby defining an ordered spatial address. We use a restricted randomization to randomly order the addresses, so that systematic sampling along the randomly ordered linear structure results in a spatially well-balanced random sample. Variable inclusion probability, proportional to an arbitrary positive ancillary variable, is easily accommodated. The basic technique selects points in a two-dimensional continuum, but is also applicable to sampling finite populations or one-dimensional continua embedded in two-dimensional space. An extension of the basic technique gives a way to order the sample points so that any set of consecutively numbered points is in itself a spatially well-balanced sample. This latter property is extremely useful in adjusting the sample for the frame imperfections common in environmental sampling.

KEY WORDS: Environmental sampling; Imperfect sampling frame; Monitoring; Non-response; Spatial sampling; Survey design; Systematic sampling.

1. INTRODUCTION

Environmental studies invariably involve populations distributed over space. Traditionally, such studies tended to focus on relatively small and well-delimited systems. However, some of the environmental issues that we face today, such as global warming, long-range transport of atmospheric pollutants, or habitat alteration, are not localized. Understanding and quantifying the extent of symptoms of widespread concern requires large-scale study efforts, which in turn needs environmental sampling techniques and methodology that are formulated to address regional, continental, and global environmental issues. Stehman and Overton (1994) gave an overview of some statistical issues associated with environmental sampling and monitoring, and Gilbert (1987) gave an extensive discussion of sampling methods for monitoring environmental pollution.

Several generic situations arise when sampling environmental resources spread over large spatial extents. Many resource populations may be represented as collections of points, lines, or areas; that is, as zero-, one-, or two-dimensional objects. For sampling purposes, the major distinctions occur between finite (pointlike, zero-dimensional), linear (one-dimensional), and areal (two-dimensional) populations. Finite populations are those with discrete, identifiably distinct units that occupy fixed locations within a bounded area. Examples are studies of the basal area of trees within a forest and the eutrophication status of lakes within the United States. Treating the lakes as points in a two-dimensional domain is appropriate if the purpose of the sample is to determine an attribute of each sampled lake and estimate characteristics of the lake population. The point associated with a lake could be any uniquely defined location in the lake, for example, the lake centroid. Linear resources are populations such as streams or rivers that are present only on a linear network within a bounded area. Attributes are defined at all points of a stream or river network, for example, water chemistry. Linear resources are often sampled as finite populations by breaking them into discrete units; say by taking fixed-length intervals beginning at the mouth, headwaters, or domain boundary. The division into units is often arbitrary, because the resource does not have well-defined natural units. Such a discretization ignores the essential nature of linear resources as one-dimensional continua embedded in two-dimensional space. Conceptualizing them as linear networks and sampling at points along the network retains the continuous nature of such populations. From this viewpoint, the population is an uncountably infinite collection of points. An areal resource is a continuous population that is present everywhere within a bounded area. Areal resources extend over large regions in a more or less continuous and connected fashion, although they may comprise disconnected polygons. As for a linear resource, an areal resource does not have distinct natural units and is viewed as an infinite point set; for example, all forested land in the United States, the Puget Sound estuary, and large wetlands such as salt marshes or the Everglades fall into this category.

A consideration that frequently arises in designing an environmental resource sample is that some population elements are perceived to be more important than others. For example, in sampling lakes, one might wish to select large lakes with a greater probability than small lakes because large lakes are less numerous than small or because they contribute disproportionately to total surface area, total water volume, or total recreational usage. For a second example, one might wish to increase the sampling rate for lakes in an arid region of the population domain to get enough samples to reliably describe lake characteristics for the region. These two examples illustrate two very different scenarios for which variable probability sampling might be required. In the first, the probability varies elementwise and depends on an attribute (in this case, size) of the element. In the second case, the probability varies on a geographical region basis, but may be the same for every element

Don. L. Stevens, Jr. is Senior Research Associate Professor, Department of Statistics, Oregon State University, Corvallis, OR 97331-4501 (E-mail: *stevens@stat.orst.edu*). Anthony R. Olsen is Mathematical Statistician, U.S. Environmental Protection Agency, NHEERL Western Ecology Division, Corvallis, OR 97333. We appreciate the willingness of the Indiana Department of Environmental Management to let us use their state monitoring program as an example: special thanks go to Stacey Sobat, who provided the biological data. Comments by the Associate Editor and referees helped improve the clarity of the manuscript. The research described in this article was funded by the U.S. Environmental Protection Agency. This document was 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. and though cooperative agreement CR82-9096-01 to Oregon State University; it was subjected to Agency review and approved for publication. The conclusions and opinions are solely those of the authors and are not necessarily the views of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

© 2004 American Statistical Association
Journal of the American Statistical Association
March 2004, Vol. 99, No. 465, Theory and Methods
DOI 10.1198/016214504000000250

within the region. Moreover, the two scenarios can occur in combination, so that we have a need to conform to both elementwise and regionwise variation in inclusion probability.

A practical complication frequently encountered in environmental sampling is the difficulty of obtaining an accurate sampling frame. In many instances, available sampling frames include a substantial portion of nontarget elements. For example, we could use the National Hydrography Dataset (NHD) available from the U.S. Geological Survey (USGS) as a sample frame for perennial streams (USGS 1999). Although attributes within NHD can be used to identify a subset of NHD that more closely matches the target population of perennial streams, the subset still includes many ephemeral or intermittent streams, or long-dry channels, especially in the more arid sections of the western United States. Another problem is that much of the resource we might like to sample is inaccessible, because of physical location, safety, or lack of access permission from the landowner. In some cases, it is possible to lose 50% or more of potential sample points because of lack of access. That is, significant nonresponse can be an issue in environmental surveys. Both problems result in fewer samples being collected than planned. If estimates of the percentage of the sampling frame that is nontarget or the percentage of inaccessible sites are available, then a common practice is to increase the planned sample size. Our experience is that such estimates are not available or at best are poorly known.

Some of the attributes of resource populations that influence sampling design are spatial pattern in the measured or observed response, uneven spatial distributions of the population, and difficulty in obtaining an adequate frame. Spatial pattern in the response arises because nearby units interact with one another and tend to be influenced by the same set of natural and anthropogenic factors. For example, neighboring trees in a forest interact by competing for energy and nutrients, and are influenced by the same set of physical and meteorological conditions, the same level of air- or water-borne pollutants, and the same set of landscape disturbances. The pattern in the response may show up either as a gradient or as a mosaic. A number of studies have concluded that regularly spaced design points (e.g., systematic designs) are optimal for a variety of reasonable spatial correlation functions (see, e.g., Cochran 1946; Quenouille 1949; Das 1950; Matérn 1960; Dalenius, Hájek, and Zubrzycki 1961; Bellhouse 1977; Iachan 1985).

The concept that some degree of spatial regularity should be used for sampling for environmental populations is well established. Accordingly, there are numerous paradigms for incorporating the spatial aspect of an environmental population into a sample. Area sampling partitions the domain of the population into polygons, which can be treated either as strata or as population units themselves. Systematic sampling using a regular grid is often applied (Bickford et al. 1963; Messer et al. 1986; Hazard and Law 1989), as are several variants that perturb the strict alignment (Olea 1984). Spatial stratification is also frequently used, with regular polygons, natural boundaries, political boundaries, or arbitrary tessellations as strata. Maximal stratification, that is, one or two points per stratum, has been viewed as the most efficient. To this end, Munholland and Borkowski (1996) used a Latin square with a single additional independent sample to achieve a spatially balanced sample. Breidt (1995) used a Markov process to generate a one-unit-per-stratum spatially distributed sample. Both of these techniques select cells in a regular grid. Another approach is to use space to order a list frame of the (finite) population and then use the order of the list to structure the sample, say, by defining strata as successive segments of the ordered list or by systematic random sampling. For example, Saalfeld (1991) drew on graph theory to define a tree that leads to a spatially articulated list frame, and the National Agricultural Statistics Service has used serpentine strips (Cotter and Nealon 1987) to order primary sample units within a state. A related idea that originated in geography is the general balanced ternary (GBT) spatial addressing scheme (Gibson and Lucas 1982). The concept behind a GBT address is related to the concept of space-filling curves, such as first constructed by Peano (1890), or the Hilbert curve (Simmons 1963). Stevens and Olsen (1999) used a similar concept, recursive partitioning, together with hierarchical randomization, to distribute sample points through space and time. Wolter and Harter (1990) used a construction similar to Peano's to construct a "Peano key" to maintain the spatial dispersion of a sample as the underlying population experiences births or deaths. Saalfeld (1991) also used the Peano key to maintain spatial dispersion of a sample.

The foregoing cited methods all do reasonably well at getting a spatially balanced sample under favorable circumstances, but have difficulties with some aspect of environmental populations. For example, spatial stratification can be applied to finite, linear, and areal populations. However, defining strata for finite or linear populations with variable probability and substantial variation in spatial density can be difficult: maximal efficiency is obtained for one or two samples per stratum. To do so, we need some means to split the population into spatially contiguous strata. We could simply adopt equal-sized strata that tile the population domain, which usually results in a variable number of samples per stratum and noninteger expected samples sizes. (We illustrate this approach in Sec. 3.) Alternatively, we could try to develop unequal-area strata with the same, or nearly so, expected number of samples (total inclusion probability) in each stratum. For a small finite population, the strata could be developed by inspection. For a large population, say the 21,000 lakes in the northeastern United States, an automated stratification procedure is necessary. Developing such a procedure is a nontrivial task. Small sample sizes per strata are good for efficiency, but cause the greatest loss of efficiency in the presence of nonresponse. Suppose we have a two-sample-per-stratum design with a moderate rate of nonresponse, say 25%. We are almost certain to lose both samples from some strata. If we replace both sample points, we double the inclusion probability for those strata. Moreover, it is quite possible that the replacement points will also be nonresponse points, so we end up tripling or quadrupling the inclusion probability. The result is an unintentional imbalance in inclusion probability; those strata with high nonresponse get less weight in the analysis. Deviation from the intended inclusion probability that introduces more variation in the weight of the sample points results in loss of efficiency. Similar arguments can be made for the other methods for achieving spatial balance.

Sampling the gamut of natural resources requires a technique that can select a spatially balanced sample of finite,

linear, and areal resources with patterned and possibly periodic responses, using arbitrarily variable inclusion probability with imperfect frame information, in the presence of substantial nonresponse. In the design discussed here, we generalize the concept of spatial stratification to create a very powerful and flexible technique for selecting a spatially well-distributed probability sample that works under all of the preceding circumstances. The technique is based on creating a function that maps two-dimensional space into one-dimensional space, thereby defining an ordered spatial address. We use a restricted randomization called hierarchical randomization (HR) (Stevens and Olsen 2000) to randomly order the address and then apply a transformation that induces an equiprobable linear structure. Systematic sampling along the randomly ordered linear structure is analogous to sampling a random tessellation of two-dimensional space and results in a spatially well-balanced random sample. We call the resulting design a generalized random-tessellation stratified (GRTS) design. We develop the design in a general setting that applies to finite, linear, and areal resources, and that accommodates arbitrary inclusion probability functions. A particularly favorable feature is that we can dynamically add points to the sample as we discover nontarget or inaccessible points, at the same time maintaining a spatially well-balanced sample. Features of the design are demonstrated with a simulation study and are illustrated with an application to rivers and streams in Indiana.

2. GENERALIZED RANDOM–TESSELLATION STRATIFIED DESIGN

Before presenting the theoretical development of the GRTS design, we give a heuristic overview of the process. Assume that the sample frame consists of N points located within a geographic region. Assign each point a unit length and place each point in some order (say randomly) on a line. The line has length N units. Select a systematic sample of size n from the line by dividing the line into N/n length intervals, randomly select a starting point between $(0, N/n]$, say k, and then take every $(k + iN/n)$th point for $i = 1, \ldots, n-1$. If the point occurs within one of the units, then that unit is selected (Brewer and Hanif 1983). For a linear resource, use the actual length of the units to construct the line. For an areal resource, randomly place a systematic grid over the region, randomly select a point in each grid cell, and then proceed as in the point case. A GRTS sample results when a process termed hierarchical randomization is used to place the points on the line. First, randomly place a 2×2 square grid over the region and place the cells in random order in a line. For each cell, repeat the same process, randomly ordering the subcells within each original cell. This second step results in 16 cells in a line. Continue the process until at most one population point occurs in a cell. Use the random order of the cells to place the points on the line. This hierarchical randomization process maps two-dimensional space into one-dimensional space while preserving spatial relationships as much as possible. The combination of hierarchical randomization to create the line and systematic sampling with a random start results in a spatially balanced equal probability sample. Unequal probability sampling is implemented by giving each point a length proportional to its inclusion probability.

Stevens (1997) derived inclusion and joint inclusion functions for several grid-based designs that were precursors to GRTS designs and share some of their properties. The designs are all generalizations of the random-tessellation stratified (RTS) design (Dalenius et al. 1961; Olea 1984; Overton and Stehman 1993). The RTS design selects random points in space via a two-step process. First, a regular tessellation coherent with a regular grid is randomly located over the domain to be sampled, and, second, a random point is selected within each random tessellation cell. The RTS design is a variation on a systematic design that avoids the alignment problems that can occur with a completely regular systematic design. Like a systematic design, a RTS design does not allow variable probability spatial sampling. Stevens (1997) introduced the multiple-density, nested, random-tessellation stratified (MD-NRTS) design to provide for variable spatial sampling intensity. The geometric concept underlying the MD-NRTS was the notion of coherent intensification of a grid, that is, adding points to a regular grid in such a way as to result in a denser regular grid with similarly shaped but smaller tessellation cells. We have since extended the same notion by generalizing to a process that creates a potentially infinite series of nested, coherent grids. In the limit, the process results in a function that maps two-dimensional space into one-dimensional space.

We can cover finite, linear, and areal populations with the same development if we work in the context of general measure and integration theory. Let R be the domain of the population we wish to sample, that is, the set of points occupied by elements of the population. We require that R be a bounded subset of $\mathbb{R}^2$. Thus, R can be enclosed in a bounded square, so that by scaling and translation, we can define a 1–1 map from R into $(0, 1/2] \times (0, 1/2]$, the lower left quadrant of the unit square. (We map to the lower left quadrant so that we can add a random offset to the image of R and stay within the unit square. The random offset guarantees that the points from any pair can end up in different quadrants.) Clearly, every point in the image is associated with a unique point in R and vice versa, so henceforth, we identify R with its image in the unit square.

2.1 Random Quadrant-Recursive Maps

The heart of the GRTS sample selection method is a function f that maps the unit square $\Im^2 = (0, 1] \times (0, 1]$ onto the unit interval $\Im = (0, 1]$. To be useful in achieving a spatially balanced sample, f must preserve some proximity relationships, so we need to impose some restrictions on the class of functions to be considered. Mark (1990), in studying discrete two- to one-dimensional maps, defined a property called quadrant recursive, which required that subquadrants be mapped onto sets of adjacent points. To define the continuous analog, let

$$Q^n_{jk} = \left(\frac{j}{2^n}, \frac{j+1}{2^n}\right] \times \left(\frac{k}{2^n}, \frac{k+1}{2^n}\right], \qquad j, k = 0, 1, \ldots, 2^n - 1,$$

and let

$$J^n_m = \left(\frac{m}{4^n}, \frac{m+1}{4^n}\right], \qquad m = 0, 1, \ldots, 4^n - 1.$$

A function $f : \Im^2 \to \Im$ is *quadrant recursive* if, for all $n \geq 0$, there is some $m \in \{0, 1, \ldots, 4^n - 1\}$ such that $f(Q^n_{jk}) = J^n_m$.

We can view a quadrant-recursive function as being defined by the limit of successive intensifications of a grid covering the unit square, where a grid cell is divided into four subcells, each of which is subsequently divided into four subsubcells, and so on. If we carried this recursion to the limit and paired grid points with an address based on the order in which the divisions were carried out, where each digit of the address represented a step in the subdivision, then we would obtain a quadrant-recursive function. For example, suppose we begin with a point at (1, 1), and replace it with four points $p_3 = (1, 1)$, $p_2 = (1, 1/2)$, $p_1 = (1/2, 1)$, and $p_0 = (1/2, 1/2)$. The next step of the recursion replaces each of the first four points $p_0, \ldots, p_3$ with $p_i - \{(0,0), (0,1), (1,0), (1,1)\}/2^2$. Thus the point $p_2 = (1, 1/2)$ is replaced with the four points $p_{23} = (1, 1/2)$, $p_{22} = (1, 1/4)$, $p_{21} = (3/4, 1/2)$, and $p_{20} = (3/4, 1/4)$. The nth step replaces each of the 4^n points $p_{i_1 i_2 \cdots i_n}$ with $p_{i_1 i_2 \cdots i_n} - \{(0,0), (0,1), (1,0), (1,1)\}/2^{n+1}$.

A spatially referenced address can be constructed following the pattern of the partitioning, with each new partition adding a digit position to the address. Thus, in the preceding example, the four points in the first group are assigned the addresses 3, 2, 1, and 0, where 3 is the original point at (1, 1). The successor points to point 2 get the addresses 23, 22, 21, and 20, and so forth. The addresses induce a linear ordering of the subquadrants. Moreover, if we carry the process to the limit, and treat the resulting address as digits in a base-4 fraction [e.g. treat $22103\cdots$ as the base-4 number $(.22103\cdots)_4$], then the correspondence between grid point and address is a quadrant-recursive function.

Recursive partitioning generates a nested hierarchy of grid cells. The derived addressing has the property that all successor cells of a cell have consecutive addresses. Thus, a path from cell to cell, following the recursive partitioning address order, will connect all successor cells of cell 0 before reaching any successor of cell 2 (Fig. 1).

A 1–1 continuous mapping of $\Re^2$ onto $\Re$ is not possible, so quadrant-recursive functions are not continuous. However, they do have the property that all points in a quadrant are mapped onto an interval, all points in any one of the four subquadrants of a quadrant are mapped onto an interval, and so on ad infinitum. This property tends to preserve proximity relationships, that is, if s is "close to" t, then $f(s)$ should "tend to be close to" $f(t)$. In Appendix A, we make this statement more precise by showing that if the origin is located at random and s is chosen at random from $\Re^2$, then $\lim_{|\delta| \to 0} E[|f(s) - f(s+\delta)|] = 0$. Intuitively, two elements that are close together will tend to fall in the same randomly located cell of a size that decreases as the distance between points decreases. Because the two elements are covered by the same cell, their addresses match to the level of that cell and thus, in expectation, their addresses will be close.

A fundamental 1–1 quadrant-recursive map is defined by digit interweaving. Let $s = (x, y)$ be a point in $\Re^2$. Each of the coordinates has an expansion as a binary fraction of the form $x = .x_1 x_2 x_3 \cdots$, $y = .y_1 y_2 y_3 \cdots$, where each x_i and y_i is either 0 or 1. Define $f_0(s)$ by alternating successive digits of x and y, that is, $f_0(s) = .x_1 y_1 x_2 y_2 \cdots$. Clearly, f_0 would be 1–1 except for different expansions of the same number. For example, .1 and $.011111\cdots$, where the 1s continue indefinitely, are two

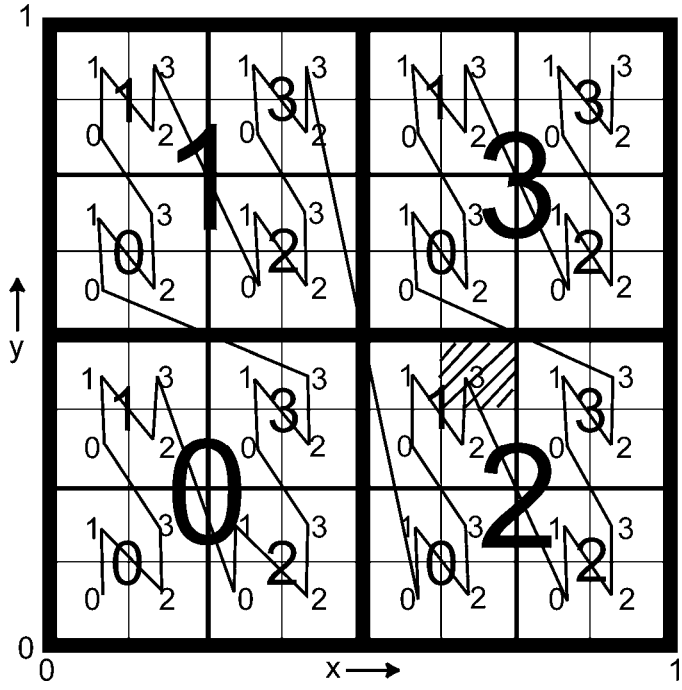

Figure 1. First Four Levels of a Quadrant-Recursive Partitioning of the Unit Square. The address associated with the cross-hatched cell is 213.

representations of the number 1/2. If we always use the binary representation with an infinite number of 1s, then f_0 is 1–1. Moreover, every point in $\Re$ is the image of a point in $\Re^2$, which is obtained by "digit splitting." That is, if $t = .t_1 t_2 t_3 \cdots$ is in $\Re$, then $s = f_0^{-1}(t) = (.t_1 t_3 t_5 \cdots, .t_2 t_4 t_6 \cdots)$ is the preimage of t. Both f_0 and f_0^{-1} are 1–1 if we always use the representation with an infinite number of 1s (Hausdorff 1957, p. 45). To show that f_0 is quadrant recursive, note that for $s \in Q^n_{jk}$, the first 4^n digits of $f_0(s)$ are fixed, so $f_0(s) \in J^n_m$, where m is defined by the first 4^n digits. Conversely, the preimages of every $t \in J^n_m$ have the same first 2^n digits, and so must be in the same Q^n_{jk}.

Figure 1 shows the first four levels of the recursive partitioning of the unit square. The address of the cross-hatched subquadrant is, as a base-4 fraction, $(.213)_4$ and the associated grid point is at (3/4, 1/2), the upper right corner of the subquadrant. Following the convention of having an infinite number of 1s in the expansion, we have $(3/4, 1/2) = (.11, .1)_2 = (.1011111\cdots, .0111111\cdots)_2$. Digit interweaving gives the image $(.10011111\cdots)_2 = (.2133333\cdots)_4$, of which the first three digits are the subquadrant address. If we carried the recursive partitioning to the limit, every point in the subquadrant would be assigned an address beginning with $(.213)_4$.

The class of all quadrant-recursive functions can be generated from the function f_0, which is defined by digit interweaving, by permuting the order in which subquadrants Q^n_{jk} are paired with the intervals J^n_m. For example, for $n = 1$, $f_0(Q^1_{jk}) = J^1_{2j+k}$. We obtain a different quadrant-recursive function by permuting the subscripts $\{0, 1, 2, 3\}$ of the image intervals. Thus, under the permutation $\tau = \{2, 1, 3, 0\}$, we get a function such that $f_\tau(Q^1_{jk}) = J^1_{\tau(2j+k)}$, so that $f_\tau(Q^1_{00}) = J^1_2$, $f_\tau(Q^1_{01}) = J^1_1$, $f_\tau(Q^1_{10}) = J^1_3$, and $f_\tau(Q^1_{11}) = J^1_0$. To see that the class of all quadrant-recursive functions is generated by such permutations, express each number in $\Re$ as a base-4 number, that

is, as $t = .t_1 t_2 t_3 \cdots$, where each digit t_i is either a 0, 1, 2, or 3. A function $h_p : \Im \to \Im$ is a *hierarchical permutation* if $h_p(t) = .p_1(t_1) p_{t_1 2}(t_2) p_{t_1 t_2 3}(t_3) \cdots$, where $p_{t_1 t_2 \cdots t_{n-1} n}(\cdot)$ is a permutation of $\{0, 1, 2, 3\}$ for each unique combination of digits $t_1, t_2, \ldots, t_{n-1}$. Again, we ensure that h_p is 1–1 by always using the expansion with an infinite number of nonzero digits. Any quadrant-recursive function can be expressed as the composition of f_0 with some hierarchical permutation h_p, because the associations $f(Q^n_{jk}) = J^n_m$ determine the series of permutations and the permutations define the associations.

If the permutations that define $h_p(\cdot)$ are chosen at random and independently from the set of all possible permutations, we call $h_p(\cdot)$ a *hierarchical randomization function* and call the process of applying $h_p(\cdot)$ *hierarchical randomization.*

2.2 Sample Selection With Probability Proportional to Arbitrary Intensity Function

We assume that the design specifications define a desired sample intensity function $\pi(s)$, that is, the number of samples per unit measure of the population. For example, if the population were a stream network, $\pi(s)$ might specify the number of samples per kilometer of stream at s. For a discrete population, $\pi(s)$ has the usual finite-population-sampling interpretation as the target inclusion probability of the population unit located at s. We call $\pi(s)$ an intensity function, because we have not yet introduced a probability measure. In Appendix B, we develop the details of a sample selection method that yields an inclusion-probability function equal to $\pi(s)$. The concept behind the method is the composition of a hierarchical randomization function with a function that assigns to every interval in $f(R)$ a weight equal to the total of the intensity function of its preimage in R. In effect, we stretch the image interval via a distribution function F, so that its total length is equal to the sample size M. We pick M points by taking a systematic sample with a unit separation along the stretched image and we map these points back into the domain R via the inverse function to get the sample of the population. We show in Appendix B that this procedure does indeed give a sample with an inclusion-probability function equal to the intensity function $\pi(s)$.

The technique of randomly mapping two-dimensional space to a line segment, systematically sampling from the range of the distribution function, and then mapping back to the population elements always produces a sample with the desired first-order inclusion-probability function as long as f is 1–1 and measurable. We required that f be quadrant recursive and claim that this is sufficient to give a spatially balanced sample. This claim follows from the fact that the map $f^{-1} \circ (F/M) \circ f$ transforms the unequal intensity surface defined by π into an equiprobable surface. The quadrant-recursive property of f guarantees that the sample is evenly spread over the equiprobable surface (in the sense that each subquadrant receives its expected number of samples) to the resolution determined by the sample size M.

2.3 Reverse Hierarchical Ordering

The sample points selected by mapping the systematic points along $(0, M]$ back to the population domain will be ordered in a way that follows the quadrant-recursiveness of f, tempered by an allowance for unequal probability selection. Thus, the first quarter of the points all will come from the same "quadrant" of the equiprobable domain and all will be approximately neighbors in the original population domain. It follows that four points, one picked from each quarter of the sample points ordered by the systematic selection, will be a spatially balanced sample. Because the random permutations that define the hierarchical randomization are selected independently of one another, it makes no difference, from a distributional standpoint, whether we pick the points systematically from each quarter, or make random selections from each quarter. Therefore, we lose no randomness by picking the points that occupy positions that correspond to being at the beginning, one-quarter, one-half, and three-quarters of the way through the ordered list of sample points.

Within each quarter of the list, the points are again quadrant-recursively ordered, so points picked at the beginning, one-quarter, one-half, and three-quarters of the way through each quarter of the list will be spread out over the corresponding quadrant, and so on down through the sequence of subquadrants. We can utilize these properties by reordering the systematically selected list so that, at *any* point in the reordered list, the samples up to that point are well spread out over the population domain.

The order is most conveniently expressed in terms of a base-4 fraction, where the fraction expresses the relative position in the systematically ordered list. Thus, the first four points correspond to the fractions $(.0, .1, .2, .3)_4 = (0, 1/4, 1/2, 3/4)_{10}$. Stepping down a subquadrant level corresponds to adding a digit position to the base-4 fraction, which we fill in such a way as to spread the sequence of points over the population domain. The pattern for the first 16 points is shown in Table 1. Note that the order corresponds to the ranking obtained by reversing the sequence of base-4 digits and treating the reversed sequence as a base-4 fraction.

We can continue this same pattern of adding digit positions through as many positions as necessary to order the entire sample. The resulting order is called *reverse hierarchical order.* It remains to show that reverse hierarchical order does indeed give a spatially well-balanced sample for any $m \leq M$. Clearly this is the case for $m = 4^k$, because the reduced sample can be viewed as a sample selected from a complete GRTS design. Stevens (1997) derived an analytic expression for the pairwise inclusion density for some special intermediate cases. Here, we investigate the spatial balance properties using simulation.

Table 1. Generation of Reverse Hierarchical Order

Order	Base 4	Reverse base 4	Order	Base 4	Reverse base 4	Order	Base 4	Reverse base 4	Order	Base 4	Reverse base 4
1	.00	.00	5	.01	.10	9	.02	.20	13	.03	.30
2	.10	.01	6	.11	.11	10	.12	.21	14	.13	.31
3	.20	.02	7	.21	.12	11	.22	.22	15	.23	.32
4	.30	.03	8	.31	.13	12	.32	.23	16	.33	.33

3. SPATIAL PROPERTIES OF GRTS SAMPLE POINTS

In this section, we investigate the spatial balance, or regularity of the sample points produced by a GRTS design. We noted in the Introduction that, generally, the efficiency of an environmental sample increases as spatial regularity increases. A design with regularity comparable to a maximally stratified sample should have good efficiency. Choosing a suitable statistic to describe regularity is nontrivial because the population domain itself is likely to have some inherent nonregularity (e.g., variation in spatial density for a finite or linear population) and because of the need to account for variable inclusion probability. The measure of regularity needs to describe regularity over the inclusion-probability-weighted irregular population domain. Various statistics to assess the regularity of a point process have been proposed in the study of stochastic point processes. One class of descriptive statistics is based on counts of event points within cells of a regular grid that covers the process domain. The mean count is a measure of the process intensity, and the variance of the counts is a measure of the regularity. The usual point process approach is to invoke ergodicity and take expectation over a single realization. In the present case, the expectation should and can be taken over replicate sample selections.

We illustrate this approach using an artificial finite population that consists of 1,000 points in the unit square with a spatial distribution constructed to have high spatial variability, that is, to have voids and regions with densely packed points. Variable probability was introduced by randomly assigning 750 units a relative weight of 1, 200 units a weight of 2, and 50 units a weight of 4. The inclusion probability was obtained by scaling the weights to sum to the sample size. We divided the unit square into 100 square cells with sides .1 units. Fifty-one of the cells were empty. The expected sample sizes (the sum of the inclusion probability for each cell) for the 49 nonempty cells ranged from .037 to 4.111.

We compared the regularity of three sampling designs: independent random sampling (IRS), spatially stratified sampling (SSS), and GRTS sampling. For each sampling scheme, we selected 1,000 replicates of a sample of 50 points and counted the number of sample points that fell into each of the 49 nonempty cells defined in the previous paragraph. For the IRS sample, we used the S-PLUS (Insightful Corporation 2002) "sample" function with "prob" set to the element inclusion probability.

As we noted in the Introduction, there is no general algorithm for partitioning an arbitrary finite spatial population with variable inclusion probability into spatial strata with equal expected sample sizes. For this exercise, we chose to use equal-area strata with variable expected sample sizes. For simplicity, we chose square strata. We picked a side length and origin so that (1) the strata were not coherent with the $.1 \times .1$ cells used for regularity assessment and (2) about 50 stratum cells had at least one population point. The strata we used were offset from the origin by (.03, .03), with a side length of .095. Exactly 50 stratification cells were nonempty, with expected sample sizes ranging from .037 to 4.111. Figure 2 shows the population with the stratification cells overlaid.

We selected the stratified sample in two stages. The fractional parts of the expected sample sizes will always sum to an integer, in this case, 21. The first step in the sample selection was to select which 21 of the 50 strata would receive an "extra" sample point. For this step, we again used the S-PLUS "sample" function, this time with "prob" set to the fractional part of the expected sample size. The second step in the sample selection was

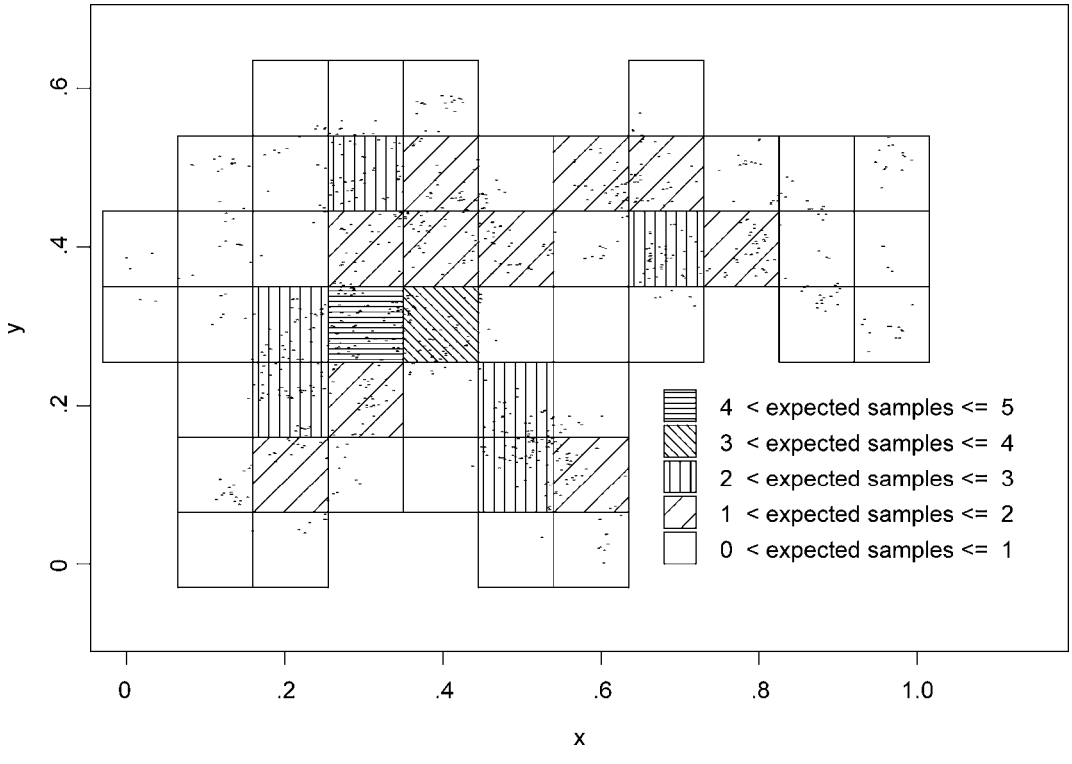

Figure 2. Finite Population Used in Spatial Balance Investigation, Overlaid With Grid Cells Used for Stratification. Cell cross-hatching indicates the expected sample size in each cell.

to pick samples in each of the 50 strata, that had a sample size equal to the integer part of the expected sample size, plus 1 if the stratum was selected in stage 1. Again, the sample was selected with the "sample" function, this time with "prob" set to the element inclusion probability. This two-stage procedure always selects exactly 50 samples with the desired inclusion probability.

In Figure 3, we plot the variance of the achieved sample size in each of the evaluation cells versus the expected sample size, with lowess fitted lines. Of the three designs, the IRS has the largest variance and the GRTS has the smallest; the SSS design is approximately midway between. Stratification with one sample per cell would likely have about the same variance as the GRTS.

Another common way to characterize a one-dimensional point process is via the interevent distance, for example, the mean interevent time for a time series measures the intensity of the process and the variance measures the regularity. An analogous concept in two-dimensions is that of Voronoi polygons. For a set of event points $\{s_1, s_2, \ldots, s_n\}$ in a two-dimensional domain, the Voronoi polygon Ψ_i for the ith point is the collection of domain points that are closer to s_i than to any other s_j in the set. Note that in the case of a finite population, the Voronoi "polygons" are collections of population points, and for a linear population, they are collections of line segments.

We propose using a statistic based on Voronoi polygons to describe the regularity of a spatial sample. For the sample S consisting of the points $\{s_1, s_2, \ldots, s_n\}$, let $v_i = \int_{\Psi_i} \pi(s)\, d\phi(s)$, so that v_i is the total inclusion probability of the Voronoi polygon for the ith sample point, and set $\zeta = \mathrm{Var}(\{v_i\})$. For a finite population with variable inclusion probability, v_i is the sum of the inclusion probability of all population units closer to the sample point s_i than to any other sample point. Because $\sum_i |\Psi_i| = |R|$ and $\int_R \pi(s)\, d\phi(s) = n$, $E[v_i] = 1$. We note that for an equiprobable sample of a two-dimensional continua, ζ is equal to the variance of the area of the Voronoi polygons for the points of S multiplied by the square of the inclusion probability.

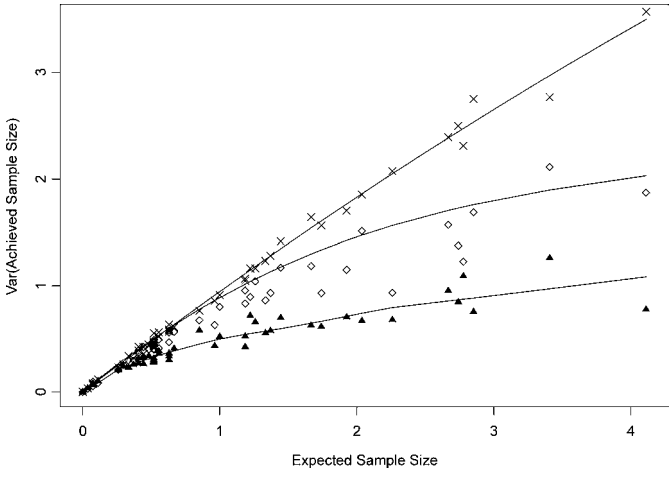

Figure 3. Comparison of the Regularity of GRTS, SSS, and IRS Designs. Results are based on the mean of 1,000 samples of size 50. The achieved sample size is the number of samples that fell into .1 × .1 square cells that tiled the population domain. Lines were fitted with lowess. (▲ generalized random tessellation stratified sampling; × independent random sampling; ◇ spatially stratified sampling.)

For the kinds of applications that we have in mind, the spatial context of the population is an intrinsic aspect of the sample selection. For a finite population, the spatial context simply comprises the locations of the population units, for a linear population, the spatial context is the network, and for an areal resource, the spatial context is described by the boundary of the resource domain, which may be a series of disconnected polygons. The effect of the interplay of sampling design and spatial context on properties of the sample cannot be ignored. For small to moderate sample sizes, or for highly irregular domains, the spatial context can have a substantial impact on the distribution of ζ. Because of the spatial dependence, the derivation of a closed form for the distribution of ζ does not seem feasible, even for simple sampling designs such as IRS. However, for most cases, it should be relatively easy to simulate the distribution of ζ under IRS to obtain a standard for comparison. The regularity of a proposed design can then be quantified as the ratio $\zeta(\text{proposed design})/\zeta(\text{IRS})$, where ratios less than 1 indicate more regularity than an IRS design.

We evaluated spatial balance using the ζ ratio under three scenarios: (1) a variable probability sample from a finite population, (2) an equiprobable point sample from an areal population defined on the unit square, and (3) an equiprobable point sample from the same extensive population, but with randomly located square holes to model nonresponse and imperfect frame information.

For the finite population study, we drew 1,000 samples of size 50 from the previously described finite population for both the GRTS and the IRS designs. To illustrate the ability of the GRTS design to maintain spatial regularity as the sample size is augmented, we ordered the GRTS points using reverse hierarchical ordering. We then calculated the ζ ratio, beginning with a size of 10 and adding one point at a time following the reverse hierarchical order. We also drew 1,000 samples of size 50 using the previously discussed spatial stratification. Because there is no sensible way to add the stratified sample points one at a time, we can compute the ζ ratio only for the complete sample of 50 points. Figure 4 is a plot of the ζ ratio for GRTS and IRS versus sample size. The single ζ ratio for SSS(50) is also shown. For the GRTS design, the ζ ratio has a maximum value of .587 with 10 samples and gradually tapers off to .420 with 50 samples. Although it would be difficult to prove, we suspect that the gradual taper is due to lessening edge effect with increasing sample size, that is, fewer of the Voronoi polygons cross the void regions in the population domain. We note that the valleys in the ζ ratio occur at multiples of 4, with the most extreme dips occurring at powers of 4. This is a consequence of quadrant-recursive partitioning: maximum regularity occurs with one point from each of the four quadrants. We also note that the SSS(50) value of the ζ ratio is .550 compared to the corresponding value of .420 for the GRTS design. Inasmuch as the GRTS is analogous to a one-sample-per-stratum SSS, we would expect the GRTS to be as efficient as a maximally efficient SSS.

For our extensive population study, we selected 1,000 samples of size $M = 256$ from the unit square using the GRTS

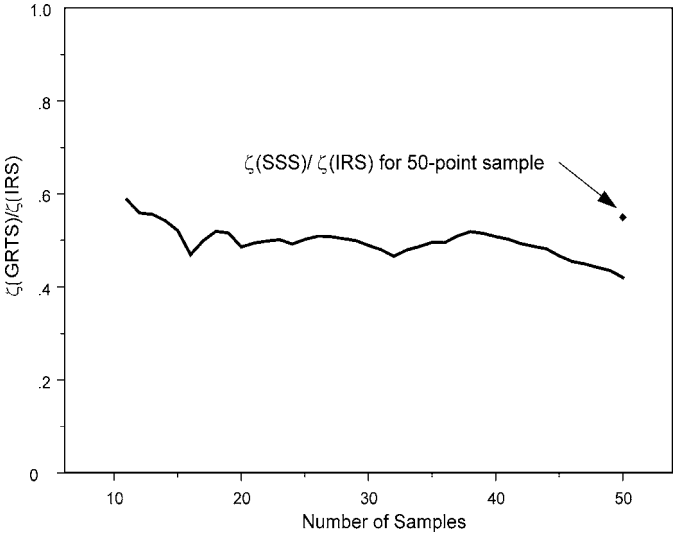

Figure 4. The ζ Ratio as a Function of Sample Size Based on 1,000 Replicate Samples From an Artificial Finite Population. Sample points were added one point at a time up to the maximum sample size of 50, following reverse hierarchical order for the GRTS sample. The ζ ratio for a spatially stratified sample is also indicated on the plot.

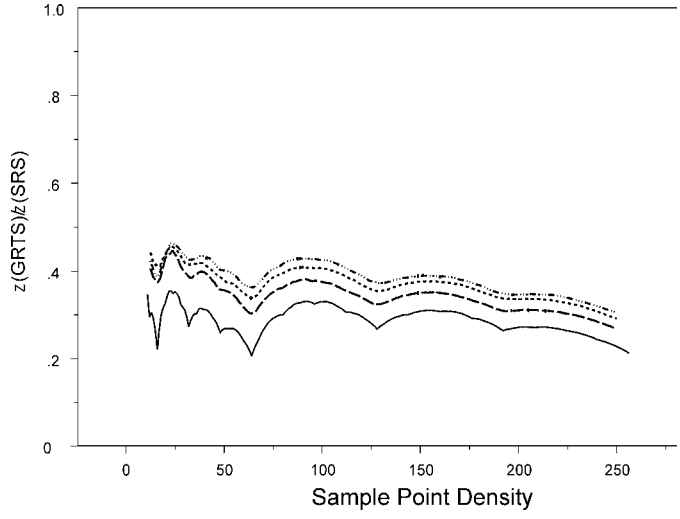

Figure 6. The ζ Ratio as a Function of Point Density Based on 1,000 Replications of a Sample of Size 256 (—— continuous domain with no voids; – – –– exponentialy increasing polygon size; - - - - - linearly increasing polygon size; · · —· · — constant polygon size).

design and ordered the samples using the reverse hierarchical order. As for the finite population study, we calculated the ζ ratio as the points were added to the sample one at a time, beginning with point number 10. The holes represent nontarget or access-denied elements that were a priori unknown. Sample points that fell in the holes were discarded, resulting in a variable number of sample points in the target domain. As for the complete domain, we ordered the points using reverse hierarchical ordering and then calculated the ζ ratio as the points were added one at a time. Because the sample points that fall into the nontarget areas contribute to the sample point density but not to the sample size, the ζ ratio was plotted versus point density.

We used three different distributions of hole size: constant, linearly increasing, and exponentially increasing. In each case, the holes comprise 20% of the domain area. Figure 5 shows the

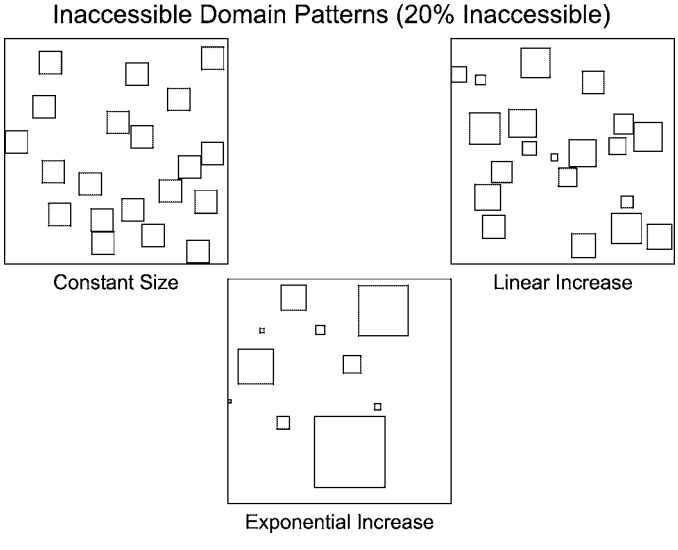

Figure 5. Void Patterns Used to Simulate Inaccessible Population Elements.

placement of the holes for each scenario, and Figure 6 shows the ζ ratio for all four scenarios: no voids, exponentially increasing, linearly increasing, and constant size.

In every scenario, the variance ratio is much less than 1. Except for small sample sizes, the ratio stays in the range of .2 to .4. The gradual decrease as the sample size increases is due to the decreasing impact of the boundary: as the sample size increases, the proportion of polygons that intersect the boundary decreases. A similar effect is seen with the different inaccessibility scenarios: even though the inaccessible area is constant, the scenarios with greater perimeter cause more increase in variance.

4. STATISTICAL PROPERTIES OF GRTS DESIGN

4.1 Estimation

The GRTS design produces a sample with specified first-order inclusion probabilities, so that the Horvitz–Thompson (Horvitz and Thompson 1952) estimator or its continuous population analog (Cordy 1993; Stevens 1997) can be applied to get estimates of population characteristics. Thus, for example, an estimate of the population total of a response z is given by $\hat{Z}_T = \sum_{s_i \in R}(z(s_i))/(\pi(s_i))$. Stevens (1997) provided exact expressions for second-order inclusion functions for some special cases of a GRTS. These expressions can also be used to provide accurate approximations for the general case. Unfortunately, the variance estimator based on using these approximations in the usual Horvitz–Thompson (HT) or Yates–Grundy–Sen (YG; Yates and Grundy 1953; Sen 1953) estimator tends to be unstable. The design achieves spatial balance by forcing the pairwise inclusion probability to approach 0 as the distance between the points in the pair goes to 0. Even though the pairwise inclusion density is nonzero almost everywhere, any moderate-sized sample will nevertheless have one or more pairs of points that are close together, with a correspondingly small pairwise inclusion probability. For both the HT and YG variance estimators, the pairwise inclusion probability appears as a divisor. The cor-

responding terms in either HT or YG variance estimators will tend to be large, leading to instability of the variance estimator.

Contrast-based estimators of the form $\hat{V}_{\mathrm{Ctr}}(\hat{Z}_T) = \sum_i w_i y_i^2$, where y_i is a contrast of the form $y_i = \sum_k c_{ik} z(s_k)$ with $\sum_k c_{ik} = 0$, have been discussed by several authors (Yates 1981; Wolter 1985; Overton and Stehman 1993). For an RTS design, Overton and Stehman also considered a "smoothed" contrast-based estimator of the form $\hat{V}_{\mathrm{SMO}}(\hat{Z}_T) = \sum_i w_i (z_i - z_i^*)^2$, where z_i^*, called the smoothed value for data point z_i, is taken as a weighted mean of a point plus its nearest neighbors in the tessellation.

Stevens and Olsen (2003) proposed a contrast-based estimator for the GRTS design that bears some resemblance to the Overton and Stehman smoothed estimator. The single contrast $(z_i - z_i^*)^2$ is replaced with an average of several contrasts over a local neighborhood analogous to a tessellation cell and its nearest neighbors in the RTS design. A heuristic justification for this approach stems from the observation that the inverse images of the unit-probability intervals on the line form a random spatial stratification of the population domain. The GRTS design, conditional on the stratification, is a one-sample-per-stratum spatially stratified sample. Recall that $\hat{Z}_T = \sum_{s_i \in R} (z(s_i))/\pi(s_i))$, where $z(s_i)$ is a sample from the ith random stratum. The selections within strata are conditionally independent of one another, so that

$$V(\hat{Z}_T) = \sum_{s_i \in R} E\left[V\left(\frac{z(s_i)}{\pi(s_i)} \middle| \text{strata} \right) \right].$$

The proposed variance estimator approximates $E[V((z(s_i))/(\pi(s_i))|\,\text{strata})]$ by averaging several contrasts over a local neighborhood of each sample point. The estimator is

$$\hat{V}_{\mathrm{NBH}}(\hat{Z}_T) = \sum_{s_i \in R} \sum_{s_j \in D(s_i)} w_{ij} \left(\frac{z(s_j)}{\pi(s_j)} - \sum_{s_k \in D(s_i)} w_{ik} \frac{z(s_k)}{\pi(s_k)} \right)^2,$$

where $D(s_i)$ is a local neighborhood of the s_i. The weights w_{ij} are chosen to reflect the behavior of the pairwise inclusion function for GRTS, and are constrained so that $\sum_i w_{ij} = \sum_j w_{ij} = 1$. Stimulation studies with a variety of scenarios have shown the proposed estimator to be stable and nearly unbiased. Applications with real data have consistently shown that our local neighborhood variance estimator produces smaller estimates than the Horvitz–Thompson estimator when IRS is assumed to approximate for the joint inclusion probabilities.

4.2 Inverse Sampling

The reverse hierarchical ordering provides the ability to do inverse sampling, that is, to sample until a given number of samples are obtained in the target population. The true inclusion probability in this case depends on the spatial configuration of the target population, which may be unknown. However, one can compute an inclusion probability that is conditional on the achieved sample size in the target population being fixed. For example, suppose we want M sample points in our domain R. We do not know the exact boundaries of R, but are able to enclose R in a larger set R^*. We select a sample of size $M^* > M$ from R^* using an inclusion density π^* scaled so that $\int_{R^*} \pi^*(s)\, d\phi(s) = M^*$. The inclusion density for the k-point reverse hierarchical ordered sample is $\pi_k^*(s) = (k/M^*)\pi^*(s)$. Using the inclusion density π_k^*, the expected number of samples in R is

$$M_k = \int_R \pi_k^*(s)\, d\phi(s) = \int_{R^*} I_R(s) \pi_k^*(s)\, d\phi(s).$$

We cannot compute M_k because the boundary of R is unknown, but an estimate is

$$\hat{M}_k = \sum_i \frac{I_R(s_i)\pi_k^*(s_i)}{\pi_k^*(s_i)} = \sum_i I_R(s_i).$$

We pick $\tilde{k}$ so that $\hat{M}_k = M$ and base inference on $\pi_{\tilde{k}}^*$. Thus, for example, an estimate of the unknown extent of R is $|\hat{R}| = \sum_i (I_R(s_i))/(\pi_{\tilde{k}}^*(s_i))$.

We illustrate this using the same inaccessibility scenarios as for the spatial balance simulation. Results are summarized in Table 2. In each case, the true area of R is .8, so that the estimator using $\pi_{\tilde{k}}^*$ is either unbiased or nearly so.

Table 2. Domain Area Estimates Using Conditional Inclusion Probability

Target sample size	Mean estimated domain area		
	Exponential	Constant	Linear
25	.8000979	.7969819	.8010589
50	.7995775	.7979406	.8005739
100	.7994983	.7980543	.8002237
150	.7994777	.7997587	.7995685

4.3 Statistical Efficiency

As discussed in the Introduction, sampling designs with some degree of spatial regularity, for example, systematic, grid-based, or spatially stratified designs, tend to be more efficient for sampling natural resources than designs with no spatial structure. The GRTS design takes the concept of spatial stratification, carries it to an extreme, and gives it flexibility and robustness. The basis for these claims is that for the case of an equiprobable sample of an areal resource over a continuous, connected domain, a GRTS sample with size $n = 4^k$ is a spatially stratified sample with one sample point per stratum. In this case, the strata are square grid cells with a randomly located origin. Generally, the efficiency of a spatially stratified sample increases as the number of strata increases (samples per stratum decreases), so maximal efficiency is obtained for a one-point-per-stratum-design. Thus, in this restricted case, the GRTS has the same efficiency as the maximally efficient spatial stratification.

The spatial regularity simulation studies provide some insight into less restrictive cases. First, the "no-void" case of the continuous domain study shows that the spatial regularity is not seriously degraded for sample sizes that are not powers of 4, so that even for intermediate sample sizes, the GRTS efficiency should be close to the efficiency of maximal spatial stratification. Second, the "holes" cases show that for irregularly shaped domains, GRTS maintains spatial regularity. In this case, GRTS with $n = 4^k$ is again a one-point-per-stratum

design, but the strata are no longer regular polygons. Nevertheless, GRTS should have the same efficiency as maximal stratification.

An example of circumstances where efficiency is difficult to evaluate is a finite population study, with variable probability and irregular spatial density. In these circumstances, spatial strata can be very difficult to form and, in fact, it may be impossible to form strata with a fixed number of samples per stratum. A GRTS sample achieves the regularity of a one-sample-per-stratum stratification and so should have the same efficiency.

The overwhelming advantage of a GRTS design is not that it is more efficient than spatial stratification, but that it can be applied in a straightforward manner in circumstances where spatial stratification is difficult. All of the pathologies that occur in sampling natural populations (poor frame information, inaccessibility, variable probability, uneven spatial pattern, missing data, and panel structures) can be easily accommodated within the GRTS design.

5. EXAMPLE APPLICATION TO STREAMS

The Indiana Department of Environmental Management (IDEM) conducts water quality and biological assessments of the streams and rivers within Indiana. For administrative purposes, the state is divided into nine hydrologic basins: East Fork White River Basin, West Fork White River Basin, Upper Illinois River Basin, Great Miami River Basin, Lower Wabash River Basin, Patoka River Basin, Upper Wabash River Basin, Great Lakes Basin, and Ohio River Basin. All basins are assessed once during a 5-year period; typically, two basins are completed each year. In 1996, IDEM initiated a monitoring strategy that used probability survey designs for the selection of sampling site locations. We collaborated with them on the survey design. In 1997, a GRTS multidensity design was implemented for the East Fork White River Basin and the Great Miami River Basin. In 1999, another GRTS multidensity design was implemented for the Upper Illinois Basin and the Lower Wabash. These designs will be used to illustrate the application of GRTS survey designs to a linear network.

The target population for the studies consists of all streams and rivers with perennially flowing water. A sample frame, River Reach File Version 3 (RF3), for the target population is available from the U.S. Environmental Protection Agency (Horn and Grayman 1993). The RF3 includes attributes that enable perennial streams and rivers to be identified, but results in an overcoverage of the target population due to coding errors. In addition, Strahler order is available to classify streams and rivers into relative size categories (Strahler 1957). A headwater stream is a Strahler first-order stream, two first-order streams joining results in a second-order stream, and so on. Approximately, 60% of the stream length in Indiana is first order, 20% is second order, 10% is third order, and 10% is fourth and greater (see Table 3). In 1997, IDEM determined that the sample would be structured so that approximately an equal number of sites would be in first order, second and third order, and fourth + order for the East Fork White River and the Great Miami River basins. In 1999, the sample was modified to have an equal number of sites in first, second, third, and fourth + order categories for the Lower Wabash and Upper Illinois basins.

Table 3. Sample Frame Stream and River Length by Basin and Strahler Order Category

		Strahler order category length (km)		
Basin	*Total length (km)*	*1*	*2*	*3+*
E Fork White	6802.385	3833.335	2189.494	779.556
Great Miami	2270.018	1501.711	621.039	147.268
L Wabash	7601.418	4632.484	1331.228	1637.706
U Illinois	5606.329	4559.123	500.188	547.018

The GRTS multidensity survey designs were applied. In both years, six multidensity categories were used (three Strahler order categories in each of two basins). Although four Strahler order categories were planned in 1999, the stream lengths associated with the third and fourth + categories were approximately equal, so a single category that combined the sample sizes was used. To account for frame errors, landowner denials, and physically inaccessible stream sites, a 100% oversample was incorporated in 1999. The intent was to have a minimum of 38 biological sites with field data in 1999; this was not done in 1997. Table 4 summarizes the number of sites expected and actually evaluated, as well as the number of nontarget, target, nonresponse, and sampled sites. Almost all of the nonresponse sites are due to landowner denial. In 1999, the sites were used in reverse hierarchical order until the desired number of actual field sample sites was obtained. The biological sites were a nested subsample of the water chemistry sites and were taken in reverse hierarchical order from the water chemistry sites. Figures 7–10 show the spatial pattern of the stream networks and the GRTS sample sites for each of the four basins by Strahler order categories. Although this is an example of a single realization of a multidensity GRTS design, all realizations will have a similar spatial pattern. Prior to statistical analysis, the initial inclusion densities are adjusted to account for use of oversample sites by recalculating the inclusion densities by basin.

Indiana determined two summary indices related to the ecological condition of the streams and rivers: the IBI score, which is a fish community index of biological integrity (Karr 1991) that assesses water quality using resident fish communities as a tool for monitoring the biological integrity of streams, and the QHEI score, which is a habitat index based on the Ohio Environmental Protection Agency qualitative habitat evaluation index (see IDEM 2000 for detailed descriptions of these indices).

Table 4. Survey Design Sample Sizes for Basins Sampled in 1997 and 1999

Basin	*Expected sample size*	*Evaluated sample size*	*Nontarget sites*	*Target sites*	*Nonresponse sites*	*Water chemistry sites*	*Biological sites*
E Fork White	60	60	5	55	9	35	34
Great Miami	40	40	12	28	5	19	19
L Wabash	128	91	11	80	9	71	39
U Illinois	128	85	8	77	5	72	41

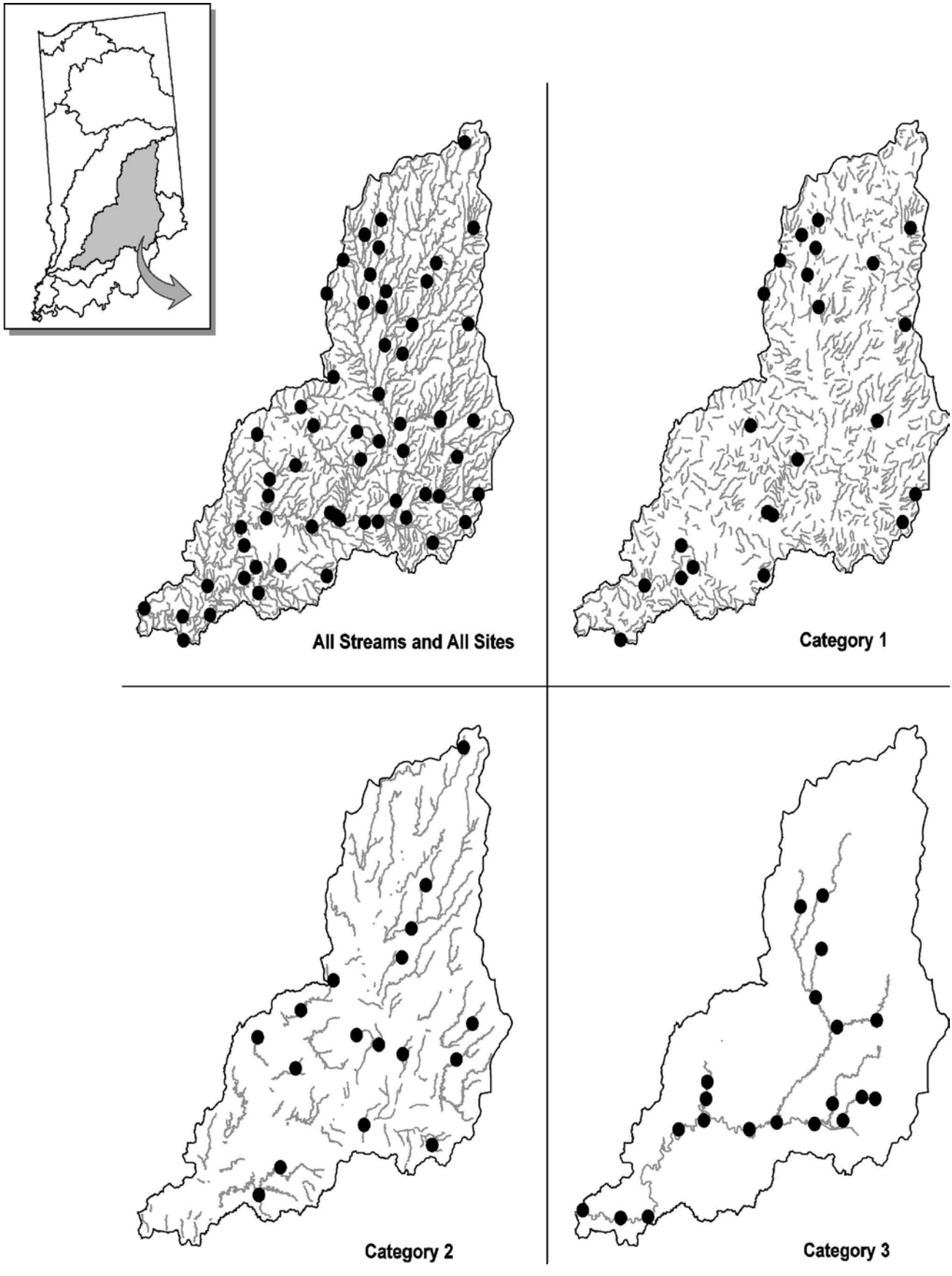

Figure 7. East Fork White River Basin Sample Sites by Multidensity Categories.

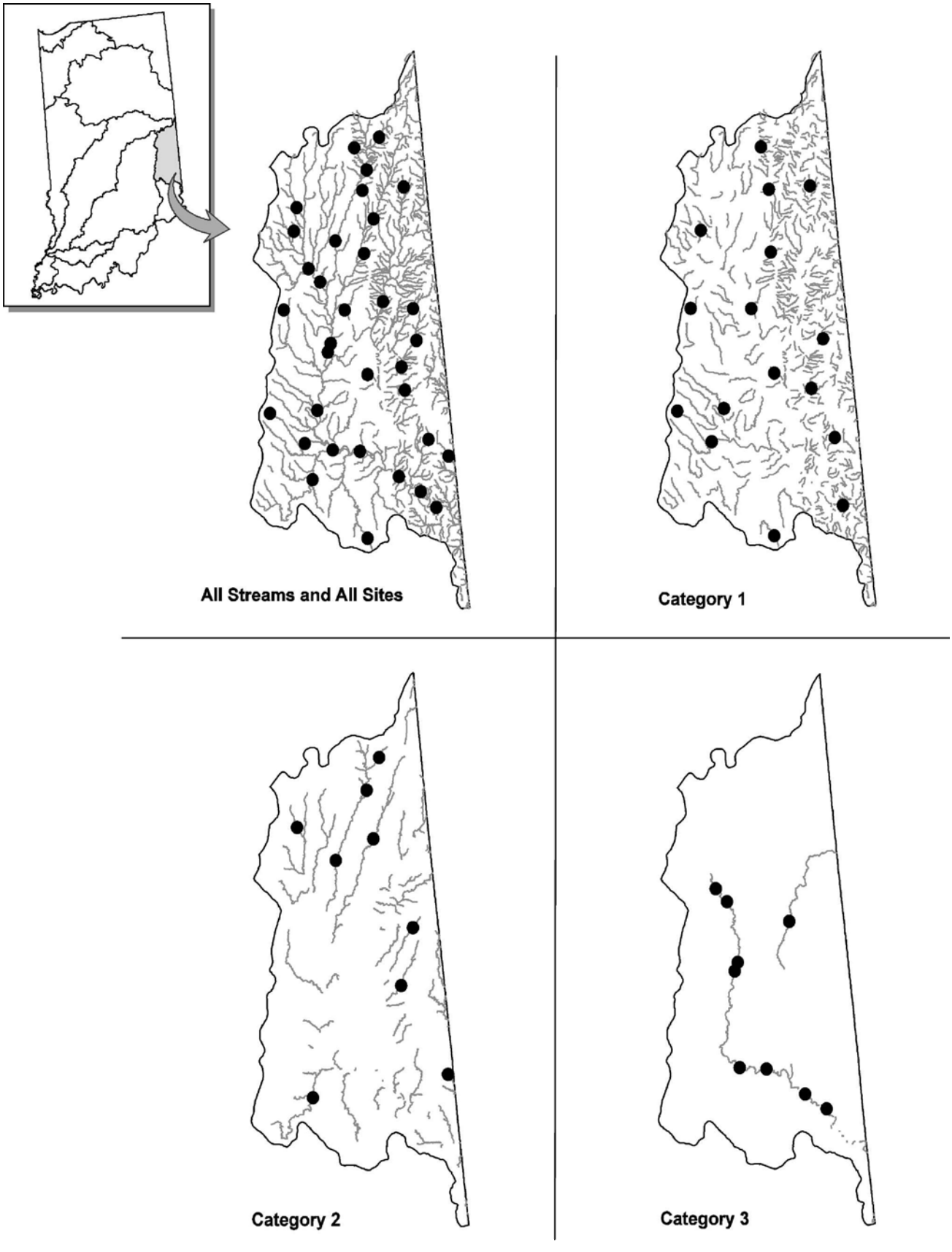

Figure 8. Great Miami River Basin Sample Sites by Multidensity Categories.

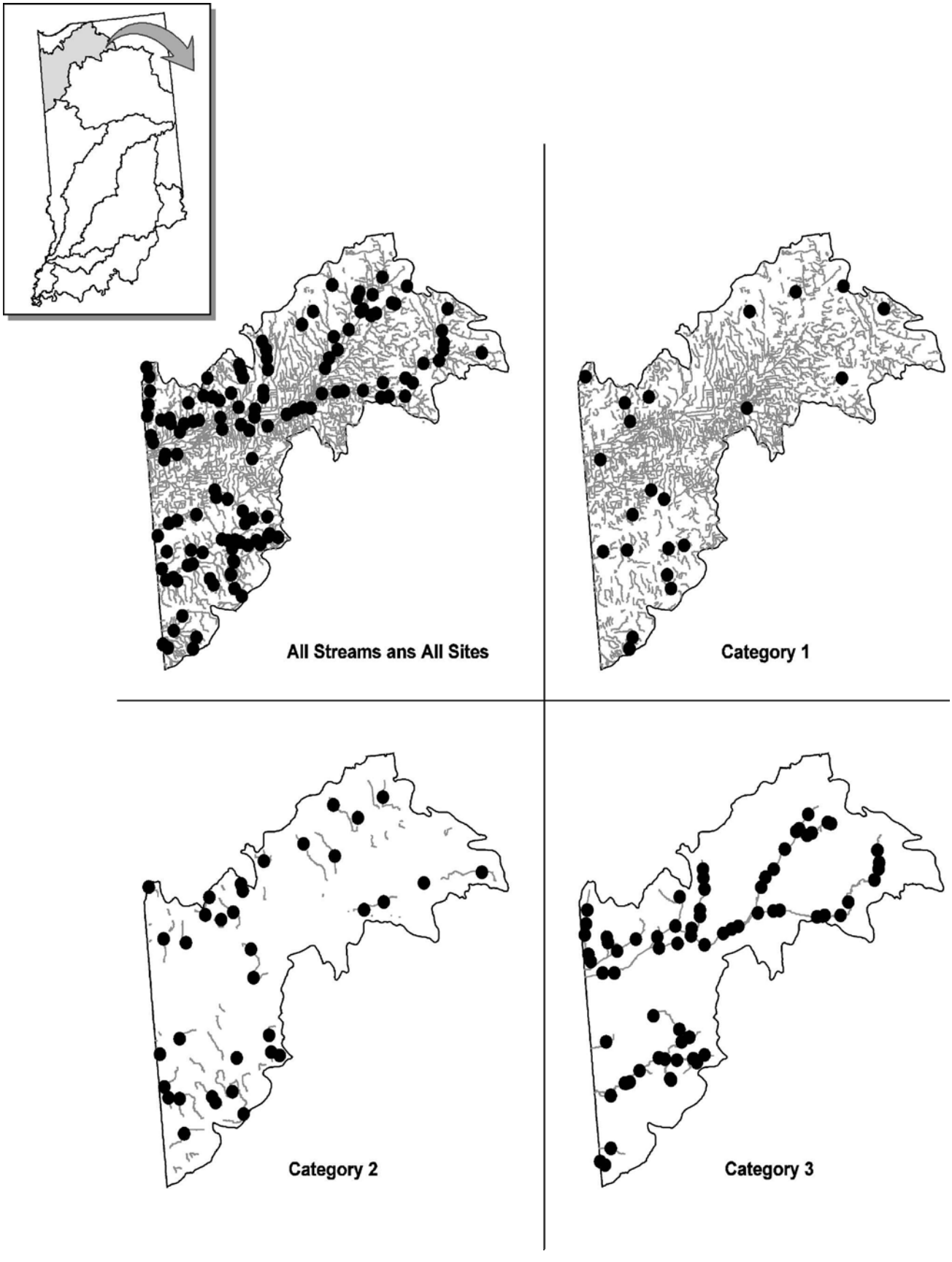

Figure 9. Upper Illinois River Basin Sample Sites by Multidensity Categories.

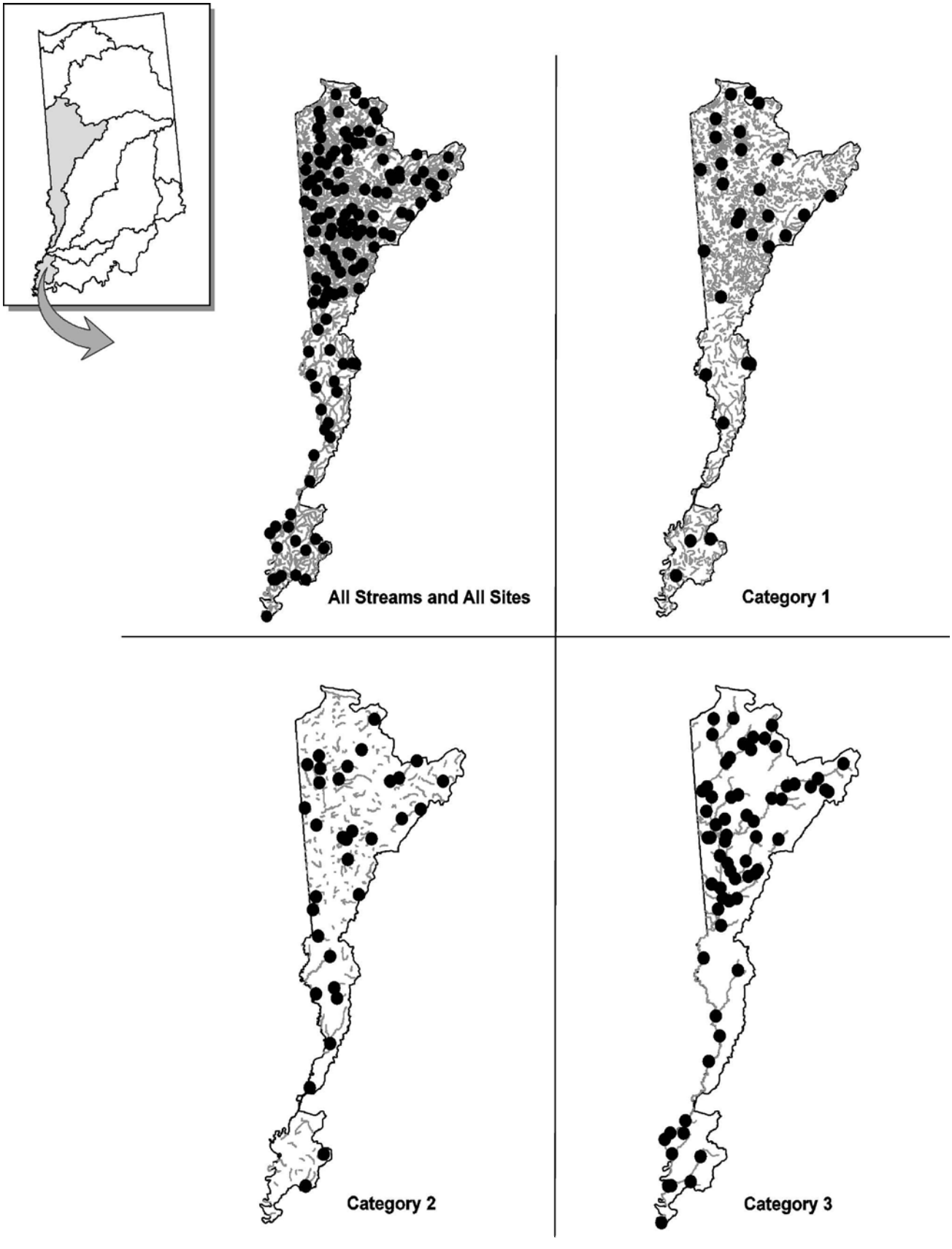

Figure 10. Lower Wabash River Basin Sample Sites by Multidensity Categories.

Table 5. Population Estimates With IRS and Local Variance Estimates

Subpopulation	Indicator score	N sites	Mean	IRS std. err.	Local std. err.	Difference (%)
L Wabash	IBI	39	36.1	2.1	1.4	−56.8
U Illinois	IBI	41	32.5	1.7	1.3	−44.8
E Fork White	IBI	32	35.1	1.3	1.2	−22.3
Great Miami	IBI	19	40.8	2.7	2.2	−33.5
L Wabash	QHEI	39	55.6	2.3	1.6	−52.2
U Illinois	QHEI	41	43.3	2.1	1.6	−39.5
E Fork White	QHEI	34	54.3	2.1	1.5	−45.9
Great Miami	QHEI	19	67.8	2.2	1.9	−26.0

Table 5 summarizes the population estimates for IBI and QHEI scores for each of the four basins. The associated standard error estimates are based on the Horvitz–Thompson ratio variance estimator, assuming an independent random sample, and on the local neighborhood variance estimator described in Section 4.1. On average, the neighborhood variance estimator is 38% smaller than the IRS variance estimator. Figure 11 illustrates the impact of the variance estimators on confidence intervals for cumulative distribution function estimates for the Lower Wabash Basin.

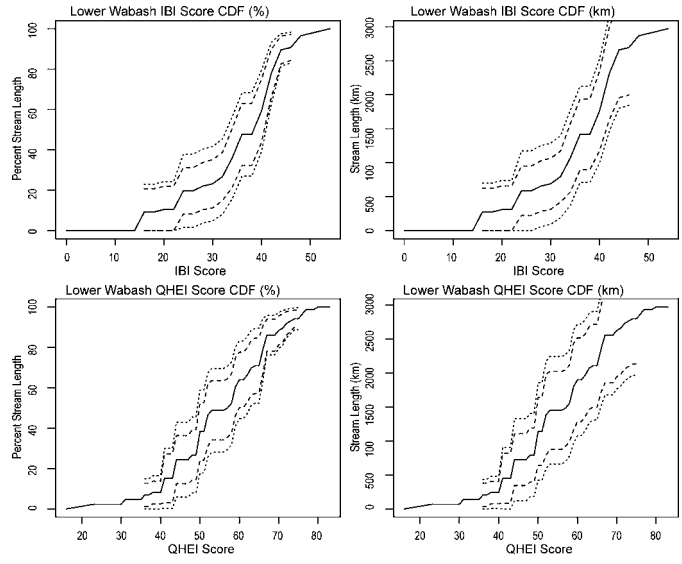

Figure 11. Stream Network and Sample Site Spatial Patterns by Multidensity Category for the Lower Wabash Basin (—— CDF estimate; – – – – – 95% local confidence limits; · · · · · · · · 95% IRS confidence limits).

6. DISCUSSION

There are a number of designs that provide good dispersion of sample points over a spatial domain. When we applied these designs to large-scale environmental sampling programs, it quickly became apparent that we needed a means (1) to accommodate variable inclusion probability and (2) to adjust sample sizes dynamically. These requirements are rooted in the very fundamentals of environmental management. The first requirement stems from the fact that an environmental resource is rarely uniformly important in the objective of the monitoring: there are always scientific, economic, or political reasons for sampling some portions of a resource more intensively than others. Two features of environmental monitoring programs drive the second requirement. First, these programs tend to be long lived, so that even if the objectives of the program remain unchanged, the "important" subpopulations change, necessitating a corresponding change in sampling intensity. Second, a high-quality sampling frame is often lacking for environmental resource populations. As far as we know, there is no other technique for spatial sampling that "balances" over an intensity metric instead of a Euclidean distance metric or permits dynamic modification of sample intensity.

Adaptive sampling (Thompson 1992, pp. 261–319) is another way to modify sample intensity. However, there are some significant differences between GRTS and adaptive sampling in the way the modification is accomplished. Adaptive sampling increases the sampling intensity *locally* depending on the response observed at a sample point, whereas the GRTS intensity change is global.

The GRTS first-order inclusion probability (or density) can be made proportional to an arbitrary positive auxiliary variable, for example, a signal from a remote sensing platform or a sample intensity that varies by geographical divisions or known physical characteristics of the target population. In some point and linear situations, it may be desirable to have the sample be spatially balanced with respect to geographic space rather than with respect to the population density. This can be achieved by making the inclusion probability inversely proportional to the population density. Although the development of GRTS has focused on applications in geographic space, it can be applied in other spaces. For example, one application defined two-dimensional space by the first two principal components of climate variables and selected a GRTS sample of forest plots in that space.

The computational burden in hierarchical randomization can be substantial. However, it needs to be carried out only to a resolution sufficient to obtain no more than one sample point per subquadrant. The actual point selection can be carried out by treating the subquadrants as if they are elements of a finite population, selecting the M subquadrants to receive sample points, and then selecting one population element at random from among the elements contained within the selected subquadrants, according to the probability specified by π.

Reverse hierarchical ordering adds a feature that is immensely popular with field practitioners, namely the ability to "replace" samples that are lost due to being nontarget or inaccessible. Moreover, we can replace the samples in such a way as to achieve good spatial balance over the population that is actually sampleable, even when sampleability cannot be determined prior to sample selection. Of course, this feature does not eliminate the nonresponse or the bias of an inference to the inaccessible population. It does, however, allow investigators to obtain the maximum number of samples that their budget will permit them to analyze.

Reverse hierarchical ordering has other uses as well. One is to generate interpenetrating subsamples (Mahalanobis 1946). For example, 10 interpenetrating subsamples from a sample size of 100 can be obtained simply by taking consecutive subsets of 10 from the reverse hierarchical ordering. Each subset has the same properties as the complete design. Consecutive subsets can also be used to define panels of sites for application in surveys over time, for example, sampling with partial replacement (Patterson 1950; Kish 1987; Urquhart, Overton, and Birkes 1993).

APPENDIX A: PROOF OF LEMMA

Lemma. Let $f : \Re^2 \to \Re$ be a 1–1 quadrant-recursive function and let $s \sim \mathrm{U}(\Re^2)$. Then $\lim_{|\delta|\to 0} E\{|f(s) - f(s+\delta)|\} = 0$.

Proof. If, for some $n > 0$, s and $s+\delta$ are in the same subquadrant Q^n_{jk}, then $f(s)$ and $f(s+\delta)$ are in the same interval J^n_m, so that $|f(s) - f(s+\delta)| \le 1/4^n$. The probability that s and $s+\delta$ are in the same subquadrant is the same as the probability of the origin and $\delta = (\delta_x, \delta_y)$ being in the same cell of a randomly located grid with cells congruent to Q^n_{jk}. For $\delta_x, \delta_y \le 1/2^n$, that probability is equal to $|Q^n(0) \cap Q^n(\delta)|/|Q^n(0)| = 1 - 2^n(\delta_x + \delta_y) + 4^n\delta_x\delta_y$, where $Q^n(x)$ denotes a polygon congruent to Q^n_{jk} centered on x. For $D(s,\delta) = |f(s) - f(s+\delta)|$, then, we have that $P(D \le 1/4^n) \ge 1 - 2^n(\delta_x + \delta_y) + 4^n\delta_x\delta_y$. Thus, the distribution function F_D of D is bounded below by

$$F_D(u) \ge \begin{cases} 0, & u \le \dfrac{1}{4^n} \\ 1 - 2^n(\delta_x + \delta_y) + 4^n\delta_x\delta_y, & u > \dfrac{1}{4^n}. \end{cases}$$

Because D is positive and bounded above by 1,

$$\begin{aligned} E[D(\delta)] &= 1 - \int_0^1 F_D(u)\,du \\ &\le 1 - \left\{\frac{0}{4^n} + \left(1 - \frac{1}{4^n}\right) - 2^n(\delta_x + \delta_y) + 4^n\delta_x\delta_y\right\}. \end{aligned}$$

For fixed n, we have that

$$\lim_{|\delta|\to 0} E[D(\delta)] \le \frac{1}{4^n},$$

but this holds for all n, so that

$$\lim_{|\delta|\to 0} E[D(\delta)] = 0.$$

APPENDIX B: PROOF THAT THE PROBABILITY INCLUSION FUNCTION EQUALS THE TARGET INTENSITY FUNCTION

We need the measure space $(\mathfrak{X}, \mathfrak{B}, \phi)$, where $\mathfrak{X}$ is the unit interval $\Re = (0, 1]$ or the unit square $\Re^2 = (0, 1] \times (0, 1]$, and the relevant σ fields are $\mathfrak{B}(\Re)$ and $\mathfrak{B}(\Re^2)$, the σ fields of the Borel subsets of $\Re$ and $\Re^2$, respectively. For each of the three types of populations, we define a measure ϕ of population size. We use the same symbol for all three cases, but the specifics vary from case to case. For a finite population, we take ϕ to be counting measure restricted to R, so that for any subset $B \in \mathfrak{B}(\Re^2)$, $\phi(B)$ is the number of population elements in $B \cap R$. For linear populations, we take $\phi(B)$ to be the length of the linear population contained within B. Clearly, ϕ is nonnegative, countably additive, defined for all Borel sets, and $\phi(\varnothing) = 0$, so ϕ is a measure. Finally, for areal populations, we take $\phi(B)$ to be the Lebesgue measure of $B \cap R$.

We begin by randomly translating the image of R in the unit square by adding independent, U(0, 1/2) offsets to the xy coordinates. This random translation plays the same role as random grid location does in an RTS design; namely, it guarantees that pairwise inclusion probabilities are nonzero. In particular, in this case it ensures that any pair of points in R has a nonzero chance of being mapped into different quadrants.

Let $\pi(s)$ be an inclusion intensity function, that is, a function that specifies the target number of samples per unit measure. We assume that any linear population consists of a finite number m of smooth, rectifiable curves, $R = \bigcup_{i=1}^m \{\gamma_i(t) = (x_i(t), y_i(t)) | t \in [a_i, b_i]\}$, with x_i and y_i continuous and differentiable on $[a_i, b_i]$. We set $\pi(s)$ equal to the target number of samples per unit length at s, for $s \in L$ and equal to zero elsewhere. For example, if the linear population were a stream network, $\pi(s)$ would specify the desired number of samples per kilometer of stream at the point s. Finally, an areal population is a finite collection of closed polygons. In this case, $\pi(s)$ specifies the target intensity as number of samples per unit area. Note that for one- and two-dimensional resources, $\pi(s)$ could be a continuous, smoothly varying function. Formally, we require $\pi(s)$ to be bounded and measurable, strictly positive on R and zero elsewhere, and scaled so that $M = \int_R \pi(s)\,d\phi(s)$. From these definitions of $\pi(\cdot)$ and $\phi(\cdot)$, it follows that $w(B) = \int_B \pi(s)\,d\phi(s)$ is a measure and that $w(B)$ is the target number of samples in B. In particular, $M = w(\Re^2)$ is the target sample size. In the following discussion, we assume that M is an integer; the noninteger case is a simple extension.

Let $f(\cdot)$ be a quadrant-recursive function that maps $\Re^2$ into $\Re$. Because $\mathfrak{B}(\Re)$ can be generated by sets of the form J^n_m and $\mathfrak{B}(\Re^2)$ can be generated by sets of the form Q^n_{jk}, both f and f^{-1} are measurable. Because f is measurable, $f^{-1}(B)$ is measurable for $B \in \mathfrak{B}(\Re)$, so that $\tilde{F}(x) = \int_{f^{-1}((0,x])} \pi(s)\,d\phi(s)$ exists. In fact, $\tilde{F}$ is a distribution function, that is, nonnegative, increasing, and right continuous. For linear and areal resources, $\tilde{F}$ is a continuous, increasing function, but for finite resource populations, $\tilde{F}$ is a step function with jumps at the images of population elements. We can modify $\tilde{F}$ to obtain continuity in the finite case via linear interpolation, that is, let x_i, $i = 1, \ldots, N$, be the ordered jump points of $\tilde{F}$, set $x_0 = 0$, $x_{N+1} = 1$, and, for $x_i < x \le x_{i+1}$, set $F(x) = \tilde{F}(x) + (\tilde{F}(x_{i+1}) - \tilde{F}(x_i))/(x_{i+1} - x_i)(x - x_i)$. If we set $F = \tilde{F}$ for the linear and areal case, then in every cases we have that F is a continuous distribution function with range $(0, M]$.

In the finite case, F^{-1} is single-valued, so that $G(y) = \min(x_i | F^{-1}(y) \le x_i)$ is well defined. In the linear and areal cases, F^{-1} may not be single-valued. Points that are in the unit square but not in R lead to flats in F that correspond to regions in the unit square with $\pi(s) = 0$. However, $F^{-1}(y)$ always will be closed and bounded, so that $G(y) = \min\{x | x \in F^{-1}(y)\}$ is well defined. In all cases, the intensity function π is positive at $s = f^{-1}(G(y))$, that is, there is a population element at s. Thus, $f^{-1} \circ G$ maps $(0, M]$ onto the target population, that is, $f^{-1} \circ G$ associates every point in $(0, M]$ with a unique element in the population.

It follows that selecting a sample from $(0, M]$ also selects population elements via the mapping $f^{-1} \circ G$. To get a sample with an inclusion function equal to the target inclusion density, we select a sample from $(0, M]$ by splitting the range into M unit-length intervals $(0, 1]$, $(1, 2], \ldots,$ $(M-1, M]$ and picking one point in each interval. Because of hierarchical randomization, we gain no additional "randomness" by picking the points independently, so we use systematic sampling with a random start and a unit-length selection interval. The selection procedure defines an inclusion probability density function on $(0, M]$ with a corresponding measure $P_M(\cdot)$. Note that P_M coincides with Lebesgue measure on $(0, M]$; in particular, the measure of a subinterval of $(0, M]$ is its length. We induce a measure P_1 on $\Re$ via $P_1(B) = \int_{G^{-1}(B)} dP_M$, and in turn induce a measure P_2 on $\Re^2$ via $P_2(B) = \int_{f^{-1}(B)} dP_1$. The measure P_2 is an inclusion probability measure on $\Re^2$, and $P_2(B) = w(B)$, so the sample selection method does give an inclusion probability function equal to the target sample intensity function.

[Received August 2002. Revised September 2003.]

REFERENCES

Bellhouse, D. R. (1977), "Some Optimal Designs for Sampling in Two Dimensions," *Biometrika*, 64, 605–611.

Bickford, C. A., Mayer, C. E., and Ware, K. D., (1963), "An Efficient Sampling Design for Forest Inventory: The Northeast Forest Resurvey," *Journal of Forestry*, 61, 826–833.

Breidt, F. J. (1995), "Markov Chain Designs for One-per-Stratum Sampling," *Survey Methodology*, 21, 63–70.

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

Cochran, W. G. (1946), "Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations," *The Annals of Mathematical Statistics*, 17, 164–177.

Cordy, C. (1993), "An Extension of the Horvitz–Thompson Theorem to Point Sampling From a Continuous Universe," *Probability and Statistics Letters*, 18, 353–362.

Cotter, J., and Nealon, J. (1987), "Area Frame Design for Agricultural Surveys," U.S. Department of Agriculture, National Agricultural Statistics Service, Research and Applications Division, Area Frame Section.

Dalenius, T., Hájek, J., and Zubrzycki, S. (1961), "On Plane Sampling and Related Geometrical Problems," in *Proceedings of the 4th Berkeley Symposium on Probability and Mathematical Statistics*, 1, 125–150.

Das, A. C. (1950), "Two-Dimensional Systematic Sampling and the Associated Stratified and Random Sampling," *Sankhyā* 10, 95–108.

Gibson, L., and Lucas, D. (1982), "Spatial Data Processing Using Balanced Ternary," in *Proceedings of the IEEE Computer Society Conference on Pattern Recognition and Image Processing*, Silver Springs, MD: IEEE Computer Society Press.

Gilbert, R. O. (1987), *Statistical Methods for Environmental Pollution Monitoring*, New York: Van Nostrand Reinhold.

Hausdorff, F. (1957), *Set Theory*, New York: Chelsea.

Hazard, J. W., and Law, B. E. (1989), *Forest Survey Methods Used in the USDA Forest Service*, EPA/600/3-89/065, Corvallis, Oregon: U.S. Environmental Protection Agency, Office of Research and Development, Environmental Research Laboratory.

Horn, C. R., and Grayman, W. M. (1993), "Water-Quality Modeling With EPA River Reach File System," *Journal of Water Resources Planning and Management*, 119, 262–274.

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.

Iachan, R. (1985), "Plane Sampling," *Statistics and Probability Letters*, 50, 151–159.

IDEM (2000), "Indiana Water Quality Report 2000," Report IDEM/34/02/001/2000, Indiana Department of Environmental Management, Office of Water Management, Indianapolis, Indiana.

Insightful Corporation (2002), "S-PLUS 6 for Windows Language Reference," Insightful Corporation, Seattle, WA.

Karr, J. R. (1991), "Biological Integrity: A Long Neglected Aspect of Water Resource Management," *Ecological Applications,* 1, 66–84.

Kish, L. (1987), *Statistical Design for Research*, New York: Wiley.

Mahalanobis, P. C. (1946), "Recent Experiments in Statistical Sampling in the Indian Statistical Institute," *Journal of the Royal Statistical Society*, 109, 325–370.

Mark, D. M. (1990), "Neighbor-Based Properties of Some Orderings of Two-Dimensional Space," *Geographical Analysis,* 2, 145–157.

Matérn, B. (1960), *Spatial Variation*, Stockholm, Sweden: Meddelanden från Statens Skogsforskningsinstitut.

Messer, J. J., Arsiss, C. W., Baker, J. R., Drousé, S. K., Eshleman, K. N., Kaufmann, P. R., Linthurst, R. A., Omernik, J. M., Overton, W. S., Sale, M. J., Schonbrod, R. D., Stambaugh, S. M., and Tuschall, J. R., Jr. (1986), *National Surface Water Survey: National Stream Survey, Phase I-Pilot Survey*, EPA-600/4-86/026. Washington, D.C.: U.S. Environmental Protection Agency.

Munholland, P. L., and Borkowski, J. J. (1996), "Simple Latin Square Sampling + 1: A Spatial Design Using Quadrats," *Biometrics*, 52, 125–136.

Olea, R. A. (1984), "Sampling Design Optimization for Spatial Functions," *Mathematical Geology*, 16, 369–392.

Overton, W. S., and Stehman, S. V. (1993). "Properties of Designs for Sampling Continuous Spatial Resources From a Triangular Grid," *Communications in Statistics, Part A—Theory and Methods*, 22, 2641–2660.

Patterson, H. D. (1950), "Sampling on Successive Occasions With Partial Replacement of Units," *Journal of the Royal Statistical Society*, Ser. B, 12, 241–255.

Peano, G. (1890), "Sur Une Courbe, Qui Remplit Toute Une Aire Plane," *Mathematische Annalen*, 36, 157–160.

Quenouille, M. H. (1949), "Problems in Plane Sampling," *The Annals of Mathematical Statistics*, 20, 335–375.

Saalfeld, A. (1991), "Construction of Spatially Articulated List Frames for Household Surveys," in *Proceedings of Statistics Canada Symposium 91, Spatial Issues in Statistics*, Ottawa, Canada: Statistics Canada, pp. 41–53.

Sen, A. R. (1953), "On the Estimate of the Variance in Sampling With Varying Probabilities," *Journal of the Indian Society of Agricultural Statistics*, 7, 119–127.

Simmons, G. F. (1963), *Introduction to Topology and Modern Analysis*, New York: McGraw–Hill.

Stehman, S. V., and Overton, W. S. (1994), "Environmental Sampling and Monitoring," in *Handbook of Statistics*, Vol. 12, eds. G. P. Patil and C. R. Rao, Amsterdam, The Netherlands: Elsevier Science, pp. 263–305.

Stevens, D. L., Jr. (1997), "Variable Density Grid-Based Sampling Designs for Continuous Spatial Populations," *Environmetrics*, 8, 167–195.

Stevens, D. L., Jr., and Olsen, A. R. (1999), "Spatially Restricted Surveys Over Time for Aquatic Resources," *Journal of Agricultural, Biological, and Environmental Statistics*, 4, 415–428.

——— (2000), "Spatially-Restricted Random Sampling Designs for Design-Based and Model-Based Estimation," in *Accuracy 2000: Proceedings of the 4th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Delft, The Netherlands: Delft University Press, pp. 609–616.

——— (2003), "Variance Estimation for Spatially Balanced Samples of Environmental Resources," *Environmetrics*, 14, 593–610.

Strahler, A. N. (1957), "Quantitative Analysis of Watershed Geomorphology," *Transactions of the American Geophysical Union*, 38, 913–920.

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

Urquhart, N. S., Overton, W. S., and Birkes, D. S. (1993), "Comparing Sampling Designs for Monitoring Ecological Status and Trends: Impact of Temporal Patterns," in *Statistics for the Environment*, eds. V. Barnett and K. F. Turkman, New York: Wiley, pp. 71–86.

USGS (1999), "The National Hydrography Dataset," Fact Sheet 106-99, U.S. Geological Survey.

Wolter, K. (1985), *Introduction to Variance Estimation*, New York: Springer-Verlag.

Wolter, K. M., and Harter, R. M. (1990), "Sample Maintenance Based on Peano Keys," in *Proceedings of the 1989 International Symposium: Analysis of Data in Time*, Ottawa, Canada: Statistics Canada, pp. 21–31.

Yates, F. (1981), *Sampling Methods for Censuses and Surveys* (4th ed.), London: Griffin.

Yates, F., and Grundy, P. M. (1953), "Selection Without Replacement From Within Strata With Probability Proportional to Size," *Journal of the Royal Statistical Society*, Ser. B, 15, 253–261.