Mapping of Bird Distributions from Point Count Surveys[1]

John R. Sauer, Grey W. Pendleton, and Sandra Orsillo[2]

Abstract: Maps generated from bird survey data are used for a variety of scientific purposes, but little is known about their bias and precision. We review methods for preparing maps from point count data and appropriate sampling methods for maps based on point counts. Maps based on point counts can be affected by bias associated with incomplete counts, primarily due to changes in proportion counted as a function of observer or habitat differences. Large-scale surveys also generally suffer from regional and temporal variation in sampling intensity. A simulated surface is used to demonstrate sampling principles for maps.

Bird distributions are of great interest to both amateur birdwatchers and professional ornithologists. Range maps published in field guides and other sources provide a large-scale view of approximate range and relative abundance that have obvious uses for determining if species are likely to be seen in an area (Robbins and others 1986, Root 1988). They are also used to evaluate more subtle questions about ecological aspects of bird distributions (Repasky 1991). Because of the importance of assessing changes in bird ranges in association with global climate change and other large-scale environmental changes, existing range maps take on added importance as standards from which we can evaluate future changes in ranges. But range maps published in field guides generally contain many biases associated with the anecdotal nature of the observations.

Maps generated from extensive bird survey data sets such as the North American Breeding Bird Survey (BBS) (Droege 1990) and the Audubon Christmas Bird Count (CBC) (Butcher 1990) provide a reasonable source of systematically-collected information on bird distributions, and several recent publications have used these data to generate distribution maps (Robbins and others 1986, Root 1988, Sauer and Droege 1990). Because information from these surveys is now used in Geographic Information Systems (GIS) to address many management-oriented questions (e.g., analysis of the potential for bird-aircraft collisions or evaluation of bird species presence in existing patches of forest for county planning), it is of interest to evaluate the potential for error in these maps and review how sampling procedures can bias our maps of bird distribution.

Home-range estimation methods provide another example of spatial mapping procedures. In this case, the map must be formed on the basis of density of points because only presence data exist for each point. In bird surveys, these data can result from presence-absence counts, such as those obtained from miniroute stations or atlas blocks. These methods require uniform sampling density to avoid distortion in the map.

A fundamental problem with creating and assessing the efficiency of maps estimated from any sample data is that we do not have complete information on the number of birds at all points in the region (the actual surface of the map) for comparison with the estimated map surface. It is, therefore, difficult to assess error in the interpolated portion of the map. A much greater difficulty exists with maps generated using point count samples. For these data, we do not even have point estimates of the number of birds at any location on the real surface. The maps are based upon counts, which are related to the actual numbers of birds by an unknown probability of detection p (Barker and Sauer in this volume). In this paper, we review methods for developing contour maps of bird distributions from data collected at discrete points and discuss how sampling constraints associated with point counts can bias and create error in the maps. We develop a measure of bias and efficiency for maps and use simulation to show how different sampling strategies can change the efficiency of maps from point count data.

Procedures for Mapping

Early maps from BBS data were prepared by a skilled ornithologist using average counts at each survey location. Using his knowledge of bird distributions and bias in the coverage of the survey, the observer drew contours that used both the existing data and "expert opinion" for areas where survey data did not exist (D. Bystrak, personal communication). Examples of these maps appear in Robbins and others (1986).

Recently, use of statistical methods for smoothing data has become popular for bird survey mapping. Let $\mathbf{m}_i$ be the location of point i in two dimensions (e.g., $\mathbf{m}_i = \{X_i, Y_i\}$), and let $Z(\mathbf{m}_i)$ be the count at point i. These procedures take the counts at points at known locations $\mathbf{m}_i$ and estimate counts at all points that were not sampled in the region. In practice, many programs (e.g., SURFER [Golden Software 1987]) use a smoothing procedure to estimate the predicted counts for a uniform grid of points spaced over the area to be mapped. They then either plot out the counts at these grid points or use some algorithm to estimate a contour map based on the uniform grid points.

We illustrate this process using a square region, which we call point count land (PCL), with a simulated surface with height $Z' = a(X + Y)$, where X and Y are locations of the point in the X, Y plane and a is a scaling factor to make the maximum value of $Z' = 20$. The actual surface (which is not observed in real life) can be thought of as an actual bird distribution map (*fig. 1a*). The counts at randomly located sampling points are shown in *figure 1b*. The gridding process based on the Z values at the randomly selected points is shown in *figure 1c*, and the smoothed topographic map from the sample is shown in *figure 1d*. This simulated surface will be used later in the paper to demonstrate certain aspects of

USDA Forest Service Gen. Tech. Rep. PSW-GTR-149. 1995

151

sampling for distribution maps. We use program SURFER (Golden Software 1987) to estimate maps from point data.

Mapping Methods

Several methods of estimating the values at the systematically-located grid nodes from data collected at random points are available. Most mapping methods were developed for geological applications, where they are used to estimate the shape of underground strata from a series of samples taken at specified locations (Isaaks and Srivastava 1989). An extensive literature has developed on smoothing methods such as kriging (Isaaks and Srivastava 1989), variants of
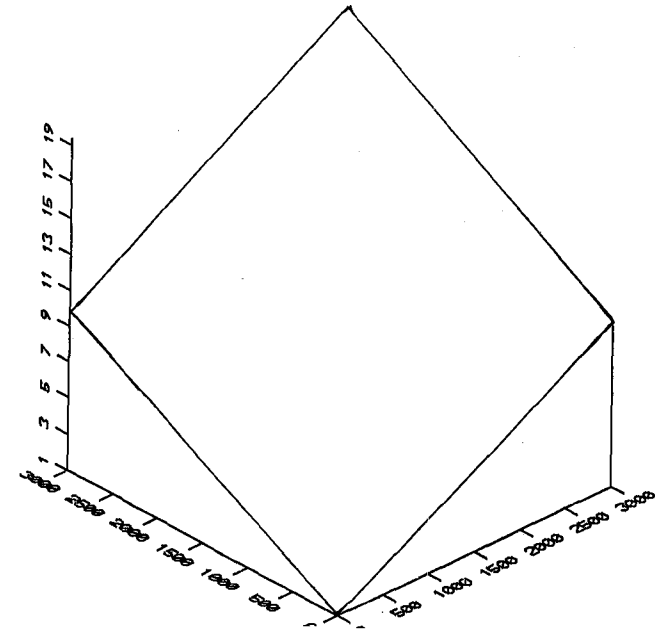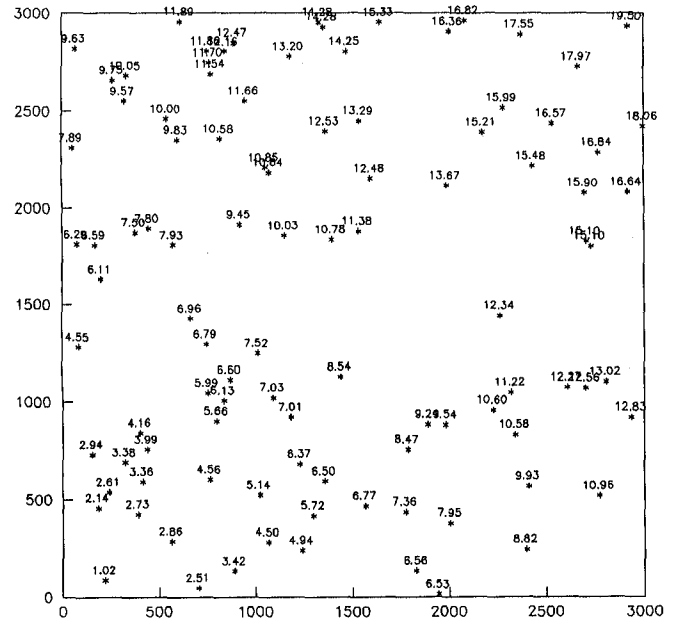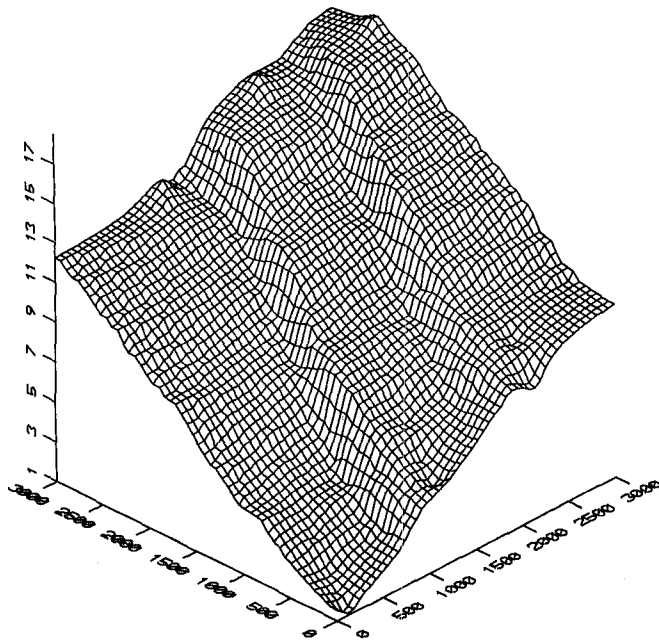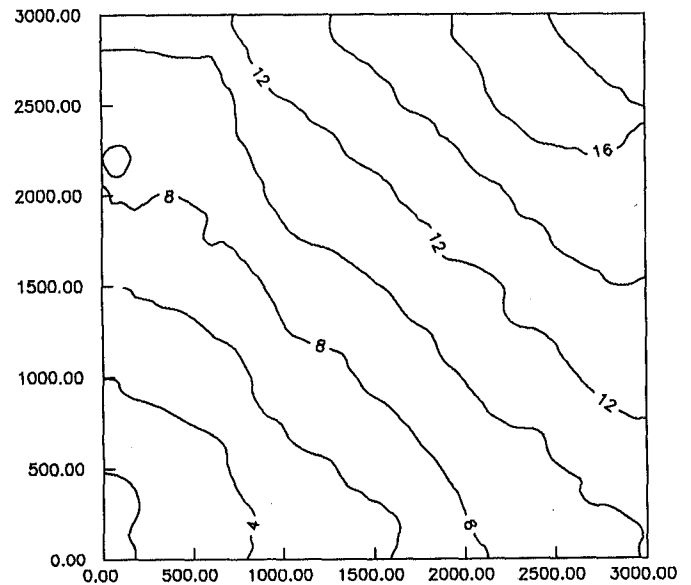
Figure 1--(a) A hypothetical surface that varies in height from 0 at the 0,0 point to 20 at the 3000,3000 point. (b) A sample of 100 randomly selected points, listed with counts. (c) A grid of counts estimated from the counts at the 100 randomly selected points. (d) A contour map based on the grid illustrated in *figure 1c.*

152

USDA Forest Service Gen. Tech. Rep. PSW-GTR-149. 1995

which have become quite popular. We briefly discuss two of these methods, inverse distancing and kriging.

Inverse Distancing

In this procedure, the count at a point at location m_i is estimated as a weighted average of points within a neighborhood of the point of interest, or

$$z(m_i)\frac{\sum_j \left(\frac{1}{h_j}\right)Z(m_j)}{\sum_j \left(\frac{1}{h_{i,j}}\right)}$$

In this average, the j is an index for all sample points which fall within a preselected neighborhood (or circle) of the location m_i, and the weights are the Euclidean distances between m_i and m_j or $h_{i,j}$, defined as:

$$h_{i,j} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$$

Often, a function of h such as h^2 is used as the weighting factor, and it is clear that the choice of both the size of the neighborhood and the choice of the function h can influence the estimated count $z(m_i)$. We present an example of a bird relative abundance map produced from BBS data using inverse distancing (*fig. 2a*).

Kriging

Kriging is a well-known statistical procedure that fits a best linear unbiased estimator to sample data. A kriging estimate of $z(m_i)$ is also a weighted linear combination of the existing sample data points, or

$$z(m_i) = \sum_j w_j Z(m_i)$$

The weights w_j. must sum to 1.0 and minimize the error variance. In practice, the weights are estimated from the covariance structure of known sample points. To do this, we must estimate the covariance among the sample points ($c_{j,k}$) and define a matrix **C**:

$$\begin{bmatrix} c_{11} & c_{12} & . & . & . & c_{1n} & 1 \\ c_{21} & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ c_{n1} & . & . & . & c_{nn} & 1 \\ 1 & 1 & . & . & . & 1 & 0 \end{bmatrix}$$

Also, the vectors $\mathbf{w'} = \{w_1, w_2, ..., w_n\mu\}$ and $\mathbf{D'} = \{c_{l,i}, c_{2,i}..., c_{n,i}, 1\}$ must be defined where i represents the point to be estimated. Note that the additional parameter μ is included as a mean term, which corresponds to the 1 and 0 values in the other matrices. The vector of weights $\mathbf{w}$ is estimated using Lagrange multipliers as $\mathbf{C\ w} = \mathbf{D}$, hence $\mathbf{w} = \mathbf{C^{-1}D}$, which is called the ordinary kriging system (Isaaks and Srivastava 1989).

In practice, the kriging system is often defined in terms of variograms, which are easier to estimate than covariances. A variogram is defined as:

$$2\gamma(m_i, m_j) = \text{Var}(z(m_i) - z(m_j)).$$

$\gamma(m_i, m_j)$ is called the semivariogram. The variogram is similar to a covariance function, but is inverse (a large covariance implies a small variogram). Furthermore, simplifying assumptions about the underlying distribution of counts must be made to estimate components of C and D. A major assumption is that the value of the covariance (and variogram) between points depends only on the distance between the points (h). Consequently, we can plot the value of the variogram as a function of h (fig. 2b), and we can model this relationship using a variety of linear, exponential, Gaussian, logarithmic, or other functions. Using this model, we can estimate the value of the variogram for any value h, which means that we can construct C and D from knowledge of the model and the distances between points, A contour map based upon an estimated variogram is presented in figure 2c.

The estimation of the variogram is a critical component of spatial analyses and has received a great deal of attention in the geostatistical literature (Armstrong 1989). Variogram analyses assume a constant covariance structure, and if this does not exist, the kriging estimates will be inappropriate. One common departure from the required consistency occurs when the covariance structure differs depending on direction as well as distance.

How Do We Evaluate the Quality of a Map?

There has been no research into the validity of applying kriging and other smoothing methods to bird survey data. When an automated procedure is used in mapping, there is a tendency to treat the analysis as a black box in which we vary the input parameters in an attempt to get a good picture. Unfortunately, to judge a "good picture," we use both other knowledge (often anecdotal) of what the map should look like and information from the data. Both of these sources are often flawed. All surveys are judged by how well they display people's "common knowledge" of populations. Is this an appropriate criterion? All maps are conditional on the existing data, but the information from the survey data contains many possible biases and errors, many of which are difficult to evaluate using the data.

Two statistical attributes, bias and variance, can be used to evaluate how good a map is or, in fact, how good any survey is.

Bias

Bias is a measure of how different the expected value of an estimator is from the underlying (true) parameter value. In point counts, the parameter is population size, but the estimate is the count. In a map, bias is $\mathbf{E}(z(m_i) - \mathbf{Z}(m_i))$: the distance from a point on the expected surface developed from the counts to the "real" surface of the bird distribution.
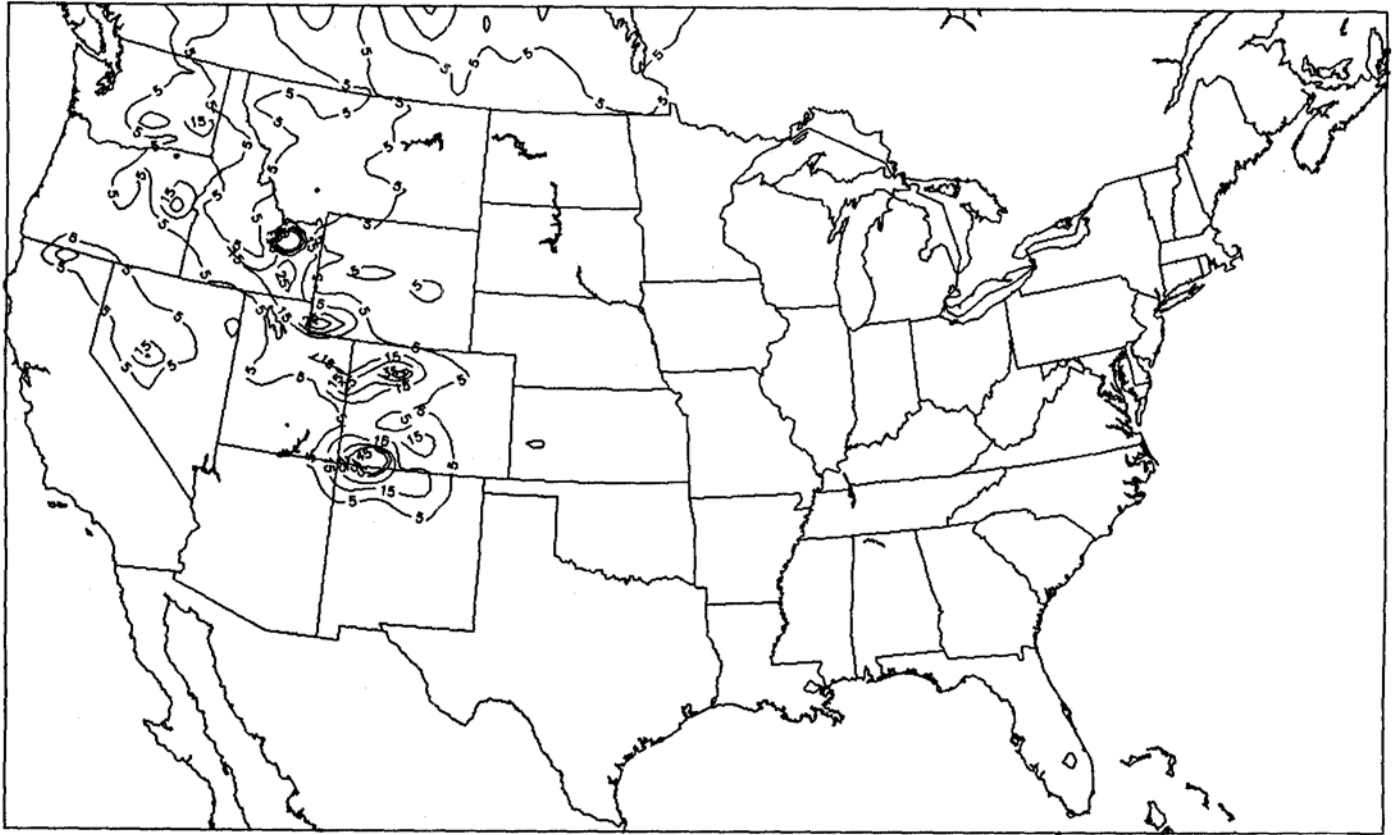
a
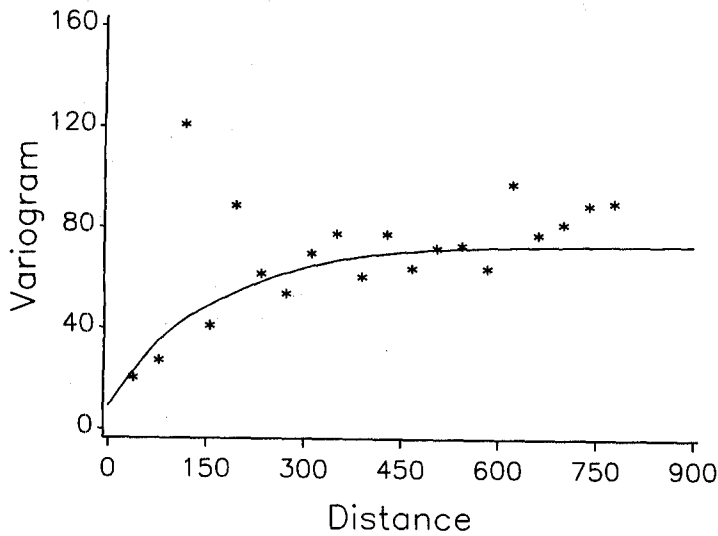
b

Figure 2--(a) BBS data on Black-billed Magpie (*Pica pica*), used as examples of mapping procedures. Data are averages of counts from the interval 1981 to 1990 from BBS routes. (a) A relative abundance map produced by using inverse distancing to estimate counts at nodes of a 100 x 100 grid over the map, and then contouring over the grid. (b) A sample variogram estimated for the Black-billed Magpie data. The smooth line represents a model fit to the variogram.
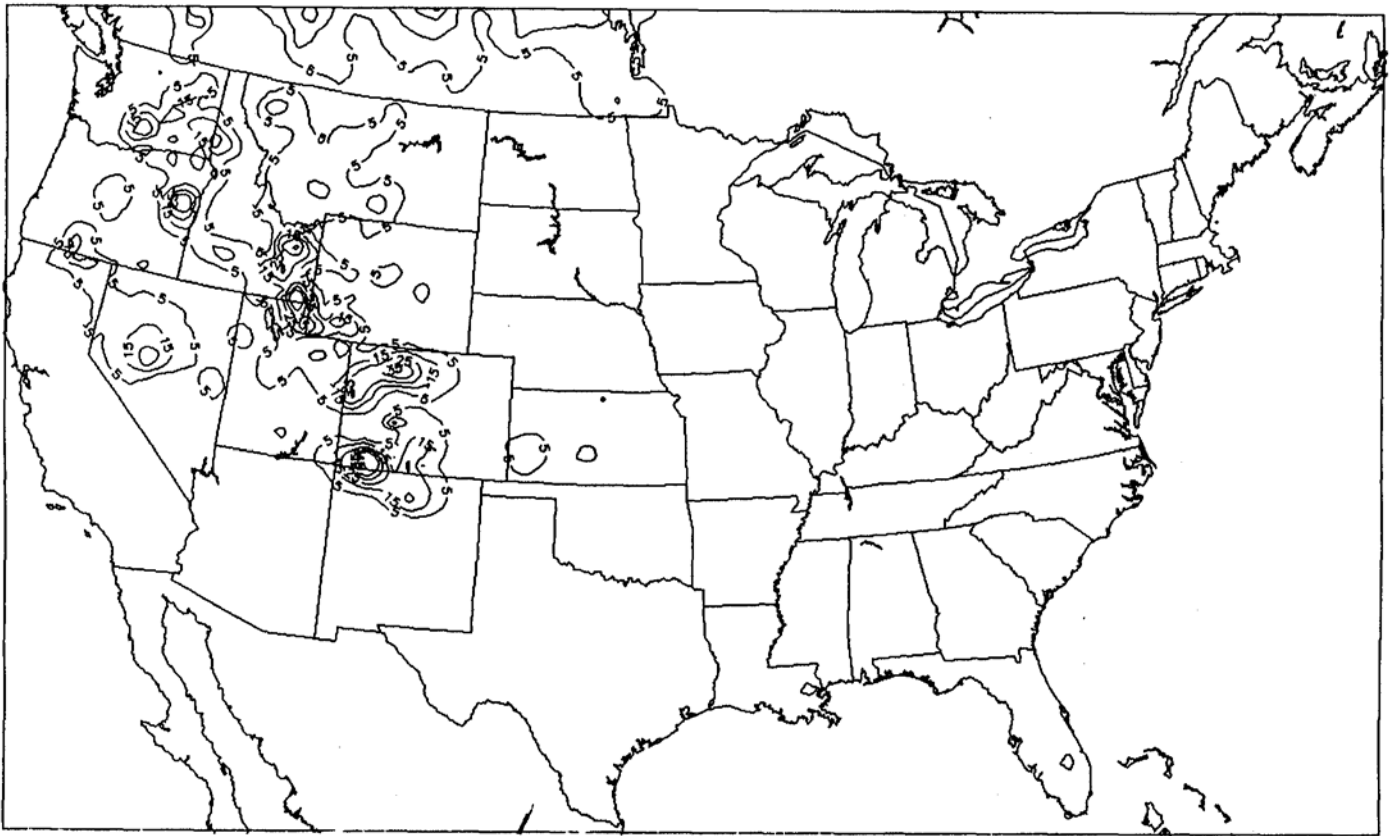
c

Figure 2 *(cont.)*--(c) A contour map developed from a kriging function using the sample variogram presented in *figure 2b*.

As with other analyses of point count data, the statistical properties of the proportion of birds sampled at a point (*p*, the ratio of the number of birds counted to the number of birds present at a point) are a major determinant of bias in mapping bird distributions. In our PCL example, this means that due to this *p* our point estimates of *Z*, that is *z*, are not unbiased. We can ***never*** observe the actual abundance of birds at any point with point count data. However, other attributes associated with sampling such as the roadside nature of counts and distortions due to topographic features can also bias smoothed maps of bird distributions.

Variance

Because we never measure the actual abundance of birds, the counts we derive from point counts are measured with error. A map made with point count data captures both error associated with incomplete counts and real variation in populations.

One reasonable measure of efficiency of a map is the mean square error, or MSE, which combines bias and variance as:

$$MSE = Bias^2 + variance.$$

How Can We Evaluate Bias and Precision in Maps?

We present two approaches to assessing possible difficulties with developing relative density maps from bird survey data.

First, we try to decipher some of these issues from existing data. Unfortunately, evaluation of bias in estimates from point count data is difficult because we infrequently know the real values. Validation of bird surveys generally involves comparison with alternative data sets that often contain similar bias in their estimates, and agreement or differences in estimates between surveys do not provide sufficient information to judge which is less biased. There are several examples, however, where we can reasonably assume that the estimates from comparative data are less biased (generally through collection using less biased methods), which can provide us with insights into bias associated with point count data.

Second, we can simulate maps and look at effects of our sampling methods on the mapping process. The advantage of this approach is that it allows us to evaluate the exact extent of bias for various sampling schemes. We, therefore, can avoid the conceptual problems that arise in comparing two surveys, each of which is of uncertain validity. Unfortunately, simulations are never completely representative of the vagaries of sampling and tend to provide idealized views of the world. We will use simulations to provide some insight into the effects of several sampling decisions on resulting maps.

Looking for Biases in Existing Data

There are many potential biases associated with large-scale surveys such as the BBS. Some of the biases are directly related to the vagaries of point counts, but others are a consequence of the constraints imposed by the necessity of collecting counts along roadsides using volunteers. The challenge in using large-scale survey data is in documenting the existence of potential biases and, if possible, modifying the analysis to accommodate them. In this section, we review some of the possible biases in surveys that could influence maps produced from survey data and, if possible, document their existence using survey data.

Point Count Biases

Point count methods are the only feasible way of monitoring birds on a large geographic scale. Unfortunately, by not explicitly modeling p at each site, changes in the count data among sites are confounded with factors that affect p. Therefore, changes in counts at points can be a function of changes in (1) observer efficiency, (2) regional or local habitat, and (3) population density.

Observer Efficiency

All observers count birds differently and differ in their ability to perceive birds. These differences are evident both from field studies (Bart and Schoultz 1984) and from analysis of survey data (Sauer and Bortner 1991, Sauer and others, 1994).

Regional or Local Habitat

It is also easy to document habitat effects on observability of birds. Birds are less observable in dense vegetation. An example of this occurs in the USDI Fish and Wildlife Service Mourning Dove (*Zenaida macroura*) call-count survey, in which data for birds seen are recorded separately from number of birds heard. As expected, distinct regional variation occurs in the relative size of these indices. In the Eastern United States, more birds are heard than seen, but in the Central and Western United States more birds are seen than heard. This suggests that the proportion of birds detected is changing for both variables. Furthermore, there is no reason to expect that variation in detectability between the two indices is consistent, so even their sum may not be a valid index of abundance. Unfortunately, with bird species composition and abundance and detection probabilities all varying among habitats and regions, associations among count data and habitats may not be accurate reflections of actual bird use of habitats.

Biases Associated with Population Density

It has been documented that a smaller proportion of birds are counted as the total number of birds at a stop increases (Bart and Schoultz 1984). This tends to lower p in regions with many birds. It has also been observed, however, that some bird species call more frequently at high population densities (Gates 1966). This increase in p with population size also would invalidate the index.

Other Survey Biases

In addition to the biases associated with the point counting technique discussed above, many other aspects of survey design can also bias maps from survey data. Any large-scale survey is constrained by logistical details such as availability of surveyors and ability to reach locations of sampling sites. These details include (1) variable sampling intensity, (2) temporal change, (3) roadside biases, and (4) appropriate analysis scale.

Regional Differences in Sampling Intensity

It is well known that all large-scale surveys for passerine birds contain extensive regional differences in sampling intensity. The BBS, for example, has a disproportionate number of routes in the Eastern United States and has few samples in northern and intermountain west regions. This bias is also obvious in surveys such as the Audubon Christmas Bird Count, and the Breeding Bird Censuses (Sauer and Droege 1990). This suggests that the validity of maps will differ depending on the region of interest. If maps are used to evaluate year-to-year changes in bird populations, these differences in precision will cause a perception of more predicted shifts in distributions and regional changes in counts in regions with lower sampling intensity.

Temporal Biases

Large-scale surveys tend to sample larger or smaller areas over time in response to changes in participation by volunteers. In particular, both the Audubon Christmas Bird Counts and the BBS have increased in range and participation over time, leading to both more consistent coverage of routes within regions and more routes established on the periphery of the survey. These changes in effort lead to extreme biases in trend estimators based upon regional average counts (Geissler and Noon 1981) and have led to the development of trend estimation procedures that model trends on consistently surveyed areas (Geissler and Sauer 1990). It is also evident that in the BBS, number of species and total counts tend to increase over time, suggesting increases in observer quality and participation (B.G. Peterjohn, personal communication). Maps based upon counts will display these biases.

Roadside Biases

It has been suggested that surveys such as the BBS, in which observers count birds along roadsides, provide a biased view of bird populations because many species are either attracted or repelled by roads (Droege 1990). Also, habitats that do not occur along roads are not sampled. It is clear that habitats are often missed along BBS routes and, therefore, marginal populations of birds near the edges of their ranges are not well sampled by the BBS. If habitats not sampled by surveys do contain population densities that differ from sampled habitats, maps can be distorted.

Bias and Scale of Analysis

The biases discussed above do not necessarily invalidate maps made from point count data. In fact, maps made from

BBS data appear to provide a reasonable view of regional abundances of many species (Robbins and others 1986). We believe that many large-scale geographic questions can be addressed using BBS data. We suggest consideration of the following guidelines, however, for analysis of maps from surveys:

(1) Extrapolations of counts between data points should be viewed with caution. Because *p* can differ between survey locations, differences in counts between routes may not accurately reflect changes in population size, and smoothed values may reflect sampling error rather than real regional variation.

(2) Regional variation in sampling intensity can create the appearance of greater variability in bird populations. Maps created from different time intervals may indicate more variation in bird populations in certain regions as a consequence of fewer samples or poor quality data.

(3) Phenomena that occur at scales smaller than the survey cannot be accurately modeled using survey data. Rare species or species sampled at the edge of their ranges will be poorly mapped. Because of the emphasis on marginal populations in evaluations of changes in ranges, edges of distributions receive special emphasis in biogeographic analysis. Unfortunately, sampling in many extensive surveys is coarse-grained, and the local patches of acceptable habitat in which marginal populations occur are often poorly sampled or missed completely. Edges of range as estimated from surveys are extremely variable, reflecting the poor sampling characteristics of low-density populations.

(4) Bird population "surfaces" are a composite of real populations and differences in sampling attributes of the population. By treating the discrete survey points as continuous functions and modeling a density surface for a species, all of the sampling problems discussed above are incorporated into the estimation. Trend analysis procedures that are structured to accommodate spatial variation in sampling intensity (through area weightings), changes in observers (through covariables), and missing data (by estimating changes over time at individual points) may provide a more reliable view of bird population changes within regions. Maps are conditional on counts, or mean counts, and methods to adjust for these biases do not exist.

Sampling for Maps

In designing any survey to estimate parameters of bird populations, choices must be made about the number of points to be sampled and the dispersion of points. Other papers in these Proceedings have examined allocation of the number of samples (e.g., Barker and Sauer, in this volume), but the dispersion of sampling locations becomes important for sampling for mapping. Geostatisticians have addressed the issue of allocating additional samples to minimize map error when pilot data have been used to define a preliminary kriging model (Barnes 1989). It is clear from this work that it is difficult to make generalizations about sampling for maps, as additional sample locations are dependent upon the model used for the pilot data.

A basic distinction exists between sampling for maps and sampling for other population attributes. When sampling for maps, a model is defined for the covariance structure of the surface, and additional samples (e.g., count locations) are selected to better define attributes of the model. Because the sampling is model-based, optimal sampling for models will introduce bias in the sample if it is used to estimate other attributes that are not model-based (such as population means), which are unbiased only if all locations have an equal chance of occurring in the sample. de Gruijter and ter Braak (1990) review this distinction between design-based and model-based sampling and suggest that design-based sampling is more likely to provide robust estimates of statistical attributes of the population. Because mapping of bird distribution is probably not the principal goal of most surveys, we suggest that model-based sampling procedures such as those suggested by Barnes (1989) not be used for allocation of additional samples in bird surveys. Steps can be taken to minimize error in mapping, however, that do not bias standard sampling.

How Can Point Count Surveys Be Designed to Provide Acceptable Information for Mapping Procedures?

In this section, we demonstrate some of the basic principles of sampling for maps. To give some insights into how sampling affects maps, we will use the simulated surface (PCL) presented in *figure 1*. The actual surface is a tilted plane that has height 0 at *X,Y* coordinates of (0,0) and has height $20(X + Y)/6000$ at point (X,Y). The constant 20 is the maximum height at the coordinates (3000,3000). To illustrate how sampling can affect maps, we conducted a simulation in which we (1) sampled from the surface by taking counts at (X,Y) locations under various conditions, (2) used mapping procedures to estimate a systematic grid and topographic map from the sampled counts, and (3) plotted the maps to provide a visual comparison of the consistency of the estimated maps.

Examples of Effects of Sampling Design on Map Error

Systematic versus Random Sampling

Random sampling is a traditional method of ensuring an unbiased sample. Systematic sampling ensures consistent coverage over a region that may not occur by chance in random sampling with small sample sizes. We illustrate this by simulating 900 sample points on PCL, using both a systematic grid and random points (*fig. 3*). Under these conditions, it is clear that a more consistent map is produced by systematic sampling. Exceptions to this are noted below.

Sample Points

The number of points sampled has an obvious effect on the estimation of any statistical attribute of a population. Comparison of the maps presented in *figure 3* with a map prepared with only 100 points (*fig. 1d*) illustrates the effects of decreased sample sizes on the efficiency of maps.

Detection Probabilities

Point counts do not provide unbiased estimates of the actual number of birds present at a point, because only a proportion of the birds are sampled. We evaluated the effects

USDA Forest Service Gen. Tech. Rep. PSW-GTR-149. 1995

157

of this by considering the counts at a point (Z') to be a binomial random variable, with parameters Z, the predicted height at point (X, Y), and p, the detection probability. To illustrate this, we set p at two levels: 0.8 and 0.5 (*fig. 4*). Compare these results with *figure 1d,* which has the same sampling intensity, but with $p = 1$. As expected, the surface becomes more biased (i.e., differs more from the true surface) and more variable as p gets smaller (*fig. 4*). Variation in detection probabilities over a surface can create serious biases in a map (*fig. 5*).

Replication

When $p < 1.0$, the counts are no longer measured without error at a point. In this case, there may be some advantage to replication at the point, as the mean of several counts is a "better" (i.e., more precise) estimate of Z' than is a single count. We demonstrate by averaging 20 independent "replicates" of Z' at each point (*fig. 6*) for comparison with *figure 4a*. Replication does not eliminate bias, in that the surface based on replicated counts never reaches the height of the real surface. In addition, if p varies within the area of interest, the observed surface is not only proportionately lower than the true surface, but is also distorted.

### Sampling Must Occur at the Appropriate Scale for Detection of the Phenomena of Interest

In nature, no surface is smoothly increasing or declining as is modeled by our PCL surface. Instead, areas of large populations are intermixed with areas of small populations as a function of both biological and geographic features.

Obviously, the more complex the distribution, the larger a sample is needed to describe it adequately. We demonstrate the effects of scale of measurement using PCL with an additional surface feature, a small area with much higher counts than the region around it (*fig. 7a*). Widely spaced sample points might not detect this feature (*fig. 7b*). One solution is to increase the sample size. If a systematic sample is used, and the spacing between sample points is less than the shortest axis of the area of interest, at least one sample point will be within the feature. Alternatively, if small features (*fig. 7a*) are known to exist, stratified sampling can be used and these small areas can be sampled with a higher density of sample points (*fig. 7c*).

Conclusions

Because of the incomplete nature of count data and deficiencies in the design of large-scale bird surveys, it is likely that maps from survey data contain significant biases. These biases should be considered in analyses of ecological attributes of the ranges of birds, and are most likely to be important at small geographic scales.

Maps are useful descriptions, and we believe that they should be produced from survey data. They have great potential for evaluation of large-scale changes in bird distributions over time. However, their deficiencies must always be made explicit. We suggest that maps of bird distributions be treated in the same way that Isaaks and Srivastava (1989:42) treat contour maps of geological data, "as helpful qualitative displays with questionable quantitative significance."
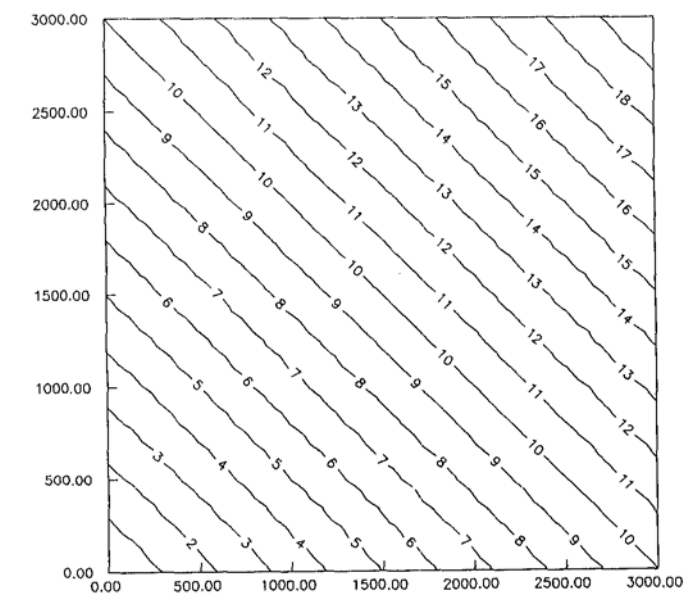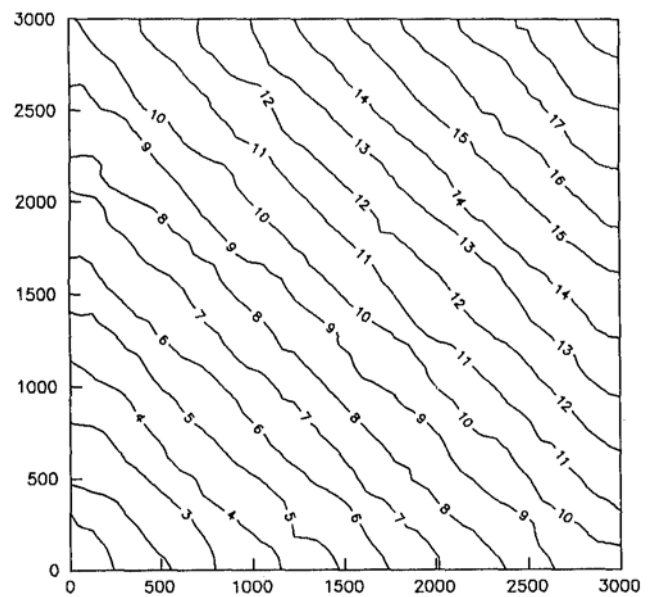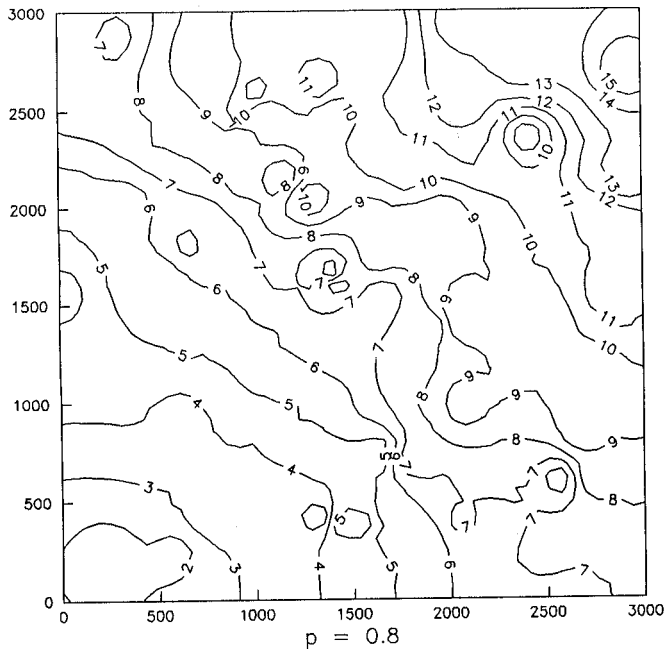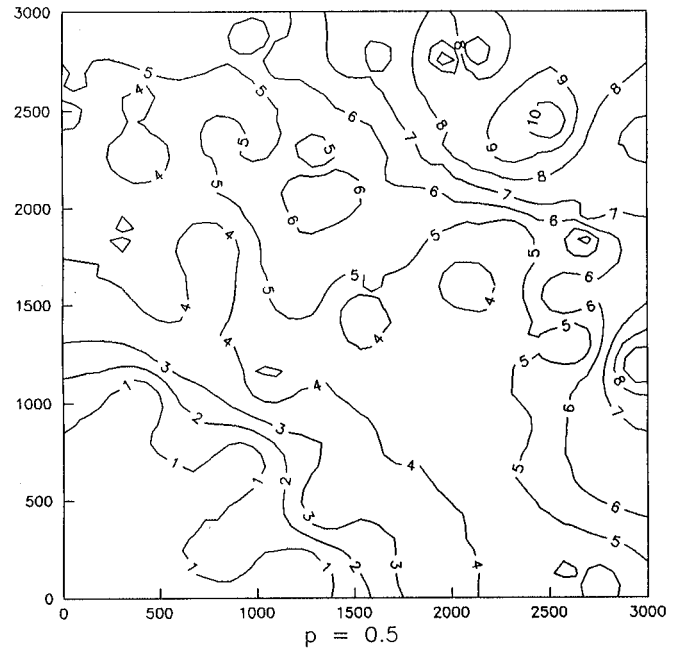

Figure 3--The effects of systematic versus random sampling on maps. (a) A sample contour map based on a systematic sample of 900 points. (b) A sample contour map based on a random sample of 900 points.

158

USDA Forest Service Gen. Tech. Rep. PSW-GTR-149. 1995

a

b

Figure 4--The effects of varying detection probabilities on maps. Both maps were generated from the same 100 randomly located points, but differed in *p*. (a) *p* = 0.8. (b) *p* = 0.5. Compare the surface of these maps with *figure 1d*.
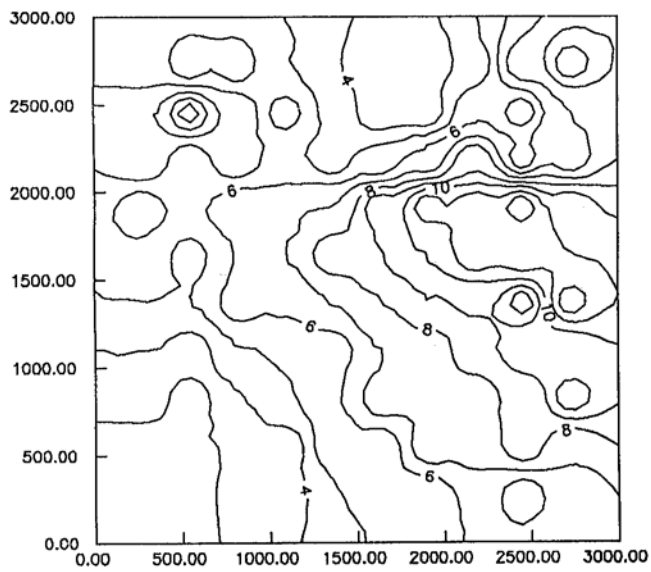

Figure 5--PCL with a systematic sample of 100 points, and *p* = 0.8 on the portion of the map below 2000 on the y-axis. Above 2000, *p* = 0.4.
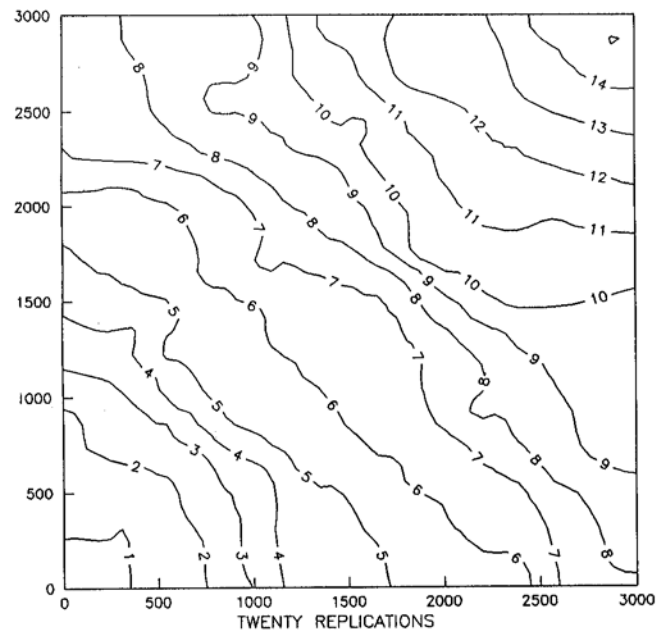

Figure 6--A map based upon similar conditions as in *figure 4a,* but the counts at each point are the average of 20 independent replicates.
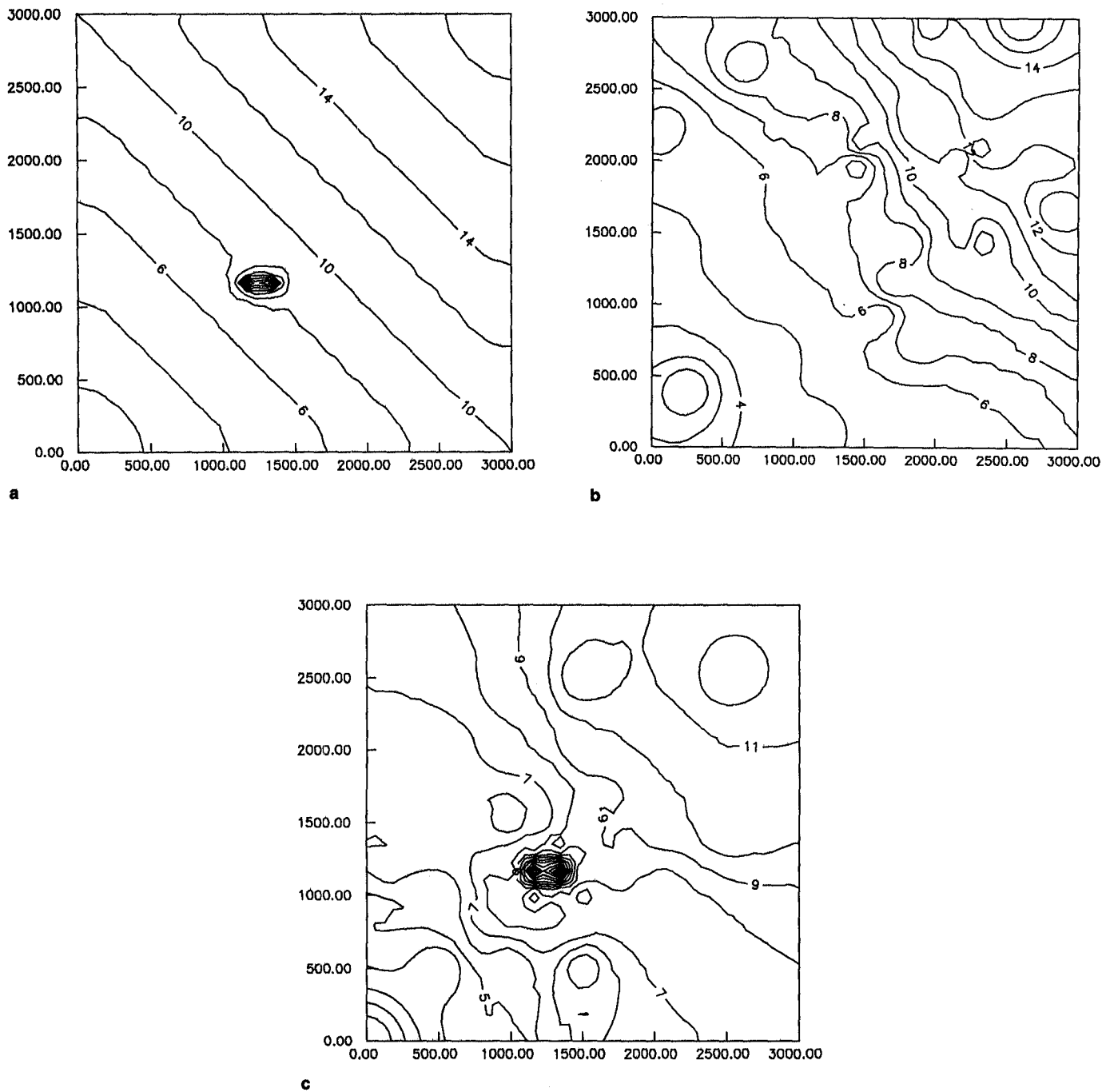
Figure 7--PCL with a raised region that is 3 times the height of the surface. (a) Detail of the surface, as shown by a 900-point systematic sample. (b) A surface produced by a low intensity sample (a 49-point sample), which misses the feature entirely. (c) An example of a stratified sample in which the surface, excluding the raised area, is sampled with 25 points uniformly located, but an additional sample of 20 points are uniformly located around the raised area.