Systematic and Stratified Sampling Designs In Long-Term Ecological Monitoring Studies

Trent L. McDonald[1] and Paul H. Geissler[2]

5 March 2004

DRAFT

Abstract

In this paper, we present an overview of systematic and stratified spatial sampling designs that we feel are useful for long-term national park ecological monitoring. Each design is described in layman's terms, estimators for the mean of a parameter are given, and the characteristics of each are discussed. A hypothetical national park is used to illustrate ideas and certain computations.

Systematic, or grid, sampling with a random start is a relatively simple, robust, and all purpose spatial sampling design that assures all areas of a population are represented in proportion to their size. Under systematic sampling, analysis and post-stratification are straightforward, and the design has reasonable precision for all responses. A disadvantage of systematic sampling is its inability to adequately sample rare habitats unless sample size is very large. Stratified sampling designs, which utilize systematic or simple random samples within stratum, provide the opportunity to optimize costs of the design for a single response by taking into account travel time and costs. Stratified sampling, unlike systematic, can guarantee collection of 1 or more samples in every strata and is therefore useful for sampling rare habitat if its location is known. Monitoring networks may wish to consider placing areas that are too dangerous to sample in one (unsampled) stratum, with other areas being placed in stratum based on their access costs or presence of rare habitats. Areas that are too dangerous to sample, and areas that are very costly to sample, should not be included in the same stratum because overall costs can be controlled by the sample allocation method. Disadvantages of stratified sampling include slightly increased analysis complexity, especially if park-wide estimates are desired for habitats that span several stratum, the inability of change strata boundaries even if such a change would be advantageous, and the inability of optimize costs for multiple response variables.

[1] West Inc., 2003 Central Ave., Cheyenne, WY 82001 Tel. 307-634-1756 FAX: 307-637-6981, TMcDonald@West-Inc.com

[2] U. S. Geological Survey, 12100 Beech Forest Road, Laurel, MD 20708-4038, Tel. 301-497-5780, FAX 301-497-5784, Paul_Geissler@usgs.gov

Contents

Introduction 3
Systematic (Grid) Sampling 5
Estimators 7
Characteristics of Systematic Sample Designs 8
Stratified Sampling 10
Allocation Example Involving Accessibility Costs and Habitat 12
Stratified Estimators 13
Characteristics of Stratified Sample Designs 14
Domain Estimation 16
Summary 18
References 18

Introduction

The primary difficulty in designing a long-term, scientifically valid environmental survey is balancing the wide range of competing factors that require consideration and compromise. These factors include the need to produce data and results quickly, the statistical precision of those results, the complexity of the analysis, the need for flexibility when conditions change, the need to estimate both trends and status, the need to let study plots recover from sampling, the need to estimate multiple parameters, the need to stay within budgetary limits, the need to safely access field sites, length of the study, etc. The consequences of failing to or only partially satisfying these factors is difficult to quantify and can leave a study designer paralyzed. Extensive planning is required to design and implement a simple long-term monitoring plan that accomplishes each of these goals, even if only partially.

In this report we present a few fundamental sampling concepts that translate into sample design options available to survey designers. This report was prepared for the National Park Service's Vital Signs monitoring program and has park network needs in mind. Two basic types of spatial sample designs are covered that, in our opinion, are among the best options for long-term monitoring plans. From these recommendations, it is hoped that individual monitoring networks will be equipped to make informed decisions regarding sample site dispersion and frequency of sampling, and ultimately be able to implement high quality monitoring programs. This report targets overall design concepts that are applicable to a wide range of environmental response variables and that are fairly general in nature. As such, questions related to field methods and specific responses will not be addressed.

For simplicity, we focus on the spatial sampling design required to sample a 2-dimensional landscape. One-dimensional and 3-dimensional landscapes, such as rivers or air, are not specifically treated. The temporal sampling design, which dictates *when* sample sites are visited, is vitally important to the success of long-term monitoring programs, but is too lengthy of a topic to discuss here (see McDonald, 2003, and references contained therein). For purposes of presenting the options for spatially sampling a landscape, we think of the monitoring program as lasting only one sample occasion. We realize that all monitoring programs will last multiple occasions, but this view simplifies the presentation significantly.

The 2 types of spatial survey design options we recommend are systematic sampling and stratified sampling. For each of these designs, we describe the first level of estimation techniques, how difficult to reach areas are handled, and other factors. Throughout, we compare and contrast the designs with regard to their strengths and weaknesses on key design factors. We also indicate the conditions under which each design is advantageous.

In choosing the designs to include here, we restricted attention to designs that contained some type of probability sample. Judgment designs (Edwards, 1998; Olsen and Smith, 1999; McDonald, 2003) wherein experts choose the location of study plots to be "representative" were excluded from consideration as being scientifically invalid and never appropriate for long-term monitoring projects. Haphazard designs (Olsen and Smith, 1999; McDonald, 2003) wherein study plots are located without the aid of a formal statistical design, or are located wherever it is convenient, were also excluded as inappropriate. It is our view that some type of probability sample will always be required in order for a monitoring plan to be defensible and statistically valid.

To present the designs, we use a hypothetical National Park (Figure 1) to describe the spatial designs. The hypothetical National Park, acronym HYNP, contains two habitat types, a common habitat inside the large square and a rare habitat inside the ovals. HYNP has an area that cannot be sampled (grey rectangle) because it is too dangerous to access or access has not been granted. It is assumed that this "too dangerous" area was known to study designers prior to implementing the study plan. It should be recognized that statistical inferences cannot be made to this "too dangerous" area and that every effort should be made to keep these areas as small as possible. Locations in the population that are difficult or expensive to access should not be included in the "too dangerous" area because costs associated with access will be considered later as part of the design. Separating "dangerous" and "costly" areas not only allows inference to a larger proportion of the park, reallocation of sampling effort is easier if access costs change but safety concerns do not. Access to all parts of HYNP is from an east-west road near the southern boundary.

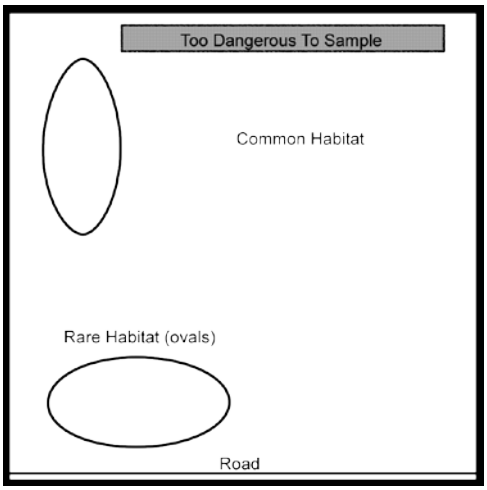

Figure 1: Hypothetical National Park (HYNP) used to illustrate sampling designs.

In what follows, we defined the *population* to be the entire park minus the "too dangerous" areas. We assume that one of the fundamental goals of the monitoring program is to make statistical inferences to the entire population. In addition, we define the sample *frame* to be a concrete representation of the (theoretical) population. Sampling frames are necessary in order to actually draw the sample. For example, population we use here is HYNP (minus dangerous areas), and the frame is a map of HYNP. It is important to realize that the frame may not be a perfect representation of the population. Frame errors occur when the frame includes areas that are not in the population, or when the population includes areas that are not in the frame.

Systematic (Grid) Sampling

Systematic, or grid, samples are the simplest and most straightforward of the designs we recommend. Systematic samples also form the basic building blocks of more complicated stratified designs and therefore deserve considerable attention. In this section, we start by assuming that travel time and costs are equal for all parts of the population, and that, apart from the "too dangerous" area, it is possible to sample a point in all parts of the population. Under this assumption, we describe the basic systematic design and estimators. Later, in the stratified design section, we relax this assumption and consider the case when travel costs are significantly different among sampling points.

Drawing a systematic sample is relatively easy. The key decision necessary to select a systematic sample of points is determining the grid's resolution, or the distance between points. Grid spacing usually determines the smallest possible sampling unit, but more importantly determines the total number of sampled points. Choosing distance between grid points is synonymous with choosing the "sampling intensity", and the study's budget should therefore be considered carefully.

Appropriate grid spacing is often related to the revisit or temporal sampling plan envisioned for the monitoring plan. For example, the temporal sampling plan may dictate that a different set of 6 sites be visited every year for 5 years. Then, in year 6, sampling is to be repeated on the first set of 6; in year 7, the second set of 6 are visited, and so on (i.e., a [1-4] rotating panel design (McDonald, 2003)). In many, but not all, situations it will be useful for estimation to distribute the entire set of $6\times5 = 30$ sites uniformly over the population. One way to do this using a grid sample is to "over-sample" the population by setting the grid spacing 5 times smaller than necessary to achieve 6 points in a single year. In other words, the grid's spacing is set to obtain 30 points. Once drawn, these 30 points are divided into 5 sets of 6 in a systematically fashion (i.e., 1st, 7th, 13th, etc points are visited year 1; 2nd, 8th, 14th, etc. points are visited year 2; and so on). This type of sampling assures uniform coverage of the population during each year and at the end of 5 years.

In many programs, researchers could be prevented from collecting data at a location due to adverse weather, logistics problems, or unforeseen access issues, and it is possible to replace the point with another. In other situations, researcher may have the resources to collect data from more sites than originally propose due to increased funding, increased manpower, or overestimated logistical costs. In both of these situations, it would be convenient to have a list of additional sample sites which can be included in the sample without destroying the statistical properties of the original sample. Such a list of additional "contingency" sites can be constructed by "over-sampling" the population. For example, the grid's spacing could be set to select 2 times the number of sites

actually needed (this may require non-square grid spacing). Researchers would then target every other selected location for visitation. If more sites are needed, or a particular site needs to be replaced, researchers would sample the next nearest point not originally scheduled. If “contingency” points are sampled, the locations of all sampled points will not be a regular grid, but this does not cause any real difficulty.

If, due to weather or unforeseen logistical problems, some grid points will likely not be visited nor replaced, it is desirable to randomize the order in which points are visited as much as possible. Travel costs will increase if the order of visits is randomized, but at the same time randomization will reduce the likelihood of large expanses of the population receiving no sampling effort if sampling is cut short. For example, suppose researchers at HYNP first visit points in the grid closest to the road, then proceed to points farther away from the convenient access afforded by the road (i.e., visit points in the southernmost row first, then the next row to the north, and so on, Figure 2). Suppose further that personnel at HYNP underestimated the amount of time necessary to sample a single point and that a wild fire broke out in the park half way through their field season. Given these events, the field crew at HYNP was able to visit only 7 out of the 16 points that were planned, and all these were in the lower half of the park due to the order of visitation. Although doing so would have cost more, it would have been desirable for the 7 points that were sampled to be dispersed randomly throughout the park. In all sampling efforts, careful thought and realistic assessments of the likelihood of a foreshortened field season are necessary when making decisions about the order of visitation.

For HYNP, we chose an initial systematic sample with a sampling intensity that was appropriate for common habitats in accessible areas. This grid spacing, labeled Δ, yielded 16 data points. Once spacing is selected, the systematic sample was drawn by first choosing two random numbers between 0 and Δ. Let these random numbers be labeled m_1 and m_2. Assuming the origin of the geographic coordinate system is the lower left-hand corner of HYNP, the lower left-most point of the grid should then be placed at the point (m_1, m_2), which effectively shifts the entire grid north and east by some (random) amount. All others points in the grid are placed exactly Δ north and east from their immediate neighbors (dots in Figure 2).

One option for researchers at HYNP is to visit 1/4th of the 16 sampled sites each year. Under this option, sites labeled 1 in Figure 2 would be visited in year 1, sites labeled 2 would be visited in year 2, and so on. In this case, the 16 sample points represent a 400% over-sample of the yearly number of desired sites. Rather than complicate the presentation by assuming a [1-3] revisit plan, we will assume researchers visit all 16 sample points each year (i.e., [1-0] revisit).

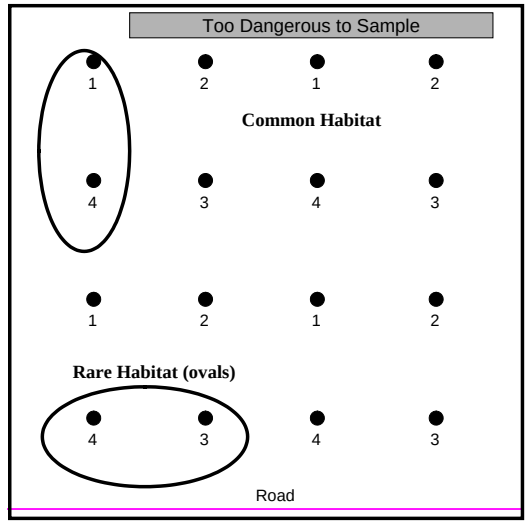

Figure 2: The hypothetic population overlaid with a systematic sample of 16 points. One revisit option calls for sites labeled '1' to be visited year 1, sites labeled '2' to be visited year 2, and so on.

Estimators

Having described selection of a systematic sample, we present formulas for estimating the mean and variance of a measured variable from data collected by systematic sampling. In what follows, we generically refer to the measured variable as a *response.* Examples of responses include biomass, stems per hectare, density of small mammals, number of species detected, temperature, soil type, etc. The key characteristic of responses is that they be measurable at every point in the sample. By no means are these formulas the only ones necessary to carry out a full analysis of long-term monitoring data. We present them hoping that understanding of the design will increase, and because estimation of the mean of a parameter is a common task.

Assume that a single response, y_i, is measured at each location in a systematic sample. An estimate of the mean of parameter y in the population is,

$$\bar{y}=\frac{\sum_{i=1}^{n} y_i}{n}$$

where n is the number of sample points (16 in our example).

The usual estimate of the standard error of $\bar{y}$ is,

$$se(\bar{y})=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n(n-1)}}\ . \qquad (1)$$

Approximate 95% confidence intervals for the true mean of *y* are usually constructed as $\bar{y} \pm 1.96se(\bar{y})$.

Characteristics of Systematic Sample Designs

Now that systematic samples have been described, we discuss some of desirable and undesirable characteristics of a systematic design. To do this, we must compare certain characteristics (like precision) of systematic sampling to another type of design--simple random sampling. We would not generally recommend simple random sampling for long-term monitoring, but it is a standard design with which many people are familiar and it provides a good "baseline" for comparison. Simple random designs independently and randomly locate points within the study area. Figure 3 contains one realization of a simple random sample in HYNP. Note the relatively large areas in HYNP that are not sampled, and the relatively small areas with multiple points. This "clumping" of sample locations is a general characteristic of simple random samples when sample sizes are small to moderate, and is the primary reason simple random samples are not recommended. If sample sizes are large enough, voids will generally not be an issue, but clumping (i.e., oversampling of certain areas) will likely remain as an inefficient characteristic.

If points close together in space tend to have similar responses (i.e., positive spatial autocorrelation exists), a systematic sample is much more precise than a simple random sample for estimating the mean of a parameter. However, to utilize this gain in precision, non-standard and complicated variance estimators must be used to accurately estimate the smaller variance of the mean estimator. Typically, the simple variance estimator appropriate for simple random sampling (Equation (1)) is applied to data collected by systematic samples, and this estimator is known to overestimate the variance of a mean when data are collected using systematic sampling and responses are positively correlated in space (Lohr, 1999:60). In many settings, the positive bias of variance estimators is negligible, and overestimated variances are acceptable. In some settings, this overestimate of variance can mask or under represent the increased precision afforded by the systematic sample. As a result, variance reduction is not the most compelling argument in favor of a systematic sample.

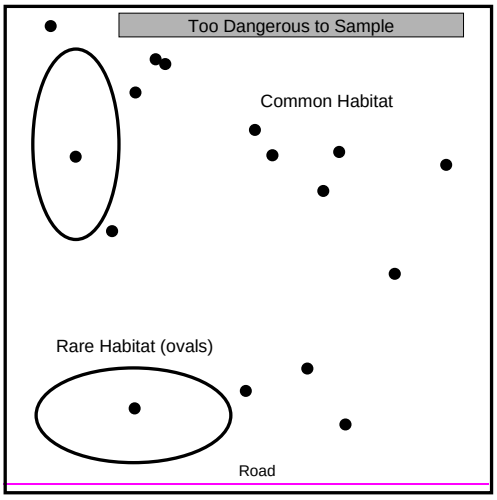

Figure 2: Map showing locations of 16 points in HYNP chosen by simple random sampling.

The most compelling argument for systematic samples is that they spread the sample over the population, thus insuring that all sections and sub-regions of the population are uniformly covered. Uniform coverage is advantageous for a number of reasons. First, it reduces the variance of location estimates and assures that the design is unbiased, even if those gains in precision are not realistically estimated (see previous paragraph). Second, many types of map-drawing and surface estimation exercises are easier and enhanced when data are collected with uniform density due to the grid's efficient description of spatial patterns (Cole et al. 2001). Third, the grid designs are suitable for investigating spatial patterns at a range of scales. Fourth, post-stratification of the sample points is simplified and enhanced because proportional allocation is assured in all possible subsets of the population. Proportional allocation means that areas of the population are sampled in proportion to their size. Finally, if GPS is not used to locate points in the field, another minor advantage of the grid sample is that sample site location is usually easier if sites are laid out in a regular pattern.

While proportional allocation is an advantage for some purposes, proportional allocation is also the prime disadvantage of (unstratified) systematic and simple random samples. Due to proportional allocation, grid and simple random samples frequently miss or under sample rare habitats. Riparian areas, for example, are difficult to sample with a grid because they typically occupy very small areas. If it is important to monitor these areas, simple grid sampling will probably not be the best option.

Another disadvantage of (unstratified) systematic designs is that they do not consider the differing costs associated with traveling to all points in the sample. Further, these samples also do not treat the "too dangerous" areas differently than any other area, and consequently sample locations that fall in these areas have to be discarded. Realized sample size is a random variable in this case, and planning may be more complicated.

In practice, questions will arise regarding grid sample points that fall so close to the park's boundary that field methods call for collection of data from areas outside the park. For example, if field methods require that a vegetation plot be placed 500m north of the sample point, all sample points within 500m of the

northern boundary of the park will result in vegetation plots outside the park. At this point, questions will arise as to whether or not an equivalent number of points should be added to the sample in the interior of the park. In practice, points close to the boundary are usually moved away from the border a distance sufficient for field methods to remain in the park. Technically, moving sample sites away from study area boundaries changes the inclusion probability of points close to the border, but in practice these changes are small and usually ignored.

Stratified Sampling

Stratified sampling (Lohr, 1999, p. 95-118) is the second type of design recommended for long-term monitoring. Stratified sampling is recommended provided 1 of 2 conditions is satisfied. The first condition under which statification is recommended is when estimates of parameters are desired for each strata during the life of the monitoring project. For example, strata 1 could be defined as a subsection of HYNP that is vitally important to management, and separate or high-precision estimates for this area are needed during the life of the program. If relatively high precision estimates are needed in areas that are considered rare habitats, and the location of the rare habitats is known before the sample is drawn, stratification can be very useful because it will guarentee a certain number of samples are collected in the rare habitat areas.

The second condition under which stratitication is recommended occurs when strata are defined by travel costs that will not change in the forseeable future. In this case, we assume the following 3 conditions are met: (1) sampling costs vary markedly between two or more areas (such as “accessible” and “inaccessible” strata), (2) costs increase linearly with the number of locations visited, and (3) an *a priori* estimate or model for variance in each stratum is available. Stratification in this case will allocate overall sampling effort into the expensive and inexpensive strata, and the resulting mean estimator will theoretically have the lowest possible variance per unit cost. We emphasize, however, that variance per unit cost will be minimized for a single response variable only. If other studies or responses utilize the same stratification scheme, the allocation procedure outlined below will not yield optimized variances unless the responses in question are highly correlated with the response used in the allocation procedure. We also emphasize that at least 2 points must be selected in all areas, including “difficult-to-access” areas, to provide unbiased population-wide estimates. We caution that if travel technologies or costs change, stratification based on travel costs will make adaptation of the spatial sample design complicated and ackward.

Stratification based on a biological auxillary variable for the sole purpose of reducing variance is not recommended for monitoring programs. Stratification simply to reduce variance is not recommended for 3 reasons. First, even though, in theory, stratification has tremendous potential to reduce variance, in practice strata are rarely known well enough *a priori* to actually achieve a substantial

reduction. Furthermore, well-intended stratification can actually inflate variance if strata are not defined appropriately. Second, even if statification reduces the variance of estimates for 1 parameter, it is unlikely to reduce the variance of other estimators. Third, it is extremely complicated and ackward to redefine strata boundaries or post-stratify if the biological auxillary variable changes. Monitoring changes is one of the fundamental objectives of most programs, and it is counterproductive to design a survey that has difficulty adapting to change when they occur.

In this section, we illustrate the stratified design by stratifying HYNP based on both rare habitats and access. It would have been sufficient for our purposes to stratify on only one attribute, but we preferred the more realistic situation. The key topic here is allocation of sampling effort, but we also provide the stratified estimators for a mean. If collection of at least a few samples in rare habitat is the motivation for stratification, and costs are not an issue or are relatively equal among strata, allocation of sampling effort to rare and non-rare habitat strata will be driven by project objectives only. For example, if at least 5 samples are needed from rare habitat areas and costs are not an issue, at least 5 samples should be allocated to the rare habitat stratum, with the remaining allocated to non-rare habitat. In this case, the subsequent discussion and formula for optimal allocation of sample effort does not apply and can be ignored.

Allocation of total sample size among strata focuses on one response and its variance within each stratum. To minimize costs per unit of variance for population-wide estimates, the number of sample points in a stratum should be set to,

$$n_h \propto \frac{N_h S_h}{\sqrt{c_h}} \tag{2}$$

where n_h is sample size in stratum h, N_h is the size of the stratum, S_h is the standard deviation of responses in the strata (i.e., standard deviation of y_i in strata h), and c_h is the cost of sampling a single point in stratum h (Lohr, 1999, p. 106-113). Here, the size of stratum h is the number of sampling units in stratum h, not necessarily the area (hectares) of stratum h. If points are the units being sampled, there are an infinite number of sampling units in each strata and N_h should be set to the area of stratum h. Total sample size is $n = \Sigma n_h$, and assuming initial startup (fixed) costs of c_0, cost of the entire sample is $c_0 + \Sigma n_h c_h$.

The most difficult part of this approach is often estimating the variance of responses in each strata (i.e., estimating S_h). Fortunately, in many situations at least an initial allocation of effort can be made by assuming a model for the S_h. If, for example, it is reasonable to assume that standard deviations are similar in all strata, S_h can be dropped from (2) and the number of sample points in each strata can be set to,

$$n_h \propto \frac{N_h}{\sqrt{c_h}}. \tag{3}$$

In another situation, where counts are collected at each sample point, it may be reasonable to assume that the variance of the counts is proportional to their mean. In this case, the square root of the expected count can be substituted for S_h in (2). Assuming $\bar{y}_h$ is a guess or estimate of the mean count in stratum h, sample size in stratum h can be set as,

$$n_h \propto \frac{N_h \sqrt{\bar{y}_h}}{\sqrt{c_h}}.$$

Once total sample size is allocated, separate and independent samples must be drawn from each stratum. At this point, the researcher must again choose between systematic and simple random samples. For the same reasons as above, we recommend that systematic samples be drawn from each stratum to reduce variance of the strata mean estimates, even if that reduction is not accounted for in the estimates. Note that it is not necessary to maintain the same grid spacing in all strata. Note also that the sample from each stratum should be randomized separately by drawing new random starting coordinates each time.

Allocation Example Involving Accessibility Costs and Habitat

In this section, allocation in and selection of a stratified systematic sample is illustrated in an example. In the example, 3 strata are defined on HYNP. The first stratum consists of common habitat in accessible areas. The second stratum consists of common habitat in inaccessible areas. The third stratum consists of areas considered rare habitats (ovals) (Figure 4). We assume stratums 1 and 2 are 6,000 hectares in size, while stratum 3 is 4,000 hectares. We also assume the cost of visiting a point in stratum 2 is 16 times the cost of visiting a point in stratum 1, and that the costs of visiting a point in stratum 3 are 4 times the costs in stratum 1.

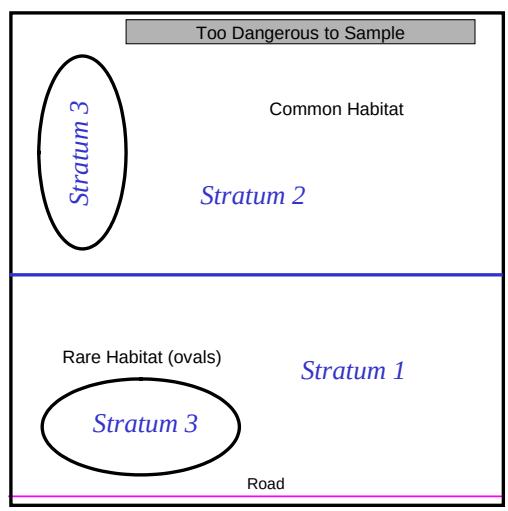

Figure 4: HYNP stratified into 3 strata based on access and rare habitats. Strata 1 is common habitat in easily accessible terrain. Strata 2 is common habitat in difficult to access terrain. Strata 3 is rare habitat whose location is known.

Assume that a target of 16 sample points is to be allocated among the 3 strata defined on HYNP. The minimum sample size allowed in any

strata is 2, because 2 points are required to estimate variance. Assuming variances are similar in all strata and utilizing (3), the number of locations to sample in each strata is,

$$n_h = round\left[n \left(\frac{N_h / \sqrt{c_h}}{\sum_{i=1}^{n} N_h / \sqrt{c_h}} \right) \right],$$

where $round[x]$ indicates that x should be rounded to the nearest integer, provided that the integer is ≥ 2. If $n_h < 2$ after rounding, n_h should be set to 2 and the sample size in other strata should be adjusted downward accordingly. Appling this equation to the example (Table 1), we find that $n_1 = 10$, $n_2 = 3$, and $n_3 = 3$ is the allocation of 16 sample locations that minimizes variance per unit cost. Total cost of this sample is 10(1) + 3(16) + 3(4) = 70 units.

Table 1: Calculations necessary to allocate 16 sample locations to 3 strata in HYNP, assuming the variance of responses is the similar in all strata.

Stratum	Area (N_h)	Relative cost (c_h)	$\sqrt{c_h}$	$N_h / \sqrt{c_h}$	Sample Size (n_h)
1—Accessible Common Habitat	6	1	1	6	10
2—Less Accessible Common Habitat	6	16	4	1.5	3
3—Rare Habitat	4	4	2	2	3
Sum				9.5	16

Stratified Estimators

Assuming that a single response, y_{hi}, is measured at each location in a stratified sample, an estimate of the mean of parameter y in stratum h is,

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}.$$

An estimate of the mean of y in the entire population is,

$$\bar{y} = \frac{\sum_{h=1}^{H} N_h \bar{y}_h}{N} = \frac{\sum_{h=1}^{H} \sum_{i=1}^{n_h} \frac{N_h}{n_h} y_{hi}}{N}$$

where H is the number of strata, and $N = \sum_h N_h$. The estimated variance of $\bar{y}$ is,

$$se(\bar{y}) = \sqrt{\frac{1}{N^2}\sum_{h=1}^{H}\sum_{i=1}^{n_h}\frac{N_h(y_{hi}-\bar{y}_h)^2}{n_h(n_h-1)}}$$

Given these equations, it is interesting to compare the standard error of $\bar{y}$ and costs under proportional allocation and the allocation scheme in (3). Under proportional allocation, $n_1 = n(N_h)/N$ = 16(6)/16 = 6 sample points would be allocated to stratum 1, n_2 = 16(6)/16 = 6 sample units would be allocated to stratum 2, and n_3 = 16(4)/16 = 4 sample units would be allocated to stratum 3. Assuming the variance of y was estimated to be s^2 in all stratum of HYNP, the standard error of $\bar{y}$ under proportional allocation is,

$$se_p(\bar{y}) = \sqrt{\frac{1}{16^2}\sum_{h=1}^{3}\frac{N_h s^2}{n_h}} = \sqrt{\frac{1}{16^2}\left(\frac{6}{6}+\frac{6}{6}+\frac{4}{4}\right)}s = 0.1083s\,.$$

The standard error of $\bar{y}$ under the allocation scheme dictated by (3) (i.e., "optimum" allocation) is,

$$se_o(\bar{y}) = \sqrt{\frac{1}{16^2}\sum_{h=1}^{3}\frac{N_h s^2}{n_h}} = \sqrt{\frac{1}{16^2}\left(\frac{6}{10}+\frac{6}{3}+\frac{4}{3}\right)}s = 0.1240s\,.$$

The survey under proportional allocation cost a total of c_p = 1(6) + 16(6) + 4(4) = 118 units. The survey under optimum allocation cost a total c_o = 1(10) + 16(3) + 4(3) = 70 units. Cost per unit of standard error under proportional allocation is $c_p/se_p(\bar{y})$ = 1090 / s, while cost per unit of standard error under optimum allocation is $c_o/se_o(\bar{y})$ = 565 / s. This exercise illustrates that the optimum allocation scheme in (3) does not yield the minimum standard error, but does yield the minimum cost per unit of variance. In other words, the same standard error is achievable with less money if the optimum allocation is chosen over proportional allocation.

Characteristics of Stratified Sample Designs

The primary advantages of stratified sampling have already been listed. The first advantage was that stratified sampling has the ability to guarantee samples from certain regions or locations. This is a major advantage if estimates are absolutely required for these areas. The second advantage is that if the cost of sampling points varies markedly between strata, stratified sampling and optimum allocation can be used to minimize the amount of money spent to achieve a particular precision level. This second advantage is achievable only if specific assumptions regarding costs and variances are correct.

One of the primary drawbacks of stratified sampling arises when and if survey designers change strata boundaries. Some reasons that designers might want to change strata boundaries include: (1) rare habitats are not actually present in the "rare habitat" strata, (2) the rare habitat in the "rare habitat" strata dies or moves, (3) access costs are not as anticipated, and (4) access costs change. Once drawn, the strata definitions and boundaries must remain fixed forever.

Changing the strata definitions results in an entirely new survey, and it is complicated to compare parameters before and after the strata changes. Such comparisons will likely involve unequal probability analyses, and should be conducted by a qualified statistician. For this reason, we recommended defining strata based on unchanging features and not a vegetation map, which is likely to change. Even though we recommend stratification based on access or rare habitat, designers must think carefully about this option because access costs may change in the future, and rare species might not be present or might emigrate from a certain stratum.

Another difficulty with stratification is perceptual, but nonetheless results in pressure to either change strata membership or boundaries. In reality, strata are artificial constructs used to control and distribute sampling and may include a mix of habitat types despite efforts to the contrary. However, because the same name is often given to both the habitat type and the stratum, it is easy to perceive habitat and strata as synonymous. This perception can result in a misguided desire to change stratum membership of sample units if habitat is not in the strata where it was expected. If habitat changes, or if sample units in certain strata do not share a characteristic with other units in the same stratum, the stratum membership of the affected units cannot be changed. Doing so will bias estimates because similar "corrections" cannot be made for the points that were not in the sample. This means that if researchers define a stratum to include rare habitat, but upon arrival at a sample point in this stratum, the point is determined to be in common habitat, the point must not be changed or relabeled into the common habitat stratum. Estimation domains (see the next section) should be used to make estimates for habitat types, whenever habitats do not completely match stratum boundaries.

Another disadvantage of stratified sampling is that stratum usually cannot be defined to improve estimation of all parameters. For example, stratification appropriate for the vegetation component of monitoring is probably not appropriate for the large undulate or glacier monitoring components. If strata are based on access, different components of the overall monitoring project will likely have different access issues. For example, field work for the small mammal monitoring component may require more equipment than the bird monitoring component, and this fact may restrict the areas accessible to the small mammal component relative to the bird monitoring program. In these cases, basing a single stratification scheme on access will be at a minimum difficult, and at a maximum, counterproductive for certain components.

If stratification is employed, we caution researchers against what we call over-stratification. Over-stratification occurs when more than 3 to 5 strata are defined. More than 3 to 5 strata are a problem for a couple reasons. During analysis, 1 degree of freedom is lost for each stratum, effectively reducing total sample size by 1 unit for each stratum. The potential for stratification “errors” (i.e., miss-classification) is also higher when many strata are defined. When many “errors”

occur, pressure to change strata boundaries increases and the analysis is continually complicated. In most situations researchers do not know enough about responses to effectively stratify in a way that increases precision. If pilot data are available, it may be possible to investigate different stratification schemes and their effects on both the complexity and precision of results.

Domain Estimation

Domains are areas like strata where estimates are needed. Domains are typically defined after the sample is taken, and are sometimes called *post-stratification.* In the context of this paper, we envision domains as "strata" that overlap one or more of the original stratum used to distribute the sampling effort. In HYNP, one domain might consist of the southern half of the park, including all of strata 1 and one oval of stratum 3. One difference between strata and domains is that strata are the fixed areas we used to select sample points, while domains are habitat types (or areas) observed in the field. In this section, we describe domain estimation, and then work a simple example involving HYNP.

In domain estimation, the number of points in a particular habitat type (domain) is a random variable. To write the equations, we introduce the indicator variable, I_{dhi} = 1 if sample unit (point) *i,* originally in strata *h*, is in domain *d*, I_{dhi} = 0 if unit *i* of strata *h* is not in domain *d.* Under this definition, $n_{dh} = \Sigma_i I_{dhi}$ is the number of sample units in strata *h* that are in domain *d.* An estimate of the mean in domain *d* from stratum *h* is,

$$\bar{y}_{dh} = \frac{\sum_{i=1}^{n_h} I_{dhi} y_{hi}}{n_{dh}}.$$

The overall population mean estimate for domain *d* is,

$$\bar{y}_d = \frac{\sum_{h=1}^{H} \frac{N_h n_{dh}}{n_h} \bar{y}_{dh}}{\sum_{h=1}^{H} \frac{N_h n_{dh}}{n_h}} = \frac{t_d}{N_d},$$

where t_d is the estimated total of response *y* in the domain, and N_d is the estimated total number of sample units in domain *d*. Assuming independent simple random samples in each stratum, the estimated standard error of t_d is,

$$se(t_d) = \sqrt{\sum_{h=1}^{H} N_h^2 \frac{(n_{dh}-1)s_{dh}^2 + n_{dh}\left(1 - \frac{n_{dh}}{n_h}\right)\bar{y}_{dh}^2}{n_h(n_h-1)}}$$

were

$$s_{dh}^2 = \frac{\sum_{i=1}^{n_h} I_{dhi}(y_{hi} - \bar{y}_{dh})^2}{n_{dh}-1}$$

(Särndal et al., 1992:392). The estimated standard error of N_d, the total number of units in domain *d,* is,

$$se(N_d) = \sqrt{\sum_{h=1}^{H} N_h^2 \frac{n_{dh}(n_h - n_{dh})}{n_h^2(n_h - 1)}} .$$

The estimated standard error of $\bar{y}_d$ is,

$$se(\bar{y}_d) = \sqrt{\frac{1}{N_d^2}\sum_{h=1}^{H} N_h^2 \left(\frac{(n_{dh}-1)s_{dh}^2 + n_{dh}\left(1 - \frac{n_{dh}}{n_h}\right)(\bar{y}_{dh} - \bar{y}_d)^2}{n_d(n_d - 1)} \right)}$$

(Särndal et al., 1992:394).

Suppose that strata were defined as before (Figure 4), and that the sample allocation resulted in the points displayed in Figure 2 (realized points in Figure 5). This non-optimal allocation scheme resulted in n_1 = 6, n_2 = 6, and n_3 = 4. Suppose further that 2 points in Stratum 1 of HYNP were in rare habitat instead of common habitat as expected (Figure 5). After discovering these additional points, we wish to make an estimate for "rare habitat" that obviously does not agree with the boundaries of Stratum 3. For the new rare habitat domain, n_{d1} = 2, n_{d2} = = 0, and n_{d3} = 4. Assuming the mean response in Stratum 3 (the original rare habitat stratum) is $\bar{y}_{d3}$ = 20, and the mean of the response from the 2 additional rare habitat points in Stratum 1 is $\bar{y}_{d1}$ = 25, the mean estimate for rare habitat in the entire park is,

$$\bar{y}_d = \frac{\frac{6(2)}{6}(25) + 0 + \frac{4(4)}{4}(20)}{\frac{6(2)}{6} + 0 + \frac{4(4)}{4}} = \frac{130}{6} = 21.67 .$$

Assuming the estimated variance of responses in Stratum 3 is s_{d3}^2 = 4, and that s_{d1}^2 = 4, the estimated standard error of $\bar{y}_d$ is,

$$se(\bar{y}_d) = \sqrt{\frac{1}{6^2}\left(6^2\left[\frac{(2-1)4 + 2(1-2/6)(25-21.67)^2}{6(6-1)}\right] + 4^2\left[\frac{(4-1)4 + 4(1-4/4)(20-21.67)^2}{4(4-1)}\right] \right)}$$

$$= \sqrt{\frac{22.5778 + 16}{36}} = 1.0352$$

Summary

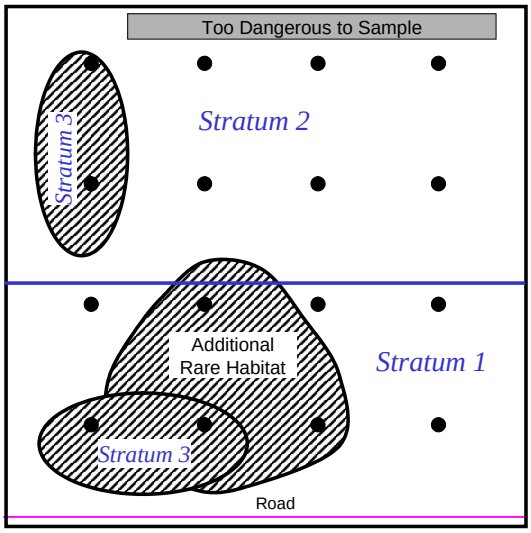

Figure 5: Map of HYNP showing 3 strata, 16 sample points, and a domain of estimation (hatched areas) constructed after 2 points in stratum 1 were discovered to be in rare habitat.

Two relatively simple spatial sampling designs that have high utility for long-term monitoring were described, along with the characteristics of each. Systematic, or grid, sampling was found to be relatively simple to implement and analyze, it provided uniform coverage of the entire population, and it was relatively easy to make estimates for domains of interest (post-stratification). All of these characteristics make (unstratified) systematic sampling a good general purpose design, offering reasonable precision for nearly all responses. The primary drawback of systematic sampling was its inability to adequately sample rare habitats unless sample sizes were very large. Stratified sampling, which utilizes either a systematic or simple random sample in each stratum, was found to be slightly more complex to design and analyze, could guarantee collection of data in certain habitats, and could be optimized to account for travel costs and thus minimize dollars spent per unit of standard error. These characteristics make stratified sampling a highly efficient design for specific responses, but not for all responses at once. Disadvantages of stratified sampling include complexity if domains of estimation span several strata, the potential for non-proportional allocation in certain domains, and an inability of change strata boundaries.

References

Cole, R.G., T.R. Healy, M.L. Wood, and D.M. Foster. 2001. Statistical analysis of spatial pattern: a comparison of grid and hierarchical sampling approaches. *Environmental Monitoring and Assessment* 69:85-99.

Edwards, D. 1998. Issues and themes for natural resources trend and change detection. *Ecological Applications.* 8, 323-325.

Lohr, S.L. 1999. Sampling: Design and Analysis. Duxbury Press.

McDonald, T. L. 2003. Review of environmental monitoring methods: survey designs. *Environmental Monitoring and Assessment.* 85, 277-292.

Olsen, A. R., and Smith, E. P. 1999. Introduction to the special issue on surveys over time. *Journal of Agricultural, Biological, and Ecological Statistics.* 4, 328-330.

Särndal, C. E., B. Swensson, and J. Wretman. 1992. Model Assisted Survey Sampling. Springer-Verlag.