USGS
science for a changing world

Sampling Large Landscapes with Small-Scale Stratification—User's Manual

Open-File Report 2011–1247

Sampling Large Landscapes with Small-Scale Stratification—User's Manual

By Jonathan Bart

U.S. Department of the Interior
KEN SALAZAR, Secretary

U.S. Geological Survey
Marcia K. McNutt, Director

U.S. Geological Survey, Reston, Virginia: 2011

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment, visit http://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit http://www.usgs.gov/pubprod

To order this and other USGS information products, visit http://store.usgs.gov

Contents

Figures

Tables

This page left intentionally blank

Sampling Large Landscapes with Small-Scale Stratification—User's Manual

By Jonathan Bart

Introduction

This manual explains procedures for partitioning a large landscape into plots, assigning the plots to strata, and selecting plots in each stratum to be surveyed. These steps are referred to as the "sampling large landscapes (SLL) process." We assume that users of the manual have a moderate knowledge of ArcGIS and Microsoft® Excel. The manual is written for a single user but in many cases, some steps will be carried out by a biologist designing the survey and some steps will be carried out by a quantitative assistant. Thus, the manual essentially may be passed back and forth between these users. The SLL process primarily has been used to survey birds, and we refer to birds as subjects of the counts. The process, however, could be used to count any objects.

Steps in the Sampling Large Landscapes Process

The process is divided into seven steps, one of which (step 5) is optional (table 1). Each step is described in this manual followed by a detailed example.

Table 1. Steps in the sampling large landscapes process.

Step 1. Define populations
Step 2. Define regions and habitats
Step 3. Make a plots shapefile
Step 4. Assign plots to strata
Step 5. Define clusters or clusters and zones (optional)
Step 6. Choose sample sizes within strata
Step 7. Select plots to survey

Step 1.—Define Populations

"Populations" in this context refers to areas, not birds or other objects being counted. We distinguish three populations: the study area, the inferential population, and the sampled population (table 2). Each population is delineated in a shapefile. The study area is simply the outer boundary of the area being studied. Statistical inferences are not made outside of this area. The "inferential population" is the population that will be subdivided into plots and to which statistical inferences will be made. The inferential population often excludes areas that are completely unsuitable, such as open water when the objective is to record nests of landbirds. The sampled population is the plots that have positive probability of being included in the survey.

Table 2. Three spatial populations.

Population	Description
Study area	Outer border; no surveys or statistical inferences made outside this area.
Inferential	Area to which statistical inferences will be made; subdivided into plots.
Sampled	Plots with positive probability of being surveyed.

Unsuitable areas could be subdivided into plots and we could then assert on the basis of biological reasoning (for example, landbirds do not nest in open water) that all plots in such areas have zero birds present. If this assertion is not made, then the unsuitable areas have to be subdivided into plots and assigned to strata, and the strata have to be included in tables, and so on, even though no surveys are actually made in these areas.

In many studies, some parts of the inferential population are inaccessible due to legal constraints or logistic issues but are similar enough to areas that can be surveyed, that investigators want to include inaccessible areas in the inferential population and extrapolate findings from surveyed areas to these inaccessible areas. Such areas are therefore included in the inferential population but are excluded from the sampled population. We discuss how statistical inferences are made to inaccessible areas in Step 4—Assign Plots to Strata, but, in brief, inaccessible areas are assigned to strata in which some parts are surveyed. Stratum means (for example, densities) are then calculated and the means are assumed to be the same in the part of the strata that was excluded from the sampled population. For example, in a study of the Lower Colorado River, some areas were considered unsafe to survey, yet our habitat maps included the entire area so we could assign unsafe areas to habitats and strata. We therefore placed the unsafe areas in similar strata that we could survey. In estimating densities and population totals, we assumed that the areas we could survey were similar, on average, to the unsafe areas.

It is often tempting to exclude some areas from the sampled population because these areas are expected to have few birds. An alternative, however, is to define these areas as strata and then sample them with low intensity. This approach avoids the assumption that few birds are present yet little effort is consumed in surveying these areas.

If non-surveyable areas may be quite different from the most similar areas that are surveyed, then it is best to exclude them completely. In the Lower Colorado River study, a portion of the Virgin River, which flows into Lake Meade from the north, was inside the study area. This area was difficult to survey and was eventually excluded from the sampled population. We considered including this area in the inferential population but the closest similar habitat was more than 100 mi away and we did not know if populations were similar in these two areas. Therefore, a decision was made to exclude the Virgin River from the inferential population and to acknowledge that we were not able to make statistical inferences to that part of the study area.

Once decisions have been made on how to define the study area, inferential population, and sampled population, shapefiles are prepared for each population. Areas outside the inferential population must be completely outside the polygons depicting the inferential population. For example, if open water is to be excluded then there should not be any open water polygons in the shapefile. The reason for this is that a fishnet of plot-sized cells will be created and a union of it and the population will then be created as a way to begin subdividing the inferential population into plots. If excluded areas like open water occur in the shapefile of the inferential population, then the union may include polygons completely covered by excluded areas.

This step thus requires the following actions:

1. Prepare a shapefile that depicts the outer boundary of the study area. The shapefile normally has just one polygon but will have more than one polygon if the study area is disjunctive.

2. Prepare a shapefile that depicts the inferential population and thus excludes areas within the study area that are considered totally unsuitable for birds or are inaccessible and potentially quite different from surveyed areas.

3. Prepare a shapefile that depicts the sampled population, defined as all areas that have positive probability of being surveyed. The shapefile may be identical to, or smaller than, the inferential population but never extends outside the inferential population.

Step 2.—Define Regions and Habitats

The approach used here, and implemented in other software we have produced such as the analysis program DS, assumes that strata will be created and will be defined using one or two variables. In most studies, we have encountered, the variables are region and habitat so we use these two terms. Any two variables, however, may be used to define strata. For example, Region could be defined solely on the basis of access (for example, good, medium, poor) and Habitat could be based solely on importance from a management perspective (for example, high, medium, low). Defining only one region and/or one habitat also is permissible.

The first task in this step is thus to decide how the terms region and habitat will be defined. An advantage of forming many strata is that one can easily obtain stratum-specific estimates using DS or other analytic programs. Two plots must be surveyed in each stratum, however, so that stratum-specific variances can be calculated. The number of strata thus cannot exceed one-half the number of plots. In addition, when many strata are defined in poor habitat then too much time may be spent in these habitats. Furthermore, having dozens or hundreds of strata, even if the sample size is large enough to make this possible, is usually cumbersome especially if the sample sizes in most strata are so small that the within-stratum estimates are not very useful (due to poor precision). In our experience, Fewer than about 30 strata is usually best for programs that cover an area the size of a Western State.

Defining regions and habitats is usually an iterative process and the first attempt often results in too many strata. Some of the possible combinations may not exist on the ground. For example, there might be 10 regions numbered 1 to 10 and three habitats (good, medium, and poor) but region 7 might not have any good habitat. If this occurs, then the process described below will automatically exclude the stratum "7-good." Nonetheless, it is quite common to define far too many strata. For example, in one study, regions were first defined by Bird Conservation Region, elevation category, and ownership and habitat was defined using several categories. We thus ended up with hundreds of permutations for a study with 180 plots. We used GIS methods to determine how much of the study area would be assigned to each stratum and then combined small strata. This process resulted in 20 strata.

Once decisions have been made on the values for region and habitat, a shapefile with the habitats should be prepared. It is not necessary, at this stage, to make a shapefile for regions.

This step thus requires the following actions:

a. Decide how regions will be defined.

b. Decide how habitats will be defined.

c. Assess whether the number of strata (region-habitat combinations) is too high or too low.

d. Repeat steps "a" through "c" as needed.

e. Make a habitats shapefile with the reduced number of habitats.

Step 3.—Make a Plots Shapefile

Plots are made by subdividing the inferential population. This can be done in two ways depending on how regions are defined. If regions are defined using strict borders, for example legal property lines, then it is usually best that plots do not cross these lines. In this case, the first step in making the plots shapefile is to subdivide the inferential population layer using a shapefile of the region boundaries. However, if the regions are broad ecological zones, then use the undivided inferential population layer, define the plots, and then assign each plot to a region based on the location of the plot's centroid. With this approach, the regions shapefile thus becomes the set of plots assigned to each region. This results in a regions map with "stair-step" boundaries but the plots have north-south or east-west borders that are often much easier to follow in the field than borders that follow other bearings. The first task in this step is to decide whether to divide the inferential shapefile using the regions layer. We use "inferential layer" below to refer to the inferential population regardless of whether it has been subdivided using the regions shapefile.

It is highly desirable that all plots be fully acceptable to surveyors. A sample will be selected and surveyed, and the density on surveyed plots in each stratum will be assumed to equal the density throughout the stratum apart from the effects of sampling error. Consequently, if a sample of plots is selected but the borders are changed for logistical reasons, then concern will arise that the surveyed plots may not actually be representative of the strata they are in.

Delineating the plots is often a complex and time-consuming process. We have tried many different approaches and work in this area continues. Typically, the goals are (a) to assign good habitat to as few plots as possible so that surveys can be concentrated there, and (b) to produce plots that are feasible to survey. We illustrate goal (a) with a study of shorebirds in northern Alaska. The good habitat occurred around wetlands that covered a small portion of the landscape and were separated by habitat with very few birds. Many of these wetlands were small enough to survey completely. The birds would often flush when the surveyor was far away and would then circle overhead while the surveyor conducted the survey. If the plot border bisected the wetland, determining which birds were nesting inside the plot was difficult. Detecting that the bird was nesting somewhere within the wetland, however, was often quite easy. It was thus highly desirable that plot borders followed wetland borders and a substantial amount of time was expended developing GIS methods that would accomplish this. With respect to goal (b), the plots need to be of reasonable size (neither too small nor too large), the plots should not occur on both sides of uncrossable boundaries such as irrigation channels, and borders should be distinguishable and easy to follow in the field.

The first step in delineating plots is to impose a grid of cells of the average plot size and shape on the inferential population. This can be done easily using the fish net command in X Tools or using ArcGIS commands. The union of the plots and inferential population is then created producing a first "plots shapefile." Columns should be added to the attribute table labeled *Region*, *Habitat*, *Stratum*, and *Surveyed*.

The borders of the inferential population are nearly always irregular whereas the borders of the fishnet cells are straight lines oriented north-south and east-west. When the two layers are intersected, the result includes many polygons that are too small to survey (fig. 1). The first step in producing a final plots layer is thus to absorb these plots into surrounding larger plots subject to the restriction that if regional boundaries are included in the inferential population shapefile, then plots cannot be formed that cross region borders.
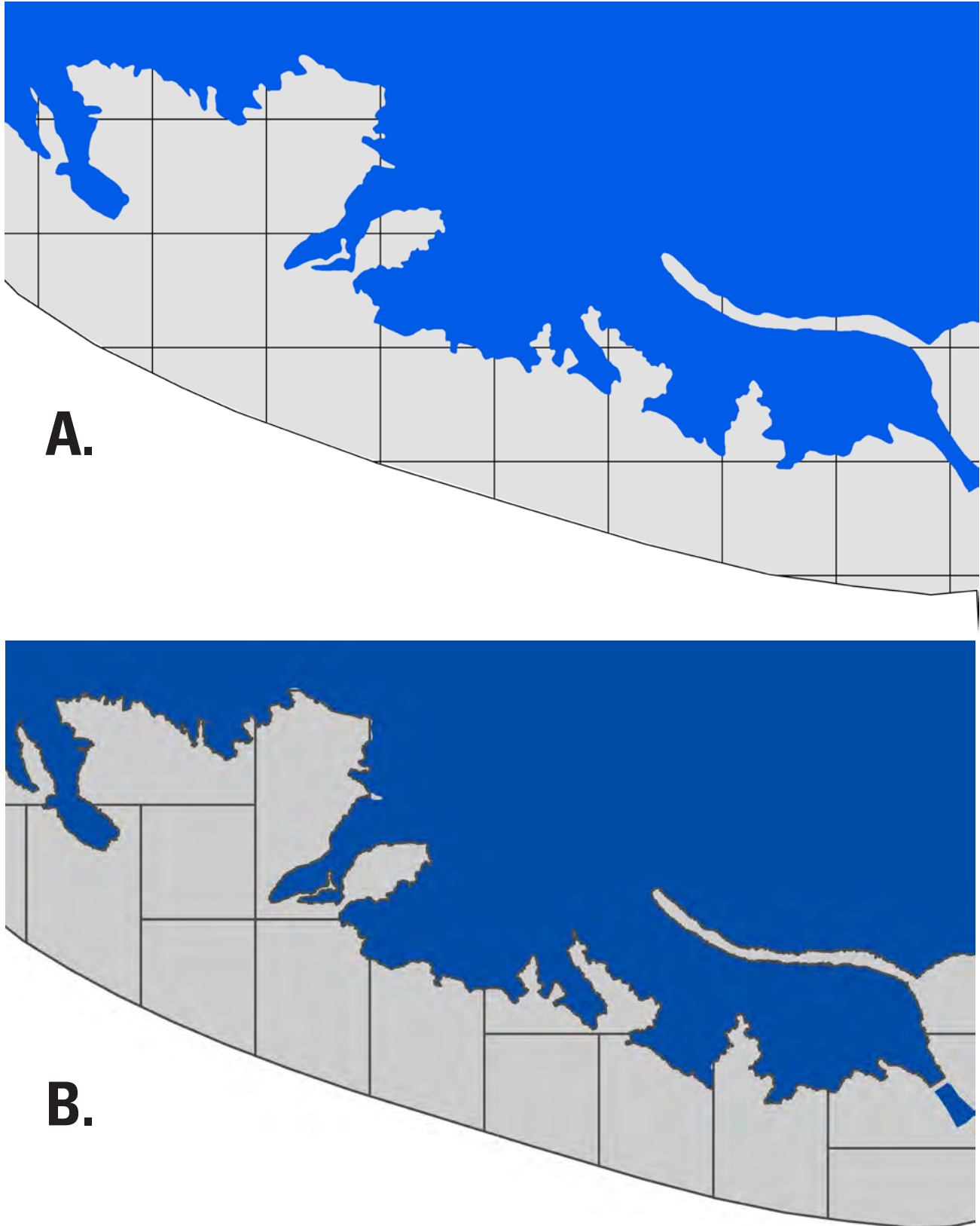
Figure 1. Example of (*A*) initial superposition of square cells on a study area and (*B*) results after merging small polygons with surrounding larger ones.

In some studies, certain habitats (usually the best ones) need to be concentrated into as few plots as possible so that survey efforts can be concentrated in this habitat. One example of this, for wetlands in Alaska, was described above. Another example occurred in the Lower Colorado River study. The best habitat ("tall cottonwood-willow") had the great majority of the birds the survey was designed to detect, yet over much of the study area these habitats occurred in small patches that covered a small fraction of the landscape (<1 percent). We therefore wanted the patches of good habitat to be within plots rather than having them bisected by plot borders (fig. 2).

Many procedures can be imagined for consolidating good habitat into the smallest possible number of plots and we must rely, in large measure, on the GIS Specialist to devise solutions. The best method to use depends in part on whether the habitat map is raster or vector. In northern Alaska, the habitat layer was raster. We used the **Majority Filter** command many times to convert isolated pixels to the values of the surrounding pixels. We then used the **Region Group** and other **Clean** commands to convert the patches of identical-valued cells to polygons and absorb small polygons into surrounding larger ones. We then used the original raster data to characterize upland and wetland polygons and subdivide the polygons into a few types.

Despite the most creative GIS work, final work on the plots layer will need to be done manually. Standard GIS methods can be used to identify plots that should be examined. For example, the smallest or largest plots may be worth looking at or all the plots with a rare but high-quality habitat might be examined. We have found that final border adjustments can be made quite quickly by subdividing a plot and then merging small parts with surrounding plots. We recommend that a few weeks (at least) of GIS Specialist time be budgeted for making final corrections to the plot borders. Doing this well will be helpful in all future stages of the project.
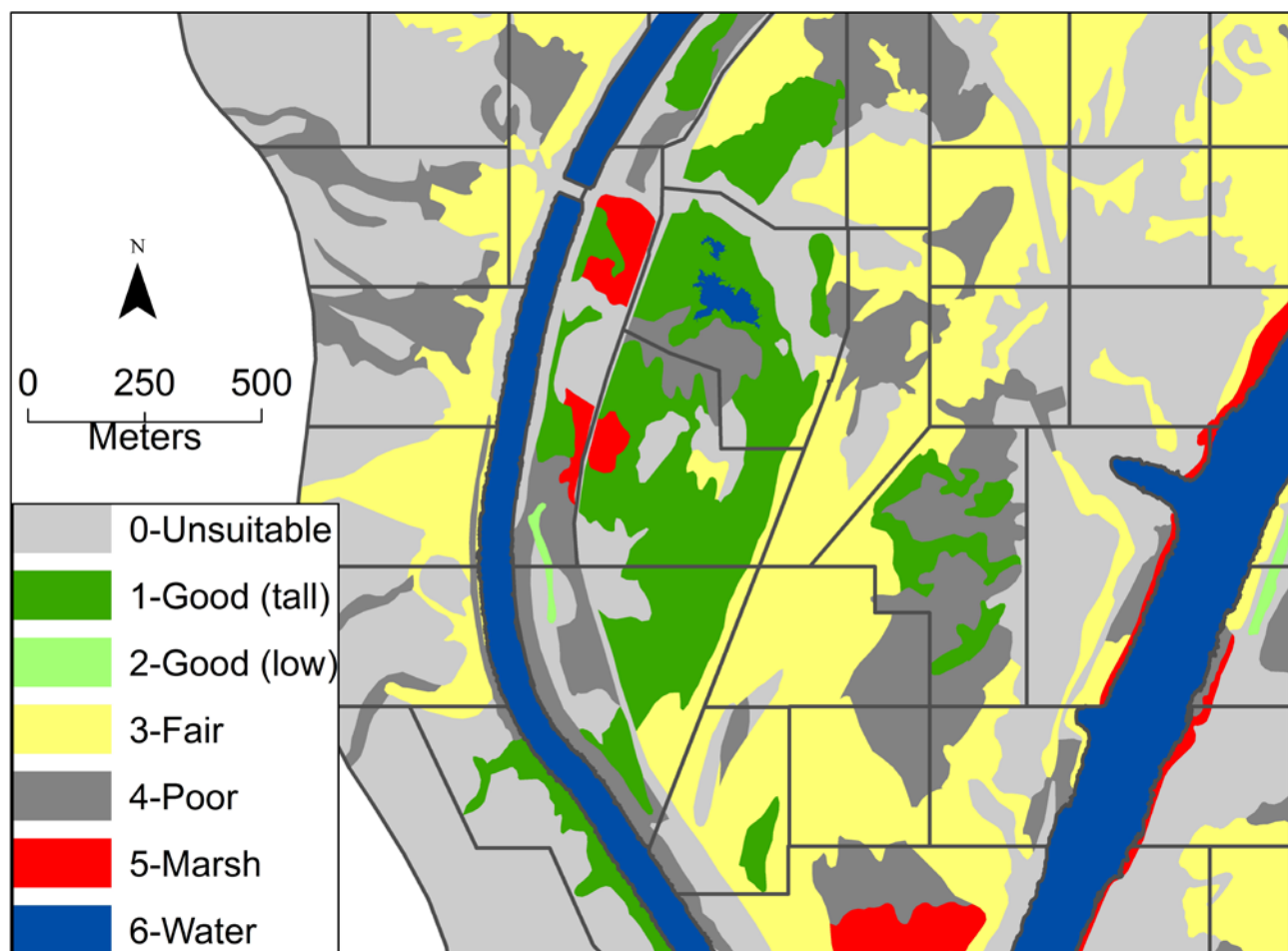

Figure 2.   Example of plots configured to concentrate good habitat.

This step thus requires the following actions:

a. Make the initial plots file using a fishnet or other similar command. Add columns to the attribute table labeled Region, Habitat, Strata, and Surveyed.

b. Intersect it with the inferential population layer.

c. Merge small polygons with adjacent ones.

d. Re-define plot borders as needed using one of the approaches described above.

Step 4.—Assign Plots to Strata

Plots are assigned to strata by assigning them to regions and habitats using the definitions adopted in Step 2. Final adjustments to the strata then sometimes have to be made due to small sample sizes in some strata or too many strata.

As mentioned in Step 2, the shapefile depicting the sampled area may or may not include region boundaries. If the shapefile does include region boundaries then the plot definition procedure automatically keeps plot borders from crossing region borders so it is easy to assign plots to regions. If the inferential population shapefile does not include region borders then the easiest way to assign plots to regions is by establishing a centroid for each plot and then determining which region the centroid is in. This information is then entered in the *Region* column of the attribute table of the plots shapefile.

Assigning plots to habitat is complicated and warrants careful thought. As a first step, we carefully define the schemes used to describe habitat. The SLL process usually makes use of three schemes. The first is a land cover, rather than habitat, map produced by a program such as GAP or Land Fire or produced specially for the project area. These schemes nearly always have too many categories to be used directly. This scheme is thus condensed, in Step 2 above, to a few (usually 4–7) categories, some of which are unsuitable habitat, for use in the SLL process. We will refer in the future to the original categories as "land cover" categories and the reduced set of them as "habitat types" (table 3). A raster or vector "habitat shapefile " should be made with the habitat types entered as values (rasters) or used to create new polygons (vector layers).

Once the habitat shapefile has been made, each plot must be assigned to a single habitat, even though many plots will include more than one habitat type. We use the terms "habitat type" for the condensed list of categories based on the land cover map (table 3) and "habitats" as the categories to which plots are assigned. We recommend using the same category names for habitat types and habitats. This is not strictly necessary but it makes the habitat

Table 3. Example of land cover categories and habitat types used in a study of shorebirds in Alaska.

Land cover categories	Habitat type
Water, clear	Unsuitable
Water, turbid	Unsuitable
Ice	Unsuitable
Aquatic, Carex aquatilus (50–80 percent water)	Wet
Aquatic, Arctophila (50–80 percent water)	Wet
Flooded, low-center polygons (25–50 percent water)	Wet
Flooded, non-patterned (25–50 percent water)	Wet
Wet, Carex aquatilus (10–20 percent water)	Moist
Moist, sedge-grass meadow (10–20 percent water)	Moist
Moist, tussock tundra	Moist
Moist, moss and lichens	Moist
Shrub, dwarf	Upland
Shrub, medium	Upland
Shrub, tall	Upland
Barren, dry dunes	Unsuitable
Barren, sparse vegetation	Unsuitable
Barren, other	Unsuitable

assignment process much easier and we see no gain in having different schemes. We assume below that they are the same.

The first task in assigning plots to habitats is to calculate the proportion of each plot covered by each habitat type and to enter this information in the attribute table of the plots shapefile. The procedures for doing this differ depending on whether the habitat layer is raster or vector. We assume the GIS Specialist can accomplish this task.

Defining the rule set to assign plots to habitats often warrants careful consideration. A very simple rule set, often appropriate is "use the most common habitat type," perhaps excluding unsuitable areas (table 4). For example if the habitat types were *Good*, *Fair*, and *Poor*, and they covered 50, 40, and 10 percent of the plot, respectively, then the plot would be assigned to habitat *Good*. This does not mean we ever assume all of the habitat type in the plot is good. We are just using the label *Good* to describe the plot.

Table 4. Example of assigning plots to habitat using the majority rule.

[In this example, P(unsuitable) was ignored in making the habitat assignments]

Plot	Area	P(unsuitable)	P(wet)	P(moist)	P(dry)	Habitat
1	0.10	0.15	0.53	0.21	0.11	Wet
2	0.11	0.00	0.25	0.43	0.32	Moist
3	0.13	0.00	0.03	0.37	0.60	Dry
4	0.09	0.43	0.21	0.17	0.19	Wet
5	0.13	0.35	0.21	0.44	0.00	Moist
6	0.12	0.00	0.05	0.34	0.61	Dry

In many studies, the best habitat types only rarely cover a majority of the plot. In such cases, the "majority rule" above does not work very well. A series of questions may then work best. In the Lower Colorado River study area, the rule set was:

1. If tall woody polygons cover more than 5 percent of the plot, then classify the plot as tall woody; otherwise

2. If tall woody and low woody polygons cover more than 50 percent of the plot, then classify the plot as low woody; otherwise

3. If herbaceous polygons cover more than 30percent of the plot, then classify the plot as herbaceous; otherwise

4. If unvegetated polygons cover more than 30 percent of the plot, then classify the plot as unvegetated; otherwise assign the plot to the type that covers the largest proportion of the plot.

If pilot study data are available, then it may be useful to design an Microsoft® Excel workbook to assign plots to habitats and thus to strata and to calculate densities for each stratum and the fraction of the population covered by each stratum. We did this in the Lower Colorado River study after collecting data during a 3-year pilot study. A first worksheet contained one row for every plot in the study area and columns for each of the original land cover categories. The entries were the proportion of the plot covered by the land cover category. A second worksheet also contained one row for every plot. The columns were the habitat types. Embedded rules calculated the proportion of the plot covered by the habitat type. A third worksheet with rows for every plot had Region and Habitat columns and contained rules for assigning plots to habitats. This worksheet also identified which plots were surveyed and recorded the number recorded of each species of interest. A final worksheet calculated stratum sizes and the densities and their variances for each of species of interest. We were then able to explore the implications of different rules for (a) consolidating land cover types into habitat types and (b) rules for assigning plots to habitats on stratum size and densities. This led us to recommend substantial changes in how strata were defined for the long-term survey.

This step thus requires the following actions:

a. Assign plots to regions.

b. Develop a rule set for assigning plots to habitats.

c. Assign plots to strata using their region and habitat.

d. Revise the procedures used in "a" to "c" as needed to achieve good stratum sizes and variation in density between strata.

Step 5.—Define Clusters or Clusters and Zones (optional)

Clusters and zones are groups of plots that are selected together (fig. 3). If the design includes selecting locations and then visiting several plots or points in close proximity, then we have cluster sampling. If, in contrast, a "simple random sample" of plots is selected—meaning that each location is selected without reference to any other locations—then we have (stratified) random sampling and neither clusters nor zones need be created.
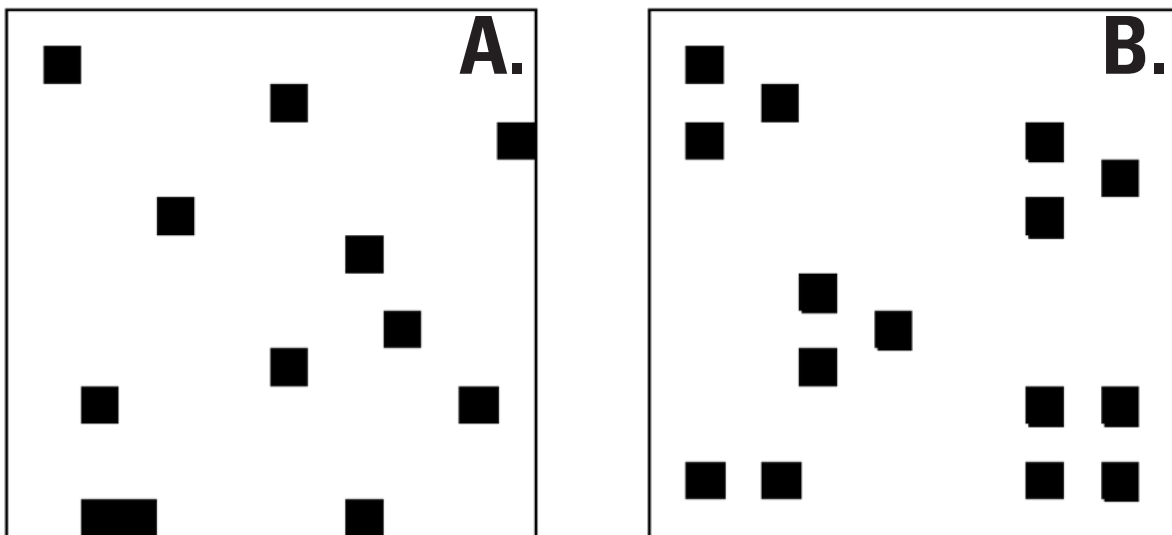

Figure 3. Typical results from (*A*) simple random sampling and (*B*) cluster sampling.

In standard survey sampling methods, plots in a cluster are assigned to the same stratum. Cluster sampling is usually used for logistical reasons. For example surveyors might be able to survey five plots or points in a morning if the locations are close together but if the points are selected independently then travel time between them may be too great to do five locations. This is a common occurrence familiar to nearly anyone who designs bird surveys. If cluster sampling (without zones) is used, then a cluster shapefile should be prepared that defines the plots in each cluster. Cluster numbers must be unique within regions but they can be repeated between regions. Statistical methods are well defined for cluster sampling and are implemented by program *DS*.

Zones are groups of plots selected together whose members are assigned to more than one stratum. Consider an analysis of BBS data in which habitats are defined and each location for a 3-minute count is assigned to a habitat. A given route, with 50 locations, will nearly always include counts in different habitats. This causes a problem in error estimation because in stratified sampling, we normally assume independent selection between strata. This is more than an academic problem as shown by the following example. Suppose 6 groups of plots were defined and we selected 10 plots in each. The habitats are *Good*, *Fair*, and *Poor*. Furthermore, suppose that each group had some plots of each habitat type. If we ignored the lack of independence, then the equations used to calculate variances would assume that we surveyed 18 different clusters, 6 with *Good* plots, 6 with *Fair* plots, and 6 with *Poor* plots. If we had actually done that, coverage probably would have been quite a bit better than we achieved with only 6 groups. The lack of independence thus needs to be acknowledged. In this example, our 6 groups of plots are zones because they each contain good, fair, and poor plots. Program *DS* contains routines to analyze designs that include zones.

In summary, if plots are going to be selected in groups, usually for logistical reasons, then delineate the group boundaries in a shapefile and assign each plot to exactly one group. Next assign cluster numbers to each plot being careful not to duplicate cluster numbers within regions and keeping in mind the definition of clusters—groups of plots in the same stratum. If groups contain plots from different strata, then also assign numbers to the groups (that is, zones). For example, the zones might be numbered 1, 2, 3... and the clusters within zones might be numbered according to their habitat, 1=good, 2=fair, and 3=poor for example. The full cluster numbers might then be constructed as "zone number, decimal, habitat number." Thus, 23.2 would mean fair plots in zone 23. The output of this step is thus that the attribute table for the plots Shapefile is given a new column for cluster number and a column for zone number if zones were defined.

If clusters, or clusters and zones, were defined, then Program DS needs a table reporting how many plots are in each cluster and the total area they cover. These tables have the following columns:

Clusters only: region, habitat, cluster, number of plots, area covered

Clusters and zones: region, habitat, zone, cluster, number of plots, area covered

See the *DS* User's Manual for more details. The needed information is easily tabulated using ArcGIS and Microsoft® Excel commands.

This optional step (needed only if clusters were defined) thus requires the following actions:

  a.   Define and number the groups of plots.

  b.   If plots within groups are in more than one stratum then number the clusters within each group.

  c.   Add the cluster number, and zone number if zones were defined, to the attribute table for the plots shapefile.

  d.   Make a table of clusters (and zones if relevant) with sizes and areas.

Step 6.—Choose Sample Sizes within Strata

Choosing the number of plots to survey in each stratum usually requires decisions that seem arbitrary. Here, we present a process designed to remove some of the arbitrary nature of the process and, most of all, to enable users to move rapidly through this step.

The first action is to make a table with the number of plots in each stratum. Second, select the total sample to be surveyed. Usually this decision is based on the available resources. Sometimes, managers will ask the biologists or analysts how large a sample they "need." Responding rigorously to this question is difficult because estimates are required of quantities that usually are poorly known.

Once the total sample size has been set, we recommend calculating the proportion of the plots that are in each stratum and then assigning the same proportion of the plots to be surveyed. To take a simple example, if there were just three strata and the proportion of plots in them were 20, 50, and 30 percent, and if 100 plots were to be surveyed, then the initial allocation of plots to strata would be 20, 50, and 30 for the three strata, respectively. Once the initial allocation is made, the results can be inspected and adjustments made, while maintaining the total sample size. Adjustments usually

increase sample size in strata with high-quality habitat or regions of high interest or in strata with very few plots in the initial allocation. Addressing these issues—using common sense and experience with the study area—is important, especially when some strata receive very few plots in the initial allocation yet may contain abundant birds. If sample sizes are not increased in these strata (and decreased in others), results may be very unstable.

This step thus requires the following actions:

a. Display the number of plots in each stratum in a table.

b. Decide the total number of surveys.

c. Allocate them proportional to the number of plots in each stratum.

d. Adjust the results to increase sample size in high-interest strata and strata with very few surveys under the initial allocation.

Step 7.—Select Plots to Survey

The procedure for plot selection depends on which sampling plan is followed within strata: simple random sampling, cluster sampling, or cluster sampling with zones (Step 5). Regardless of the sampling plan, however, we prefer to distribute the surveyed plots throughout the strata rather than using a procedure that might result in surveyed plots being concentrated largely or entirely in one small part of the stratum. This is usually easy to accomplish by taking a systematic sample of the FID numbers of the plots, clusters, or zones. To do this, transfer the FID numbers to a spreadsheet, sort them, and proceed as follows within each stratum. Divide the FID numbers into as many groups as the number of plots to be selected, keeping the group sizes as similar as possible. For example, if a stratum has 93 plots and the desired sample size is 5, then the group sizes would be 19, 19, 19, 18, and 18. The groups would be plots 1-19, 20-38, 39-57, 58-75, and 76-93. A number between 1 and 19 inclusive would be chosen randomly and that plot would be selected. Then 19, 19, 18, and 18 would be added to that number to select the remaining plots. This procedure ensures that non-existent plots do not get identified for inclusion in the sample.

In cluster sampling, the procedure is similar except that it is carried out on the clusters and in each selected cluster, a fixed number of plots are randomly selected. Ideally, the plots are selected within clusters completely at random but if this results in excessive travel time between plots, then groups of plots—all fairly close together—may be defined and one of the groups can be selected at random. This procedure does introduce some selection bias due to the plots being close together, but this may be a reasonable trade off with the costs of plots being too far apart.

Sample selection with zones and clusters is complicated and advice from a statistician may be needed. The most rigorous approach is to select the desired number of plots in each zone completely at random. The problem with this approach, however, is that it does not control the sample size within strata and rare habitats—which may be especially important to survey well—may hardly ever be included. It may be possible to overcome this problem by defining regions in such a way that they largely contain plots assigned to just one habitat, or at least that the habitats of greatest concern are concentrated into one or a few regions. If this can be done, then random selection of plots within regions—and selecting more plots in the regions of greatest interest—may still result in the desired number of plots in each habitat (although it will not be exactly the desired number because the habitat is determined by random selection). More experience is needed with how to use zone and cluster sampling to obtain desired numbers of plots in each stratum.

If plots will be covered during several years, or other periods, then it is important that in each period the surveyed plots cover the entire study area as evenly as possible. Otherwise strong year effects can easily produce apparent variation across the study area that is due to transitory effects, such as poor weather in 1 year.

This step thus requires the following actions:

a. Number the plots, clusters, or zones sequentially within regions.

b. Arrange them in sequential order using the FID numbers.

c. Select the desired number systematically as described above.

d. If clusters are defined, selected the desired number in each cluster randomly or by defining groups and selecting one of the clusters.

e. If zones are defined, either select plots completely at random or devise a method—with help from a statistician—that will yield essentially unbiased estimates of density in each stratum.

Example: Snake River Birds of Prey Area

A detailed example of the SLL process is presented in this section. We use this example in workshops—presenting the background for each step and then asking participants to divide into groups and discuss how they would carry out each task. Each group then presents the approach they have devised and, afterwards, we discuss the approach we followed. We then proceed to the next step, using the materials that we have previously prepared based on the approach we adopted. We encourage you to participate in this process and have presented the material below so that you can study the information at the beginning of each step and then decide how you would carry out the task before reading the rest of the section to see how we proceeded. The introductory material and the description of the choices we made are separated by a line of asterisks.

The Snake River Birds of Prey National Conservation Area (NCA) in southern Idaho was created in part as a refuge for birds of prey during the summer and winter (fig. 4). The canyon along the Snake River is especially valuable for birds

of prey, in part as nesting sites, but the other areas also are used extensively. The NCA is owned by the Department of Defense and is used by them for training, especially in the Orchard Training Area (OTA). For this example, we assume that a long-term survey of breeding birds is being designed for the NCA.

The goals of the survey are to:

1. Estimate trends in abundance for species of concern including raptors, threatened or endangered species and other species on "of concern" lists.

2. Help assess impact of training and other activities on migratory birds.

3. Provide data useful in studying habitat and other relations.

We will assume that the survey will be conducted during the day by two, 2-person teams, each with a vehicle for 3 months excluding training. In addition, one person-month by a supervisory biologist will be available for hiring, training, analysis, and reporting.
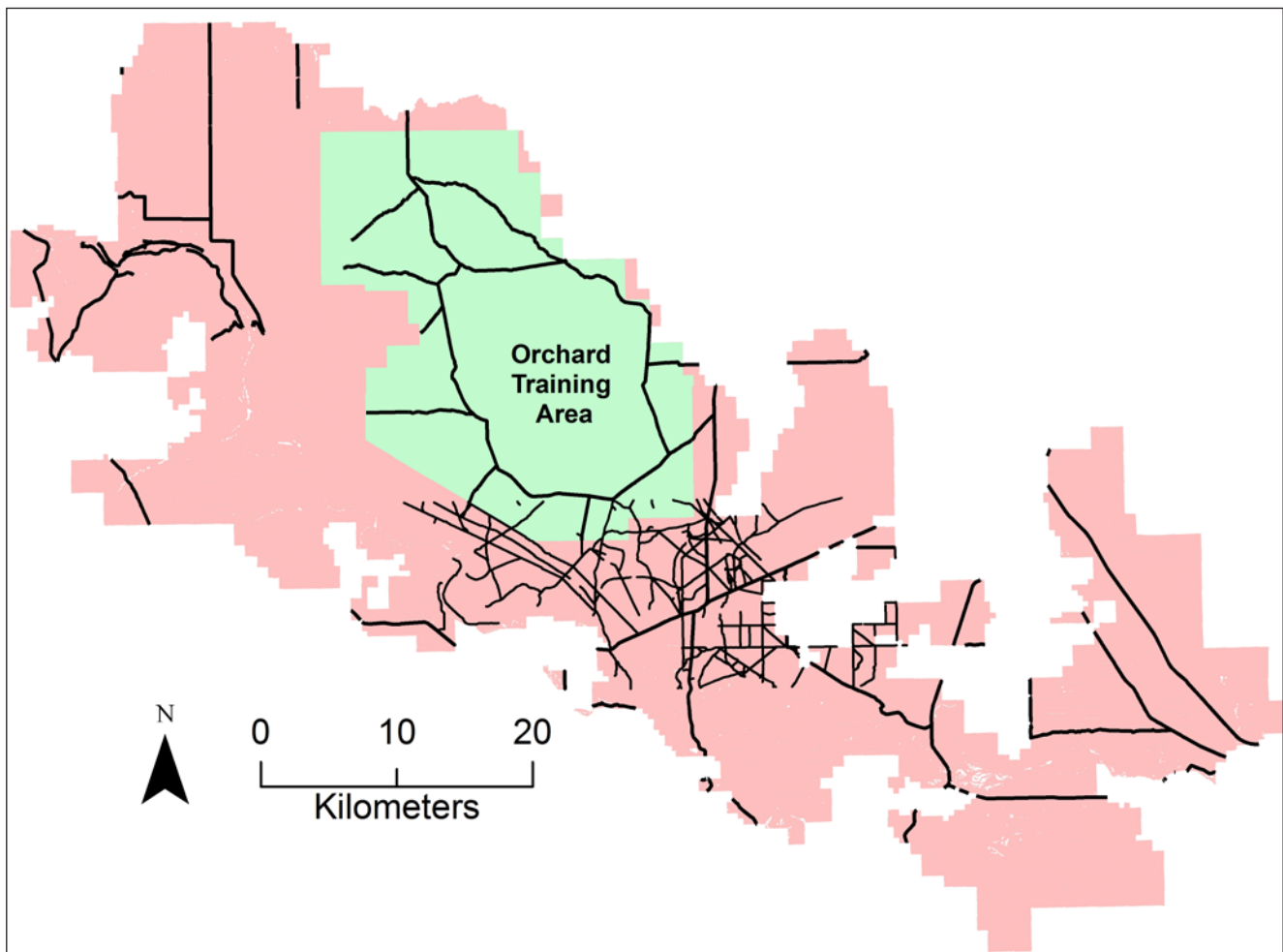

Figure 4.   Snake River Birds of Prey NCA showing major and minor roads and the Orchard Training Area.

Step 1.—Define Populations

Introduction: Recall that this step requires you to define the study area, the inferential population, and the sampled population. Give some thought to how you would define populations before continuing.
****

Our decisions: The study area was defined for us as the NCA. The borders shown in figure 4 thus delineate the study area. We considered removing the Snake River from the inferential population but did not do so because aquatic birds frequently use the Snake River and would be counted on surveys. We also decided that the area was small enough that all parts of the area could be surveyed, although areas far from roads might be lightly sampled, and that no areas warranted exclusion because no birds would be present. We therefore defined both the inferential and sampled populations as including all parts of the study area.

Step 2.—Define Regions and Habitats

Introduction: A detailed land cover map, with 26 types in the NCA, was available (table 5).

Table 5.   Land cover types in the Snake River Birds of Prey National Conservation Area.

Land cover types	Hectares
Inter-Mountain Basins Playa	6
Inter-Mountain Basins Active and Stabilized Dune	739
Quarries, Mines. and Gravel Pits	18
Inter-Mountain Basins Cliff and Canyon	291
Columbia Plateau Steppe and Grassland	15,956
Columbia Basin Foothill and Canyon Dry Grassland	1,895
Inter-Mountain Basins Semi-Desert Grassland	4,769
Introduced Upland Vegetation - Annual Grassland	91,332
Recently burned grassland	743
Introduced Upland Vegetation - Forbland	115
Pasture/Hay	2,800
Cultivated Cropland	13,315
Columbia Plateau Ash and Tuff Badland	2,217
Inter-Mountain Basins Greasewood Flat	5,529
Inter-Mountain Basins Mixed Salt Desert Scrub	15,767
Inter-Mountain Basins Semi-Desert Shrub-Steppe	10,075
Rocky Mountain Lower Montane-Foothill Shrubland	3
Inter-Mountain Basins Big Sagebrush Steppe	2,427
Inter-Mountain Basins Big Sagebrush Shrubland	65,800
Columbia Plateau Western Juniper Woodland and Savanna	38
Rocky Mountain Aspen Forest and Woodland	1
Developed, Open Space	3,314
Developed, Low Intensity	463
Developed, Medium Intensity	222
Developed, High Intensity	18
Open Water	4,255

In this step, you define habitat types based on the land cover types or any other information you wish to use, and regions again defined using any variables you wish to use. Consider the information above and then construct your list of habitats and regions. Remember that each region-habitat combination becomes a stratum (unless it has no plots) and that two plots must be selected in each stratum. Recall that at this stage you also decide whether regions will be defined now and intersected with the shapefile of the inferential population or whether plots will first be defined and assigned to regions that implicitly defines the regions. This example provides an illustration of why this can be an important decision to make correctly.
****

Our decisions: We felt that three regions were fairly obvious: the canyon, because it is of great importance as a nesting site; the OTA because it is used for training and one of the goals was assessing impacts of training on the birds; and the rest of the area that must be included because the inferential population includes the entire study area. We also were concerned about travel time to reach plots far from roads and about the fact that small roads were concentrated in only one part of the study area. We therefore decided to define a second set of regions based on distance from a major road. We initially defined three buffers using 3 km and 6 km as the divisions (fig. 5), but we were concerned about having too many strata and the more than 6 km buffer covered only a small area so we decided to use just two buffers, 0–3 km and more than 3 km.

Regarding the issue of how and when to delineate regions on the inferential population shapefile, the OTA was a political subdivision so it seemed important that plots not cross its boundaries. More importantly, however, the canyon walls were essentially uncrossable during a survey so it was important that plots not cross them. We therefore defined the rim of the canyon as the border of the canyon region and then intersected the inferential population shapefile with the regions shapefile.

The number of land cover codes obviously had to be reduced to just a few. We initially defined seven types (table 6). Birds in agriculture, grassland, and shrubsteppe would be different and that estimates would be wanted for them. We would have liked to keep the rest of the habitats but that would have given us too many strata. We decided that the "canyon" habitat would be captured by the "canyon" region and that it was surrounded by shrubsteppe so we combined it with shrubsteppe. We decided that the "developed" category was particularly interesting and it was surrounded with agriculture so we combined it with that category. We combined playa, dunes and woodland with shrubsteppe because they covered such tiny areas and were surrounded by shrubsteppe. We thus defined three habitat types: agriculture, grassland, and shrubsteppe.
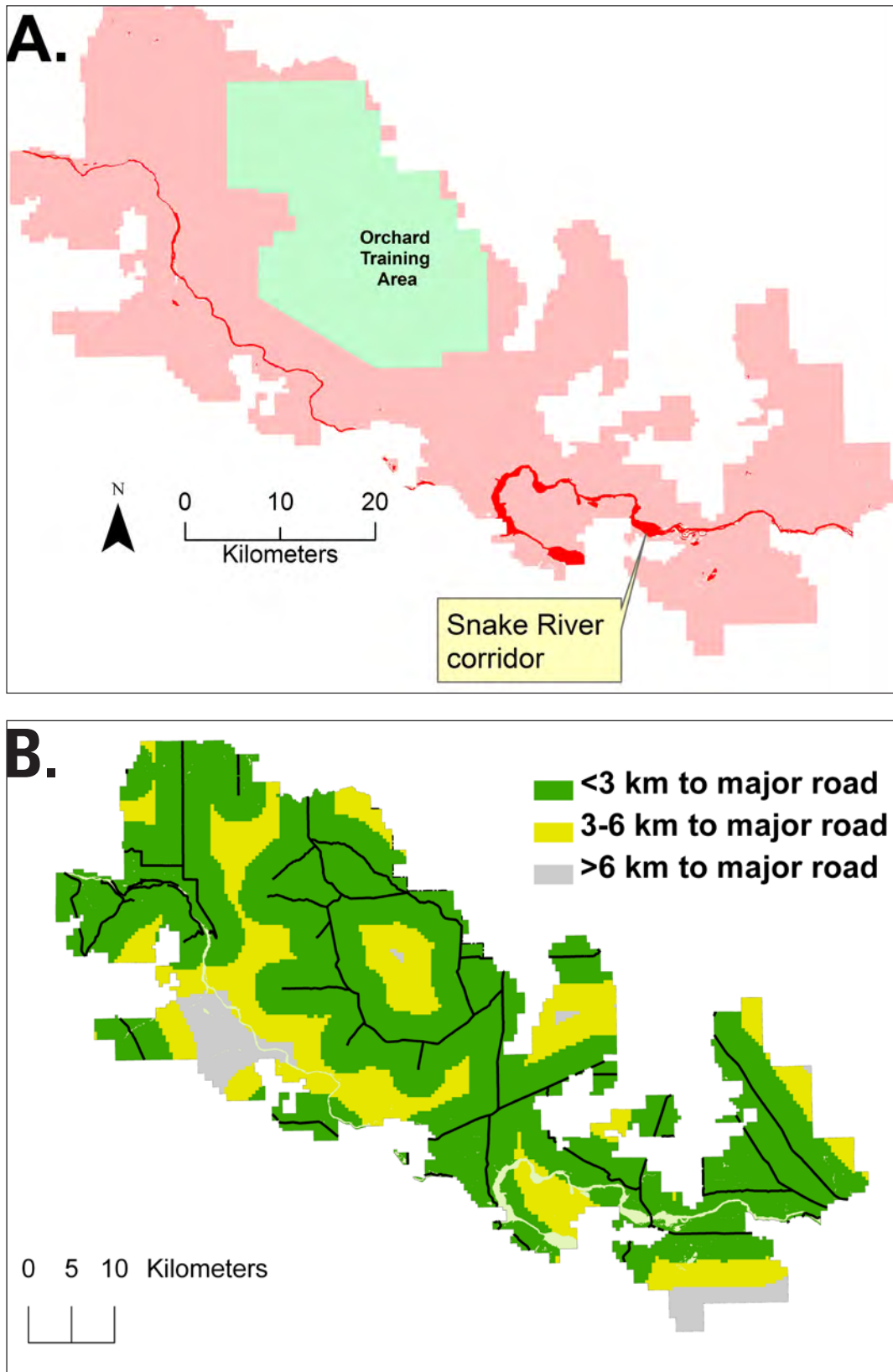
Figure 5.   Regions defined for the Snake River Birds of Prey National Conservation Area. *A.* Snake River corridor, Orchard Training Area, rest of area; *B.* buffers defined by distance to roads). In the final design, only two buffers, 0–3 and more than 3 km were used.

Table 6. Initial set of habitat types.

Habitat type	Hectares	Percent
Agriculture	16,115	7
Canyon	2,508	1
Developed	4,017	2
Grassland	114,811	48
Playa dunes	745	<1
Shrubsteppe	99,601	42
Woodland	38	<1

Step 3.—Make a Plots Shapefile

In this step we imposed a grid of square cells on the inferential population. We used 16-ha squares because the squares can easily be surveyed in a morning in open country with fairly sparse birds. This approach results in fewer plots than doing several plots per morning, but such groups of plots have to be defined as clusters or zones so the effective sample size is well below the number of plots. Furthermore, by doing one plot per morning we do not lose time during the survey period to travel between plots, less time is spent trying to decide whether birds on the edge are in or out of the plot (because the edge-to-area ratio is more favorable), and we do not have to worry about the complexities of sampling with clusters or zones and clusters.

Step 4.—Assign Plots to Strata

Introduction: In this step, we assigned plots to regions and habitats and checked that the resulting strata had appropriate numbers of plots. The issues to consider are how to assign plots to regions and how to define habitats.
****
Our decisions: Because the inferential population had region borders embedded within it, the plot delineation process automatically prevented plots from crossing region borders. It was therefore easy to determine which region each plot was within. We considered assigning habitats using a series of questions as described earlier in this report but decided that a simple majority rule was sufficient.

Step 5.—Define Clusters or Clusters and Zones (optional)

We did not carry out this step because we used simple random sampling within strata.

Step 6.—Choose Sample Sizes within Strata

Introduction: Review the description of this step above to be sure you understand the tasks. Only the second one can be done without access to the detailed data. In this study, as in most, the resources were fixed so sample size can be estimated from the resources available and the field methods described immediately above.
****
Our decisions: The project description said two, 2-person teams would each work for 3 months. This works out to about 240 person-days assuming a 5-day work week and that a few days are lost to weather or other problems. We said, under step 3 above, that each person would survey one plot/day so we assume that a maximum of 240 plots can be surveyed. The reality might be closer to 200 plots if logistical problems arise, but it is safest to select the maximum number of plots and then be sure —as the field work proceeds—that effort is evenly distributed across strata so that at least two plots are surveyed in every stratum even if the number of plots surveyed turns out to be fewer than the optimistic initial estimate.

We initially distributed sample sizes across the strata proportional to stratum size (table 7). We then increased sample sizes where necessary to insure that each stratum had at least 10 plots. If agriculture had covered a larger area, we would have decreased sampling there, because it was less interesting habitat for this survey, but it actually covered so little area that we had to increase the number of plots in agriculture in each region to meet the "minimum of ten" rule.

Table 7. Initial and final sample sizes (SS) in each stratum.

[Region: OTA, Orchard Training Area]

Region	Habitat type	Area (ha)	Prop.	Sample sizes (SS)	
				Initial	Final
Canyon	Agriculture	2,056	0.009	2	10
Canyon	Grassland	12,334	0.052	12	12
Canyon	Shrubsteppe	3,548	0.015	3	10
OTA	Agriculture	8,234	0.035	8	10
OTA	Grassland	43,566	0.183	43	40
OTA	Shrubsteppe	32,884	0.138	33	31
Other	Agriculture	9,842	0.041	9	10
Other	Grassland	58,911	0.248	59	57
Other	Shrubsteppe	66,460	0.279	67	60
Totals		237,835	1	236	240

Step 7.—Select Plots to Survey

Plots would be selected using simple random sampling.

This page left intentionally blank