Appendix G

Sampling Design Issues for Section 305(B) Water Quality Monitoring

By
Stephen L. Rathbun

Sampling Design Issues for Section 305(b) Water Quality Monitoring

by

Stephen L. Rathbun

Abstract

State 305(b) water quality monitoring programs typically employ judgment sampling designs, in which sample sites are selected according to a number of often vaguely defined criteria. The resulting data are likely to yield biased estimates of parameters such as the percent of the water resources that are satisfactory for their designated uses (e.g., swimming, drinking, fishing, etc.). Moreover, there is no statistically justifiable method for combining such data across states as mandated by Section 305(b) of the Clean Water Act. This paper describes how probability-based sampling designs can be implemented to sample water resources. A diverse variety of probability-based sampling designs are available, the scientific judgment of the investigator can be taken into account during the selection of strata, and multiple-stage designs can be used to reduce sampling costs. Data resulting from probability-based sampling

designs can be used to obtain unbiased estimates of such quantities as the percent of water resources meeting environmental criteria for designated uses, and the total mass of a chemical contaminant is a state's water resources. Moreover, data from the various states can be easily combined even if different states use different probability-based sampling designs. Despite these advantages, managers of state water quality monitoring programs are reluctant to implement probability-based sampling designs. Much of this reluctance stems from the fear that information from the historical data base will be lost. A procedure for combining data from probability-based and judgment sampling designs is demonstrated. This procedure exploits spatio-temporal correlation among the observations from both data bases to back predict what data would have been obtained had a probability-based sampling design been implemented from the very beginning of the monitoring program.

Contents

1 Introduction

Section 305(b) of the 1972 Federal Water Pollution Control Act (usually known as the Clean Water Act) mandates that each state submit a surface water quality assessment report to the Environmental Protection Agency (EPA) every two years, and that the EPA submit a comprehensive assessment of the condition of the nation's waters to Congress every two years. The latter requires the combining of data obtained by the former as well as various native American tribes. However, current state monitoring efforts present a number of obstacles to combining data at a national level in a statistically defensible manner. Many of the obstacles arise from differences in the objectives among the states and between the states and the EPA. While the EPA is required to report on the condition of the totality of all of the nation's aquatic resources, states and tribes often select monitoring stations based on local purposes

(305(b) Consistency Workgroup, 1996).

The combination of data across states and tribes would be straightforward provided all states and tribes used probability-based sampling designs, provided that there is some consistency in what variables are measured and how they are measured, and provided that there is consistency in the definitions of target populations and sample units. It is not necessary that all states employ the same probability-based design, and so, states are free to implement designs tailored to their local requirements. However, few states or tribes employ probability-based sampling designs, and for most states and tribes, the sample population covers less than 100% of their water resources. Consequently, the representativeness of the current monitoring stations must be questioned. Statistical inference is limited to statements, for example, regarding the percentage of sample sites showing impaired conditions, and not the percentage the state's water resources that show impaired conditions. Efforts to combine data across states and tribes are also impaired by variation among states and tribes in site selection criteria, definition of target populations and strata, what variables are measured, sampling protocols, and analytical laboratory procedures. Some limitations are biological: there is considerable natural variation among states in the composition of their biota (i.e., what species are present) irrespective of anthropogenic effects. Moreover, different states have different types of water resources (e.g., estuaries in coastal states, mountain streams in states having mountains, etc.), and different

resource types are likely to respond differently to environmental insults.

Recent years have seen new efforts to improve the quality of Section 305(b) water resource monitoring. In 1992, the Intergovernmental Task Force on Water Quality (ITFWQ) was established in response to Office of Management and Budget Memorandum 92-01. Cochaired by the EPA and the United States Geological Service (USGS), the ITFWQ is charged with the review and evaluation of national water quality monitoring efforts, and to recommend improvements. They have recommended that Section 305(b) change from the current 2-year reporting cycle to a 5-year reporting cycle. This would help states achieve better coverage of their water resources through the implementation of rotating panel and similar designs (see Section 3.3) under which 1/5 of the sample sites are monitored each year. The EPA has also established a 305(b) Consistency Workgroup, which as its name implies is tasked with improving the consistency of Section 305(b) water quality monitoring among the states and tribes. The 305(b) Consistency Workgroup is also exploring the implementation of probability-based designs.

This paper considers issues regarding the replacement of current judgment sampling designs used by most state water resource monitoring programs with probability-based sampling designs. This together with efforts to improve consistency among state and tribal programs in their sampling protocols, analytical laboratory procedures, definitions of target populations, etc., would facilitate future efforts to combine

data across states and tribes. Of particular concern is how we might replace current sampling designs with minimal loss of historical monitoring information. Methods are developed for combining historical data with new probabilistic data to obtain predictions of what data would have been obtained had a probability-based design been implemented in the very beginning of the monitoring program. Although it is intended that sampling at judgment sample sites be discontinued at some point in the future, sampling at a subset of such sites could continued to address site specific questions and for purposes of model building. This paper does not consider methods for combining judgment sample data with probability sample data collected during the same time interval to improve estimates at that time interval. For a discussion of such methods, see Overton, Young, and Overton (1993) and Cox Pieogorsch (1996).

After describing the available data in Section 2, Section 3 provides a general discussion of survey designs including those for sampling over space and time. The current status of 305(b) water quality monitoring efforts is discussed in Section 4; this includes a response to the concerns raised by the 305(b) Consistency Workgroup regarding the replacement of current judgment sampling designs by probability based sampling designs, and a discussion of how data may be combined across state under probability-based sampling. Section 5 gives some specific design alternatives for sampling lakes and streams, including designs that involve sampling at access points. Methods for combining historical judgment data with new probabilistic data are con-

sidered in Section 6.

2 Available Data

The Regional Environmental Monitoring and Assessment Program (REMAP), and the Clean Lakes Program provide data on Secchi depth from lakes in the Savannah River Basin. Secchi depth is a measure of water clarity. It is obtained by dropping a Secchi disk over the side of a boat and measuring the depth at which the disk is no longer visible.

2.1 Savannah River Initiative (REMAP)

The Savannah River Initiative of the Regional Environmental Monitoring and Assessment Program is sponsored by the Environmental Protection Agency. Data on chlorophyll A and Secchi depth was collected in July 17-21, 1995 and June 24 to July 5, 1996. Each year, 37-40 sites were sampled from the embayments of large lakes in the Savannah River Basin, including Russell, Thurmond, Hartwell, Keowee, Jocassee, and Burton. Sample sites were selected according to the two-tiered sampling design. A $7 \times 7 \times 7$ fold enhancement of the EMAP base grid was placed over the Savannah River Basin. Each grid point is circumscribed by a 1.86 km^2 hexagon; 7 of these hexagons form a 13 km^2 hexal, and 7 hexals form a 635 km^2 EMAP hexagon. The tier 1 sample is comprised of 3 randomly selected 13 km^2 hexals from each of the

635 km^2 EMAP hexagons covering the Savannah River Basin. All embayments were enumerated within each of the selected hexals. The tier 2 sample of embayments to be sampled each year was then selected using the procedure of Larsen and Christie (1993).

2.2 Clean Lakes Program

The Clean Lakes Program is sponsored by the South Carolina Department of Health and Environmental Control (SC-DHEC) and the Environmental Protection Agency. This program involves the collection of data used to evaluate the quality of lake water in South Carolina. Secchi depth was observed at 17 sites located in five large lakes in South Carolina. These sites were selected according to a judgment sampling design favoring the main channels of each lake. At each site, 0-2 monthly observations were collected between April and October of each year. The length of the data records depends on the sample site. This study was initiated at 10 sample sites scattered throughout lakes Russell, Hartwell, Keowee, and Jocassee in April 1991. Three additional sample sites were added in May 1992, one in Lake Russell and two in Lake Keowee. Three sites in Broadway Lake were sampled only in 1994, and one site in Lake Hartwell was sampled in 1993. In addition to Secchi depth, chlorophyll A was measured occasionally, but records of this variable were too sparse to warrant further analysis.

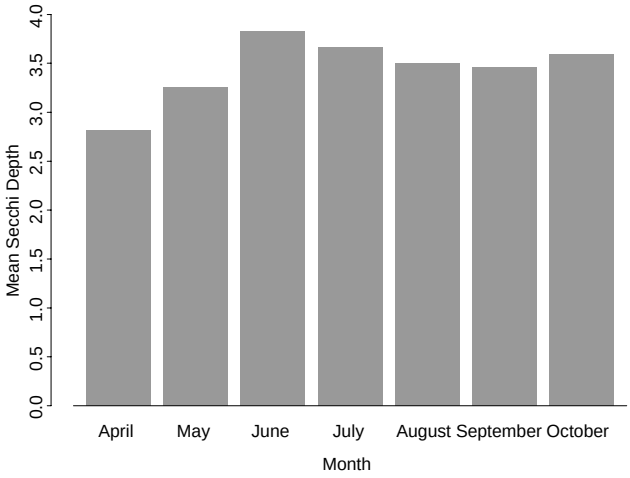

Figure 1: Mean secchi depth by month.

The seasonal pattern of variation in Secchi depth is illustrated in Figure 1. Monthly means were adjusted to take into account variation among years in what sites were included the sample. Mean Secchi depth was lowest in April at 2.82 m, increased to a maximum of 3.83 m in June, then decreased to an asymptote of approximately 3.5 m thereafter.

3 General Design Issues

Environmental monitoring programs should be designed within the context of their objectives in such a way as to optimize the amount of information they yield about the resource of interest. The objectives may call for the selection of specific sites of interest, for example sites near point sources of environmental contamination. For

the latter, pairs of sites are often employed, one immediately downstream and the second upstream of the point source. In such cases, inferences are restricted to the environmental conditions that occur at those specific sites, and may include comparisons between upstream and downstream sites. When interest is restricted to specific sites, sufficient monitoring resources should be made available to sample all of these sites. If, however, the objectives call for inferences regarding the status of the environment on a regional scale, sufficient monitoring resources are not available to census all of the waters in that region. In such cases, a sample of sites must be selected. To guarantee unbiased estimates of status, this sample of sites must be selected using a probability-based sampling design. Probability-based designs involve some method of random selection of sample sites, but are not restricted to simple random sampling. Probability-based sampling designs may be used to estimate the mean value of an environmental parameter in the lakes of a region of interest, the percent of stream miles that have impaired environmental conditions, the total mass of a contaminant in the estuaries in a study region, or the percent of the area of lakes showing improving environmental conditions. Probability-based sampling designs are most appropriate for investigating nonpoint sources of environmental contamination and can also be used to select reference sites for the investigation of the impact of point sources of environmental contamination.

Nonprobability-based sampling designs must rely on the judgment of the investi-

gator. Such judgment sampling designs are not likely to yield a representative sample, and hence, can lead to biased estimates of population parameters. Unbiased estimation of environmental parameters under judgement sampling requires the assumption that the population or region of interest is homogeneous, an assumption that seems unlikely to be tenable in nature.

3.1 Statistical Inference

Two types of statistical inference can be distinguished, design-based and model-based. Design-based inference requires that data be obtained under a probability-based sampling design. Under design-based inference, the values of the variable of interest in the population or region of interest are assumed to be fixed and nonrandom. Here, the source of random variation comes from the random selection of sample sites. Since the sampling design is specified by the investigator, and hence is known, no model assumptions are required. Design-based inferences are made on the actual population or region from which the sample was drawn, and not on the parameters of some assumed model. Such inferences may include unbiased estimates of the mean value of an environmental parameter in the lakes of a region of interest, the percent of stream miles that have impaired environmental conditions, the total mass of a contaminant in the estuaries of a region of interest, or the percent of the area of lakes showing improving environmental conditions. Standard errors and confidence intervals are

available for all of these population parameters. Design-based hypothesis testing procedures test how likely that a sample with the observed data could have been drawn from a population with the null parameter value(s). Since inferences are restricted to the population from which the sample was drawn, design-based inference cannot be used to predict future observations or data at unsampled sites.

Under model-based inference, it is assumed that the data are realized from some random model. In multiple regression, for example, the variable of interest is assumed to be a linear function of some explanatory variables plus a random error. Further assumptions may include the homoskedasticity of the errors, and that the data are uncorrelated and normally distributed. However, we may wish to assume that data are spatially and temporally correlated, in which case, assumptions are required regarding the specific correlation structure. Instead of making inferences about the region from which the data were obtained, model-based inferences are made on model parameters. Such inferences may include estimates of the model parameters, together with their corresponding standard errors, as well as predictions of future observations and data at unsampled sites. Model-based hypothesis testing procedures test whether or not the data are compatible with a null model. Although model-based inferences are available for both probability-based and nonprobability-based sampling designs, parameter estimates can be biased under the latter. Typically, model-based inferences ignore variability due to random selection of sample sites.

3.2 Examples of Probability-Based Designs

A wide variety of probability-based sampling designs are available. The *simple random sampling design* is the most basic method for selecting sample sites from a region. For rectangular study regions, a simple random sample is obtained by random and independent selection of X, Y coordinates from. For irregularly shaped regions, locations are sampled from the smallest rectangular region until a sufficient number of sites are located in the study region (Figure 2); only those sites falling in the study region are retained in the sample. Subregions will tend to be sampled in proportion to their areas; for example, if 40% of the region is in loamy soils, then we expect 40% of the sample sites to fall on loamy soils. Aside from the selection of the study region, the selection of sample sites does not involve any scientific judgment.

The selection of a probability-based design need not, and should not ignore the scientific judgment of the investigator. Under a *stratified random sampling design,* the region is partitioned into strata, often corresponding the different habitats of interest. For example, streams may be partitioned into first-, second-, and third-order streams, while lakes may be partitioned by trophic level, ecoregion, size, access (public or private), or whether they are natural or man-made. The wetlands surrounding the Carolina Bay in Figure 3 are partitioned into five "undisturbed" habitat types. Sample units are then selected from each stratum according to a some probability-based sampling design; a simple random sampling deign is used in Figure 3. Here, scientific

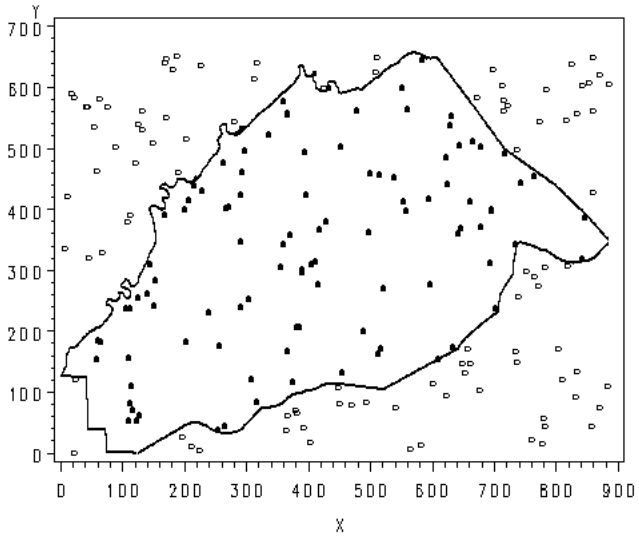

Figure 2: Simple random sample of 100 sites in Ebenezer Aquifer (closed circles). Sites falling outside the study region (open circles) are excluded from the sample.

judgment is required for optimal selection of strata. Strata should be selected in such a way that differences between strata are as large as possible, while units within strata are as uniform as possible. By controlling for differences among strata, the stratified random sampling design reduces the sampling variance and hence improves the precision of population parameter estimates. Therefore, a stratified random sampling design can achieve the same precision at a smaller sample size than a simple random sampling design, and hence reduce costs.

The optimum allocation of sampling effort among strata requires the within stratum variances of the variables of interest, information that is not likely to be available at the beginning of a new monitoring program. However, allocation proportional to stratum size works well, and sample allocation may adjusted as data are obtained. It is almost certain that different variables will yield different optimal allocation schemes, and so, some compromise allocation scheme Costs may be reduced by decreasing sampling effort in expensive strata, and increasing sampling effort in cheap strata. By adjusting the allocation of sampling effort to the different strata, we may increase the sampling effort in ecologically important strata, and ensure that an adequate sample is obtained from rare habitats.

Another way in which the cost of sampling efforts can be reduced is to employ a double sampling design. Double sampling can be used when a inexpensive ancillary variable is available as a surrogate for the variable of interest. For example, Secchi

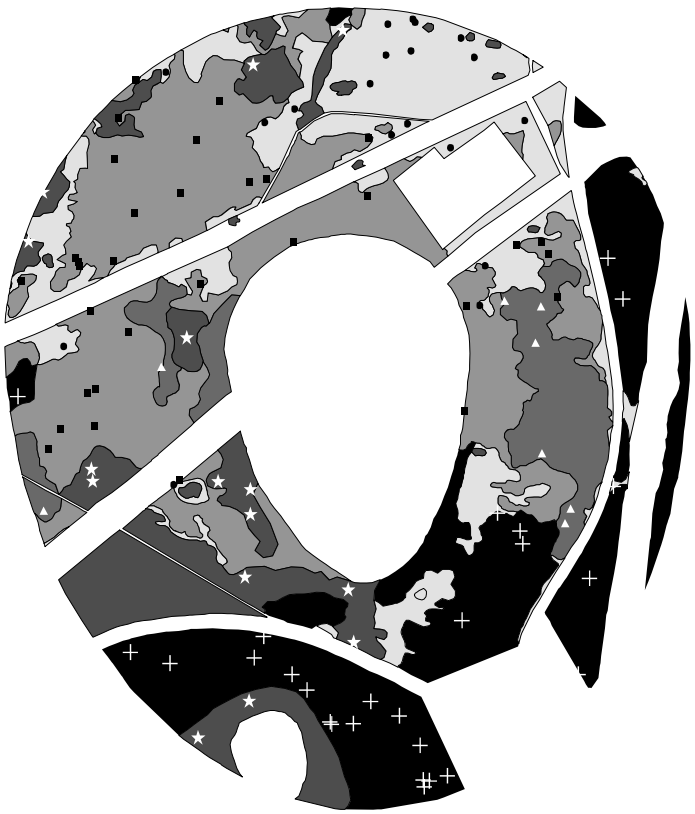

Figure 3: Stratified random sampling design in the wetlands surrounding a Carolina Bay. Circles are in grasslands, squares are in briars and shrubs, triangles are in vines and small trees, stars are in hardwoods and pines, and crosses are in pines.

depth, which is inexpensive to obtain, may be an ancillary variable for total suspended solids or Chlorophyll A, which require more expensive equipment and laboratory procedures. Under a *double sampling design,* primary sample sites are first selected according to any sampling design, then secondary sample sites are obtained by taking a simple random sample of the primary sites (Figure 4). Both the ancillary variable and the variable of interest are measured at the secondary sample sites, while only the ancillary variable is measured at the primary sample sites. Under double sampling, parameter estimation relies on the correlation between the variable of interest and the ancillary variable. The ratio of secondary over primary sample sites depends on the cost of obtaining the variable interest relative to the cost of the ancillary variable, and on the magnitude of the correlation between the two variables. As the cost of the variable of interest increases and the correlation increases, the optimal ratio of secondary over primary sample sites decreases.

The above sampling designs require maps depicting all of the state's water resources, from which a listing of all lakes, stream reaches, and estuaries may be obtained. Such information might be obtained from River Reach File Version 3 (RF3) (Horn and Grayman 1993). This file is not perfect; information on new man-made reservoirs, small lakes, and higher order streams may be missing, and it also includes some lower order ephemeral streams that may not be present if sought on the ground. In any case, the information contained in RF3 should be verified on the ground, and

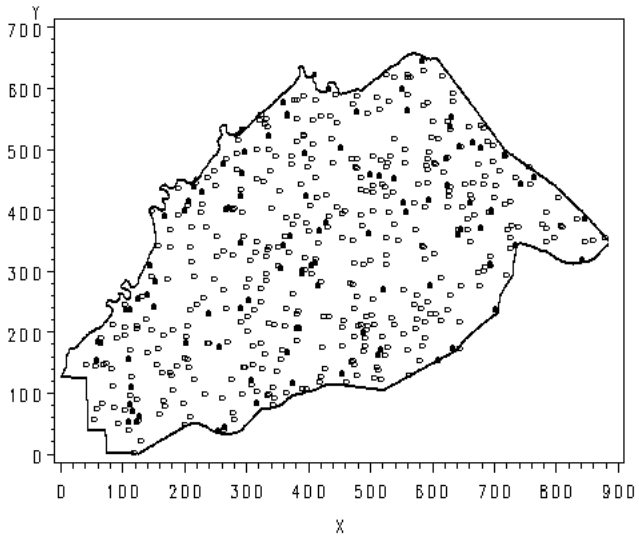

Figure 4: Double sampling design in Ebenezer Aquifer. Primary sample sites are designated by open circles, and secondary sample sites are designated by closed circles.

an effort should be made to fill in any missing information. If point sources of contamination are of concern, then a list of all point sources is required, information that is not available from RF3.

It may not be practical to be obtain such a list frame of all water resources within a state. Multiple-stage sampling designs do not require list frames of all water resources, and hence, may be more practical for water resource monitoring. Under a *two-stage sampling design* (a special case of a multiple-stage sampling design), the population is first partitioned into primary sample units, then a simple random sample of primary units is selected, and finally, a simple random sample is obtained from each of the selected primary units. Thus, water resources only need to be enumerated within each of the selected primary units. Primary units should be small enough so that all water resources within each of them can be easily enumerated. The flexibility of multiple-stage sampling designs is illustrated by the following examples:

- To investigate the trophic levels of all small lakes in a region, the state may first be partitioned into hexagons. Then a simple random sample of hexagons is selected. The small lakes are enumerated within each of the selected hexagons, and then a simple random sample of lakes is obtained from each of the selected hexagons. Finally water samples are collected from each of the selected lakes.

- To investigate point sources of environmental contamination, the state may first be partitioned into Natural Resource Conservation Service (NRCS) watershed

units. Then a simple random sample of watershed units is selected. The point sources are enumerated within each of the selected watershed units, and then a simple random sample of the point sources is obtained. Finally, water samples may be obtained upstream and downstream of the selected point sources.

- To investigate the soils of Ebenezer Aquifer, n parallel line transects may be randomly located within the aquifer, and then m soil samples may be randomly selected along the length of each transect (Figure). Here, the transects are treated as the primary sample units.

From the third example above, observe that the transect sampling design familiar to ecologists is a special case of a two-stage sampling design. Two-stage sampling designs can be extended into multiple stage designs by further partitioning each of the sampled primary units into secondary units, partitioning sampled secondary units into tertiary units, and so on. At each stage, a simple random sample of the units defined that stage is obtained. Multiple-stage sampling designs may be modified to allow stratified random sampling during any stage of the design.

Under the above conventional sampling designs, the sample selection procedure does not depend on the observations obtained during the course of the survey. Under adaptive sampling designs, however, the selection of future sample sites depends on the observations that have been obtained up to the present time. Adaptive cluster sampling designs are particularly suitable for the investigation of highly localized

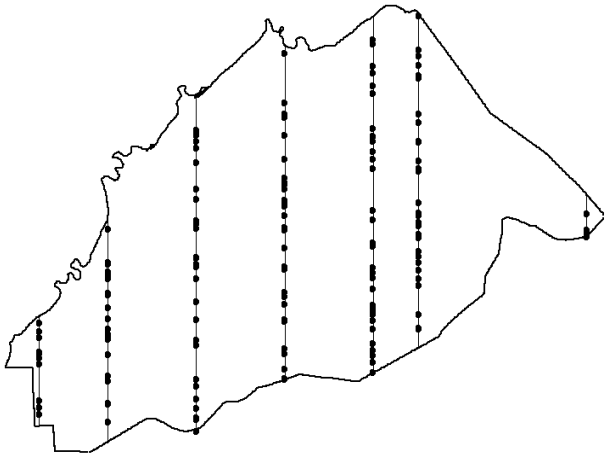

Figure 5: Line transect design in Ebenezer Aquifer.

phenomena such as clusters of a rare species or hot spots of highly contaminated environmental resources (Thompson 1990, 1992). Under an *adaptive cluster sampling design,* a simple random sample of locations is first selected (Figure 6). If a given sample site satisfies a given condition (i.e., presence of a rare species, or high levels of contamination), addition sample sites are clustered around that site. This process is repeated with the new sample locations until no new sites are added which satisfy the criterion.

The above examples illustrate just a fraction of the diversity of available probability-based sampling designs. Probability-based sampling designs can be tailored for almost any scientific situation and can be constructed in response to many budgetary and scientific constraints.

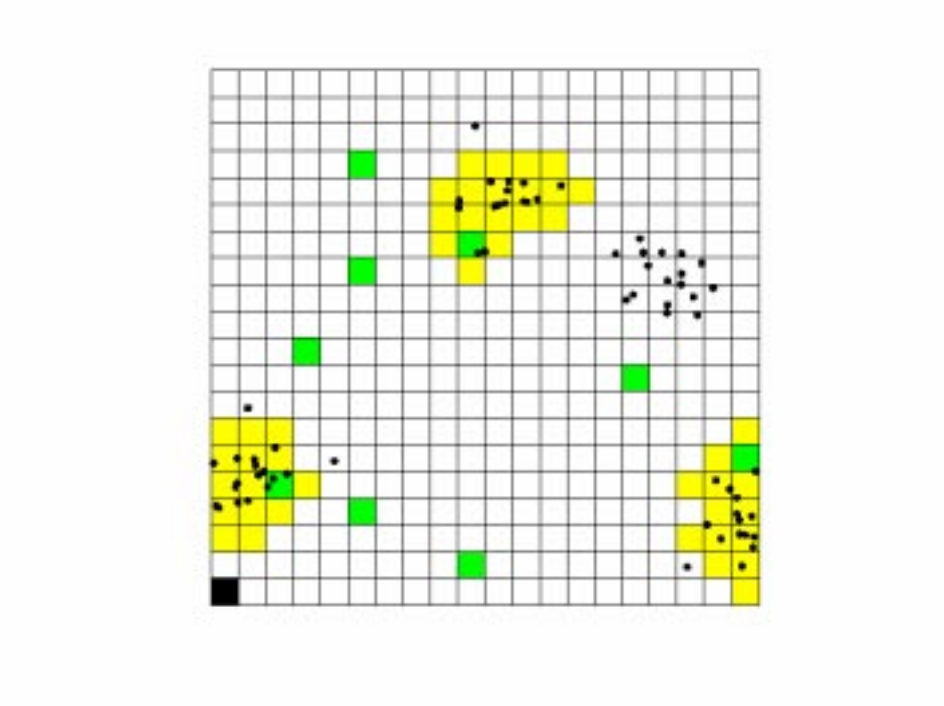

Figure 6: Adaptive cluster sampling design. First 10 sample units are selected at random (dark shaded squares). Then adjacent unit are added to the sample whenever one or more points are observed in the selected unit (light shaded squares).

3.3 Sampling over Space and Time

So far, we have only considered probability-based designs for selecting sample sites at a given point in time. Here, we shall consider the allocation of sampling effort over space and time. There are at least four approaches to sampling over space and time:

- *Permanent Stations*: A sample of n permanent sample stations are selected from some probability-based design; data are collected from each sample station during every sample interval.

- *Serially Alternating Design:* Sample stations are selected from some probability-based design and are partitioned into m sets of equal size n. Set i is then sampled during intervals $i, i+m, i+2m, \cdots$, as shown in Table 1 (Rao and Graham 1964). This design was proposed for the Environmental Monitoring and Assessment Program (EMAP) (Messer et. al 1991); here EMAP hexagons are partitioned into m sets of size n, and hexagons are sampled as described above.

- *Rotating Panel Design:* Sample stations are initially selected from some probability-based design and are partitioned into m sets of equal size n. During each sample interval, one set of sites is dropped from the sample, and is replaced by an additional set of n sites selected from the probability-based design as shown in Table 2 (Skalski 1990).

	Sampling Interval (ie., year, month, season)											
Set	1	2	3	4	5	6	7	8	9	10	11	12
1	X				X				X			
2		X				X				X		
3			X				X				X	
4				X				X				X

Table 1: Serially alternating design.

- *Ever-Changing Stations.* Under this sample design, a new and independent probability sample is obtained during each sample interval.

The latter three sample designs can be augmented by selection of additional permanent sample stations which are to be sampled during each interval (Urquhart, Overton, and Birkes 1993).

The various alternatives to spatio-temporal sampling offer a number of advantages and disadvantages with respect to design-based inference and spatio-temporal modeling and prediction. Correlation matrices are of block-Toeplitz form under permanent station and serially alternating designs, and so, computationally more efficient algorithms may be used during spatio-temporal modeling. If temporal trends are expected to depend on location, permanent station and serially alternating designs are most

	Sampling Interval (ie., year, month, season)											
Set	1	2	3	4	5	6	7	8	9	10	11	12
1	X											
2	X	X										
3	X	X	X									
4	X	X	X	X								
5		X	X	X	X							
6			X	X	X	X						
7				X	X	X	X					
8					X	X	X	X				
9						X	X	X	X			
10							X	X	X	X		
11								X	X	X	X	
12									X	X	X	X
13										X	X	X
14											X	X
15												X

Table 2: Rotating panel design.

suitable for estimating such quantities as the proportion of stream miles showing improving or degrading environmental conditions. Permanent station designs yield the smallest level of spatial coverage, while the greatest level of spatial coverage is obtained under ever-changing station designs. Repeated sampling at permant stations may have an impact on the local environments at those stations, or example, through the trampling of sensitive vegetation by observers, or through modification of the behavior of people knowing the locations of those stations.

The optimal allocation of sampling effort over space and time depends on the relative magnitude of spatial and temporal autocorrelation. This spatio-temporal autocorrelation comes from the observation that data close together in space or time are likely to be more similar than data collected far apart over space or time. Under strong temporal autocorrelation, repeated observations at a given site will contain a large amount of redundant information, and so the optimal design will sample a large number of sites at infrequent times. In contrast, when spatial autocorrelation is strong, data collected at different locations at a given point in time will contain are large amount of redundant information, and so, the optimal design will consist of a few sights that are sampled frequently. To quantify the optimal allocation of sampling effort over space and time, we require estimates of the relative magnitudes of spatial and temporal autocorrelation. The following considers Secchi depth data from two environmental monitoring programs involving the lakes of the Savannah River Basin,

the Clean Lakes Program of SC-DHEC, and the Regional Environmental Monitoring and Assessment Program sponsored by EPA.

Data from the 17 sites of the Clean Lakes Program were used to estimate magnitude of temporal correlation in Secchi depth. Observations were not collected at a sufficient number of sites to effectively model spatial correlation using this data. The Secchi depth $Z(\mathbf{s}_i, t_{jk})$ at site $\mathbf{s}_i$ and time t_{jk} (month j in year k) was fit to the linear model

$$Z(\mathbf{s}_i, t_{jk}) = \mu + \alpha_i + \tau_j + \varepsilon(\mathbf{s}_i, t_{jk})$$

where μ is the overall mean, α_i is the effect of site i, τ_j is the effect of month j, and $\varepsilon(\mathbf{s}_i, t_{jk})$ is the model error. The year of the observation did not enter significantly into the model. Temporal dependence in between the data at times t and t' at site $\mathbf{s}$ may be modeled through the temporal variogram

$$2\gamma_t(|t - t'|) = \text{var}\{Z(\mathbf{s},t) - Z(\mathbf{s}, t')\};$$

assume that the variogram depends only on the difference $|t - t'|$ between the two points in time. In general, there will be little variability (high autocorrelation) between data at times that are close together, and hence the temporal variogram will be small for short time lags. Conversely, there will be high variability (low autocorrelation) between data at times that are far apart, and hence the variogram will tend to be an increasing function of time lag. If temporal trends are adequately modeled, then the variogram will tend to approach an asymptote as the time lag increases; the

time it takes to approach that asymptote is the range of temporal correlation. Pairs of observations further apart than the range of temporal correlation are negligibly correlated.

A nonparametric estimate of the variogram can be obtained from the residuals

$$\widehat{\varepsilon}(\mathbf{s}_i, t_{jk}) = Z(\mathbf{s}_i, t_{jk}) - \widehat{\alpha}_i - \widehat{\tau}_j,$$

where $\widehat{\alpha}_i$ and $\widehat{\tau}_j$ are the ordinary least squares estimates of the parameters α_i and τ_j, respectively. Then the temporal variogram at site $\mathbf{s}_i$ may be estimated by

$$2\widehat{\gamma}_i(r) = \frac{1}{N_i(r)} \sum_{j,k} |\widehat{\varepsilon}(\mathbf{s}_i, t_{jk}) - \widehat{\varepsilon}(\mathbf{s}_i, t_{jk} + r)|^2,$$

where $N_i(r)$ is the number of pairs of observations lag r apart in time at site $\mathbf{s}_i$. A pooled estimate of the temporal variogram over all n sites may then be obtained from

$$2\widehat{\gamma}_t(r) = \frac{2\sum_{i=1}^{n} N_i(r)\widehat{\gamma}_i(r)}{\sum_{i=1}^{n} N_i(r)}.$$

Figure 7 gives the nonparametric estimate of the temporal variogram for the Clean Lakes program data (closed circles). The curved line gives the least squares fit of the Gaussian variogram model

$$2\gamma_t(r) = c_0 + c_g(1 - e^{-\alpha r^2}). \tag{1}$$

Estimates of the variogram parameters are $\widehat{c}_0 = 0.2815$, $\widehat{c}_g = 0.207$, and $\widehat{\alpha} = 0.144$. The large nugget effect of $\widehat{c}_0 = 0.2815$ suggests that there is a large amount of measurement error, or short-term variability in Secchi depth. The estimate of α corresponds to

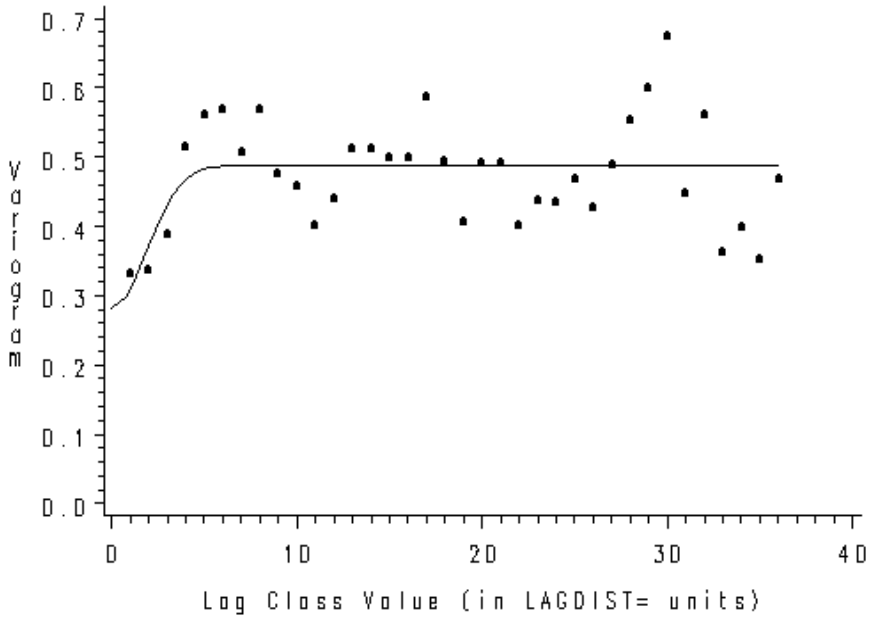

Figure 7: Temporal variogram for Clean Lakes Program data. The closed circles give the nonparametric estimates, while the curved line gives the fitted variogram model.

a range of temporal autocorrelation of $\sqrt{3/\widehat{\alpha}} = 4.6$ months; observations more than 4.6 months apart are negligibly correlated (correlations are less than $e^{-3} \cong 0.05$). Estimated monthly (Table 3) means show the same pattern as in Figure 1; Secchi depth is lowest in April, increases to a maximum in June, and then decreases to an asymptote.

Month	Mean (m)	Standard Error
April	2.831	0.092
May	3.247	0.090
June	3.839	0.093
July	3.684	0.093
August	3.499	0.096
September	3.454	0.091
October	3.583	0.094

Table 3: Mean secchi depth by month for Clean Lakes Program data

The REMAP data was used to model spatial correlation in Secchi depth. REMAP observations were not collected at a sufficient number of times to effectively model temporal autocorrelation. Moreover, the above results of analysis of the Clean Lakes Program data suggest that the range of temporal autocorrelation is only 4.6 months, which is shorter than the one-year time interval separating the REMAP observations. The Secchi depth $Z(\mathbf{s}_i, t_j)$ at location $\mathbf{s}_j$ and year t_j was fit to the linear model

$$Z(\mathbf{s}_i, t_j) = \mu + \tau_j + \varepsilon(\mathbf{s}_i, t_j),$$

where μ is the overall mean, τ_j is the effect of year j, and $\varepsilon(\mathbf{s}_i, t_j)$ is the model error. The spatial dependence between data at locations $\mathbf{s}$ and $\mathbf{u}$ at a given time t is

modeled through the spatial variogram

$$2\gamma_s(\|\mathbf{s}-\mathbf{u}\|) = \text{var}\{Z(\mathbf{s},t) - Z(\mathbf{u},t)\};$$

assume that $2\gamma_s$ depends only on the distance $\|\mathbf{s}-\mathbf{u}\|$ between the two locations. In general, there will be little variability (high spatial autocorrelation) between data at close locations, and hence the temporal variogram will be small for short distance lags. Conversely, there will be high variability (low spatial autocorrelation) between data at far apart locations, and hence the variogram will tend to be an increasing function of distance lag. If spatial trends are adequately modeled, then the variogram will tend to approach an asymptote as the time lag increases; the distance at which it approaches that asymptote is the range of spatial correlation. Pairs of observations further apart than the range of spatial autocorrelation are negligibly correlated.

A nonparametric estimate of the spatial variogram at lag distance d, and at time t_j may be obtained from

$$2\widehat{\gamma}_j(d) = \frac{1}{N_j(d)} \sum_{i,k} |Z(\mathbf{s}_i,t_j) - Z(\mathbf{s}_k,t_j)|^2 ,$$

where the sum is over all pairs of sites approximately d apart, and $N_j(d)$ is the number of such pairs of sites. A pooled estimate of the spatial variogram over all sampling intervals may then be obtained from

$$2\widehat{\gamma}_s(d) = \frac{2\sum_{j=1}^{T} N_j(d)\widehat{\gamma}_j(d)}{\sum_{j=1}^{T} N_j(d)}.$$

Figure 8 give the nonparametric estimate of the spatial variogram for the REMAP data (closed circles). The curved line gives the weighted least squares fit of the exponential variogram model

$$2\gamma_s(d) = c_0 + c_e(1 - e^{-\alpha d}). \tag{2}$$

Restricted maximum likelihood estimates of the variogram parameters are $\widehat{c}_0 = 0.726$, $\widehat{c}_e = 1.2937$, and $\widehat{\alpha} = 0.0797$. The large nugget effect of $\widehat{c}_0 = 0.726$ suggests that there is a large amount of measurement error, or microscale spatial variability in Secchi depth. The estimate of α corresponds to a range of temporal correlation of $3/\widehat{\alpha} = 37.7$ km; observations more than 37.7 km apart are negligibly correlated (correlations are less than $e^{-3} \cong 0.05$).

The results described above show that Secchi depth exhibits both strong spatial and temporal correlation in lakes of the Savannah River basin. This correlation suggests that there is some redundancy in the data. The level of redundancy may be quantified by computing the effective sample size, which is defined to be the number of independent samples required to achieve the same precision of parameter estimate as a sample of correlated observations of a given sample size. Consider, for example, model based estimation of the mean. The variance of the sample mean of n uncorrelated observations is equal to

$$V_1 = \frac{\sigma^2}{n},$$

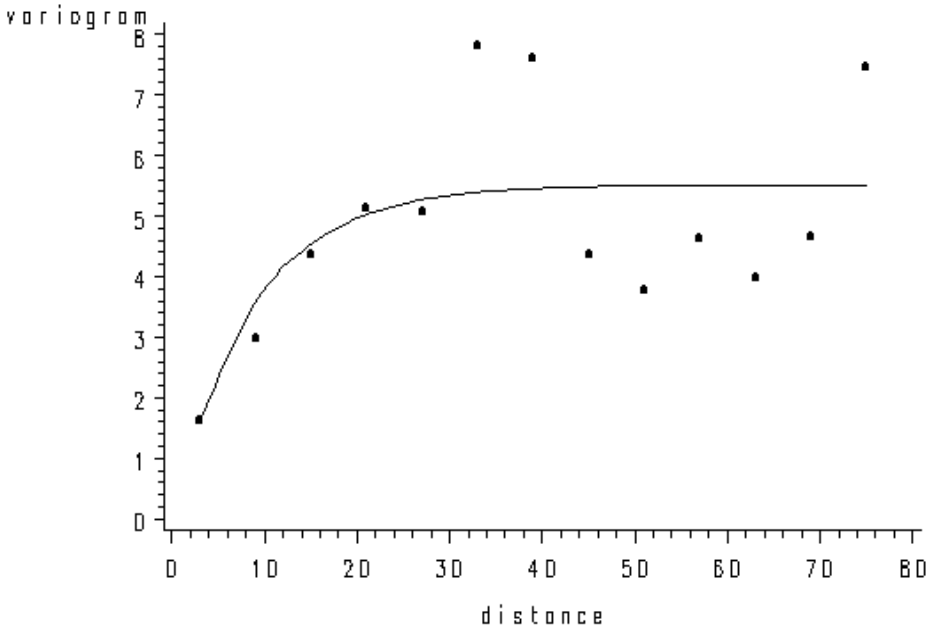

Figure 8: Spatial Variogram for REMAP data. The closed circles give the nonparametric estimates, while the curved line gives the fitted variogram model.

while the variance of the sample of n correlated observations is equal to

$$V_2 = \frac{\sigma^2}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij},$$

where σ^2 is the population variance and ρ_{ij} is the correlation between observations i and j. Then the effective sample size is equal to

$$n \times \frac{V_1}{V_2} = \frac{n^2}{\sum_{i=1}^{n} \sum_{j=1}^{n} \rho_{ij}}.$$

Table 4 gives the effective sample size for different sampling frequencies under the fitted temporal Gaussian variogram model (1). When sampling up to three times per year, the effective sample size is very close to the number of samples taken. However, as the sampling frequency increases, the redundancy in the data also increases, resulting in effective sample sizes that are a fraction of the total number of samples taken.

Table 5 shows the effective sample sizes of the two REMAP samples under the fitted exponential variogram model (2). Notice that there is considerable redundancy in the REMAP data; the effective sample size is less than a third of the number of samples taken in each of the two years.

Sample Frequency	Total Samples	Effective Sample Size
Twice per Month	240	53.5
Once per Month	120	47.4
Six times per Year	60	38.6
Four times per Year	40	32.5
Three times per Year	30	27.7
Twice per Year	20	19.9
Once per Year	10	10.0

Table 4: Effective sample size as a function of sample frequency for a 10 year study.

Year	Total Samples	Effective Sample Size
1995	42	11.3
1996	35	10.9

Table 5: Effective sample size for the two REMAP sample years

The optimal allocation of sampling effort over space and time was investigated under varying ranges of spatial and temporal correlation. Serially alternating sampling designs with varying sampling frequencies, number of sample stations per sampling interval, and numbers of cycles were investigated. Each design has an equal total sampling effort of $n = 256$ samples in a 16×16 unit region over an 8 year period. Sampling frequencies of 0.5, 1, 2, 4, and 8 times per year were considered. The number of cycles ranged from $1, 2, 4, \cdots, 8f$, where f is the sampling frequency. Note that when the number of cycles is equal to 1, we have a permanent station design, and when the number of cycles is equal to $8f$, we have an ever-changing station design. Under a k-cycle design with a sampling frequency of f, the total number of locations sampled is $m = 32k/f$. These stations were randomly located in the 16×16 unit region under the constraint that no two stations be located within $8/\sqrt{m}$ of one another.

Table 6 gives the optimum number of cycles to estimate linear temporal trend for a serially alternating under different ranges of spatial and temporal autocorrelation. Here, the data $Z(\mathbf{s},t)$ at the location $\mathbf{s}$ at time t are modeled as

$$Z(\mathbf{s},t) = \beta_0 + \beta_1 t + \varepsilon(\mathbf{s},t),$$

where the errors have exponential spatio-temporal correlation function

$$\rho(\mathbf{h},r) \quad = \quad \text{corr}\{Z(\mathbf{s},t), Z(\mathbf{s}+\mathbf{h},t+r)\}$$

$$= \exp\{-3\|\mathbf{h}\|/\alpha_s - 3r/\alpha_t\},$$

α_s is the range of spatial autocorrelation, and α_t is the range of temporal autocorrelation. The optimum design is defined to be the design under which the variance of the general least squares estimator of β_1 is minimized, and hence yields the greatest power for detecting linear temporal trends in the data. Among the designs considered, the optimal sampling frequency was 8 times per year. From Table 6, the optimal design under a range of temporal autocorrelation of 1/2 year and spatial autocorrelation of 8 units, the optimal design is an 8 cycle design. The optimal number of cycles depends on the relative ranges of spatial and temporal autocorrelation. As the range of temporal autocorrelation increases, the optimal number of cycles also increases, but as the range of spatial autocorrelation increases, the optimal number of cycles decreases.

4 Current Status of Section 305(b) Water Resource Monitoring

Although Section 305(b) of the Clean Water Act mandates that each state submit a surface water quality assessment report to the Environment Protection Agency (EPA) every two years, little guidance is given as to what specific data should be collected. Consequently, states tend to design their water quality monitoring programs to meet local priorities governing the allocation of their water resources, and in response to local sources of environmental degradation.. Most states do not monitor all of their

		Range of Spatial Autocorrelation				
		1	2	4	8	16
Range of Temporal Autocorrelation	0.0625	2	2	1	1	1
	0.125	4	2	2	2	1
	0.25	8	4	4	2	2
	0.5	8	8	8	4	4
	1.0	16	16	8	8	8
	2.0	32	16	16	16	8
	4.0	32	32	32	16	16

Table 6: Optimum number of cycles for serially alternating designs under different levels of spatial and temporal correlation.

waterbodies every two years, and do not employ probability-based sampling designs when selecting locations for sample sites. Instead sample sites are selected according to a number of criteria, that differ among states and are not always well defined. For example, the South Carolina Water Quality Monitoring Program selects 265 primary stations that are influent or effluent to sub-basins, at major streams at state lines, at the confluence of major streams, above and below major industrial and municipal areas, in major lakes, and at the mouth of major tributaries. In Maryland, the Basic Water Monitoring Program established a network of 68 sites in locations where known water quality programs exist, and in rivers or major tributaries just above the confluence with a river, but excludes areas with no serious water quality problems. In either case, the representativeness of the sample sites cannot be readily quantified, and hence estimates of the overall quality of the states' water resources are likely be biased, especially in states which avoid areas thought to contain no serious water quality problems.

In defense of state efforts, it should be pointed out that federal water resource monitoring designs have not provided leadership by employing probability-based designs themselves. The National Stream Quality Accounting Network (NAWQAN), the National Water-Quality Assessment Program (NAWQA), and the National Status and Trends Program (NS&T) all employ judgment sampling designs. It is interesting to note that, while the Biomonitoring Environmental Status and Trends Program

(BEST) uses a probability-based design to monitor presticides in starlings, it uses a judgment sampling design to monitor pesticides in fish. The Environmental Monitoring and Assessment Program (EMAP) is the only large federal program that employs a probability-based sampling design to monitor aquatic resources, but this program was only recently established and has a questionable future. In contrast, most programs that monitor terrestrial resources, including EMAP, use probability-based sampling designs (Olsen et al. 1998).

Recent years have seen attempts to improve water quality monitoring efforts. The Intergovernmental Task Force on Monitoring Water Quality (ITFM) was established in 1992 to review and evaluate national water quality monitoring efforts and to make recommendations for improvements. The ITFM has recommended that states change from a 2 year reporting cycle to a 5 or 6 year reporting cycle. By doing so, states may increase spatial coverage of their water resources through the implementation of serially alternating sampling designs.

In 1990, the EPA established the National 305(b) Consistency Workgroup to address variation in sampling protocols and reporting methods among states. In response to efforts of this workgroup, several states are exploring methods for obtaining more representative samples of their water resources. For example, South Carolina is establishing Watershed Water Quality Management (WWQM) Stations at the downstream access of every Natural Resource Conservation Service (NRCS) watershed unit.

Thus, a census of all watershed units is obtained. However, the representativeness of the resulting data depends on how watershed units are partitioned. Nevertheless, the WWQM stations provide good spatial coverage of the South Carolina's watersheds.

Some states have implemented probability-based sampling designs. The Delaware Department of Natural Resources and Environmental Conservation selected a sample of 96 sites, randomly selected from a list frame of 3200 roadway crossings of nontidal streams in the northern two counties of the state. The Maryland Department of Natural Resources randomly selected a sample of about 350 sites from a list frame of all first, second, and third order stream reaches.

4.1 Response to 305(b) Consistency Workgroup

The failure of states to adapt probability-based sampling designs in their water quality monitoring efforts may in part be due to misperceptions regarding their limitations. Many of these misperceptions can be found in the draft report of the Monitoring and Assessment Design Focus Group of the 305(b) Consistency Workgroup (1996), which lists a number of disadvantages and concerns with probability-based sampling designs. The following shall address each these by suggesting how a probability-based design that can be used to address each of these concerns. Note that these proposed designs may require some modification for specific applications.

Concern 1. Probability-based designs will not identify new problem sites unless

they happen to be selected randomly. A similar statement could be made about judgment sampling designs: Judgment sampling designs will not identify new problem sites unless they can be identified by the investigator. Thus, under a judgment sampling design, the ability to identify new problem sites is limited by the judgment of the investigator. The probability of identifying new problem sites can be increased by increasing the spatial coverage of a sampling design either through implementation of serially alternating or rotating panel designs, or through sampling a new set of sites during each sampling interval. A more efficient approach would require assumptions regarding causal mechanisms, and then information on the causal variables, preferrably over the entire population. For example, an investigator might attempt to identify all potential point sources of environmental contamination (for example, from a listing of all sewage treatment plants, or all paper mills in the state). However, sufficient resources may not be available to sample all of the potential point sources. Further information regarding the characteristics of the identified potential point sources might be used to select which ones are most likely to pose environmental hazards, but the cost of compiling such information may be prohibitive. Moreover, some potential point sources which appear to pose to no environmental hazard, and hence are not included in the sample, may in truth pose a significant environmental hazard.

A probability-based sampling design can be used to identify which potential point

sources pose a significant environmental hazard. This can be accomplished by selecting a simple random sample of potential point sources in the first year of the investigation. Selected sites that show significant environmental damage may then be sampled in each of the next years, perhaps until they meet or exceed regulatory standards. In the second year, a simple random sample of the remaining sites is selected, and again, those sites showing significant environmental damage are retained. This process is repeated in subsequent years until all potential point sources are sampled at least once.

In states where it is prohibitively expensive to identify all potential point sources of environmental contamination, a two-stage sampling design might be used to assist in the identification of point sources as follows: The state's water resources are partitioned into the NRCS watershed units. In the first year, a simple random sample of the watershed units is selected. Then the potential point sources of environmental contamination are identified within each of the selected watershed units. A simple random sample of the identified point sources may then be selected. In each of the subsequent years, a simple random sample of the heretofore unsampled watershed units is sampled until all watershed units have been sampled. After that time, the process may be repeated. Thus, in each year, only those potential point sources within the selected watershed units need by enumerated, from which a simple random sample can be selected for field sampling.

Adaptive sampling designs are particularly well suited to the identification of new problem sites under nonpoint sources of environmental contamination. Start with a simple random sample of sites. Then cluster new sample sites around each site showing a level of environmental degradation about some threshold. A response-surface model (Myers 1976) may be fit to the data, to identify locations where additional sampling is required to obtain an estimate of the location of the local maximum level of environmental degradation. Occasionally, additional sample points should be randomly selected to ensure the identification of new problem sites.

Concern 2. Probability-based designs will not determine temporal trends at priority sites. There are a number of very legitimate reasons why specific priority sites may be of interest. For example, we may wish to investigate the efficacy of environmental remediation at locations of sewage or industrial discharge, or hot spots known to show especially high levels environmental damage. To assess the efficacy of such restoration efforts, however, it may be necessary to compare temporal trends at these priority sites to temporal trends at reference sites, selected to represent conditions existing prior to environmental degradation at the priority sites. If interest lies in the levels of contaminants in the waters of a river or stream, then it may suffice to locate reference sites upstream of priority sites and a probability-based sampling design need not be considered. If, however, interest lies in the restoration of the ecological community at priority sites, then upstream sites are not guaranteed to be representative of conditions

that had existed prior to environmental degradation at priority sites, and hence, a probability-based sampling design should be used to select reference sites. To further ensure the representativeness of reference sample sites, a stratified random sampling design might be used, where the allocation of sampling effort to strata is proportional to the number of priority sites found in each stratum. Alternatively, reference sites may be located some random distance and direction from each of the priority sites, or if sufficient resources are available, two or more reference sites may be clustered around each priority site.

Concern 3. Probability-based designs are not designed to assess improvements in specific waterbodies or watersheds due to controls, enforcement, or restoration. When assessing improvements at specific waterbodies or watersheds is of interest, then each of the specified waterbodies or watersheds must be sampled. However, the question remains as to what specific locations should be sampled within those waterbodies or watersheds. If the water quality of a stream or river is of interest, it may suffice to sample at the effluent end of that stream or river. If, on the other hand, the status of the ecological community, or the quality of bottom sediments are of interest, a probability-based design is required to ensure that the sample sites are representative of the waterbody or watershed of interest. Here, individual waterbodies or watersheds can be treated as strata for a stratified random sampling design. The use of a judgment sampling design to select what specific sites are to be sampled

within each waterbody or watershed can result in biased estimates of status and temporal within that waterbody or watershed.

Concern 4. Probability-based designs respond poorly to political priorities. Without more specifics regarding what political priorities are to be considered, it is not possible to make specific recommendations as to how a probability-based sampling design may accommodate them. However, the sampling intensity can adjusted to ensure that a higher density of sample sites is obtained in high priority regions at the cost of a lower density of sample sites in low priority regions.

Concern 5. If all 305(b) assessments were based on changing probabilistic sites, States would no longer track specific waterbodies and mapping a spatial analysis would be curtailed. The use of changing probabilistic sites does not preclude temporal and spatial analysis of the data. Statistical methods for such analyses shall be discussed in Section 5.2 below. Regardless of whether a probability-based or judgment sampling design is used, the power of analysis for temporal trends within specific waterbodies will depend on how many observations are available within those waterbodies. However, if permanent sample sites are selected according to a judgment sampling design, then the only statistically justifiable inferences are with respect to those specific sites. Under a probability-based design, statistically justifiable inferences regarding temporal trends can be made regarding the waterbodies as a whole. Moreover, statistical tests for trend are also likely to be more powerful under changing probabilistic sample

sites than under a fixed station design (See Section 4.2).

Concern 6. Probability-based designs require significant up-front effort for proper design and long-term adherence to the study plan. The ability to make statistically justifiable inferences regarding the water resources as a whole should justify the added up-front effort required to obtain an appropriate probability-based design. The costs of long-term adherence to the study plan can be reduced by using a serially alternating design (see Section 4.2) instead of selecting a new set of probabilistic sample points for each sample interval.

Concern 7. Under a probability-base design, states would lose the benefits of existing sites with many years of data. In Section 6.0, a method for combining historical data from a judgment sample design with new data from a probability-based is developed. The proposed method calls for a period of overlap in which observations are collected from both designs. Then the spatio-temporal autocorrelation among the observations from both data bases is exploited to back predict what data would have been obtained had a probability-based design been used from the very beginning of the monitoring program. The resulting predictor relies heavily on the historical data base, especially for predictions many years in the past.

Concern 8. Determining sources of impairment may be beyond the capability of probability-based designs. Results of observational studies can not provide definitive evidence that a given factor or combination of factors are responsible for environ-

mental impairment. Correlations between levels of environmental impairment and alleged sources of impairment may be spurious. Moreover, the highest contaminant concentrations are not necessarily located near their sources, but may be located downstream where local site characteristics may promote adsorbtion of contaminants in the sediment or their entry into the food chain. Definitive evidence for causal relationships can only be obtained through randomized experimental manipulations of the environment. However, such manipulations may not only be impractical, but also unethical. Nevertheless, it may be possible to gain some insight through a carefully planned observational study. Sites should be selected in a factorial arrangement in which all combinations of high and low levels of each of the alleged causal factors are equally replicated. However, the information required for such a design may not be readily available. A more cost-effective approach would be to implement a probability-based design in which the alleged causal factors are measured along with the measures of impairment. Supplemental sites may then be added to provide information from factor combinations missed by the probability-based design, improving the power to separate out causal contributions.

Concern 9. If the design does not allow sampling at access points like bridges, sampling elsewhere will be difficult and expensive. The savings incurred by sampling at access points may allow larger sample sizes under tight budgetary constraints, and hence potentially more precise estimates of environmental parameters and more

statistical power for detecting trends. Probability-based sampling methods can be used to select what access points are to be included in the sample. However, to statistically justify inference to the water resource as a whole, evidence is required that the access points are representative of that water resource, or alternatively, an estimate of the bias introduced by sampling at access points. There are a number of reasons why the representativeness of access points may be questioned:

- The density of access points such as bridges will tend to be higher in regions of high human population density, and lower where human populations are sparse.
- The level of environmental impairment may vary with the suitability of locations for bridge construction. Do we really want bridge engineers to determine where we sample?
- The bridges themselves may adversely affect their local environments.

Section 5.3 discusses how each of these concerns may be addressed using probability-based designs.

Concern 10. Concern over the number of years required to determine spatial or temporal trends in a basin or state. Probability-based designs require no more years to determine spatial or temporal trends than judgment sampling designs. The power to detect such trends is a function of the sample size, and the degree of spatio-temporal correlation in the data. If probability-based designs show less spatio-temporal correla-

tion, as would often be the case, then they should be more powerful than a judgment sample of the same size. It should be kept in mind that spatial and temporal trends should be interpreted with caution. Ecological systems are inherently dynamic; so, in order to investigate the impact of management on environmental impairment, we must distinguish between trends resulting from management practices and natural environmental fluctuations. This requires an understanding of the natural fluctuations that may occur in a waterbody that might only be obtained from collecting data over a number of years.

Concern 11. Concerns over the expense of sampling sufficient sites for statistical rigor and also availability of technical support for States. Given the high cost of environmental monitoring, it is essential that the sampling design yield the strongest possible statistical inference with respect to the states' water resources. Regardless of sample size, statistically justifiable inferences can be made regarding the status of the water resources as a a whole under a probability-based sampling design. Since the only statistically justifiable inferences that can be made under a judgment sampling design are with respect to status and trends at the sample themselves, judgment sampling designs make very inefficient use of funds allocated to environmental monitoring. The EPA should be responsible for providing technical support to the states for implementing probability-based sampling designs, and analyses of the resulting data.

4.2 Combining Data Across States

In addition to the biennial water quality assessment reports that must submitted by the states, Section 305(b) of the Clean Water Act also mandates that the EPA submit a comprehensive assessment of the quality of the nation's water resources to Congress every two years. The latter requires the combining of data submitted in the states' reports. Given that most states employ judgment sampling designs, valid statistical inference is limited to statements regarding what percentage of sample stations support their designated uses (e.g., drinking water supply, fish consumption, recreation, etc.), and what percentage of stations show improving or degrading water quality. Statements regarding what percentage of water resources support their designated uses, or show improving or degrading water quality cannot be statistically justified.

The combining of data across states would be straightforward if all states were to employ probability-based sampling designs and provided that they use the same defintion for the target population, and consistent measurement protocols. Then the different states can be treated as strata, and the mean level of an environmental indicator across the 50 states can be estimated by

$$\widehat{\mu} = \frac{1}{|A|} \sum_{i=1}^{50} |A_i| \cdot \widehat{\mu}_i, \tag{3}$$

where $\widehat{\mu}_i$ is the estimated mean level of the environmental indicator in state i, $|A_i|$ is the quantity of the water resource (e.g., stream miles, total surface area of lakes

or estuaries, etc.) in state i, and $|A|$ is the total quantitative of that resource in the nation (i.e., $|A| = \sum_{i=1}^{50} |A_i|$). The precision of this estimate can be estimated through its variance

$$\text{var}(\widehat{\mu}) = \frac{1}{|A|^2} \sum_{i=1}^{50} |A_i|^2 \cdot \text{var}(\widehat{\mu}_i). \tag{4}$$

In a similar manner, the proportion of the nation's water resources showing a given condition (e.g., degraded, supporting designated uses, showing improving conditions, etc.) can be estimated by

$$\widehat{p} = \frac{1}{|A|} \sum_{i=1}^{50} |A_i| \cdot \widehat{p}_i, \tag{5}$$

where $\widehat{p}_i$ is the estimated proportion of the water resources of state i that show that condition. The corresponding variance estimate is

$$\text{var}(\widehat{p}) = \frac{1}{|A|^2} \sum_{i=1}^{50} |A_i|^2 \cdot \text{var}(\widehat{p}_i). \tag{6}$$

The above estimates do not require that the same sampling design be employed by all states; they only require that each state employ a probability-based sampling design. Estimates of state means μ_i, proportions p_i, and their corresponding variances depend on the particular sampling designs employed by each state. However, unbiased estimation across the 50 states requires some consistency among states with respect to what data are collected and how the data are obtained.

Differences among states in definitions of target populations (e.g., what orders of streams or sizes of lakes are sampled) can lead to biased estimates of the status of

the nation's water resources. For example, if some states do not sample lower order stream reaches, and such stream reaches tend to have better (lower) water quality than higher order reaches, then the overall proportion of stream miles meeting a water quality standard will be underestimated (overestimated). To avoid this source of bias, the EPA (with input from the state agencies) should provide the states a clear and concrete definition of the target population of water resources that should be sampled. Depending on their needs, individual states may elect to sample sites not included in this target population, but data from those sites should be reported separately.

Differences among states in sampling protocols (e.g., at what depth a water sample is obtained, when samples are taken, how samples are handled and stored following collection), and laboratory procedures for assaying samples may also lead to biased estimates. This bias may be reduced by having states adopt consistent sampling protocols, and laboratory procedures for assaying samples (ITFM 1995). Nevertheless, it is likely that there will remain some variation among state field crews and laboratories with respect to how sampling protocols and laboratory procedures are applied. To reduce the resulting biases, groups of states should engage in joint sampling efforts, in which field crews from the various states sample the same sites using their own sampling protocols, and their own laboratories for assaying resulting samples. The analysis of variance model

$$y_{ij} = \mu + \beta_i + \gamma_j + \varepsilon_{ij} \tag{7}$$

can then be fit to the resulting data y_{ij} at site j using the field crew from state i. Here, μ is the overall mean, β_i is the bias attributed to the methods for state i, γ_j is the effect of site j, and ε_{ij} is the model error. The bias terms β_i are not individually estimable unless further assumptions are made; for example we assume that $\sum_{i=1}^{50} \beta_i = 0$, or, alternatively, that one of the individual states uses unbiased methods (i.e., $\beta_i = 0$ for some i). The parameters of (7) can be estimated using the generalized linear model procedure (PROC GLM) of the Statistical Analysis System (SAS Institute 199?). Given estimates of the bias terms, a bias corrected estimate of the overall mean can then be obtained from

$$\widehat{\mu} = \frac{1}{|A|} \sum_{i=1}^{50} |A_i| \cdot (\widehat{\mu}_i - \widehat{\beta}_i).$$

Note that if the analysis of variance shows that there are no significant differences among the states, then no bias correction is necessary.

Note that the above does not require that all 50 states sample each site. Instead, it suffices that the data from all of the states be connected (sensu Searle 1971, pp. 319-324). To determine if all states are connected, create a table showing which state crews sampled which sites. For example, see Figure 9 in which six sites are sampled by six states; here state 'B' sampled sites 2 and 5, and site 2 was sampled by both states 'B' and 'F'. To find the connected subsets, draw horizontal and vertical line segments connecting any pair of observations on the same row or column; observations that can be connected by such line segments form a connected subset; in Figure 9,

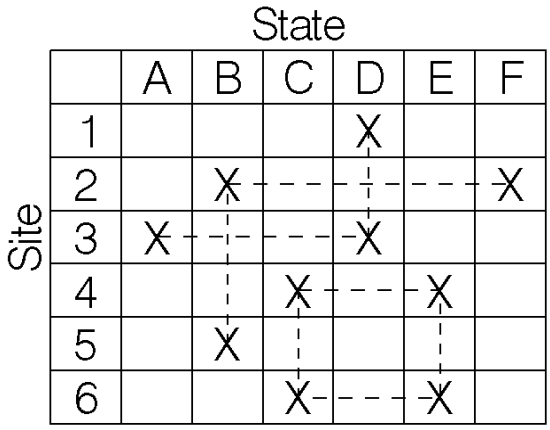

Figure 9: Connected subsets of states.

for example, states 'B' and 'F' form one connected subset, states 'A' and 'D' form a second connected subset, and states 'C' and 'E' form a third connected subset. Since there are more than one connected subsets, the data are disconnected, and hence we would not be able to estimate the relative biases of the states' data. The states would be connected if, for example, state 'B' were to sample the additional sites 3 and 4.

The above analyses also assume that there is no interaction between states and sites, so that the bias in a given state's methods does not depend on site. Tukey's procedure (Snedecor and Cochran 1980, pp. 283-285) may be used to test for this interaction. If a significant interaction is found, then the analysis variance model may be fit to log transformed data:

$$\ln y_{ij} = \mu + \beta_i + \gamma_j + \varepsilon_{ij}.$$

Then, the overall mean can be estimated by

$$\widehat{\mu} = \frac{1}{|A|} \sum_{i=1}^{50} |A_i| \cdot \widehat{\mu}_i e^{-\widehat{\beta}_i}.$$

Regardless of efforts to improve consistency among state water resource monitoring programs, it is likely that states will continue to differ with respect to what variables are measured. Moreover, it is not necessarily appropriate for states with widely different types of water resources to measure the same variables. This is especially true for biotic measurements, since there is considerable geographic variation in the composition of aquatic communities over the United States. Obviously, estimation of the overall mean level of an environmental variable across the 50 states requires that the same variable be measured in each state. On the other hand, estimation of the proportion of water resources showing a given condition (i.e., degraded, supporting a designated use, showing improving conditions), do not require that the same variables be measured across the states. However, the quality of the estimates could be improved by some general agreement with respect to definitions of what is meant by a degraded condition, or when a waterbody supports a designated use or shows improving conditions. Without such an agreement, expression (5) would only estimate what proportion of the nation's water resources were designated as showing a given condition, and not necessarily in any clearly defined way what proportion actually shows that condition.

5 Design Alternatives for Section 305(b) Water Resource Monitoring

Statistically defensible methods for combining data across the 50 states require that the states replace their current judgement sampling designs with probability-based sampling designs. The specific probability-based design to be implemented by a given state depends on the resources available from that state to support monitoring efforts, the logistical constraints under which monitoring is to be carried out, and the characteristics of that state's water resources. Therefore, detailed descriptions of specific monitoring designs are beyond the scope of this report. The following broadly outlines some alternative probability-based designs that may be implemented for water resource monitoring. For each sampling design, methods for estimating the population mean, population proportion, and the total mass of an environmental contaminant are considered.

5.1 Sampling Lakes

The recommended approach to sampling lakes depends on the monitoring objectives, the distribution of sizes and types of lakes within a state, and the information available on the population of lakes to be sampled. The objectives may call for sampling all of the larger lakes in the state, but resources are unlikely to be available for sampling all of the smaller lakes each year. For the latter, we may require a random sample.

5.1.1 Sampling Large Lakes

A stratified random sampling design can be used to sample the large lakes within a state, where each lake is treated as a stratum. Under such a design, n_i sample sites are randomly located within each of m large lakes according to a simple random sampling design (Figure 2); $i = 1, \cdots, m$. Suppose that sufficient funds are available to sample n sites during each sample interval. Then the recommended allocation of sampling effort calls for selecting

$$n_i \cong n\left(\frac{|A_i|}{\sum_{j=1}^{m}|A_j|}\right) \text{ or } n_i \cong n\left(\frac{|V_i|}{\sum_{j=1}^{m}|V_j|}\right)$$

sites from lake i, where $|A_i|$ and $|V_i|$, respectively, are the surface area and volume of lake i. Thus, lakes are sampled proportional to their sizes. The allocation scheme is optimal (minimizes sampling variance) under the assumption that the within lake variances are homogeneous (i.e., they are identical among the large lakes). If the within-lake variances are heterogeneous, then an optimal allocation scheme would call for increased allocation of sampling effort within lakes showing high variability, and decreased allocation within lakes showing low variability. Different environmental variables are likely to show different patterns of within-lake variability, so that an allocation scheme is optimal for one variable is not likely to be optimal for the remaining variables. Moreover, the within-lake variances are not likely to be known a priori, and hence, allocation proportional to lake size is recommended.

Under the stratified random sampling design described above, the mean level of an environmental variable across the surface area of lake i can be estimated by the sample mean $\overline{y}_i$ of the n_i observations in that lake. The precision of this estimate can be estimated by

$$\widehat{\text{var}}(\overline{y}_i) = \frac{s_i^2}{n_i}$$

where s_i^2 is the sample variance of the n_i observations in lake i. The overall mean across the surface of all m large lakes can be estimated by

$$\widehat{\mu}_{\text{st}} = \frac{1}{|A|} \sum_{i=1}^{m} |A_i| \cdot \overline{y}_i$$

with corresponding variance estimate

$$\widehat{\text{var}}(\widehat{\mu}_{\text{st}}) = \frac{1}{|A|^2} \sum_{i=1}^{m} |A_i|^2 \cdot \frac{s_i^2}{n_i}.$$

The proportion of the surface area of lake i showing a given condition (i.e., degraded, supporting designated uses, etc.) can be estimated by $\widehat{p}_i$, the proportion of sample sites showing that condition. The corresponding variance estimate is given by

$$\widehat{\text{var}}(\widehat{p}_i) = \frac{\widehat{p}_i(1-\widehat{p}_i)}{n_i}.$$

The proportion of the surface area of all m large lakes showing that condition can then be estimated by

$$\widehat{p}_{\text{st}} = \frac{1}{|A|} \sum_{i=1}^{m} |A_i| \cdot \widehat{p}_i$$

with corresponding variance estimate

$$\widehat{\text{var}}(\widehat{p}_{\text{st}}) = \frac{1}{|A|^2} \sum_{i=1}^{m} |A_i|^2 \cdot \frac{\widehat{p}_i(1-\widehat{p}_i)}{n_i}.$$

Suppose that the concentration of an environmental contaminant in a given water sample is expressed in terms of mass per unit volume. Then the total mass of that contaminant in lake i can be estimated by

$$\widehat{\tau}_i = \frac{c|A_i|}{n} \sum_{j=1}^{n_i} d_{ij} \cdot y_{ij},$$

where d_{ij} and y_{ij} are the water depth and concentration at site j in lake i, and the constant c is defined to achieve the appropriate units of measurement. The corresponding variance estimate is

$$\widehat{\text{var}}(\widehat{\tau}_i) = \frac{c^2|A_i|^2}{n(n-1)} \left\{ \sum_{j=1}^{n_i} d_{ij}^2 \cdot y_{ij}^2 - \frac{1}{n} \left(\sum_{j=1}^{n_i} d_{ij} \cdot y_{ij} \right)^2 \right\}.$$

Then the total mass of the contaminant across all m large lakes can be estimated by

$$\widehat{\tau}_{\text{st}} = \sum_{i=1}^{m} \widehat{\tau}_i$$

with corresponding variance estimate

$$\widehat{\text{var}}(\widehat{\tau}_{\text{st}}) = \sum_{i=1}^{m} \widehat{\text{var}}(\widehat{\tau}_i).$$

Instead of locating sample sites according to a simple random sampling design within each of the large lakes in the population, sample sites can be located according to a randomized-tessellation stratified design (Stevens 1997). Under such a design,

a grid of contiguous polygons is randomly placed over the study region, as is shown for example in Figure 10, where a hexagonal tessellation is randomly located over Lake Jocassee. Then a single site is randomly located within each of the polygons. Only sites falling in the region of interest are included in the sample. The sampling variance under the randomized-tessellation stratified design is smaller than that under the simple random sampling design, especially if the data shows strong spatial correlation. The Yates-Grundy estimator for its variance is reasonably stable under strong spatial correlation. If there is a large measurement error, or if there is large microscale variation in the data, however, the Yates-Grundy estimator for the variance can be unstable; in such cases, the randomized-tessellation stratified design cannot be recommended.

5.1.2 Sampling Small Lakes

The recommended approach to sampling small lakes depends the quality of information that is available regarding what lakes are present in a state. Ideally a listing of all small lakes in the state would be available, perhaps from USGS maps, aerial photos, or satellite images. Then a simple random sample or stratified random sample could be selected from the list frame of lakes. However, the cost of obtaining a list frame of all lakes within a state may be prohibitive. In this case, a two-stage sampling design may be required.

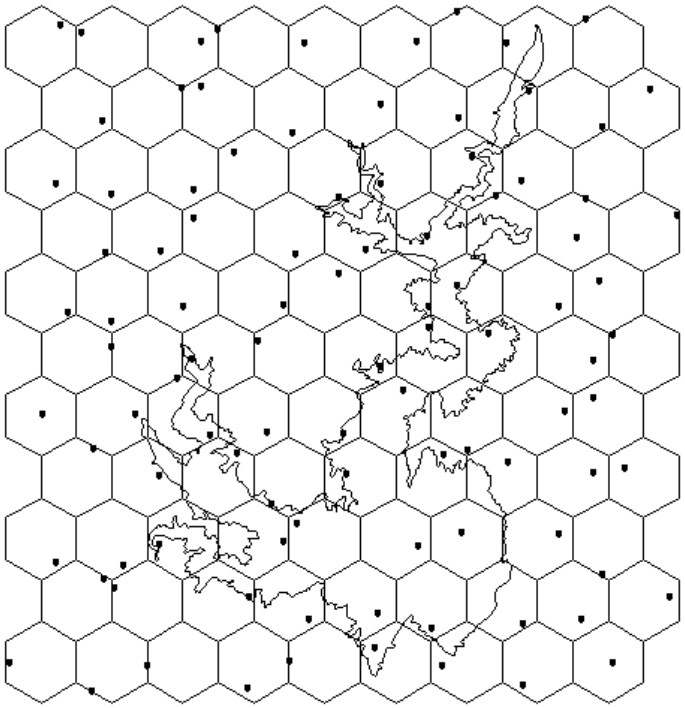

Figure 10: Randomized-tessellation stratified design for Lake Jocassee.

Simple Random Sample. Suppose that a list frame of all N lakes in a state is available. Then a simple random sample n lakes can be obtained from randomly drawing numbers between 1 and N until a sample of n unique lakes is drawn. Then one sample site is located within each of the sampled lakes. The decision as to what actual location is selected within each of the sampled lakes depends on the variable that is to be measured and the monitoring objectives. If it is desired to make inferences about the total mass of contaminants in the lakes of a given state, then sites should be selected randomly. A random sample would also be required to estimate the proportion of the volume of lake waters or surface area of lakes of a state that are impaired. If, on the other hand, it is desired to make inferences about the mean level of an environmental variable accross the population of lakes, or the proportion of lakes showing impaired conditions, random selection of sites within lakes may not be necessary. In such cases, water samples may be taken from the deepest part of the lake, or biota may be sampled in the multiple habitats around the lake in which they are found.

Under a simple random sampling design, the mean level of an environmental variable across the lakes in a state can be estimated by the sample mean $\overline{y}$, with corresponding variance estimate

$$\widehat{\text{var}}(\overline{y}) = \left(\frac{N-n}{N}\right)\frac{s^2}{n},$$

where s^2 is the sample variance. The proportion of lakes showing a given condition

(i.e., degraded, supporting designated use, etc.) can be estimated by $\widehat{p}$, the proportion of sample sites showing that condition. The corresponding variance estimate is

$$\widehat{\mathrm{var}}(\widehat{p}) = \left(\frac{N-n}{N}\right)\frac{\widehat{p}(1-\widehat{p})}{n-1}.$$

Instead of estimating the mean level of an environmental variable across the lakes, we may wish to estimate the mean level of that variable across the surface area of those lakes, or over the volume of the lakes. In such cases, sample sites should be randomly located within each of the selected lakes. The ratio estimators

$$\widehat{\mu}_s = \frac{\sum_{i=1}^{n}|A_i|\cdot y_i}{\sum_{i=1}^{n}|A_i|}, \tag{8}$$

and

$$\widehat{\mu}_v = \frac{\sum_{i=1}^{n}|V_i|\cdot y_i}{\sum_{i=1}^{n}|V_i|}, \tag{9}$$

can then be used to estimate the mean level of the variable across the surface area and volume of lakes in the population, respectively. Here, $|A_i|$ and $|V_i|$ respectively are the area and volume of lake i, and y_i is the value of the variable of interest in lake i. Thus, the data are weighted by the sizes of the lakes that were sampled. The variance estimates are

$$\widehat{\mathrm{var}}(\widehat{\mu}_s) = \frac{N(N-n)}{n|A|^2}\cdot\frac{\sum_{i=1}^{n}\left(|A_i|\cdot y_i - \widehat{\mu}_s|A_i|\right)^2}{n-1} \tag{10}$$

and

$$\widehat{\mathrm{var}}(\widehat{\mu}_v) = \frac{N(N-n)}{n|V|^2}\cdot\frac{\sum_{i=1}^{n}\left(|V_i|\cdot y_i - \widehat{\mu}_v|V_i|\right)^2}{n-1}, \tag{11}$$

where $|A|$ and $|V|$ respectively are the total surface area and volume of the N lakes in the population. If $|A|$ and $|V|$ are unknown, we may replace these quantities in the expressions above by their estimates

$$|\widehat{A}| = \frac{N}{n}\sum_{i=1}^{n}|A_i| \text{ and } |\widehat{V}| = \frac{N}{n}\sum_{i=1}^{n}|V_i|.$$

To estimate the proportion of the total surface area or volume of lakes that shows a given condition, replace y_i in the expressions above with a binary variable that takes the value 1 if sample site i shows that condition, and the value 0 if otherwise.

The total mass of an environmental contaminant can be estimated by

$$\widehat{\tau} = |V| \cdot \widehat{\mu}_v, \tag{12}$$

with corresponding variance estimate

$$\widehat{\text{var}}(\widehat{\tau}) = |V|^2 \widehat{\text{var}}(\widehat{\mu}_v). \tag{13}$$

If the total volume is unknown, total mass may be estimated by

$$\widetilde{\tau} = \frac{N}{n}\sum_{i=1}^{n}|A_i| \cdot d_i \cdot y_i \tag{14}$$

where d_i is the water depth at sample site i. The corresponding variance estimate is

$$\widehat{\text{var}}(\widetilde{\tau}) = \frac{N(N-n)}{n(n-1)}\left\{\sum_{i=1}^{n}|A_i|^2 d_i^2 y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n}|A_i| \cdot d_i \cdot y_i\right)^2\right\}. \tag{15}$$

To sample lakes over time, a serially alternating design with k cycles may be implemented by randomly partitioning the small lakes into k sets of size $n \cong N/k$.

This may be accomplished by taking a simple random sample of size n from the N lakes in the list frame to form the first set of lakes. The second set of lakes is obtained by taking a simple random sample of size n from the remaining $N - n$ lakes. This process is repeated until all lakes have been assigned to sets. Lakes in set i are then sampled at time intervals $i, i + k, i + 2k, \cdots$, as shown in Table 1 for a $k = 4$ cycle design. Observations from each time interval can be treated is though they were obtained from a simple random sample from the original population of N lakes, and so, population parameters may be estimated as described above. The proportion of lakes showing improving (deteriorating) conditions can be obtained by dividing the number of lakes showing improving (deteriorating) conditions by N. Since the entire population of lakes is sampled, this estimate has no sampling variance.

Stratified Random Sample. Suppose that in addition to a simple listing of lakes, further information is available about each lake in the list frame. For example, we may know which lakes are man made and which lakes are natural, we may have a list of oligotrophic and eutrophic lakes, or a description of the geological formation on which each lake lies. If the variable of interest depends on such characteristics, then a stratified random sampling design can be used to reduce sampling variation, and hence improve the precision of population parameter estimates. Strata may also correspond to their designated uses (i.e., drinking water, fishing, etc.). Stratified

random sampling designs can also guarantee that rare types of lakes are included in the sample, and to allocate more sampling effort to lakes that are deemed to be ecologically, economically, or sociologically important. Populations of lakes will tend to contain a very small number of larger lakes, and a very large number of small lakes, and so, a simple random sample may not pick up any of the important large lakes in the population. By stratifying by lake size, we can ensure that an adequate sample of large lakes is selected.

Under a stratified random sampling design, the list frame of lakes is first partitioned into K strata; let N_h denote the number of lakes in stratum h. Then a simple random sample of n_h lakes is obtained from stratum h, $h = 1, 2, \cdots, K$. Finally, one sample site is randomly located within each of the sampled lakes. The number of lakes sampled from each stratum may be proportional to the total number of lakes in each stratum

$$n_h = n\left(\frac{N_h}{\sum_{k=1}^{L} N_k}\right),$$

proportional to the total surface area of lakes in each stratum

$$n_h = n\left(\frac{|A_h|}{\sum_{k=1}^{L} |A_k|}\right),$$

or proportional to the total volume of lakes in each stratum

$$n_h = n\left(\frac{|V_h|}{\sum_{k=1}^{L} |V_k|}\right).$$

Using one of these sample allocation schemes as a starting point, sampling effort can

be increased in strata deemed to be more important, and reduced in strata deemed to be less important.

Since a simple random sample of lakes is obtained from each of the strata, the stratum means and proportions, the total mass of contaminant within a stratum, and their corresponding variances can be estimated using the same methods as described above for the simple random sampling design. The mean level of an environmental variable across the N lakes in the population can be estimated by

$$\overline{y}_{st} = \frac{1}{N}\sum_{h=1}^{K} N_h \overline{y}_h,$$

where

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

is the sample mean of the observations $y_{h1}, \cdots, y_{hn_h}$ from stratum h. The corresponding variance estimate is

$$\widehat{\text{var}}(\overline{y}_{st}) = \frac{1}{N^2}\sum_{h=1}^{K} N_h(N_h - n_h)\frac{s_h^2}{n_h},$$

where

$$s_h^2 = \frac{\sum_{i=1}^{n_h} y_{hi}^2 - n_h \overline{y}_h^2}{n_h - 1}$$

is the sample variance of the observations from stratum h. Similarly, the proportion of lakes showing a given condition can be estimated by

$$\widehat{p}_{\text{st}} = \frac{1}{N}\sum_{h=1}^{K} N_h \widehat{p}_h$$

where $\widehat{p}_h$ is the proportion of observations from stratum h showing that condition. The corresponding variance estimate is

$$\widehat{\text{var}}(\widehat{p}_{\text{st}}) = \frac{1}{N^2}\sum_{h=1}^{K} N_h(N_h - n_h)\frac{\widehat{p}_h(1-\widehat{p}_h)}{n_h - 1}.$$

The mean level of an environmental variable across the surface area or volume of the lakes may be estimated by

$$\widehat{\mu}_{\text{sts}} = \frac{1}{|A|}\sum_{h=1}^{K} |A_h| \cdot \widehat{\mu}_{hs},$$

and

$$\widehat{\mu}_{\text{stv}} = \frac{1}{|V|}\sum_{h=1}^{K} |V_h| \cdot \widehat{\mu}_{hv},$$

respectively, where $|A_h|$ and $|V_h|$ are the total surface area and volume of lakes in stratum h, and $|A|$ and $|V|$ are the total surface area and volume of all lakes. Here, $\widehat{\mu}_{hs}$ and $\widehat{\mu}_{hv}$ are computed from observations in stratum h using expressions (8) and (9), respectively. The corresponding variance estimators are

$$\text{var}(\widehat{\mu}_{\text{sts}}) = \frac{1}{|A|^2}\sum_{h=1}^{K} |A_h|^2 \widehat{\text{var}}(\widehat{\mu}_{hs}),$$

and

$$\text{var}(\widehat{\mu}_{\text{stv}}) = \frac{1}{|V|^2}\sum_{h=1}^{K} |V_h|^2 \widehat{\text{var}}(\widehat{\mu}_{hv}),$$

respectively, where $\widehat{\text{var}}(\widehat{\mu}_{hs})$ and $\widehat{\text{var}}(\widehat{\mu}_{hv})$ are computed from the observations from stratum h using expressions (10) and (11), respectively.

The total mass of an environmental contaminant over all lakes can be estimated by summing the estimated mass of that contaminant over the strata. That is, take

$$\widehat{\tau}_{\text{st}} = \sum_{h=1}^{K} \widehat{\tau}_h,$$

where $\widehat{\tau}_h$ is computed from the observations in stratum h using either expressions (12) or (14). The variance of $\widehat{\tau}_{\text{st}}$ can then be estimated by

$$\widehat{\text{var}}(\widehat{\tau}_{\text{st}}) = \sum_{h=1}^{K} \widehat{\text{var}}(\widehat{\tau}_h).$$

To sample lakes over time, a serially alternating design may be implemented in each of the strata as described above for the simple random sampling design. If a k cycle design is implemented in each stratum, then at each time the sample allocation is proportional to the number of lakes in each stratum. Note, however, that there is no requirement that the number of cycles k be identical among strata. By using a smaller number of cycles, more sampling effort can be made in more important strata, while larger number of cycles can be used in less important strata.

Two-Stage Sample. The implementation of the above sampling designs requires a list frame of all lakes in the target population. The cost of obtaining such a list frame can be prohibitive. These costs can be reduced by implementing a two-stage sampling design. Under a two-stage sampling design, the state is first partitioned into primary sample units, which may correspond to counties, watershed units, or a contiguous

grid of hexagonal or square quadrats. The first stage of the design is comprised of the random selection of n primary sample units from the population of N primary units. Then the lakes are enumerated within each of the selected primary sample units. The second stage of the design is comprised of the random selection of lakes within each of the selected primary units. Typically, allocation of sampling effort among primary units is proportional to the number of lakes in each of the selected primary units. Thus, if a total of m lakes are to be sampled, select

$$m_i = m\left(\frac{\quad_i}{\sum_{j=1}^{n} \quad_j}\right),$$

from primary unit i, where $\quad_i$ is the number of lakes in the i-th selected primary unit. Note that for variance estimation, we require $m_i \geq 2$ (unless a particular primary unit only contains one or two lakes).

To sample lakes over time, a serially alternating design with k cycles may be implemented by randomly partitioning the N primary units into sets of size $n \cong N/k$. Primary units in set i are then sampled at time intervals $i, i+k, i+2k, \cdots$, as shown in Table 1 for a $k = 4$ cycle design. In each time interval, the lakes are enumerated within each member of the appropriate set of primary units, from each of which, a simple random sample of lakes is drawn. Thus, after k time intervals, all of the lakes within the state will have been enumerated.

Within a given time interval, the mean level of an environmental variable across

the population of lakes in the state can be estimated by

$$\widehat{\mu}_{II} = \frac{N}{n} \sum_{i=1}^{n} \quad {}_i\overline{y}_i, \tag{16}$$

where is the total number of lakes in the state, and $\overline{y}_i$ is the sample mean value of the environmental variable among the selected lakes in primary unit i. The variance of $\widehat{\mu}_{II}$ may be estimated by

$$\widehat{\mathrm{var}}(\widehat{\mu}_{II}) = \left(\frac{N}{\quad}\right)^2 \left(\frac{N-n}{N}\right) \frac{s_u^2}{n} + \frac{N}{n \quad {}^2} \sum_{i=1}^{n} \quad {}_i(\quad {}_i - m_i)\frac{s_i^2}{m_i},$$

where s_i^2 is the sample variance of lakes selected from primary unit i, and

$$s_u^2 = \frac{\sum_{i=1}^{n} \quad {}_i^2\overline{y}_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} \quad {}_i\overline{y}_i\right)^2}{n-1}.$$

Although the total number of lakes in the state will be known after the first k time intervals of the serially alternating design described above, this quantity may not be known beforehand, or if this serially alternating design is not implemented. The total number of lakes in the state may however be estimated by

$$\widehat{\quad} = \frac{N}{n} \sum_{i=1}^{n} \quad {}_i.$$

Substituting $\widehat{\quad}$ into expression (16), we obtain the ratio estimator for the population mean:

$$\widehat{\mu}_R = \frac{\sum_{i=1}^{n} \quad {}_i\overline{y}_i}{\sum_{i=1}^{n} \quad {}_i},$$

whose variance may be estimated by

$$\widehat{\mathrm{var}}(\widehat{\mu}_R) = \left(\frac{N}{\widehat{\quad}}\right)^2 \left(\frac{N-n}{N}\right) \frac{\widetilde{s}_u^2}{n} + \frac{N}{\widehat{\quad}^2} \sum_{i=1}^{n} \quad {}_i(\quad {}_i - m_i)\frac{s_i^2}{m_i},$$

where

$$\widetilde{s}_u^2 = \frac{\sum_{i=1}^{n} \quad {}_i^2(\overline{y}_i - \widehat{\mu}_R)^2}{n-1}.$$

Similarly, the proportion of lakes showing a given condition (i.e., degraded, supporting designated use, etc.) may be estimated by

$$\widehat{p}_{II} = \frac{N}{n} \sum_{i=1}^{n} \quad {}_i\widehat{p}_i$$

if the total number of lakes is known, or by the ratio estimator

$$\widehat{p}_R = \frac{\sum_{i=1}^{n} \quad {}_i\widehat{p}_i}{\sum_{i=1}^{n} \quad {}_i}$$

if the total number of lakes is unknown. Here, $\widehat{p}_i$ is the proportion of lakes sampled in stratum i that satisfy that condition. The corresponding variances are

$$\widehat{\mathrm{var}}(\widehat{p}_{II}) = \left(\frac{N}{\quad}\right)^2 \left(\frac{N-n}{N}\right) \frac{s_p^2}{n} + \frac{N}{n \quad {}^2} \sum_{i=1}^{n} \quad {}_i(\quad {}_i - m_i) \frac{\widehat{p}_i(1-\widehat{p}_i)}{m_i - 1},$$

where

$$s_p^2 = \frac{\sum_{i=1}^{n} \quad {}_i^2\widehat{p}_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} \quad {}_i\widehat{p}_i\right)^2}{n-1},$$

and

$$\widehat{\mathrm{var}}(\widehat{p}_R) = \left(\frac{N}{\widehat{\quad}}\right)^2 \left(\frac{N-n}{N}\right) \frac{\widetilde{s}_p^2}{n} + \frac{N}{n \widehat{\quad} {}^2} \sum_{i=1}^{n} \quad {}_i(\quad {}_i - m_i) \frac{\widehat{p}_i(1-\widehat{p}_i)}{m_i - 1}$$

where

$$\widetilde{s}_p^2 = \frac{\sum_{i=1}^{n} \quad {}_i^2(\widehat{p}_i - \widehat{p}_R)^2}{n-1}.$$

The mean level of an environmental variable across the surface area of the lakes may be estimated by

$$\widehat{\mu}_s = \frac{\sum_{i=1}^{n} \frac{i}{m_i} \sum_{j=1}^{m_i} |A_{ij}| \cdot y_{ij}}{\sum_{i=1}^{n} \frac{i}{m_i} \sum_{j=1}^{m_i} |A_{ij}|},$$

where y_{ij} and $|A_{ij}|$ are the observation from and surface area of lake j in primary unit i. The variance of $\widehat{\mu}_s$ may then be estimated by

$$\begin{aligned}\widehat{\mathrm{var}}(\widehat{\mu}_s) &= \frac{1}{|\widehat{A}|^2} \left\{ \frac{N(N-n)}{n(n-1)} \sum_{i=1}^{n} \quad {}_i^2 \left| \frac{1}{m_i} \sum_{j=1}^{m_i} |A_{ij}| \cdot y_{ij} - \frac{\widehat{\mu}_s}{m_i} \sum_{j=1}^{m_i} |A_{ij}| \right|^2 \right. \\ &\left. + \frac{N}{n} \sum_{i=1}^{n} \frac{{}_i(\quad {}_i - m_i)}{m_i(m_i-1)} \sum_{j=1}^{m_i} \left| \left(|A_{ij}| \cdot y_{ij} - \frac{1}{m_i} \sum_{k=1}^{m_i} |A_{ij}| \cdot y_{ij} \right) - \widehat{\mu}_s \left(|A_{ij}| - \frac{1}{m_i} \sum_{k=1}^{m_i} |A_{ij}| \right) \right|^2 \right\}\end{aligned}$$

where

$$|\widehat{A}| = \frac{N}{n} \sum_{i=1}^{n} \frac{i}{m_i} \sum_{j=1}^{n} |A_{ij}|$$

is the estimated total surface area of the lakes in the population. The proportion of the surface area satisfying a given condition can be estimated by replacing y_{ij} in the expressions above with a binary variable that takes the value 1 if that condition is satisfied in lake j of primary unit i, and takes the value 0 if otherwise. The mean level of the environmental variable across the volume of the lakes may be estimated by replacing the lake areas $|A_{ij}|$ by the corresponding volumes $|V_{ij}|$.

The total mass of an environmental contaminant may be estimated by

$$\widehat{\tau} = \frac{N}{n} \sum_{i=1}^{n} \frac{i}{m_i} \sum_{j=1}^{m_i} |A_{ij}| \cdot d_{ij} \cdot y_{ij}$$

where d_{ij} is the water depth at the sample site in lake j in primary unit i. The

corresponding variance estimate is

$$\begin{aligned}\widehat{\mathrm{var}}(\widehat{\tau}) &= \frac{N(N-n)}{n(n-1)}\sum_{i=1}^{n}\left(\frac{{}_i}{m_i}\sum_{j=1}^{m_i}|A_{ij}|\cdot d_{ij}\cdot y_{ij} - \frac{1}{n}\sum_{k=1}^{n}\frac{{}_k}{m_k}\sum_{j=1}^{m_i}|A_{kj}|\cdot d_{kj}\cdot y_{kj}\right)^2 \\ &\quad + \frac{N}{n}\sum_{i=1}^{n}\frac{{}_i(\quad{}_i - m_i)}{m_i(m_i-1)}\sum_{j=1}^{m_i}\left(|A_{ij}|\cdot d_{ij}\cdot y_{ij} - \frac{1}{m_i}\sum_{k=1}^{m_i}|A_{ik}|\cdot d_{ik}\cdot y_{ik}\right)^2.\end{aligned}$$

5.2 Sampling Rivers and Streams

Rivers and streams are unique among natural resources in that, except for regions under tidal influence, the waters flowing past a given point originate from upstream of that point. Thus, observations of the water quality at the effluent end of a watershed are in some sense representative of the waters flowing through that watershed. This observation has led many water quality monitoring programs to target sampling at the effluent ends of watersheds. For example, South Carolina's Watershed Water Quality Management (WWQM) program targets sites at the downstream access of every National Resource Conservation Service (NRCS) watershed units. Note that not all NRCS watersheds units are watersheds unto themselves, but are subwatersheds. A subwatershed is a subset of a watershed obtained by subtracting out those regions covered by other watershed units in the collection. A mass balance model can be constructed from WWQM sample stations provided sufficient information is available. The total mass of a contaminant passing by a sample station can be computed by the product of the concentration of that contaminant in a water sample times the volume of water flowing past that station per unit time. Then the contribution of the

watershed unit to that mass can be obtained by subtracting the mass of contaminants input into that watershed unit by upstream watershed units from the mass of contaminants effluent from the watershed unit. However, such computations require the assumption that no contaminants are lost due to adsorption onto bottom substrates, uptake in organisms, or evaporation.

Unless a mass balance model or other mechanistic modeling effort is planned, there is very little reason to target sampling at confluences of waterways. Moreover, since representativeness of such sample site is not known, such targeted efforts are not appropriate for sampling the bottom substrate, or biotic communities. Only a probability sampling design can be used to obtain unbiased estimates of the mean level of an environmental contaminant across the length of rivers and streams, the proportion of stream and river miles that support designated uses, or the total mass of an environmental contaminant in the streams and rivers of a state.

The following considers three broad design alternatives for sampling rivers and streams within a state. The choice of design depends on what information is available on the population of streams and rivers, and the resources available for planning sampling efforts.

Simple Random Sampling. A simple random sampling design requires a digitized map of all rivers (and streams) within the state. Such a design can be constructed by

first partitioning the rivers into river segments, defined to be any length of river containing no branches. The river segments are then laid out end to end in any arbitrary order. Finally, n sample points are obtained by random selection of locations between 0 and L, the total length of the river segments. Since there tend to be more miles of first-order streams, than higher-order streams, a simple random sample will tend to be dominated by first-order stream sites. Therefore, it is generally recommended that streams be stratified by stream order (see below).

Parameter estimation under the simple random sampling design is straightforward: The mean level of an environmental variable across the length of the river system can be unbiasedly estimated by the sample mean $\overline{y}$, with corresponding variance estimate

$$\widehat{\text{var}}(\overline{y}) = \frac{s^2}{n},$$

where s^2 is the sample variance. The proportion of river miles showing a given condition (i.e., degraded, supporting designated uses, etc.) can be estimated by $\widehat{p}$, the proportion of sample sites showing that condition. The corresponding variance estimate is

$$\widehat{\text{var}}(\widehat{p}) = \frac{\widehat{p}(1-\widehat{p})}{n-1}.$$

Finally, the total mass of an environmental contaminant in the rivers of the state can be estimated by

$$\widehat{\tau} = \frac{L}{n}\sum_{i=1}^{n} |A_i| \cdot y_i, \tag{17}$$

where y_i is the concentration of the contaminant in a water sample collected at site i, and $|A_i|$ is the cross-sectional area of the river at that site. The variance of $\widehat{\tau}$ may then be estimated by

$$\widehat{\text{var}}(\widehat{\tau}) = \frac{L^2}{n} \frac{\sum_{i=1}^{n} |A_i|^2 y_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n} |A_i| \cdot y_i \right)^2}{n-1}. \tag{18}$$

Stratified Random Sample. A stratified random sampling design may be implemented to improve the precision of parameter estimates, to facilitate comparisons among strata, and ensure adequate sampling effort in rare strata. Here, strata may correspond to stream orders, or designated uses (i.e., swimming, drinking water, fishing, etc.). Under a stratified random sampling design, the list of river segments is first partitioned into K strata. Then a simple random sample of n_h sites is selected from each stratum h; $h = 1, 2, \cdots, K$, as described above. The number of sites sampled from each stratum may be proportional to the total length L_h of river segments in each stratum:

$$n_h = n \left(\frac{L_h}{\sum_{k=1}^{K} L_k} \right).$$

Using this sample allocation scheme as a starting point, additional sampling effort can be designated in strata deemed to be more important, while reduced sampling effort can be designated in strata deemed to be less important.

Since a simple random sample design is obtained from each stratum, the stratum means and proportions, the total mass of a contaminant within each stratum, and

their corresponding variances can be obtained using the same methods as described above for simple random sampling. Then mean level of an environmental contaminant across the lengths of all rivers in the population can be estimated by

$$\overline{y}_{\text{st}} = \frac{1}{L}\sum_{h=1}^{K} L_h \cdot \overline{y}_h,$$

where

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

is the sample mean of the observations $y_{h1}, \cdots, y_{hn_h}$ from stratum h, and L is the total river miles in the population of rivers. The corresponding variance estimate is

$$\widehat{\text{var}}(\overline{y}_{\text{st}}) = \frac{1}{L^2}\sum_{h=1}^{K} L_h^2 \frac{s_h^2}{n_h},$$

where

$$s_h^2 = \frac{\sum_{i=1}^{n_h} y_{hi}^2 - n_h \overline{y}_h^2}{n_h - 1}$$

is the sample variance of the observations from stratum h.

Similarly, the proportion of rivers miles showing a given condition can be estimated by

$$\widehat{p}_{\text{st}} = \frac{1}{L}\sum_{h=1}^{K} L_h \cdot \widehat{p}_h,$$

where $\widehat{p}_h$ is the proportion of sample stations from stratum h showing that condition. The corresponding variance estimate is

$$\widehat{\text{var}}(\widehat{p}_{\text{st}}) = \frac{1}{L^2}\sum_{h=1}^{K} L_h^2 \frac{\widehat{p}_h(1-\widehat{p}_h)}{n_h - 1}.$$

The total mass of an environmental contaminant across the lengths of all rivers in the population can be estimated by summing the estimated mass of that contaminant over the strata. That is, take

$$\widehat{\tau}_{\text{st}} = \sum_{h=1}^{K} \widehat{\tau}_h,$$

where $\widehat{\tau}_h$ is computed from observations in stratum h using expression (17). The variance of $\widehat{\tau}_h$ can then be estimated from

$$\widehat{\text{var}}(\widehat{\tau}_{\text{st}}) = \sum_{h=1}^{K} \widehat{\text{var}}(\widehat{\tau}_h).$$

Two Stage Sample. The implementation of the above sampling designs requires a digitized map of all rivers and streams in the target population. The cost of obtaining such a map can be prohibitive. These costs may be reduced by implementing a two-stage sampling design. Under this design, the state is first partitioned into N primary sample units, which may correspond to counties, NRCS watershed units, or a contiguous gird of hexagonal or square quadrats. The first stage of the design is comprised of the simple random selection of n primary sample units from the population of N primary units. Then the rivers and streams are digitized within each of the selected primary units; there is no need to digitize waterways within the remaining primary units. The second stage of the design is comprised of taking a simple random sample of sites along the lengths of the waterways within each of the selected primary units. Typically, allocation among the primary units is proportional

to the number of river miles in each of the selected primary units. Thus, if a total of m sites are to be sampled, select

$$m_i = m\left(\frac{L_i}{\sum_{j=1}^{n} L_j}\right),$$

from primary unit i, where L_i is the total river miles in primary unit i. Note that for variance estimation, we require that $m_i \geq 2$.

The mean level of an environmental variable along the lengths of the rivers and streams in the population may be estimated by the ratio estimator

$$\widehat{\mu}_R = \frac{\sum_{i=1}^{n} L_i \cdot \overline{y}_i}{\sum_{i=1}^{n} L_i},$$

where $\overline{y}_i$ is the sample mean of the variable among the observations from primary unit i. The corresponding variance estimate is

$$\widehat{\text{var}}(\widehat{\mu}_R) = \left(\frac{N}{\widehat{L}}\right)^2 \frac{\tilde{s}_u^2}{n} + \frac{N}{n\widehat{L}^2} \sum_{i=1}^{n} L_i^2 \frac{s_i^2}{m_i},$$

where

$$\tilde{s}_u^2 = \frac{\sum_{i=1}^{n} L_i^2 (\overline{y}_i - \widehat{\mu}_R)^2}{n-1},$$

s_i^2 is the sample variance of observations from primary unit i, and

$$\widehat{L} = \frac{N}{n} \sum_{i=1}^{n} L_i$$

is the estimated total length of waterways in the target population.

Similarly, the proportion of river miles showing a given condition can be estimated by

$$\widehat{p}_R = \frac{\sum_{i=1}^{n} L_i \cdot \widehat{p}_i}{\sum_{i=1}^{n} L_i},$$

where $\widehat{p}_i$ is the proportion of sites from stratum i showing that condition. The corresponding variance estimate is

$$\widehat{\text{var}}(\widehat{p}_R) = \left(\frac{N}{\widehat{L}}\right)^2 \frac{\widetilde{s}_p^2}{n} + \frac{N}{n\widehat{L}^2} \sum_{i=1}^{n} L_i^2 \frac{\widehat{p}_i(1-\widehat{p}_i)}{m_i - 1},$$

where

$$\widetilde{s}_p^2 = \frac{\sum_{i=1}^{n} L_i^2 (\widehat{p}_i - \widehat{p}_R)^2}{n-1}.$$

The total mass of an environmental contaminant across the volume of the population of waterways can be estimated by

$$\widehat{\tau}_{\text{II}} = \frac{N}{n} \sum_{i=1}^{n} \frac{L_i}{m_i} \sum_{j=1}^{m_i} |A_{ij}| \cdot y_{ij},$$

where y_{ij} and $|A_{ij}|$ are the contaminant concentration and the cross-sectional area of the waterway at sample site j in primary unit i. The corresponding variance estimate is

$$\begin{aligned} \widehat{\text{var}}(\widehat{\tau}_{\text{II}}) &= \frac{N^2}{n(n-1)} \sum_{i=1}^{n} \left(\frac{L_i}{m_i} \sum_{j=1}^{m_i} |A_{ij}| \cdot y_{ij} - \frac{1}{n} \sum_{k=1}^{n} \frac{L_k}{m_k} \sum_{j=1}^{m_i} |A_{kj}| \cdot y_{kj} \right)^2 \\ &\quad + \frac{N}{n} \sum_{i=1}^{n} \frac{L_i^2}{m_i(m_i-1)} \sum_{j=1}^{m_i} \left(|A_{ij}| \cdot y_{ij} - \frac{1}{m_i} \sum_{k=1}^{m_i} |A_{ik}| \cdot y_{ik} \right)^2. \end{aligned}$$

5.3 Sampling at Access Points

The savings incurred by sampling at access points allows larger sample sizes under tight budgetary constraints, and hence potentially more precise parameter estimates and greater statistical power for detecting spatial and temporal trends. The collection of access points can be treated as the sample population from which a probability sample can be obtained. However, to statistically justify inference to the water resource as a whole, we require evidence that the access points are representative of that water resource, or alternatively, we require an estimate of the bias introduced by sampling at the access points.

There are a number of reasons why the representativeness of access points may be questioned: First, the density of bridges will tend to be higher in regions of high human population density, and lower where human populations are sparse. Thus, by taking a simple random sample of bridges, the level of environmental impairment may be over estimated. This source of bias may be reduced by weighting the data proportional to the length of the river segment comprised of all points closer to the selected bridge than any other bridge (Figure 11a). Thus, the population mean level of an environmental variable is estimated by

$$\widehat{\mu}_w = \frac{1}{n}\sum_{i=1}^{n} w_i y_i \tag{19}$$

where y_i is the data collected at the bridge i, the weight $w_i = \ell_i/L$, ℓ_i is the length

of the river segment comprised of all points closer to bridge i than any other bridge, and L is the total river miles of the target population. The precision of $\widehat{\mu}_w$ can be estimated by its sampling variance:

$$\widehat{\text{var}}(\widehat{\mu}_w) = \left(\frac{N-n}{N}\right)\frac{s_w^2}{n} \tag{20}$$

where N is the total number of bridges in the population of bridges, n is the number of bridges sampled, and

$$s_w^2 = \frac{\sum_{i=1}^{n} w_i^2 y_i^2 - n\widehat{\mu}_w^2}{n-1}.$$

The estimated mean (19) assumes that the bridge is representative of the river segment containing that bridge, and the corresponding variance (20) makes the further assumption that the variable is constant over the length of that river segment (Figure 11b). So the sampling variance is likely to be underestimated.

If the lengths of the river segments vary considerably, then the sampling variance of $\widehat{\mu}_w$ can be quite large. This sampling variance can be reduced by using an unequal probability sample of bridges: Randomly locate points along the lengths of the rivers and streams, and then select the bridge that lies closest to each of the selected points. Bridges are sampled with replacement; that is, if a given bridge is selected more than once, data collected by that bridge should be counted as many times as that bridge is selected. Again, we shall assume that each bridge is representative of all points along the length of the river closer to that bridge than any other bridge. Then the population mean can be estimated by the sample mean $\overline{y}$, with corresponding

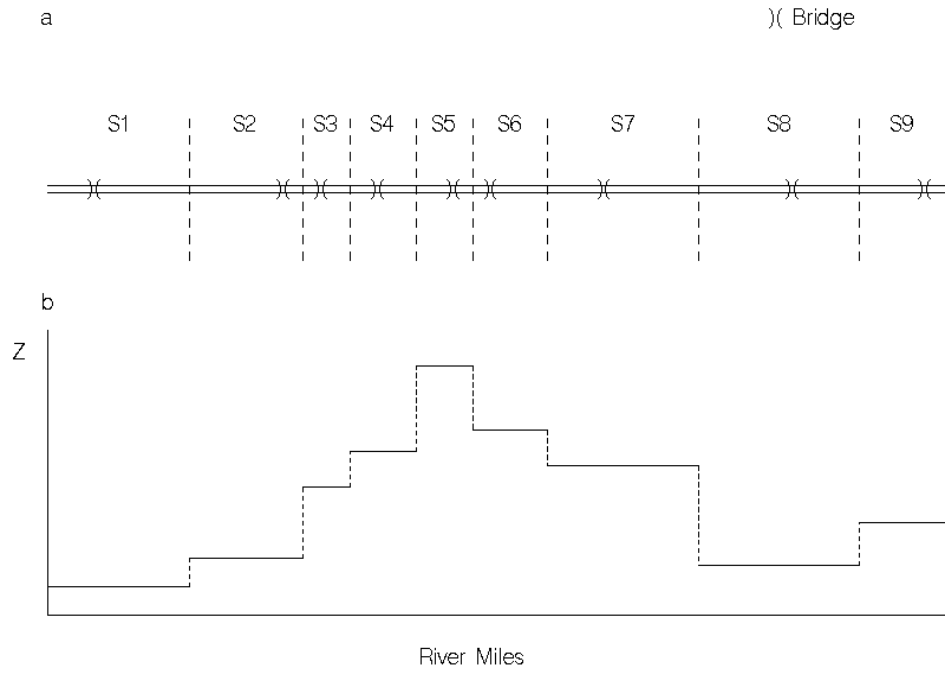

Figure 11: Sampling Bridges. (a) The locations of nine bridges along the length of a river. The river is partitioned into segments S1 to S9 as shown. A bridge will be sampled if a random point falls in that bridge's segment. (b) Assumed relationship between the variable of interest and location along the length of the river.

variance estimate $\widehat{\text{var}}(\overline{y}) = s^2/n$, where s^2 is the sample variance. This estimate of the population mean assumes that the bridge is representative of the river segment containing that bridge, and the corresponding variance estimate does not take into account variation along the length of that river segment.

The level of environmental impairment may vary with the suitability of locations

for bridge construction, also resulting in biased estimates of the mean level of an environmental variable. This source of bias may be reduced by using a stratified random sampling design: The river segments associated with the bridges are partitioned into m strata defined by their suitability for bridge construction. Thus each stratum will consist of river segments that are roughly equally suitable for bridge construction. River segments within each stratum are then laid out end to end, and n_i points are randomly selected along the total length of stratum i; $i = 1, \cdots m$. Finally, select the bridge closest to each of the selected points. Then the population mean can be estimated by

$$\widehat{\mu}_{\text{st}} = \frac{1}{L} \sum_{i=1}^{m} L_i \overline{y}_i,$$

where $\overline{y}_i$ is the sample mean of selected bridges in stratum i, L_i is the total length of river segments in stratum i, and L is the total river miles of the system. The corresponding variance estimate is

$$\widehat{\text{var}}(\widehat{\mu}_{\text{st}}) = \frac{1}{L^2} \sum_{i=1}^{m} \frac{L_i^2 s_i^2}{n_i},$$

where s_i^2 is the sample variance of selected bridges in stratum i. Again, the estimator $\widehat{\mu}_{\text{st}}$ assumes that the bridge site is representative of the river segment containing that bridge, and the corresponding variance estimator does not account for variation within river segments.

Some portion of the lengths of rivers may be completely unsuitable for bridge construction. This portion cannot be sampled at access points, and so, should be

treated as a separate stratum to be sampled using one of the methods described in Section 5.2.

The bridges themselves may have adverse effects their local environment, resulting in overestimates of environmental impairment. This source of bias might be reduced by sampling some random distance upstream from each bridge, instead of immediately below or adjacent to them.

Regardless of what design is used to select the access points to be sampled, evidence is required to demonstrate that the resulting sample yields unbiased estimates of environmental parameters. This requires data collected from a probability-based design, in which sites are selected from the water resource as a whole (e.g., using methods such as described in Section). Let $\widehat{\mu}_b$ denote the estimated population mean obtained from sampling at bridges, let $\widehat{\mu}_a$ denote the estimated population mean obtained from sampling along the water resource as a whole, and let $\text{var}(\widehat{\mu}_b)$ and $\text{var}(\widehat{\mu}_a)$ denote the corresponding variances. Then the null hypothesis that sampling at bridges yields an unbiased estimate of the population mean can be tested using the test statistic

$$t = \frac{\widehat{\mu}_b - \widehat{\mu}_a}{\sqrt{\text{var}(\widehat{\mu}_b) + \text{var}(\widehat{\mu}_a)}}.$$

Under the null hypothesis, t is approximately t-distributed with $n_b + n_a - 2$ degrees of freedom, where n_a and n_b are the number of observations from the two respective samples. If estimates from access points are not significantly different from estimates

obtained from the probability-based design over the resource as a whole, then sampling at access points suffices. If not, then the bias can be estimated by

$$\widehat{\beta} = \widehat{\mu}_b - \widehat{\mu}_a.$$

Assuming that the two samples are independent, then the variance of the estimated bias can be estimated by

$$\widehat{\text{var}}(\widehat{\beta}) = \widehat{\text{var}}(\widehat{\mu}_a) + \widehat{\text{var}}(\widehat{\mu}_b)$$

This bias correction can then be applied to future data collected exclusively from access points; that is, if $\widehat{\mu}_u$ is an uncorrected estimate of the population mean obtained from access point data, then a bias corrected estimate of the population mean is given by

$$\widehat{\mu}_c = \widehat{\mu}_u - \widehat{\beta},$$

with corresponding variance estimate

$$\widehat{\text{var}}(\widehat{\mu}_c) = \widehat{\text{var}}(\widehat{\mu}_u) + \widehat{\text{var}}(\widehat{\beta}).$$

Note that this presumes that the same sampling design was employed, and assumes that the bias does not change over time. It is recommended that the latter assumption be checked periodically using data from a probability based design including non-access points. A better approach would be to include both access and non-access points in the design during each sampling interval. The allocation of sampling effort

between access and non-access points can be determined so as to obtain the most precise estimates at minimum cost. Improved performance may also be achieved by applying different bias corrections to different strata.

6 Retaining Information from Historical Data

Despite the advantages outlined above, managers of state water quality monitoring programs are reluctant to implement probability-based sampling designs. Much of this reluctance stems from the fear that information from the historical data base will be lost. Therefore, probability-based sampling designs are not likely to be widely implemented unless statistical approaches to combining data from judgment and probability sampling designs are available. Unfortunately, methods for combining such data have received very little attention in the statistical literature. Overton, Young, and Overton (1993) use sampling frame attributes to assign judgment sites to clusters of similar probability sites. Judgment sites assigned to a given cluster are assumed to be representative of that cluster, and are treated as though they were obtained from a probability-based sampling design. However, the representativeness of the judgment sites with respect to their assigned clusters is difficult to diagnose, and if false, the combined data may yield biased estimates (Cox and Piegorsch 1996).

The following proposes an alternative approach to combining data from historical judgment sample sites with data from new probability-based sample sites. This ap-

proach requires an interval of overlap in which both historical judgment sites and new probability-based sites are sampled. Then the spatio-temporal correlation between the two sampling designs is exploited to predict what data would have been obtained had a probability-based sampling design been implemented from the very beginning of the monitoring program.

6.1 Space-Time Model

The following assumes that the data are a partial realization of a spatio-temporal random process. In particular, assume that the data $Z(\mathbf{s},t)$ at site $\mathbf{s}$ =(x,y) and time t are realized from the model

$$Z(\mathbf{s},t) = \beta_0 + \beta_1 x_1(\mathbf{s},t) + \cdots + \beta_p x_p(\mathbf{s},t) + \varepsilon(\mathbf{s},t), \tag{21}$$

where $\beta_0, \beta_1, \cdots, \beta_p$ are model parameters, and $\varepsilon(\mathbf{s},t)$ is a zero-mean error term. The explanatory variables $x_1(\mathbf{s},t), \cdots, x_p(\mathbf{s},t)$ may be functions of the spatial coordinates, time, distances to known geographic features (e.g., the mouth of the river system), or environmental variables such as water temperature, current, or turbidity.

Pairs of observations that are close together in space and time are likely to be more similar to one another than pairs of observations that are far apart. This spatio-temporal dependence can be modeled through the spatio-temporal correlation function

$$\rho\left(\|\mathbf{s}_1 - \mathbf{s}_2\|, |t_1 - t_2|\right) = \mathrm{corr}\{Z(\mathbf{s}_1,t_1), Z(\mathbf{s}_2,t_2)\},$$

which depends only on the distance $\|\mathbf{s}_1 - \mathbf{s}_2\|$ between the pair of sample sites $\mathbf{s}_1$ and $\mathbf{s}_2$, and the difference in sample times t_1 and t_2. The correlation function takes values between -1 and 1; positive values indicating positive spatio-temporal dependence, while negative values indicate negative spatio-temporal dependence. Typically, the correlation function will be a decreasing function of both $\|\mathbf{s}_1 - \mathbf{s}_2\|$ and $|t_1 - t_2|$, asymptotically approaching zero as the spatial and temporal distances between the observations increase. The rate at which the correlation function approaches zero determines the range of spatio-temporal correlation; correlation functions that rapidly approach zero characterize processes where interactions occur only between sites that are very close together, while correlations that slow approach zero characterize processes where distant sites interact. Observations have a perfect correlation of 1 with themselves so that $\rho(0,0) = 1$. However, there is often a discontinuity at zero when the correlation function is plotted against distance in space or time. This discontinuity is the so-called nugget effect, and is typically the result of measurement error or small-scale sampling variation.

Alternative measures of spatio-temporal dependence in the data include the covariance function

$$C(h,r) = \sigma^2 \rho(h,r)$$

and the variogram

$$2\gamma(h,r) \quad = \quad \mathrm{var}\{Z(\mathbf{s},t) - Z(\mathbf{s}+\mathbf{h},t+r)\}$$

$$= \quad \sigma^2(1 - \rho(h, r)),$$

where σ^2 is the variance of the data. The importance of the variogram comes from the observation that nonparametric estimates of the variogram are less biased than nonparametric estimates of the covariance or correlation functions (Cressie 1991).

Typically, the variogram is assumed to take a parametric form, such as given by the exponential model

$$2\gamma(h, r) = \begin{cases} 0; & (h, r) = (0, 0) \\ c_0 + c_e[1 - \exp\{-3(h/\kappa_s + r/\kappa_t)\}]; & (h, r) \neq (0, 0) \end{cases}.$$

The parameter c_0 is the nugget effect, κ_s is the range of spatial correlation, and κ_t is the range of temporal correlation. The nugget effect c_0 can be interpreted to be the variance due to measurement error plus microscale sampling variance. Pairs of sites located distances further than κ_s apart or observations collected at times further than κ_t are negligibly correlated. The variance $\sigma^2 = \frac{1}{2}(c_0 + c_e)$.

A variety of methods are available for estimating variogram parameters; for a review, see Cressie (1991, Section 2.6). The weighted least squares estimate requires no distributional assumptions, and is particularly well suited to fitting models to large spatio-temporal data sets. It involves the fitting of a parametric variogram model $2\gamma(h, r; \boldsymbol{\theta})$ to the method of moments estimator of the variogram

$$2\widehat{\gamma}(h, r) = \frac{1}{N_{hr}} \sum |\widehat{\varepsilon}(\mathbf{s}_i, t) - \widehat{\varepsilon}(\mathbf{s}_j, t + r)|^2, \tag{22}$$

where the sum is over all pairs of observations collected at sites approximately distance h apart and at sample times r apart, and N_{hr} is the number of such pairs of sites. The values $\widehat{\varepsilon}(\mathbf{s}_i, t)$ are residuals from a multiple regression of the data against the explanatory variables $x_1(\mathbf{s},t), \cdots, x_p(\mathbf{s}, t)$. The weighted least squares estimator of the parameter $\boldsymbol{\theta}$ is then obtained by finding $\widehat{\boldsymbol{\theta}}$ that minimizes

$$\sum_j \sum_k \frac{|\widehat{\gamma}(h_j, r_k) - \gamma(h_j, r_k; \boldsymbol{\theta})|^2}{\text{var}\{\widehat{\gamma}(h_j, r_k)\}},$$

where the sum is over all spatial and temporal lags at which $2\widehat{\gamma}(h, r)$ is computed, and

$$\text{var}\{\widehat{\gamma}(h_j, r_k)\} \cong 2\{2\gamma(h, r; \boldsymbol{\theta})\}^2 / N_{hr}.$$

6.2 Spatio-Temporal Prediction

Suppose that fixed sites $\mathbf{s}_1, \cdots, \mathbf{s}_n$ are selected according to an arbitrary judgment sampling design, and that the variable of interest is observed at those sites at time $t = 1, \cdots, T$. Thus, the judgment sample data are $\{Z(\mathbf{s}_i, t) : i = 1, \cdots, n;\ t = 1, \cdots, T\}$. At time $t = \quad < T$, a probability-based sampling design is implemented, selecting sites $\mathbf{u}_1, \cdots, \mathbf{u}_m$. Data are then collected at times $t = \quad , \quad + 1, \cdots$, so that the probability sample data are $\{Z(\mathbf{u}_i, t) : i = 1, \cdots, m;\ t = \quad , \cdots, T\}$. More generally, new probability sample sites may be selected in each sample interval, or a serially alternating design may be implemented. For ease of notation, however, we shall use a permanent station sample design here (see Section 3.3).

Our objective is to back predict what data would have been obtained had a probability-based sampling design been implemented from the very beginning of the monitoring program; that is, predict the unobserved values of $\{Z(\mathbf{u}_i, t) : i = 1, \cdots, m;$ $t = 1, \cdots, \quad - 1\}$. Kriging is perhaps the most popular method of spatial prediction (Cressie 1989), and can be easily extended to spatio-temporal prediction. This popularity owes much to its stability with respect to violations of model assumptions (e.g., Cressie and Zimmerman 1992). In particular, kriging is not sensitive to whether or not a spatial trend is included in the model (Journel and Rossi 1989), or to misspecification of the variogram model (Stein and Handcock 1989).

If the complete data base were to be used, spatio-temporal prediction would require the solution of $nT + m(T - \quad + 1) + p + 1$ linear equations for the same number of unknowns. This may not be practical for a reasonably large data set. Therefore, the following spatio-temporal predictor shall only use data from the judgment sample at time t, and data from the probability design at time to predict the unobserved values of the data from the probability sample at time t. Then the universal kriging predictor is

$$\widehat{Z}(\mathbf{u}_k, t) = \sum_{i=1}^{n} \lambda_{1i} Z(\mathbf{s}_i, t) + \sum_{i=1}^{m} \lambda_{2i} Z(\mathbf{u}_i, \quad), \tag{23}$$

where the coefficients $\lambda_{11}, \cdots, \lambda_{1n}, \lambda_{21}, \cdots, \lambda_{2m}$ are selected to minimize the mean squared prediction error subject to the constraint that the resulting predictor be unbiased for the true value. These coefficients can be obtained by solving the linear

system of $n+m+p+1$ equations

$$\sum_{i=1}^{n}\lambda_{1i}\gamma\left(\left\|\mathbf{s}_i-\mathbf{s}_j\right\|,0\right)+\sum_{i=1}^{m}\lambda_{2i}\gamma\left(\left\|\mathbf{u}_i-\mathbf{s}_j\right\|,\quad -t\right)+\xi_0+\sum_{i=1}^{p}\xi_i x_i(\mathbf{s}_j,t) = \gamma\left(\left\|\mathbf{s}_j-\mathbf{u}_k\right\|,0\right);$$
$$j = 1,\cdots,n,$$

$$\sum_{i=1}^{n}\lambda_{1i}\gamma\left(\left\|\mathbf{s}_i-\mathbf{u}_j\right\|,0\right)+\sum_{i=1}^{m}\lambda_{2i}\gamma\left(\left\|\mathbf{u}_i-\mathbf{u}_j\right\|,\quad -t\right)+\xi_0+\sum_{i=1}^{p}\xi_i x_i(\mathbf{u}_j,t) = \gamma\left(\left\|\mathbf{u}_j-\mathbf{u}_k\right\|,\quad -t\right);$$
$$j = 1,\cdots,m,$$

$$\sum_{i=1}^{n}\lambda_{1i}+\sum_{i=1}^{m}\lambda_{2i}=1,$$
$$\sum_{i=1}^{n}\lambda_{1i}x_j(\mathbf{s}_i,t)+\sum_{i=1}^{m}\lambda_{2i}x_j(\mathbf{u}_i,\quad)=x_j(\mathbf{u}_k,t);\ j=1,\cdots,p,$$

for the $n+m+p+1$ unknowns $\lambda_{11},\cdots,\lambda_{1n},\lambda_{21},\cdots,\lambda_{2m},\xi_0,\xi_1,\cdots,\xi_p$. This system of equations is called the kriging equations. The precision of the resulting kriging predictor is described by the kriging variance

$$\sigma^2(\mathbf{u}_k,t)=\sum_{i=1}^{n}\lambda_{1i}\gamma\left(\left\|\mathbf{u}_k-\mathbf{s}_i\right\|,0\right)+\sum_{i=1}^{m}\lambda_{2i}\gamma\left(\left\|\mathbf{u}_k-\mathbf{u}_i\right\|,\quad -t\right)+\xi_0+\sum_{i=1}^{p}\xi_i x_i(\mathbf{u}_k,t).$$

6.3 Simulation Model

Spatio-temporal data comprised of observations from both judgment and probability sampling designs are not available. Therefore, we must rely on simulation to assess the efficacy of the above approach to combining. In particular, data shall be simulated from the spatio-temporal random model

$$Z(\mathbf{s},t)=a_0 f(t)+a_s\mu(\mathbf{s})+a_t\alpha(t)+a_{st}\beta(\mathbf{s},t)+a_\varepsilon\varepsilon(\mathbf{s},t). \qquad (24)$$

The function $f(t)$ models the background temporal trend (Figure 12). The spatial random field $\mu(\mathbf{s})$ has unit variance and spatial correlation function

$$\rho_s(h) = \exp\{-3h/\kappa_s\}$$

with a long range of spatial dependence of $\kappa_s = 200$ km; for the current application, it can be considered to model the spatial trend in the data (Figure 13). Likewise, the temporal process $\alpha(t)$ has unit variance and temporal correlation function

$$\rho_t(r) = \exp\{-3r/\kappa_t\}$$

with a long random of temporal correlation of $\kappa_t = 3000$ years. The spatio-temporal random process $\beta(\mathbf{s}, t)$ allows the temporal trend to depend on location; it has unit variance and spatio-temporal correlation function

$$\rho_{st}(h, r) = \exp\{-3h/\kappa_s - 3r/\kappa_t\}$$

with relatively short ranges of spatial and temporal correlation set at $\kappa_s = 20$ km and $\kappa_t = 10$ years. All three of the above processes were simulated using the spectral method (Shinozuka 1971; Mejia and Rodriguez-Iturbe 1974). The error $\varepsilon(\mathbf{s}, t)$ is Gaussian white noise with unit variance, and models the effects of measurement error. It was simulated using the polar method (Ripley 1987, p. 62).

The relative influence of the four component processes on the resulting data can be fixed by varying the levels of the coefficients a_s, a_t, a_{st}, and a_ε. If we set $a_{st} = 0$,

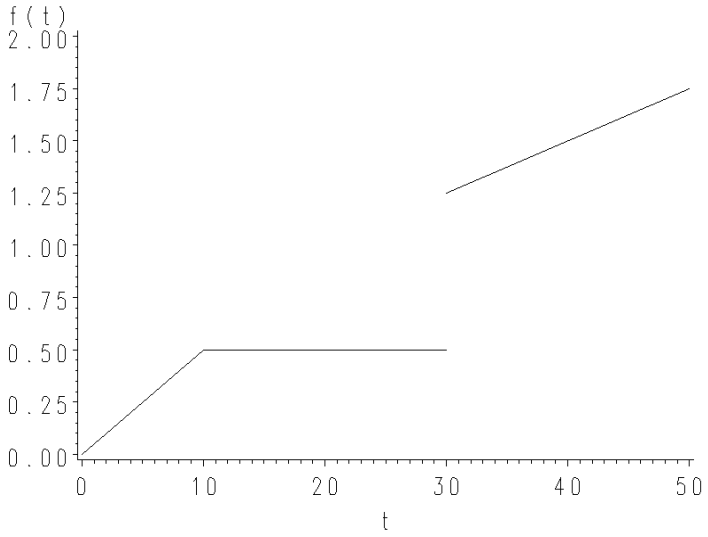

Figure 12: Temporal Trend.

then the spatial and temporal effects are additive, and the sampling bias attributed to the judgment sampling design can be simply removed by subtraction. It seems more likely that temporal trends occurring in the data may depend on location, and so, sampling bias cannot be simply removed by subtraction.

Two samples of data are generated. A total of 100 probability sites $\mathbf{u}_1, \cdots, \mathbf{u}_{100}$ are obtained a simple random sample over a 100×100 km region. For the judgment sample, an additional 100 sites $\mathbf{s}_1, \cdots, \mathbf{s}_{100}$ are independently selected from the density proportional to

$$p(\mathbf{s}) = \frac{\exp\{\beta_0 + \beta_1 \mu(\mathbf{s})\}}{1 + \exp\{\beta_0 + \beta_1 \mu(\mathbf{s})\}},$$

which depends on the realization of the first component of our simulation model (24).

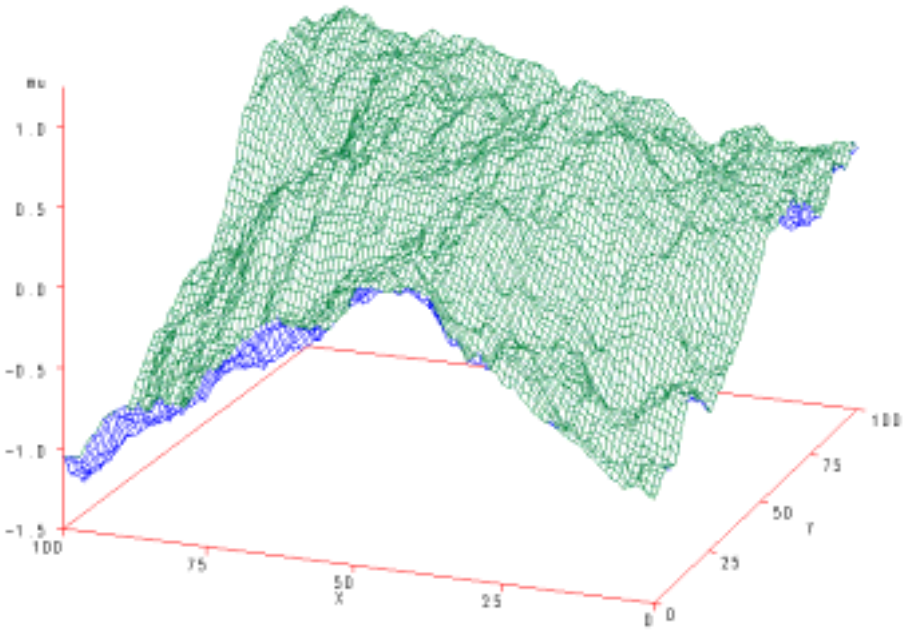

Figure 13: Spatial Trend.

Note that if $\beta_1 = 0$, then we obtain another simple random sample. For $\beta_1 > 0$, the judgment sample is biased in favor of high data values, while for $\beta_1 < 0$, the judgment sample is biased in favor of low data values. Data for both designs is generated for years $t = 1, \cdots, 50$, but it is assumed that the probability-based points are only observed for years $t = 41, \cdots, 50$.

6.4 Effect of Sampling Bias

The geostatistical methods described in Sections 6.1 and 6.2 are carried out conditional on what sites are actually included in the sample, and thus, ignore the effects of sampling variation on variogram estimates and spatio-temporal predictions. In particular, the potential effects of sampling bias in the judgment sampling design are not considered. These effects shall be thoroughly explored under the following values of the model parameters: $a_0 = 5$, $a_s = 10$, $a_t = 3$, $a_{st} = 3$, $a_\varepsilon = 0.5$, $\beta_0 = -1$, and $\beta_1 = 4$. Taking $\beta_1 > 0$ yields a judgment sampling design biased in favor of large values. In Figure 14, the sample means for both designs are plotted against time. Data from the judgment sampling design show an increasing trend over time (triangles), with a large jump in mean level occurring in year 31. The probability sites were only sampled after year 41, but, as expected given that $\beta_1 > 0$, have lower means than the judgment sites (circles). Our objective is to predict the unobserved values for the probability-based design from years 1 to 40.

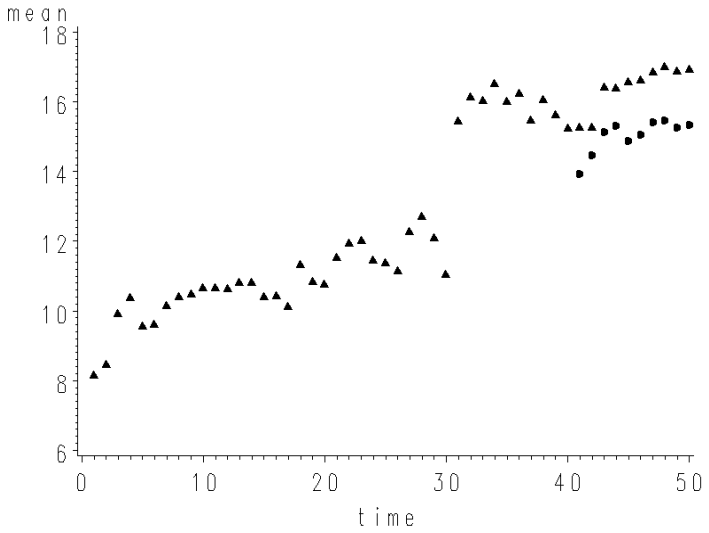

Figure 14: Annual means for judgment (triangles) and probability (circles) sample sites.

The data from the two designs were fitted separately to the planar trend model

$$Z(x, y, t) = \alpha_0 + \alpha_1 x + \alpha_2 y + \varepsilon(x, y, t),$$

where $Z(x, y, t)$ denotes the data collected at coordinates (x, y) at time t, and $\varepsilon(x, y, t)$ is the model error. Ordinary least squares estimates yield the fitted models

$$Z(x, y, t) = 15.7 + 0.0701x - 0.0155y$$

for the judgment design, and

$$Z(x, y, t) = 18.2 + 0.0867 - 0.0269y$$

for the probability-based design. Notice that the estimated partial slopes are of lower magnitude under the judgment design than under the probability-based design.

The method of moments estimator $2\widehat{\gamma}_s(h)$ for the spatial variogram $2\gamma_s(h) = 2\gamma(h,0)$ (expression (22)) was computed separately for each of the two designs. The results suggest that the biased judgment sampling design also yields a biased estimate of the variogram. For both designs, $2\widehat{\gamma}_s(h)$ increases rapidly to an asymptote with increasing h (Figure 15). However, the asymptote under the judgment sampling design appears to be larger than that under the probability-based design. Weighted least squares estimation was used to fit the exponential variogram model

$$2\gamma_s(h) = 2\sigma^2(1 - \exp\{-3h/\kappa_s\})$$

to $2\widehat{\gamma}_s(h)$ for each design, where σ^2 is the variance of the data, and κ_s is the range of spatial correlation. The two designs yielded nearly identical estimated ranges of spatial correlation; $\widehat{\kappa}_s = 31.7$ km for the probability-based sites, and $\widehat{\kappa}_s = 30.6$ km for the judgment sites. However, the judgment sites show a higher variance ($\widehat{\sigma}^2 = 12.02$) than the probability sites ($\widehat{\sigma}^2 = 10.45$).

Under the assumptions of the model, the method of moments estimator of the temporal variogram $2\gamma_t(r) = 2\gamma(0,r)$ (expression 22) remains unbiased even when a biased sample is obtained. Therefore, the estimate of the temporal variogram was obtained by pooling all of the observed data. The temporal variogram $2\widehat{\gamma}_t(r)$ increases to an asymptote with increasing time lag r (Figure 16). Weighted least squares estimation was used to fit the exponential variogram model

$$2\gamma_t(r) = 2\sigma^2(1 - \exp\{-3h/\kappa_t\})$$

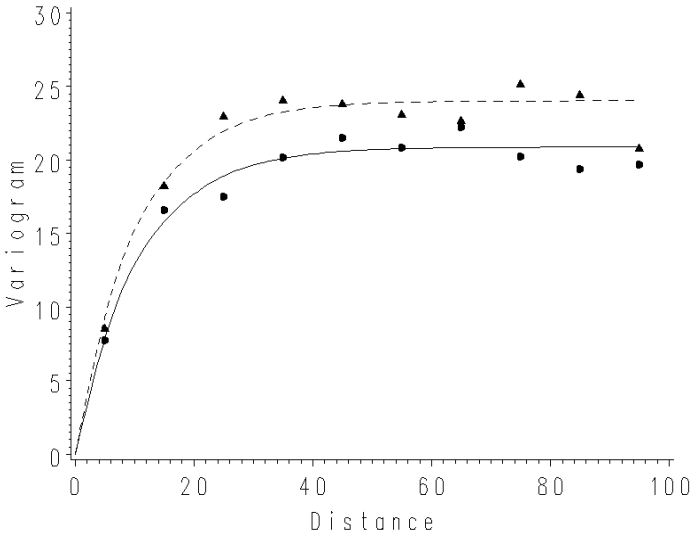

Figure 15: Fitted spatial variogram models for probability-based (solid line fit to the circles), and judgment (dashed line fit to the triangles) sample sites.

to $2\widehat{\gamma}_t(r)$, where σ^2 is the variance of the data, and κ_t is the range of temporal correlation. The estimated range of temporal correlation was $\widehat{\kappa}_t = 12.2$ years.

The universal kriging predictor (23) was computed for the unobserved data at the probability sample sites between years 1 and 40. Then, within each of these years, the mean of the predicted values was computed using

$$\widehat{\mu}_t = \frac{1}{m}\sum_{i=1}^{m} \widehat{Z}(\mathbf{u}_i, t), \tag{25}$$

where $\widehat{Z}(\mathbf{u}_i, t)$ is given by expression (23). Figure 17 compares these mean predicted values (x's) with the unobserved mean values (open circles) of the probability sites in years 1 to 40. Note that the means of the predicted values form a smoother curve

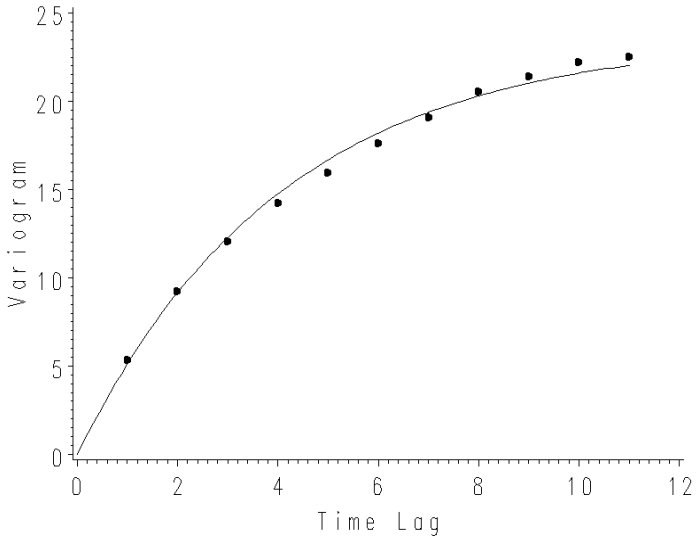

Figure 16: Fitted temporal variogram model.

than either the observed means of the judgment sample sites, or the unobserved means of the probability sample sites. This is not unexpected given that kriging is a smoothing algorithm. The means of the predicted values do tend to fall below the observed means from the judgment sample sites (triangles) indicating that the proposed procedure does reduce the bias attributed to the judgment sampling design. Moreover, the means of the predicted values successfully pick up the discontinuity in the data at year 31. However, the means of the predicted values also tend to fall above the unobserved means of the probability sample sites (open circles), which they were intended to predict. Thus, the proposed procedure still yields biased predictions.

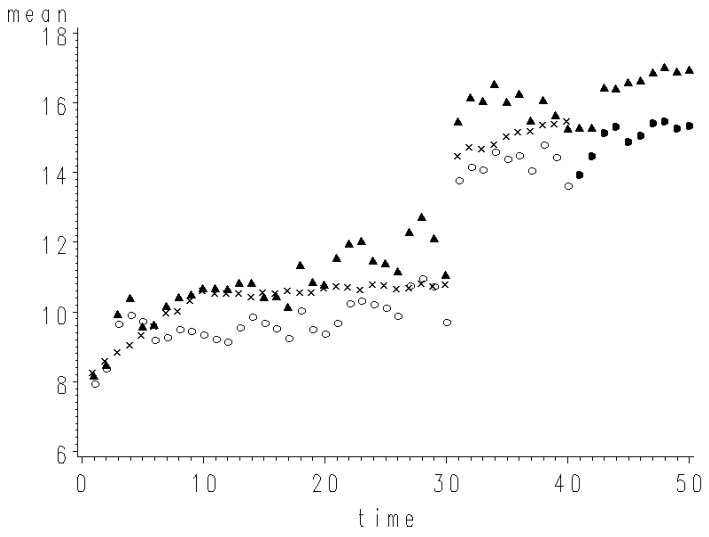

Figure 17: Comparison of predicted (x's) and unobserved (open circles) sample means for probability sites in years 1 to 40. In addition, observed annual means for judgment (triangles) and probability (open circles) sample sites are given.

6.5 Bias Reduction

The observed positive bias is not unexpected given that the judgment sample sites are biased in favor of high data values, and given the role that the judgment sample sites play in predicting unobserved past values at the probability sample sites. The magnitude of this bias can be estimated using data from those years in which observations from both designs are available. This can be accomplished by predicting the observed data from the probability based design using

$$\widehat{Z}_j(\mathbf{u}_k, t) = \sum_{i=1}^{n} \lambda_{1i} Z(\mathbf{s}_i, t) + \sum_{i=1}^{m} \lambda_{2i} Z(\mathbf{u}_i, t+j);\ t = \quad , \cdots, T-1,$$

where the coefficients $\lambda_{11}, \cdots, \lambda_{1n}, \lambda_{21}, \cdots, \lambda_{2m}$ are selected to minimize the mean squared prediction error subject to the constraint that the resulting predictor be unbiased for the true value of the data. This predictor uses data from the judgment sampling at time t, and data from the probability-based sampling design at time $t+j$ to predict the data for the probability-based design at time t. Then the prediction bias of $\widehat{Z}_j(\mathbf{u}_k, t)$ is given by

$$b_j(\mathbf{u}_k, t) = \widehat{Z}_j(\mathbf{u}_k, t) - Z_j(\mathbf{u}_k, t).$$

The subscript j is included in $\widehat{Z}_j(\mathbf{u}_k, t)$ and $b_j(\mathbf{u}_k, t)$ to take into account that the prediction bias may depend on the number of time lags j in the past we are attempting to back predict the probability-based data. The mean prediction bias in year t under

a predictor using probability based data j years in the future is then given by

$$\overline{b}_{jt} = \frac{1}{m} \sum_{k=1}^{m} b_j(\mathbf{u}_k, t).$$

Table 7 gives the mean prediction bias in year t under predictors using probability based data j time lags in the future. Notice that the prediction bias depends strongly on what year's data we are attempting to predict. However, this is of little use for estimating the mean bias in years 1 to 40. Within each year, the bias appears to increase somewhat with increasing time lag. This suggests that the magnitude of bias in the proposed predictor will increase as we attempt to back predict the probability data further into the past.

To quantify the relationship between mean bias and time lag, the general linear model

$$\overline{b}_{jt} = \mu + \alpha_j + \beta_t + \varepsilon_{jt}$$

was fit to the observations in table 7, where μ is the over mean α_j is the effect of time lag j, and β_t is the effect of year t. Then the mean bias for time lag j was adjusted to take into account variation among years using the general linear models procedure of SAS (SAS Institute 1985). The adjusted mean bias is then plotted against time lag as shown in Figure 18. Notice that the adjusted mean bias appears to increase linearly with increasing time lag, further indicating the bias in the proposed predictor increases as we attempt to back predict the probability data further into the past.

	lag j								
Year	1	2	3	4	5	6	7	8	9
41	1.428	1.488	1.515	1.515	1.517	1.542	1.567	1.566	1.579
42	1.118	1.149	1.144	1.145	1.176	1.203	1.201	1.215	
43	0.509	0.487	0.481	0.514	0.542	0.536	0.551		
44	0.518	0.517	0.561	0.596	0.589	0.606			
45	1.015	1.059	1.093	1.077	1.093				
46	0.927	0.972	0.950	0.969					
47	0.717	0.675	0.693						
48	0.691	0.725							
49	1.097								

Table 7: Mean bias as a function of year in which probability data are predicted and time lag.

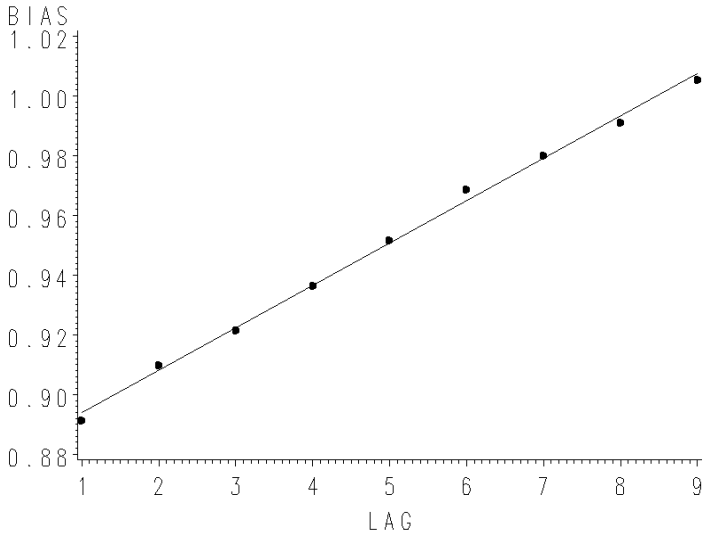

Figure 18: Adjusted mean bias plotted against time lag.

Fitting a linear model to the data in Figure 18, we obtain the following estimate for the bias at time lag j:

$$\widehat{b}_j = 0.87989 + 0.014164j.$$

Using the expression above, a bias corrected predictor for the mean of the probability sample sites in year t is given by

$$\begin{aligned} \widetilde{\mu}_t &= \widehat{\mu}_t - \widehat{b}_{-t} \\ &= \widehat{\mu}_t - 0.87989 - 0.014164 \times (41 - t). \end{aligned}$$

Figure 19 compares bias corrected predicted mean values (x's) with the unobserved mean values (open circles) of the probability sites in years 1 to 40. Comparing the results in Figure 19 with those previously obtained in Figure 17, notice that the bias

correction was successful in reducing the bias in predicted values. However, there is a suggestion of a small overcorrection, with biased corrected predictions falling slightly below the unobserved means that they are attempting to predict. That the predicted values fall well below the unobserved means in the first nine years can be attributed to the observation that the judgment sites show very little sampling bias in those years. This points to one of the shortcomings of the proposed approach to back prediction; it assumes that the sampling bias shows no temporal trends. Nevertheless, it is interesting to note that the predicted values track the trend function in Figure 12 very well.

6.6 Conclusions and Recommendations

The above approach exploits the spatio-temporal correlation with historical data from the judgment sampling design to back predict the unobserved means at probability sample sites. To compensate for the sampling bias of the judgment sample, a bias correction is required. This approach requires the careful modeling of any spatial trends that may occur over the study region, the spatio-temporal correlation structure in the data, and the bias resulting from the judgment sample design. To ensure that model assumptions are satisfied, appropriate diagnostic procedures should be implemented.

Bias correction requires a period of overlap in which observations are collected

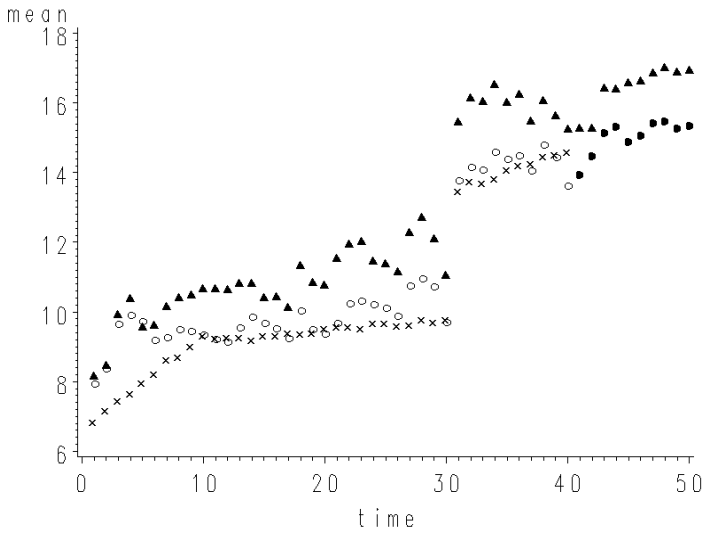

Figure 19: Comparison of bias corrected predicted (x's) and unobserved (open circles) sample means for probability sites in years 1 to 40. In addition, observed annual means for judgment (triangles) and probability (closed circles) sample sites are given.

from both sampling designs. Further research is required to determine how long that period of overlap should be. The bias correction also assumes that the sampling bias of the judgment design shows no temporal trends. In practice, it is not possible to determine that validity of this assumption. Improved predictions could potentially be obtained if sites from both the probability-based and judgment sampling designs are partitioned into strata selected to minimize sampling bias of judgment sites within strata. Such strata might be selected using the methods of Overton et al. (1993). Stratum identification can then be used as explanatory variables in the spatio-temporal model (21), not only improving the precision of predictions, but also reducing the effects of sampling bias.

References

Cox, L.H., and Piegorsch, W.W. 1996. Combining environmental information I: environmental monitoring, measurement, and assessment. *Envrionmetrics* **7,** 299-308.

Cressie, N. 1989. The many faces of spatial prediction. In *Geostatistics, Vol.* **1,** M. Armstrong, ed., Kluwer, Dordrecht, 163-176.

Cressie, N. 1991. *Statistics for Spatial Data.* Wiley, New York.

Cressie, N., and Zimmerman, D.L. 1992. On the stability of the geostatistical method. *Mathematical Geology* **24,** 45-59.

Horn, C.R., and Grayman, W.M. 1993. Water-quality monitoring with EPA reach file system. *Journal of Water Resources Planning and Management* **119**, 262-274.

Journel, A.G., and Rossi, M.E. 1989. When do we need a trend model in kriging? *Mathematical Geology* **21,** 715-739.

Larsen, D.P., and Christie, S.J. 1993. EMAP-Surface Waters 1991 Pilot Report. EPA/620/R-93/003. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.

Mejia, J.M., and Rodriguez-Iturbe, I. 1974. On the synthesis of random field sampling from the spectrum: An application to the generation of hydrologic spatial processes. *Water Resources Research* **10,** 705-711.

Messer, J.J., Linthurst, R.A., and Overton, W.S. 1991. An EPA program for monitoring ecological status and trends. *Environmental Monitoring and Assessment* **17,** 67-78.

Myers, R.H. 1976. *Response Surface Methodology.*

Olsen, A.R., Sedransk, J., Edwards, D., Gotway, C.A., Liggett, W., Rathbun, S., Reckhow, K.H., and Young, L.J. 1998. Statistical issues for monitoring ecological and natural resources in the United States. *Environmental Monitoring and*

Assessment (in press).

Ripley, B.D. 1987. *Stochastic Simulation.* Wiley, New York.

Overton, J., Young, T., and Overton, W. 1993. Using found data to augment a probability sample: procedure and a case study. *Environmental Monitoring and Assessment* **26,** 65-83.

Rao, J.N.K., and Graham, J.E. 1964. Rotation designs for sampling on repeated occasions. *Journal of the American Statistical Association* **59,** 492-509.

Searle, S.R. 1971. *Linear Models.* Wiley, New York.

Shinozuka, M. 1971. Simulation of multivariate and multidimensional random processes. *Journal of the Acoustical Society of America* **49,** 357-367.

Skalski, J.R. 1990. A design for long-term status and trends monitoring. *Journal of Environmental Management* **30,** 139-144.

Snedecor, G.W., and Cochron, W.G. 1980. *Statistical Methods, Seventh Edition.* Iowa State University Press, Ames, Iowa.

Stein, M.L., and Handcock, M.S. Some asymptotic properties of kriging when the covariance function is misspecified. *Mathematical Geology* **21,** 171-190.

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

Thompson, S.K. 1990. Adaptive cluster sampling. *Journal of the American Statistical Association* **85,** 1050-1059.

Thompson, S.K. 1992. *Sampling.* Wiley, New York.

Urquhart, N.S., Overton, W.S., and Birkes, D.S. 1993. Comparing sampling designs for monitoring ecological status and trends: impact of temporal patterns. In *Statistics for the Environment,* V. Barnett, and K.F. Turkman, eds., 71-85.