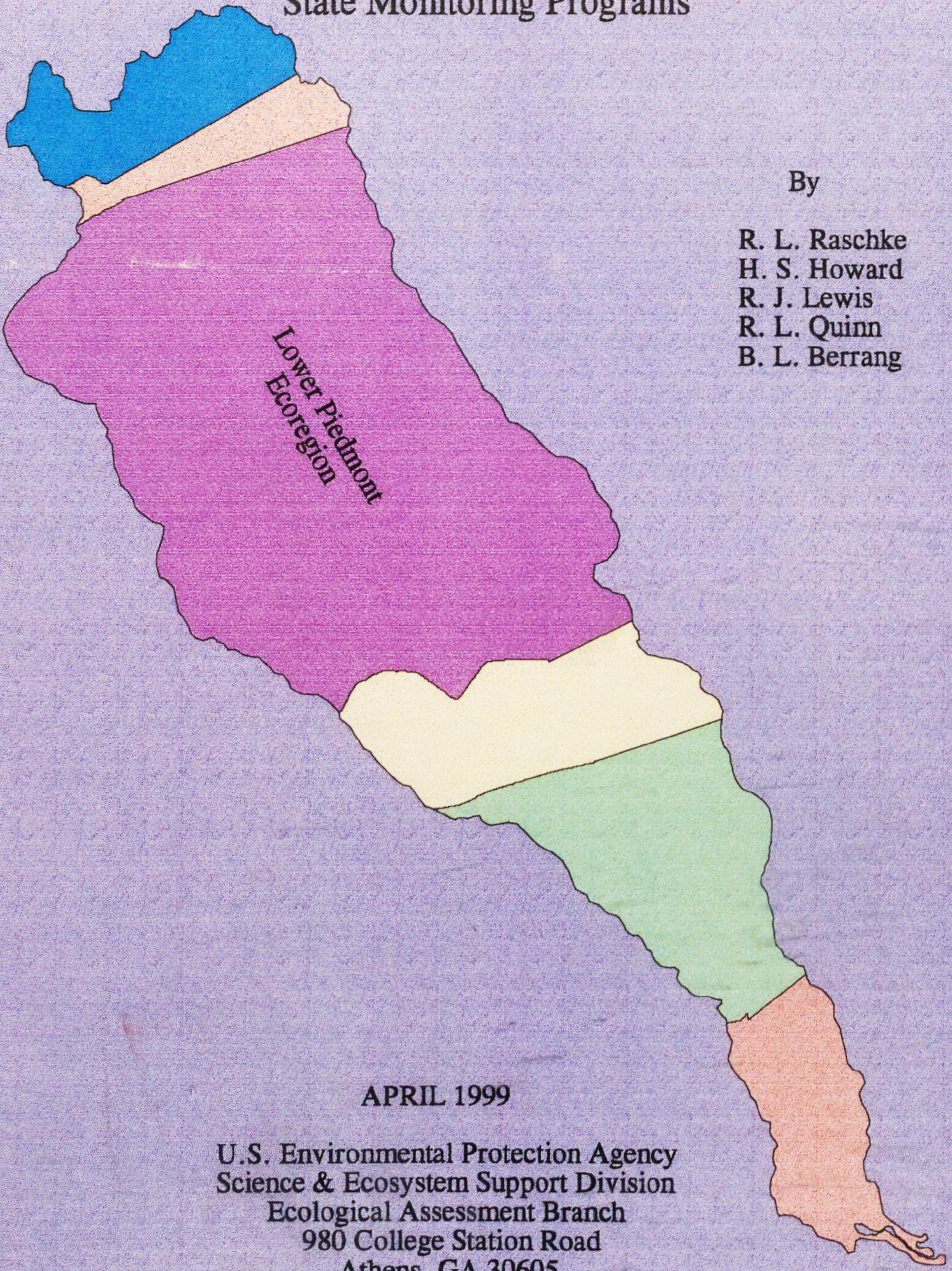
EPA/904-R-99-002

Savannah River Basin REMAP: A Demonstration of the Usefulness of Probability Sampling for the Purpose of Estimating Ecological Condition in State Monitoring Programs



By

R. L. Raschke
H. S. Howard
R. J. Lewis
R. L. Quinn
B. L. Berrang

APRIL 1999

U.S. Environmental Protection Agency
Science & Ecosystem Support Division
Ecological Assessment Branch
980 College Station Road
Athens, GA 30605

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Completion of this project was dependent upon EPA's management and the talents of numerous individuals within and outside of EPA. Special thanks is extended to EPA Region 4's Administrator, Mr. John Hankinson, for his financial support and faith in the Savannah REMAP team. Likewise, appreciation goes to Russ Wright, Director of Region 4's Science and Ecosystem Support Division (SESD), Delbert Hicks, Chief of the Ecological Assessment Branch (EAB), Antonio Quinones, Acting Chief of the EAB, and Meredith Anderson, EPA Region 4's Savannah River Basin Coordinator for their support over the life-time of the project. Trudy Stiber, Mel Parsons, Bo Noakes, and Jim Davee of the EAB contributed significantly to the graphics of this report. Jim Harrison, Dave Melgaard, Jim Greenfield, John Greis, John Deatrick, Dan Ahern and Larry Burns deserve a special thanks for their ideas and critiques of the early work. Gratitude is extended to Don Schultz and Alana Conley of the Algal Assay Laboratory who provided timely assistance. Thanks also to David Millie and Larry Boihem of USDAs Southern Regional Research Center and Lila Melendez and Arthur Burks of the Region 4 Analytical Support Branch (ASB) for their assistance in HPLC analysis. Additionally, we would like to thank Jenny Scifres, Anthony Carroll, Tom Sack, and Pam Betts of ASB for phosphorus and solids analyses. Kent Thornton, Don Stevens Jr., Steve Paulsen, Steve Weisberg, Gary Saul, and Christina Laurin of the the Office of Research and Development's EMAP team provided invaluable assistance during the planning and data evaluation stages. Special acknowledgement is due Jim Maudsley and his team who assisted in the field sampling and development of guidelines for locating and accessing the sites. Likewise, a special debt of gratitude is extended to Mike Van Den Avayle of the National Biological Survey who worked with us on an interagency agreement that provided the services of Dr. Steven Rathbun of the University of Georgia. Dr. Rathbun was invaluable in providing statistical guidance and significant work products during the study. Likewise, the expertise of Dr. Bud Freeman of the University of Georgia was most helpful during confirmation of fish species.

The cooperation of the state's of Georgia and South Carolina was the key to accomplishing numerous tasks from the planning stage through the field work and review of reports. We especially extend our thanks to Alan Halum, Mork Winn, Trish Foster, David Kamps and Bill Kennedy of the Georgia Environmental Protection Division and Lonnie Dorn of the Georgia Department of Natural Resources; Butch Younginer, Kathy Stecker, Harry Gaymon, Rick Renfrow, Jim Glover,Scott Castleberry, Chad Altman,and Dave Chestnut of the South Carolina Department of Health and Environmental Control for their steadfast assistance.

Finally, we owe a debt of gratitude to the many private homeowners, lumber/paper companies, and forest mangers who permitted access through their properties to the sites.

SUMMARY

The Environmental Protection Agency's Environmental Monitoring and Assessment Program (EMAP) is an outcome of EPA's National Eutrophication and Acid Lake Monitoring Programs of the 1980s. EMAP is a statistical sampling program that has adopted a uniform approach for national and regional monitoring assessments across ecosystem types. EMAP uses a serially alternative probability-based sampling design that systematically allocates sampling effort over space and time to ensure adequate coverage followed with randomization to ensure unbiased estimates of status throughout the life of a project. The design does not rely on assumptions of population distribution, but describes the underlying structure of the population of interest. The approach is flexible and applicable to all landscape media. It has the ability to increase or reduce sampling density down to the ecoregion level, respond quickly to environmental problems, maintain representative coverage of environmental resources, and provide for sampling of fewer sites in an area but over rotating cycles. Through this project, an interval-overlap technique is presented that minimizes the loss of monitoring data when the EMAP approach is incorporated into a fixed station (judgement) monitoring program. The technique uses a back-prediction method with a bias-corrective factor to best fit the two types of monitoring derived data.

In cooperation with EMAP's desire to transfer this monitoring approach to the EPA regions and states, Region 4 established the Regional Environmental Monitoring and Assessment Program (REMAP). Region 4 teamed with scientists and managers in EPA's Office of Research and Development and the states of Georgia and South Carolina to conduct a demonstration of the new monitoring approach, answer questions about probability sampling and analysis, and address the concerns about the ecological condition of streams and large lake tributary embayments in the Savannah River Basin.

From a basin perspective, the tributary embayments with regard to trophic condition are in good condition. At worst, only about 5% of the acreage exhibited less than desirable conditions. There appeared to be a general decline southward with respect to stream EPT Index, dissolved oxygen, and conductivity. Average stream temperatures increased southward. Water quality violations were noted for dissolved oxygen and pH. A dissolved oxygen violation was noted on an unnamed tributary to Cliatt Creek in Columbia County, Georgia. Likewise about 8% of the stream miles were less than both state's pH standard of 6.0 and 2% of the miles were greater than the allowable South Carolina standard. An examination of basin-wide stream

conditions over a two-year period indicated that up to 52% of the stream miles were in poor ecological condition.

Because of a sufficient number of reference and sampling stations in the Lower Piedmont Ecoregion, EPA scientists focused on that scale in assessing stream condition over a four-year period. Consolidating information from an EPT Index, Fish Index, and Habitat Score, scientists developed a Lower Piedmont Ecological Index (LPEI). The LPEI showed that 69% of the Ecoregion's stream miles are in fair to poor ecological condition. Most of this adverse impact is attributed to habitat degradation in the form of excessive sedimentation. One area of the landscape along the I85 corridor showed an unusually high number of poor stream sites and it is the conclusion of the scientists that this area is in need of further study.

1.0 INTRODUCTION

1.1 PURPOSE

Responding to increased population growth and demands for multiple uses of natural resources, The Environmental Protection Agency (EPA) established the Watershed Protection Approach (WPA) in 1991 (EPA, 1991; 1996). The WPA is a program for identifying and preventing environmental problems, setting priorities, and developing solutions through an open, inclusive process with the people (stakeholders) who live in a geographical setting. Consideration of economic prosperity and environmental well-being is the cornerstone of WPA. The Savannah River Basin was one of two areas selected in 1993 for the WPA in Region 4 because of its high public use, known environmental problems, susceptibility for further degradation, interest in participation by the users, and the likelihood of success. Through the WPA initiative, EPA Region 4 brought together scientists and stakeholders who developed a strategy to provide an ecological focus for resolving problems. This strategy gave birth to the Savannah River Basin Watershed Project (SRBWP) (Management Committee, 1995). The goal of the SRBWP is to develop and implement a multi-agency environmental protection management project which incorporates the authorities and expertise of all interested parties in an effort to accomplish the vision of conserving, restoring, enhancing, and protecting the Basin's ecosystems in a way that allows the balancing of multiple uses. Further details on objectives and issues within the basin can be found in Volume I of the "SRBWP Initial Assessment and Prioritzation Report" by the Management Committee (1995). Part of the SRBWP strategy included a monitoring component, The Regional Environmental Monitoring and Assessment Program (REMAP)(FTN et al., 1994).

Environmental monitoring programs have developed in response to specific needs, such as compliance monitoring by regulating agencies responsible for the condition of surface waters, or fixed-station monitoring networks that primarily address indicators of exposure and stress. Some of the monitoring programs are driven by mandates in the Clean Water Act (CWA). The reports required by Sections 305(b) and 314 of the CWA are an example. Programs that collect data on other ecosystem types have also been established. For example, the U. S. Department of Agriculture (USDA) National Agricultural Statistical Survey collects data for agricultural resources; The Forest Service's Inventory and Analysis Surveys analyze forest resources; and the U. S. Geological Survey's National Water Quality Assessment (NAWQA) program monitors water quality in selected basins. None of the programs, however, have adopted a uniform approach for national and regional assessments across and among ecosystem types. The Environmental Monitoring and Assessment Program (EMAP) and its counterpart, REMAP, is intended to fill that gap by providing the U. S. EPA Administrator, Congress, and the public with statistical data summaries and periodic interpretative reports on ecological status and trends. Because knowledge about uncertainty is important for interpreting quantitative environmental data, EMAP is designed to make rigorous uncertainty estimates as well (Larsen et al., 1991).

The REMAP was developed as a partnership between EMAP, EPA's Regional Offices, and States to promote the use of EMAP science. The objectives of REMAP follow:

1. **To evaluate and improve EMAP concepts for State and local use.**

2. **To assess the applicability of EMAP indicators and the EMAP approach at differing spatial scales.**

3. **To demonstrate the utility of EMAP for resolving issues of importance to the EPA, Regions, and States.**

The REMAP strategy lends itself to the benefits of a full partnership between states and federal agencies because both national and state monitoring needs can be met in a cost-effective manner. The EMAP approach can provide a cost-effective approach for assessing ecological data and reporting estimates of status and trends in indicators of condition with known confidence. State reporting requirements under several sections of the Clean Water Act (CWA) can be accomplished using an EMAP monitoring approach. Section 305(b) of the CWA requires states to submit biennial reports that include analysis of water quality data of all navigable waterways to estimate environmental impacts. The Clean Lakes Section 314 requires states to submit biennial reports that identify, classify, describe, and assess status and trends in water quality of publicly owned lakes. REMAP projects are being designed to provide meaningful information to decision-makers within a 1- to 2-year period.

1.2 POLICY-RELEVANT QUESTIONS

The Science and Ecosystem Support Division (SESD) of EPA Region 4 was asked by the Savannah River Watershed Project Policy Committee to implement the REMAP strategy as a demonstration project for the states of South Carolina and Georgia. These states were interested in reducing sampling frequency and analyses, having the ability to reduce or increase sampling density, responding quickly to emerging environmental problems, and maintaining representative coverage of environmental resources through a systematic-random means of sampling. Before the monitoring study, a set of questions was posed by the states of Georgia and South Carolina to provide direction for the monitoring design. The following policy-relevant questions were identified to guide the development of a plan of study and subsequent monitoring efforts.

- **What is the status of condition of the water resources of the Savannah River Basin?**
- **What proportion of the Savannah River Basin surface waters are attaining designated uses?**
- **What are the changes of ecological condition over time?**
- **What factors might be associated with changes?**
- **Is there a tendency for distribution of condition in a specific direction (spatial gradient) over the basin landscape? What are the possible reasons for these gradients?**
- **What resources are at risk in the Savannah River Basin?**

1.3 PROGRAM OBJECTIVES

In response to the needs of the states and policy-relevant questions posed, The Ecological Assessment Branch (EAB) of the SESD developed the following study objectives with the concurrence of the Policy Committee of the Savannah River Watershed Project.

- **Estimate the status and change of the condition of water resources in the Savannah River Basin;**
- **Identify water quality spatial gradients that exist within the Savannah River Basin and associate current and changing condition with factors that may be contributing to this condition and spatial gradients;**
- **Demonstrate the utility of the REMAP approach for ecoregion and river basin monitoring and its applicability for state**

monitoring programs;

- **Incorporate the REMAP approach in the formulation and accomplishment of the State River Basin Management Plans; and**

- **Provide baseline information required to conduct comparative risk assessments in the Savannah River Basin.**

1.4 DESCRIPTION OF THE SAVANNAH RIVER BASIN

The Savannah River originates in the mountains of Georgia, South Carolina, and North Carolina and flows south-southeasterly 312 miles to the Atlantic Ocean near the port city of Savannah, Georgia (Figure 1.1). The Savannah River is formed at Hartwell Reservoir by the Seneca and Tugaloo Rivers.

Headwater streams of the Seneca River are the Keowee River and Twelve-Mile Creek. The Tugaloo River is formed by the confluence of the Tallulah and Chattooga Rivers. The Savannah River flowing in a south-southeasterly direction forms the border between the states of Georgia and South Carolina. The river's entire length of 312 miles is regulated by three adjoining Corps of Engineers multipurpose reservoirs, each with appreciable storage. The three lakes, Hartwell, Russell, and Thurmond, form a chain along the Georgia-South Carolina border 120 miles long. Six power developments that are part of the Georgia Power Company hydropower network exist upstream of Hartwell Lake on the Tugaloo River system; Yonah and Tugaloo lakes on the Tugaloo River, and Tallulah Falls, Rabun, Seed, and Burton lakes on the Tallulah River. Upstream of Lake Hartwell, on the Seneca River, is Duke Power Company's Keowee-Toxaway Project. The project is composed of three adjoining reservoirs, the most downstream of which is Keowee Lake, and the other two, Jocassee and Bad Creek Lakes are pump storage projects(Figure 1.2).

The Savannah River Basin has a surface area of 10,577 square miles, of which 4,581 square miles are in South Carolina, 5,821 square miles are in Georgia, and approximately 175 square miles are in North Carolina. Like other basins of large rivers in the Southeast which flow into the Atlantic Ocean, the Savannah River Basin embraces three distinct areas: the Mountain Province, the Piedmont Province, and the Coastal Plain (Figure 1.3). The mountains and Piedmont are part of the Appalachian area. The division between the Mountain and Piedmont is an irregular line extending from northeast to southwest, crossing the Tallulah River at Tallulah Falls. The Fall Line, or division between the Piedmont Province and the Coastal Plain, also crosses the basin in a generally northeast to southwest direction, near Augusta, Georgia. Elevations within the Mountain Province of the basin vary from 1,500 feet National Geodetic Vertical Datum (NGVD) on the Tallulah River to 5,030 feet NGVD for the highest peak, Little Bald Mountain, in North Carolina along the watershed divide. The Blue Ridge is characterized by mountains covered naturally with Appalachian oak. Forests and ungrazed woodlands are the predominant land uses with some cropland and pastures. The Piedmont Province, due to its great width of over a hundred miles, is truly Piedmont only in the upper parts, and gives way to a midland area before reaching the Coastal Plain. Exclusive of river valleys, its elevation generally varies from 500 feet NGVD at the Fall Line to about 1,800 feet NGVD at its upper extremity. The Piedmont is characterized by gently sloping hills and smooth to irregular plains. This province is underlain naturally with nutrient poor soils supporting oak/hickory/pine and southern mixed forests. Land use is a mixture of crop lands, pasture, and woodlands with some urban areas. Within the Coastal Plain, elevations vary from 500 feet NGVD at the Fall Line to sea level at the Atlantic Ocean. Flat plains dominated naturally by oak/hickory/pine forests, pocasin (pine, holly) forests, southern flood plain forests (oak/tupelo, bald cypress), and southern mixed forests (beech, sweetgum, magnolia, pine and oak) are characteristic of the Coastal Plain.

Within the three physiographic provinces there exist distinct ecosystems based on the interrelationships between organisms and their environment. These distinct ecosystems are defined as ecoregions. Ecoregions are ecologically distinctive areas that result from the mesh and interplay of the geologic landform, soil, vegetative, climatic, wildlife, water and human factors which may be present (from Wilken, 1986) While physiographic provinces may prove suitable for regional or national assessments, definition of ecoregions among broad physiographic areas is necessary to accurately assess ecological condition or health. Ecoregions are distinct areas grouped by climate, soils, land forms, and vegetative cover. The Blue Ridge physiographic province stands alone as a separate ecoregion as does the Piedmont physiographic province. However, the Coastal Plains physiographic province is composed of three distinct ecoregions: the Fall Line Hills (or Sand Hills), the Southeastern Plains and Hills, and the Coastal Plains.

Land use in the basin is agriculturally oriented. Sixty-six percent of the basin is considered timberland and 34.1% is nonforested. The number of acres farmed remains constant. Between 1987 and 1992 there was little change in the total farm acreage in the basin. However, Georgia had 330 fewer farms and lesser acreage in 1992 than in 1987 while South Carolina had an increase of 931 farms and an increase of 110,134 acres in farm land. There was a shift over the same five-year period in the types of crops grown. An increase in the number of acres cultivated have occurred in corn (18%), cotton (86%), peanuts (12%), and tobacco (31%). These gains have been made with corresponding decreases in primarily wheat (-30%) and soybeans (-32%).

The Savannah River Basin contains all or part of 43 counties in Georgia, South Carolina, and North Carolina. Four of the counties are in North Carolina, thirteen in South Carolina, and twenty-six in Georgia. The population of the basin in 1990 was about 1,500,000 and is expected to grow to 1,800,000 by the year 2030. About 53% of the population resides in Georgia, 42% in South Carolina, and 5% in the headwaters located in North Carolina. Four metropolitan areas contain 62% of the basin's population. Savannah, Georgia is the largest city with 137,560 persons followed by Augusta, Georgia with a population of 44,619 (FTN et al., 1994; SRBWP, 1995; EPA, 1991; EPA, 1996).

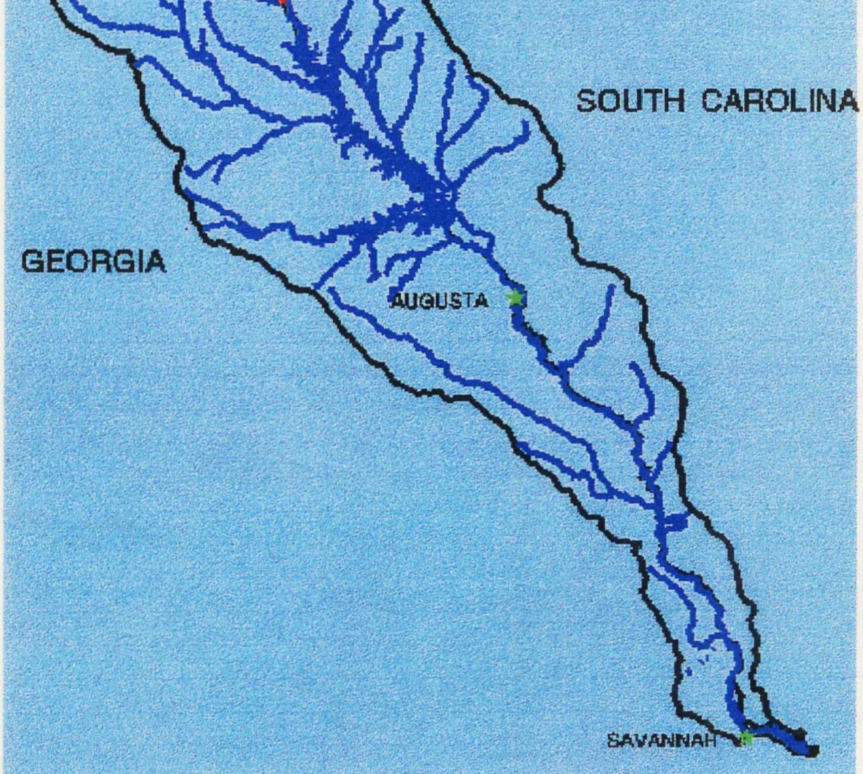


Figure 1.1 Savannah River Basin.

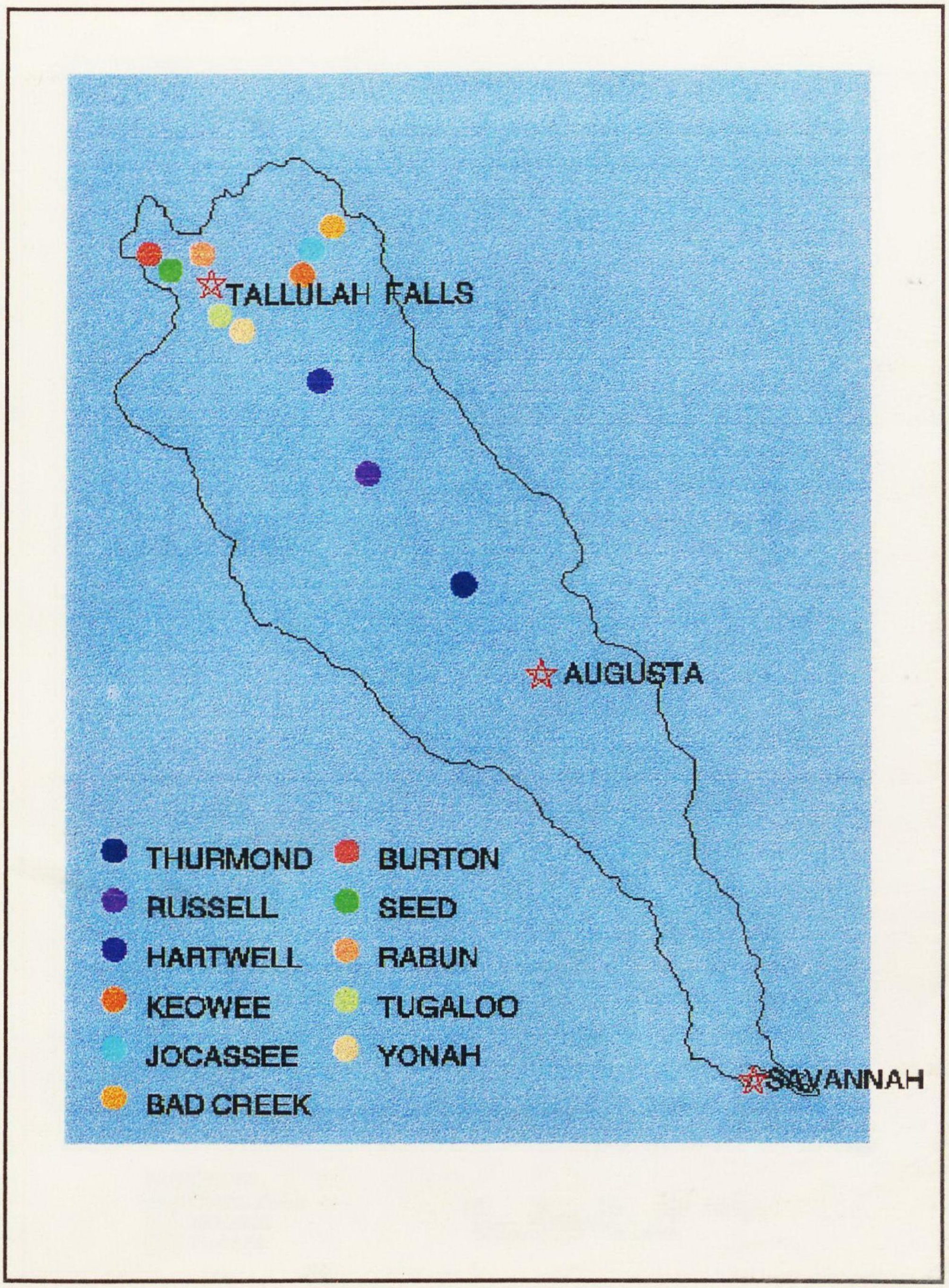


Figure 1.2 Location of Major Lakes in the Savannah River Basin.

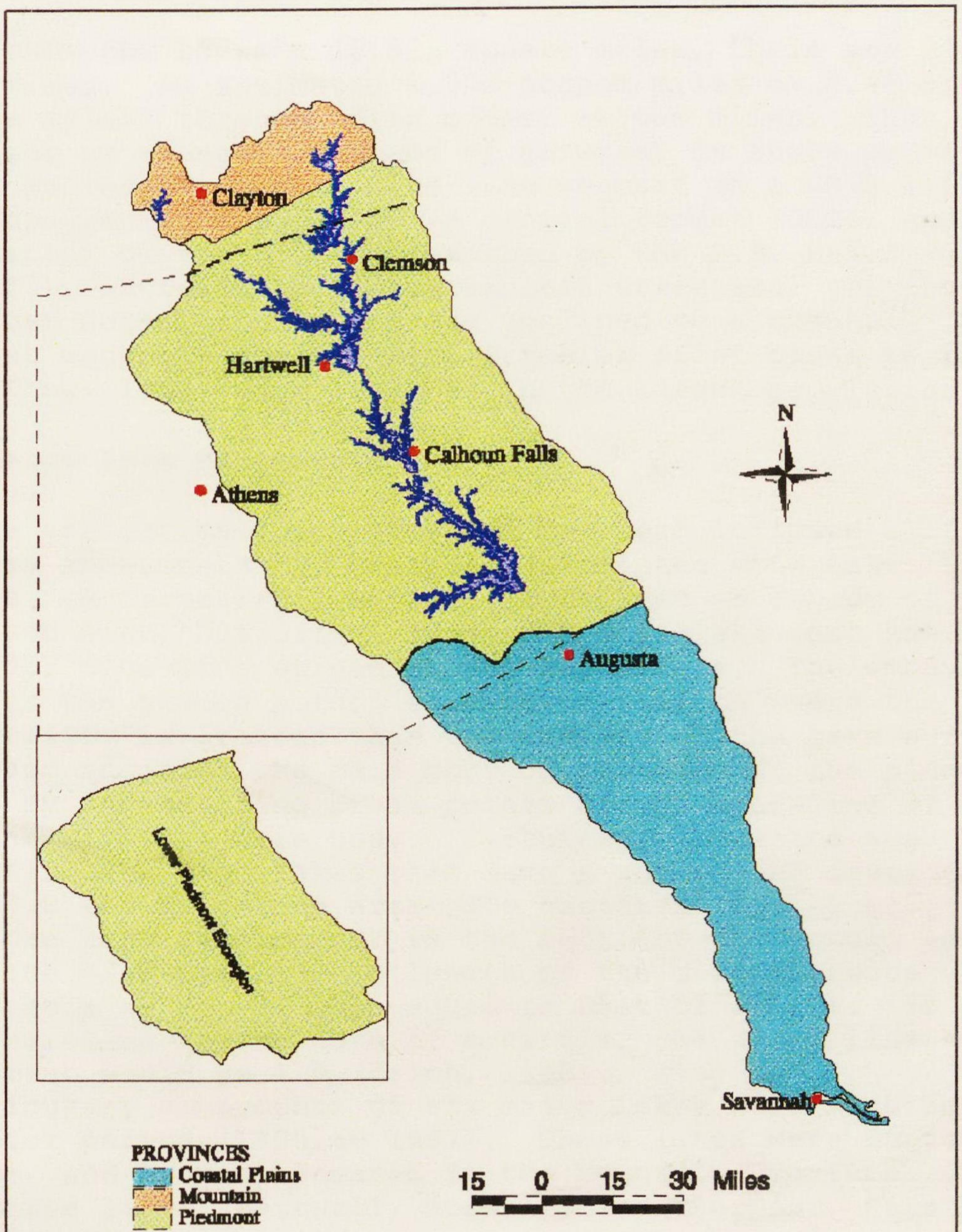


Figure 1.3. Physiographic Provinces and the Lower Piedmont Ecoregion of the Savannah River Basin.

2.0 STUDY DESIGN

2.1 Resources of Interest

2.1.1 Streams

Within the basin's 10,579 square miles, there are 17,354 stream miles. An estimated 1,503 stream miles or 5.4% are wadeable (first through third order) stream miles. The population of wadeable streams of interest is those permanent streams as indicated by a blue-line segment on a USGS 1:100,000-scale topographic map series in digital format (DLGs) and the modification of the DLGs represented by the U.S. EPA River Reach File (RF3). Streams typically exhibit unilateral gravity flow that under normal conditions are confined to a channel. All permanent wadeable streams from Strahler first order to third order (Chow, 1964) were included in the target population.

2.1.2 Large Lake Embayments

The statistical population of interest included all tributary embayments >20 hectares associated with lakes >500 hectares. A tributary embayment is defined as a body of water associated with, but offset from, the main lake that has a permanent, blue-line stream at its headwaters. The embayment begins at the plunge point, the stream stretch where the inflow water density is greater than the density of the lake surface water, and it joins the main body of the lake at the plane created by intersecting break points of the shoreline of the embayment with the main body. Tributary embayments are associated only with lakes that have a shore line development ratio >3.0 and a surface area >500 hectares (FTN _et al_., 1994).

Shore line development is the ratio of the actual length of shore line of a lake to the length of the circumference of a circle the area of which is equal to that of a lake. If a lake had a shoreline in the form of a circle, the shore line development would be 1.0 (Welch, 1948).

Tributary embayments of six major lakes were studied over a three-year period (1995 to 1997). These lakes were Burton, Jocassee, and Keowee, located in the Mountain Province. The other three lakes, Hartwell, Russell, and Thurmond, were located in the Piedmont Province.

Lake Burton, controlled by Georgia Power Company, is located near Clayton, Georgia. It is an old reservoir impounded in 1919. The lake has a shoreline length of 62 miles surrounding 2,775 acres containing 1,000,080 acre-feet of water.

Hartwell Lake is 7 miles east of Hartwell, Georgia. A dam is located at river mile 305.0. When the lake level is at elevation 660 ft. NGVD, the top of the conservation pool, the lake extends 49 miles up the Tugaloo River in Georgia, and 45 miles up the Seneca and Keowee Rivers in South Carolina, covering 55,900 acres. The shoreline at elevation 660 NGVD is about 962 miles long, excluding island areas. The lake has a total storage capacity of 2,550,000 acre-feet below elevation 660 NGVD. Hartwell dam began operation in 1963.

Russell dam is at River Mile 275.2 in Elbert County, Georgia and Abbeville County, South Carolina. The dam is 18 miles southwest of Calhoun Falls, South Carolina, and 40 miles northeast of Athens, Georgia. At the top of conservation pool elevation of 475 NGVD, the lake has a useable storage capacity of 126,800 acre-feet and a shoreline of 523 miles encompassing 26,000 acres. Operation of the project began in January 1984.

Thurmond Lake is 22 miles upstream of Augusta Georgia. At elevation 330 NGVD, at the top of the lake pool, the lake extends 40 miles up the Savannah River and about 30 miles up the Little River in Georgia. The lake has about 1,050 miles of shoreline, excluding island areas. At the top of the flood control pool (elevation 335 NGVD), the lake has an area of 78,500 acres with a total storage capacity of 2,510,000 acre-feet.

The three-project system is authorized and operated by the U.S. Corps of Engineers for fish and wildlife, flood control, hydro power, navigation, recreation, water quality, and water supply.

Duke Power Company built and controls Lakes Jocassee and Keowee. The upper lake, Jocassee, was built in 1973. It contains an area of 7,318 acres holding 1,077 acre-feet of water with a shoreline length of 75 miles. Lake Keowee, built in 1971, has a shoreline length of 300 miles encompassing 18,373 acres with a storage holding capacity of 955 acre-feet.

2.2 Statistical Sampling Design

A probabilistic sampling survey strategy was used to characterize the wadeable streams and tributary embayments of the Savannah River Basin. The sampling design was derived from the approach used in EMAP (Messer et al., 1991; Overton et al., 1990; Stevens et al., 1992).

Probability sampling designs use randomization in the sample site selection process. Probability sampling is the general term applied to sampling plans in which

- **every member of the population (i.e., the total assemblage from which individual sample units can be selected) has an equal chance of being included in the sample;**

- **the sample is drawn by some method of random selection consistent with these probabilities; and**

- **the probabilities of selection are used in making inferences from the sample to the target population (Snedecor and Cochran, 1967),**

One advantage of probability-based surveys is their minimal reliance on assumptions about the underlying structure of the population (e.g., normal distribution). In fact, one of the goals of probability-based surveys is to describe the underlying structure of the population. Randomization is an important aspect of probability-based surveys. Randomization ensures that the sample represents the population. Without probability sampling, each sample often is assumed to have equal representation in the target population, even though selection criteria clearly indicate this is not the case. Without the underlying statistical design and probability samples, the representativeness of an individual sample is unknown. Drawing inferences from samples selected without randomization and without incorporating inclusion probabilities can lead to misleading conclusions.

One can study conditions of streams in two ways. The first is by census, which entails examining every point on the streams. This method is impracticable. A more practicable approach is to examine some points systematically to ensure adequate coverage of the basin, and randomly to prevent bias in selection of stream points. For example, we would not obtain a good estimate of the percent of all students in a region with hepatitis if we polled only students in small towns of less than two thousand people. This preferential or biased sample would most likely include a much lower proportion of students with hepatitis than the general population of students. Similarly, in a stream study, preferential sampling occurs if the sample includes only sites, for example, downstream of sewage outfalls where sewage outfalls affect only a small percentage of total stream length. This kind of sampling program may provide useful information about conditions downstream of sewage outfalls, but it will not produce estimates that accurately represent conditions of the whole basin. Preferential selection can be avoided by collecting random samples.

Randomization can be thought of as a kind of lottery drawing to determine which points are included in the sample. Randomization is important. When used, it is possible to estimate condition of streams with a known degree of confidence. In REMAP, hexagons are used to add the systematic element to the design. The hexagonal grid is positioned randomly over the basin map, and sampling points from within each hexagon are selected randomly. The grid ensures spatial separation of selected sampling points. This design's sampling requirements reduce sampling locations to a logistically and economically feasible number. It allows fewer sites to be sampled annually, but provides for sampling of all randomly selected sites over a rotating year period.

2.2.1 Frame Material

A sampling frame is an explicit representation of a population from which a sample can be selected. The sampling frame for wadeable streams and tributary embayments is the USGS 1:100,000-scale map series in digital format (DLGs) and the modification of the DLGs represented by the U.S. EPA River Reach File (RF3), which established edge matching and directionality in the DLG files.

2.2.2 Sample Site Selection

The survey design follows the general design strategy proposed for EMAP (Overton et al.,1990; Messer et al.,1991). The EMAP sampling design (Overton et al., 1990) achieves comprehensive coverage of ecological resources through the use of a grid structure. White et al. (1992) describe the construction of the underlying triangular point grid and its associated tessellation of hexagonal areas.

A two stage sampling approach was used to select the sample units. The same general approach was used to select the Stage I samples of wadeable streams. A 7x7x7 fold enhancement of the random EMAP base grid was placed over the Savannah River Basin (Fig. 2.1). Each grid point was circumscribed by a hexagonal area 1.86 km^2. These 1.86-km^2 hexagons are aggregated into groups of seven, one central hexagon surrounded by six other 1.86 km^2 hexagons. These seven hexagons form a rough, crenulated hexagon, or hexal of about 13 km^2. Seven 13-km^2 hexals comprise one 90-km^2 hexagon and there are seven 90-km^2 in the EMAP base grid hexagon which covers 640 km^2 (Fig.2.1). This results in the 7x7x7 fold enhancement of the Savannah River grid over the original EMAP base grid. There are about forty-three 640 km^2 hexagons (hex) located within the Savannah River Basin.

Stage I sampling selected three 13-km^2 hexals at random within each EMAP 640-km^2 hexagon (Fig. 2.2). The process constituted a probability sample and preserved the spatial distribution of samples throughout the basin. Every stream reach within each of the selected 13-km^2 hexals was identified and designated with a unique code. These streams constituted the elements for the Stage II sample.

Stage I samples streams in direct proportion to their occurrence on the landscape. There are orders of magnitude of more small streams than there are large streams. Different weights were assigned to the streams based on stream order. If these sampling units are not weighted for size, random selection will result in a preponderance of smaller streams in the monitoring program.

The exact weighting procedure is based on the population distribution of the streams. For streams in the Savannah River Basin, a weight of 1.0 was assigned to first order streams (i.e., the smallest streams), a weight of 3.5 was assigned to second order streams and a weight of 6.0 was assigned to third order streams.

The selection process for streams illustrates the randomization and spatial distribution preservation inherent in the EMAP approach: For each stream segment located within each 13 km^2 hexal, the length (km) of the segment and its classification (e.g., first order, etc.) are transposed onto a line that constitutes the total length (km) of streams of all stream orders located within the hexal (Fig. 2.3). The individual stream length segments are then multiplied by an appropriate weight. All first order segments, all second order segments, etc. are added to this line until the line contains all segment lengths for the subject hexal. The total stream length contained within a hexal is the sum of the stream reaches in the hexal (Fig.2.3). The order of the segments on the line is randomized but the location of each uniquely identified segment is preserved. Following this same pattern, hexals within the EMAP 640-km^2 hexes are randomized (Fig. 2.4). The final line represents the total length of all wadeable streams selected in the Stage I sample. Spatial distributions are preserved through the randomization process (all stream segment lengths randomized within a hexal, hexals randomized within an EMAP 640-km^2 hex and the 640-km^2 hexes randomized). Once the sample size has been determined, the total wadeable stream length (weighted) is divided by the required sample number to derive a length interval for sample selection. A random start location on the weighted line is selected and sample sites are systematically drawn using the derived length interval. For example if the weighted line is 200 km long and the sample size is 50 (200/50=4km), then a station is selected every 4 km along the line beginning from the random start point (Fig. 2.4).

In a similar manner large lake embayment stations were selected for sampling. The hexagonal tessellation was randomonly located over the area covered by the embayment population. Within each hexagon, a point was randomonly selected. If the point fell within one of the embayments, then that point became a sample point. The selection process ensured that each location in the embayment population was equally likely to be sampled, and that the set of sites was spatially distributed throughout all embayments (Stevens, 1997).

2.3 Temporal Sampling Rationale

The EMAP has developed an approach that permits fewer sites to be sampled annually, but provides for sampling all sites over a rotating year period. Currently, this rotation period, or interpenetrating cycle, is four years for the wadeable streams and two years for the lake embayment sampling, but it can be two, three, five years etc. This approach preserves the spatial distribution of the samples throughout the Basin and randomonly assigns similar numbers of streams or embayments in each year. This reduces the sampling requirements in any year to a logistically or economically feasible number while still permitting estimates of resource condition. The design is well adapted for detecting persistent, gradual change on dispersed populations or sub populations and for representing patterns in indicators of condition. The period for rotation is based on the desired precision of estimates for any given year. For this demonstration project, precision was set at +/-10% with a 95% confidence Interval (CI).

The large lake embayment study extended over a period of three years. Two independent systematic random samples were selected - one for each year. A total of 111 embayment sample locations was selected such that 52 were allocated in 1995 and 59 in 1996. During the third year, we cycled back to the first set of samples allocated for the embayments. For the three-year period, 126 embayment stations out of 163 (77%) were sampled. Those stations not sampled were non-targets, that is, the location was on land, less than one meter deep, or inaccessible.

Sixty sites per year for a total of 240 sites over a four-year period were selected for stream sampling. Only 119 sites were sampled because of access denial, some were intermittent streams, some were ponds or embayments, some were on dry land, some were in wetlands, and a few did not meet our criteria of <½ hour to walk to the site.

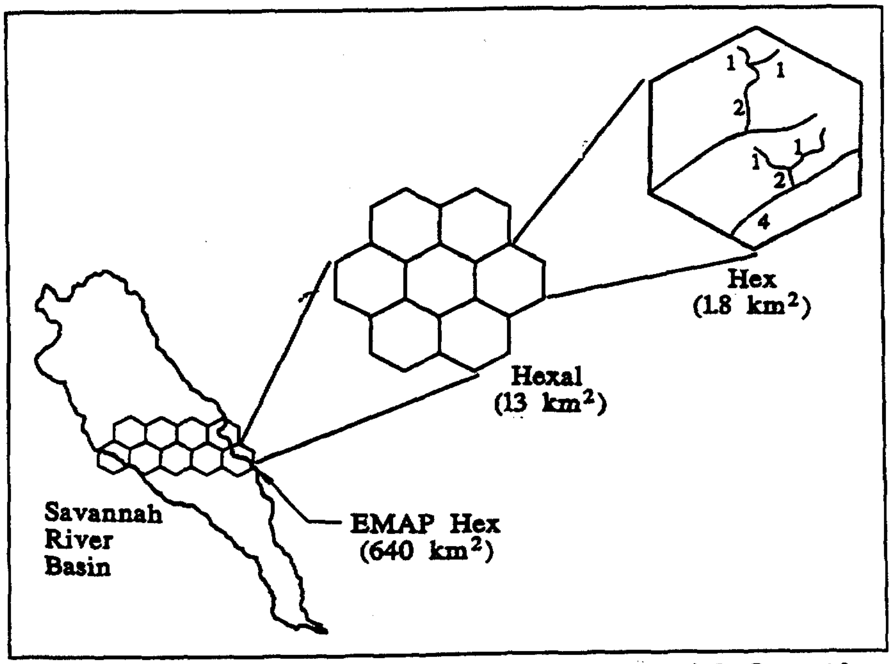


Figure 2.1. Illustration of Base Grid for the Savannah River Basin

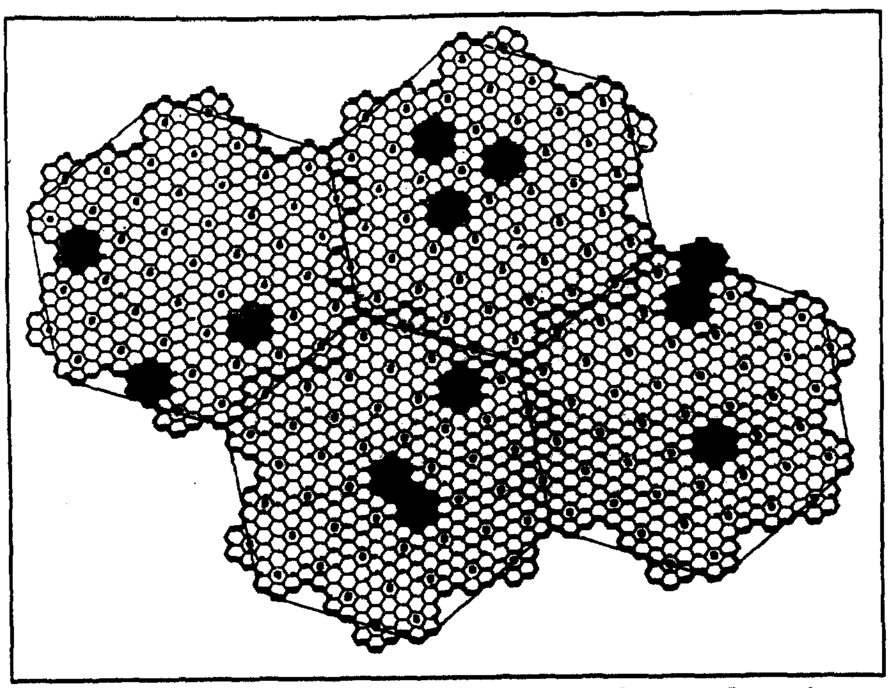


Figure 2.2. Illustration of Random Selection of Hexals from 640-km² Hexagon

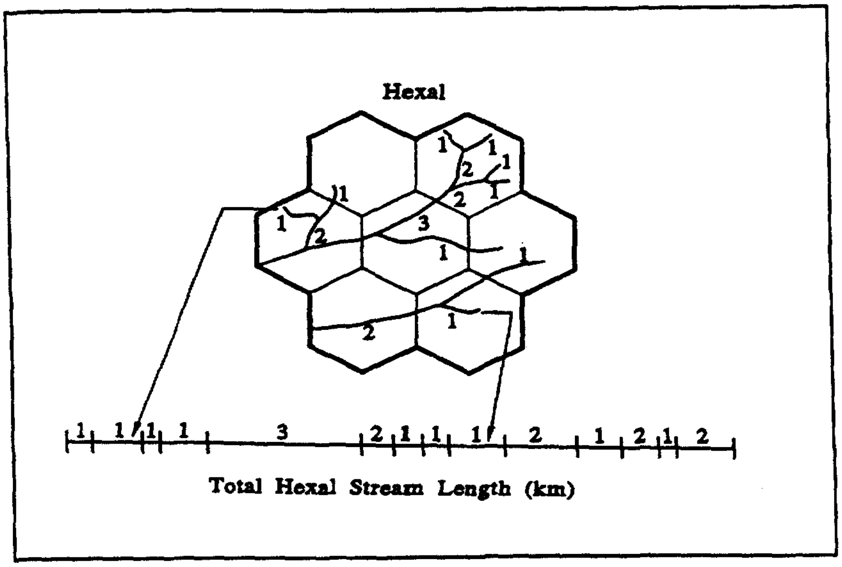


Figure 2.3. Weighted Hexal Stream Length

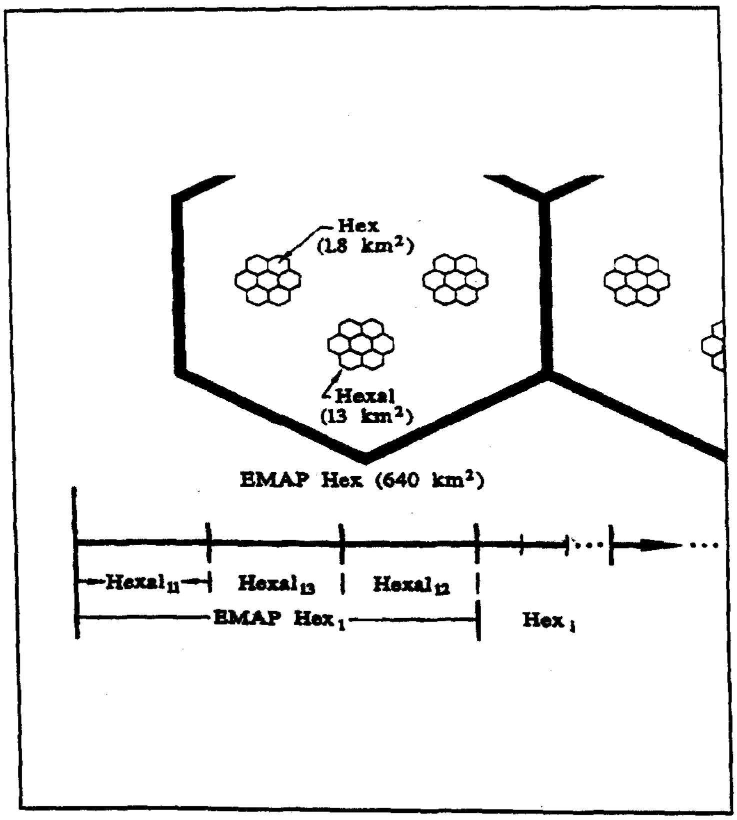


Figure 2.4. Total Weighted Stream Length Selected in Stage I Sample

3.0 INDICATORS

REMAP monitors ecological indicators to assess status, trends, and changes in the condition and extent of the Region's ecological resources (Bromberg, 1990, Hunsaker and Carpenter, 1990; Hunsaker et al., 1990). Indicators are defined as any characteristic of the environment that estimates the condition of ecological resources, magnitude of stress, exposure of a biological component to stress, or the amount of change in condition.

Ecological principles state that ecosystem responses and condition are determined by the interaction of all the physical, chemical, and biological components in the system. Because it is impossible to measure all these components, REMAP's strategy emphasizes indicators of ecological structure, composition, and function that represent the condition of ecological resources relative to societal values. The challenge is to determine which ecological indicators to monitor. One approach for selecting these indicators starts with those attributes valued by society and determines which indicators might be associated with these values.

3.1 Societal Values

To be effective, information from the monitoring program must prompt action when required. This means the information produced must be related to perceptions of aquatic health and represent issues and values of concern and importance to the public, aquatic scientists and decision makers. The selection of these societal values drives the selection of appropriate indicators. After extensive discussions with resource managers, decision makers and the scientific community by members of the EMAP - Surface Waters Resource Group (Larsen and Christie 1993), an initial set of societal values and concerns were identified for evaluation in EMAP. These values are:

- **Biological Integrity,**
- **Trophic Condition, and**
- **Fishability.**

Biological integrity can be defined as the ability to support and maintain a balanced, integrated, adaptive community with a biological diversity, composition, and functional organization comparable to those of natural lakes and streams of the region (Frey 1977; Karr and Dudley 1981) and includes various levels of biological, taxonomic and ecological organization (Noss 1990). Biological integrity incorporates the idea that all is well in the community. That is, the different groups are stable and working well with little if any external management of the community, whether it is a township, coral reef, or stream. Waters in which composition, structure and function have not been adversely impaired by human activities have biological integrity (Karr et al. 1986). Karr and others (1986) also defined a system as healthy "when its inherent potential is realized...and minimal external support for management is needed." This value or ethic differs considerably from values oriented toward human use or pollution that are traditionally assessed in water quality and fisheries programs, in which production of a particular species of game fish is the goal (e.g., Doudoroff and Warren, 1957), and may conflict with these definitions (Callicott 1991; Hughes and Noss, 1992; Pister, 1987).

Fishability is defined as the catchability and edibility of fish and shellfish by humans and wildlife (Larsen and Christie 1993). Fish represent a major human use of an aquatic ecosystem product. Protecting fish is the goal of many water quality agencies, and fish drive their water quality standards.

Trophic condition has been defined in EMAP as the abundance of production of algae and macrophytes (Larsen and Christie 1993). Trophic condition involves both aesthetic (water clarity) and fundamental ecological (production of plant biomass) components. It is a key aspect in determining both a lake's relative desirability to the public, its production of fish and its ecological character or classification by limnologists (e.g.,eutrophic or oligotrophic). Because of limited resources, a decision was made to concentrate on trophic condition indicators for lakes over a three-year period; and for streams, we emphasized integrity all four years and trophic condition (algal growth potential) only for two years.

3.2 Types and Selection of Indicators

EMAP defines two general types of ecological indicators, condition and stressor indicators. A condition indicator is any characteristic of the environment that estimates the condition of ecological resources and is conceptually tied to a value. There are two types of condition indicators: biotic and abiotic. Condition indicators relate to EMAP's first and second objectives: estimating the status, trends, and changes in ecological condition; and the extent of ecological resources.

Stressor indicators are characteristics of the environment that are suspected to elicit a change in the condition of an ecological resource, and they include both natural and human-induced stressors. Selected stressor indicators are monitored in

EMAP only when a relationship between specific condition and stressor indicators are known, or a testable hypotheses can be formulated. Monitoring selected stressor and condition indicators addresses the third EMAP objective of seeking associations between selected indicators of stress and ecological condition. These associations can provide insight and lead to the formulation of hypotheses regarding factors that might be contributing to the observed condition. These associations can provide direction for other regulatory, management, or research programs in establishing relationships.

3.2.1 Streams

In concert with the EMAP approach, the Savannah REMAP Project considered a suite of indicators to evaluate the condition of ecological resources of streams in the Savannah River basin. Selection of specific ecological indicators was based on societal values. Upon consideration of the type of streams (wadeable) to be investigated, a set of societal values were first identified. They were **biological integrity** and **trophic condition**. After identification of the values, four indicators were selected to assess biological integrity and trophic condition - benthic macroinvertebrates, fish, habitat, and algal growth potential (AGP).

Benthic macroinvertebrate insects represent the first consumer level in streams. They are important as processors of organic matter, like leaves and sewage, that find their way into a stream. By fragmenting or breaking down this organic matter, stream insects prepare it for decomposition by bacteria that attach too or colonize the organic matter. In turn, bacteria may serve as a food source for other stream insects that seek out and graze on the organic matter. Because of their limited mobility and relatively long life span, stream insects provide a "window" of cumulative impacts on ecological or resource condition. This community is sensitive to changes; they have for many years been used as a reliable barometer of water quality conditions. Some groups of insects are very sensitive to stresses, like man-made pollution, while others are tolerant. By focusing on the presence or absence of different groups of insects, an aquatic biologist is provided insight about the ecological health of a stream. Sometimes pollution effects may stem from discharges of chemicals, pesticides, or nutrients that are of a manmade origin. Often, sediments from erosion and attributable to land clearing or silviculture practices may adversely affect the stream habitat. The materials that constitute a stream bottom are very important to both fish and stream insects. For example, very fine sediments, like silt, clay, or very fine sand, are detrimental to the reproduction of fish and eliminate preferable habitat for stream insects (Plafkin et al., 1989; Barbour et al., 1998). Silt, especially, can interfere with a fish's or stream insect's ability to breathe. Assessment of the insect community was accomplished by using a standard field survey technique known as Rapid Bioassessment Protocol II(RBPII)(Plafkin et al, 1989; Barbour et al., 1998). With the RBP II protocol, most sites can be surveyed with relatively limited time and effort in the field and laboratory. Although RBP II is not the most intense level of bioassessment (RBP III is the most intense effort), it serves well the goal of the Savannah REMAP Project of characterizing the ecological health of streams in a large geographic area with a minimum of laboratory time and support coupled with efficient turn-around of study results. This is accomplished because most benthic macroinvertebrates can be identified in the field to the family level. RBP II provides a basis for ranking and prioritizing impaired sites for further study.

The biological metric of choice utilized for benthic macroinvertebrates was the family level EPT Index (Barbour et al., 1998). The EPT Index, as reported in the scientific literature (Barbour et al., 1992; Wallace 1996), is a useful and widely accepted biological metric for analysis of benthic macroinvertebrate data. The EPT Index is an approved biometric put forth in guidance documents used by state and federal resource agencies because of its ability to detect impairment and its defensibility in legal proceedings. The EPT Index is simply a summation of the total number of families at a sampling site in the generally pollution-sensitive orders of benthic macroinvertebrates. These orders are the mayflies (Ephemeroptera), stoneflies (Plecoptera), and caddisflies (Trichoptera). The EPT Index is a richness measure which is expected to decrease in response to increasing perturbation.

Habitat is important when examining the ecological condition of sites. These evaluations focus on variables like substrate (bottom sediments) characteristics, flow regimes, impacts to the stream channel (e.g., channelization, deposition), impacts to stream side vegetation, stability of the stream banks, and available cover. Ecoregion reference sites provide a basis for the best attainable conditions for all streams with similar physical dimensions for a given ecoregion. Presently, there are two reference sites per ecoregion except for the Coastal Plain ecoregion. The process of reference site identification is still ongoing in Georgia and South Carolina.

Fish were chosen primarily for their societal value and role as a top consumer in streams. Fish are relatively easy to identify and with minimal training most fish can be collected, sorted, and identified at the field site and then released unharmed. Fishes represent a variety of feeding types. Their diet can consist of food derived from both inside the stream and outside the stream. One important food source is stream insects. Changes in the stream insect community often result in a change in the fish community. Like stream insect communities, fish communities will respond to environmental change, whether it is biological, chemical or physical. Some fishes are very sensitive to environmental change while others are not. By examining all fish groups that live in a stream, the general condition of a stream can be assessed. For example, if there are only one or two groups of fish in a stream who are very tolerant to pollution, and there are no groups that are sensitive to a pollutant, then impairment is suspected because of environmental change that has eliminated the sensitive groups.

The Environmental Protection Agency's Rapid Bioassessment Protocol V (RBP V) (Barbour et al, 1998) is an index used to assess stream condition based on the fish community. The EPA RBP V (Barbour et. al., 1998)is based primarily on the Index of Biotic Integrity (IBI) (Karr,1981; Fausch et al. 1984;Karr et al. 1986). The index consists of up to twelve measures scored to assess changes in the fish community compared to a reference stream, or a stream with least impact. For example, one of the measures assesses the proportion of fishes in a stream considered to be tolerant to environmental change. If the proportion of tolerant groups are high compared to the reference stream, then this would result in a lower score for that measure. Another measure looks at the number of fish groups. If the number of fish groups collected is similar to that of the reference stream then this would result in a high score. After all twelve measures have been given a score, the scores are totaled and the condition of the fish community is then characterized as either good, fair, or poor depending on how far the total score deviates from that of a reference stream.

The primary indicator selected to address trophic condition in streams for the first two years was the algal growth potential test (AGPT) (APHA, 1995). The AGPT is based on the premise that maximum yield of plants (e.g. algae) is limited by the amount of nutrients available to the test alga. With higher algal growth concentrations (AGPT), there is good likelihood that obnoxious plant growths can occur in a stream. The test was selected as the indicator of choice to assess trophic condition primarily because of its specific sensitivity, reliability and the ease and economy of using it as a monitoring tool.

3.2.2 Large Lake Embayments

We focused on condition indicators related to trophic condition because of limited resources. The original study plan (FTN et al., 1994) proposed sampling for fishability indicators, Fish Health Index and Fish Tissue Residues; biological integrity, phytoplankton and zooplankton identification and counts; and one other trophic condition indicator, zeaxanthin, a marker pigment for blue-green algae. Work is continuing on this pigment, but the information was not sufficient for inclusion into this report.

The trophic condition indicators measured during this study were chlorophyll a, total phosphorus (TP), algal growth potential (AGP), Secchi disc transparency, and total suspended solids (TSS). These indicators were selected because they provide different insights into the condition of the embayment waters.

Chlorophyll a is commonly used to estimate the degree of phytoplankton bloom conditions that can affect aesthetics, fishing and swimming quality, taste and odor of fishes and drinking water, and the health of fish, waterfowl, and livestock. Chlorophyll is a measure of instantaneous standing crop, whereas TP and AGP indicate potential for blooms. Total phosphorus reveals insights about nutrient input and the potential for serious bloom conditions if we assume all of it is available. However, much of the TP is not normally available. The AGP can show how much of the TP is available for algal growth and the potential, under optimum conditions, for blooms. Secchi disc transparency is related to swimming conditions. Total suspended solids is related to transparency, but it also can be used to indicate effects upon fish production.

4.0 METHODS

4.1 Streams

4.1.1 Field Sampling

Benthic macroinvertebrate sampling and habitat evaluation followed basic guidelines put forth in the EPA document *"Rapid Bioassessment Protocols for Use in Streams and Rivers"* (Barbour et al., 1998). Multiple habitats (riffles, undercut banks, leaf packs, woody debris, and pools) were sampled with D-frame and A-frame biological dipnets according to the Ecological Assessment Branch's (EAB) Standard Operating Procedures (SOP) (EPA, 1998). In addition to the benthic macroinvertebrate sampling or biosurvey, the RBP II also includes in-situ water quality measurements (dissolved oxygen, pH, temperature, and conductivity). These parameters were measured with a multiparameter in-situ water quality device (HYDROLAB SCOUT) prior to the habitat evaluation phase according to EAB's SOP.

Stream fish sampling followed basic guidelines set forth in Barbour et al. (1998). A Smith-Root Type VII backpack electrofishing unit was used to collect stream fish. A single pass electrofishing run moving from downstream to upstream, thoroughly sampling each habitat type (pools, runs, riffles, eddies, undercut banks, etc.) was conducted at each stream sampling location. Equal effort was given at each location. Fish were identified at stream side and released. A few individuals of each species were preserved in 10% formalin and transported back to the lab for identification verification.

Based on the guidance provided in the EPA RBP V(Barbour et. al. 1998) document, nine metrics were utilized to evaluate the data to assess the condition of stream fish assemblages. The metrics were selected from a pool of metrics listed in the EPA RBP document and other studies that have been conducted in Georgia (DeVivo 1996). A list of metrics utilized and the scoring criteria for each are presented in Appendix C.

Habitat assessment was based on a matrix of nine parameters (EPA, 1989). These nine parameters fall into three principal categories: primary, secondary, and tertiary parameters. Primary parameters (bottom substrate, available cover, embeddedness, and flow regime) characterize the stream "microscale" habitat and are most influential to community structure. Secondary parameters (channel alteration, bottom scouring/deposition, and sinuosity) measure the "macroscale" habitat such as channel morphology. Tertiary parameters (bank stability, bank vegetation, and stream side cover) evaluate the integrity and composition of the riparian zone.

4.1.2 Analytical Methods

RBP II and V do not require analytical methods because the organism identifications usually are made in the field. When organisms need to be returned to the laboratory for identification, they are sorted by specialists and identified by an expert following protocols spelled out in the EAB's SOP (1998). Algal growth potential tests conducted the first two years followed the protocols of standard methods (APHA, 1995) as modified by Schultz (1994) (EPA, 1998).

4.2 Large Lake Embayments

4.2.1 Field Sampling

Standard operating procedures (SOP) of EAB were followed as the principle means of sample collection and measurement (EPA, 1998). All lake sampling and measurements took place the weeks of 7/17 to 7/21, 1995, 6/21 through 7/5, 1996, and 7/7 through 7/10, 1997. One hundred and twenty-four stations were sampled over the three-year period. This annual sampling window was selected because it is a time of maximum recreational use, and maximum water supply use.

Secchi disc transparency was measured according to EAB's SOP that was adopted from EPA methodology (Klessig, 1988) using a 30 cm black and white disc lowered on the shady side of the boat. Photic zone was determined by multiplying the Secchi measurement by a factor of 2.1 (Raschke, 1993).

Collection of water samples consisted of using a battery operated pump to fill a 5 gallon carboy with a composite depth integrated sample taken from the photic zone (1% light level). The water sample was mixed thoroughly and then the various individual sample containers were filled, labeled and stored on ice. Samples were collected for total phosphorus (TP), total suspended solids (TSS), algal growth potential tests (AGPT), and chlorophyll a. Field duplicates were collected at a minimum of once in every ten samples. For the field duplicate, the carboy was emptied, rinsed, and a second sample collected.

Chlorophyll a sampling followed basic guidelines set forth in Standard Methods, 19th Edition, section 10200. A 100 to 250 ml sample was filtered through a 24 mm diameter Whatman GF/F glass fiber filter. The filters were folded, blotted dry, enclosed in aluminum foil, labeled and stored in a cooler containing dry ice and returned to SESD for analyses. Samples were filtered in triplicate.

4.2.2 Analytical Methods

Total phosphorus and total suspended solids were analyzed using methods given in the EPA document " Methods for Chemical Analysis of Water and Wastes" (EPA, 1983). In 1995, Cycle 1, total phosphorus was analyzed using EPA Method 365.1. Results of most analyses were below the minimum detection level of 20 ug/L for this method. In 1996 and 1997, Cycles 2 and 3, a low detection level method was used (EPA, 1992a) that allowed for detection of phosphorus at 3 ug/L. Total suspended solids were determined by using EPA Method 160.2.

Chlorophyll samples were measured by high performance liquid chromatography (HPLC) following the basic guidelines given in Standard Methods and in EPA Method 447.0. The chlorophyll was extracted in a 90% acetone solution.

Algal growth potential test (AGPT) maximum standing crop (MSC) and limiting nutrient was determined using The Selenastrum Capricornutum Printz Algal Assay Bottle Test (Miller et al.,1978) as modified by Schultz et al.(1994).

4.3 Quality Assurance/Quality Control

Standard operating procedures of the Ecological Assessment Branch and the Analytical Support Branch of EPA's Region 4 SESD were followed as the principal means of monitoring appropriate quality assurance/quality control (QA/QC). Quality control checks were included in sample collection, physical measurements performed in the field, chemical analyses, and data gathering and processing. Data were subject to verification and validation. Verification included range checks and internal consistency checks. Validation consisted of a review of the data from a data user's perspective for consistency based on known numerical relationships.

4.3.1 Lakes

Secchi disk transparency was measured at each site to determine the photic zone for lake sampling. Prior repetitive test measurements of Secchi depth in a variety of water bodies showed that the coefficient of variation (CV) ranged from 5 to 15% among several investigators.

Water samples were collected as depth integrated samples throughout the photic zone. Samples were collected for total phosphorus (TP), total suspended solids (TSS), chlorophyll a, and algal growth potential tests (AGPT). Field duplicates were collected at a minimum of once in every ten samples. Results of precision as coefficient of variation (CV) are given in Appendix

A. In 1997, field blanks were collected along with the duplicates. In this case, each of the sample containers was filled with deionized water, preserved or filtered as appropriate, and returned to the laboratory for analyses. Results are given in Appendix A.

In 1995, (Cycle 1), TP in most of the samples was below the minimum detection level of 20 ug/L for the method used. In 1996 and 1997 (Cycles 2 and 3), a low level phosphorus method was used (EPA, 1992a). The CV for the field duplicates ranged from 0 to 71.2% with an average CV of 20.9% (Appendix A).

All of the field TSS duplicates in cycles 1 and 2 were below the laboratory's detection limit of 4.0 mg/L. For Cycle 3, ASB modified their procedure by filtering a greater volume of sample (APHA, 1995). This modification reduced the detection limit to 1.0 mg/L. The CV ranged from 0 to 23.6% with an average CV of 18.6%. Standard Methods gives the CV as 33% at a concentration of 15 mg/L TSS. Both laboratory and field precision were well within the values of Standard Methods (APHA, 1995).

Chlorophyll a and AGPT were measured to determine the trophic status of the lakes. For chlorophyll a the CV for field duplicates ranged from zero to 53.8% with an average CV of 16%. The standard method (APHA,1995) does not give any precision data for field duplicates that include a filtration step. The method does state that for multiple injections on the HPLC, the average CV for seven pigments is 10 percent.

The precision of the field duplicates for AGPT ranged from 1.3 to 53.1% with the average CV equal to 15.7%. The test gave an average CV of 26.4% for the 1.0 to 2.0 Maximum Standing Crop (MSC) level (Miller et al., 1978) which was typical for the Savannah lake samples.

4.3.2 Streams

Field measurements at each sampling station included temperature, DO, pH, and conductivity. Measurements were taken using a Hydrolab Scout. The Hydrolab was calibrated each morning and then again at the end of each day according to EAB's SOP (EPA, 1998).

Biological integrity was accomplished in part by using a standard field survey technique known as Rapid Bioassessment Protocols II (RBPII) (Barbour et al., 1998) to assess the benthic macroinvertebrate community. This is a screening procedure in which the macroinvertebrates are identified in the field to the family level. If identification is uncertain, the specimen is brought back to the laboratory for verification. No replication of sites were performed as this is a screening method.

The Rapid Bioassessment V Protocol (RBP V) (Barbour et al., 1998) was the index used to assess stream condition based on the fish community. To insure fish were properly identified during the study, all fish that were captured during the first year were preserved and sent to Dr. Byron Freeman at the Institute of Ecology at the University of Georgia for identification. In subsequent years, voucher specimens of each species collected in the field were preserved for identification verification at the US EPA SESD laboratory. At the end of the four year study, preserved fish with questionable identifications, were sent to the Institute of Ecology for verification.

The primary indicator selected to address trophic conditions in streams is the algal growth potential test. This test was also used in the lake work and the QA/QC used is the same as given in Section 4.3.1 except that limiting nutrient was not determined for the streams.

5.0 Findings

5.1 Basin Perspective

5.1.1 Large Lake Embayments

The distribution of data for each variable can be characterized by its cumulative distribution frequency (cdf). These curves show the percent of embayment acreage in the basin equal to or less than some specified measurement plus or minus a confidence level. For the purpose of this study, we have set a confidence level of 95%. This means that we are 95% sure that the acreage estimated to be equal to or less than a given measurement is within the bounds of our confidence lines on the graph (Fig. 5.1). There is a 1 in 20 chance (5% error) that the true or real percent of acreage affected at a particular measurement is not within the confidence bounds.

Chlorophyll a ranged from a low of 0.84 at Lake Hartwell to 11.56 ug/L at the most downstream lake, Lake Thurmond (Table 5.1).

Table 5.1. Range of Values for the Savannah River Lakes

Lakes	CHL. A ug/L	AGPT mg/L	Limit NUT.	TP ug/L	SD Meters	TSS mg/L
Thurmond	0.98-11.56	0.66-11.0	N+P	3-50	1.2-4.8	0.7-27
Russell	1.10-5.47	0.39-2.01	N+P	3-60	0.7-3.4	2-32
Hartwell	0.84-6.84	0.33-2.27	N+P	3-30	1.4-10	1.0-6
Keowee	0.91-2.03	0.49-5.08	N+P	3-11	2.4-5.5	0.7-5.5
Jocassee	1.35-2.59	0.66-1.95	N+P	3-10	3.3-6.0	1.2-34
Burton	1.60	1.62	N	6	2.2	2

This range of concentrations at the times of sampling exhibit trophic conditions related to classical lake classifications of oligotrophic to mesotrophic (Olem and Flock, 1990). Chlorophyll a was less than 12 ug/L over the entire basin's large lakes (Figure 5.1). Based on experience (Raschke, 1994) over the past 30 years, generally, when chlorophyll a ranges from 0 to 10 ug/L, there is no discoloration of the water and no problems. At a range of 10 to 15 ug/L, waters can become discolored and algal scums could develop. Between 20 to 30 ug/L, the water is deeply discolored, scums are more frequent, and matting of algae can occur (Raschke, 1993). EPA Region 4 (Raschke, 1993) has shown that a mean photic zone growing season average of equal to or less than 15 ug/L of chlorophyll a should satisfactorily meet multiple uses, including drinking water supply.

One of the objectives of the Savannah River REMAP is to detect trends in important environmental variables over both time and space. One means of comparison is through the testing of the null hypothesis that the population's distributions from two or more annual cycles are identically distributed. This can be accomplished through use of the Cramer von Mises test statistic

Table 5.2. Cramer-von Mises Tests for Equality of Cumulative Distribution Functions for the Savannah River Basin Embayments. Equality of Cumulative Distribution Functions Between Cycles (Years) is Tested.

Variable	W
Chlorophyll a	1.70*
Agpt	8.60*
Total Phosphorus	3.16*
Secchi Disc	0.44
Total Suspended Solids	2.84*

*Significant at alpha=.05

(W) which is founded on design-based methods of statistical inference (Appendix E). For design-based statistical inference, the source of random variation is the random selection of sample sites. This is in contrast to model-based statistical inference, where the source of random variation is in the assessed deviations from the statistical model (e.g., a regression model). Thus, designed-based statistical inference has the advantage that no model assumptions are required. The distribution of a population can be characterized through its cumulative distribution function (cdf). This is equivalent to testing the null hypothesis that the cdf's are identical. A test of cdf differences at alpha .05 (Table 5.2) using the Cramer-von Mises test statistic (W) showed that four variables, chlorophyll a, AGPT, total phosphorus (TP), and total suspended solids (TSS) had significantly different distributions from one cycle to the other. Chlorophyll Cycles 2 and 3 are intertwined and slightly different from Cycle 1 (W=1.70, k=3). The curve for Cycle 1 rises more gradually than that of Cycles 2 and 3 (Figure 5.2) culminating in a high of 11.56 ug/L thus suggesting the mean is higher for Cycle 1.

Chlorophyll a represents phytoplankton standing crop or yield at given time periods, whereas AGPT is representative of the potential phytoplankton production, given optimum conditions of sufficient nutrients, light, time and temperature. Algal growth potential ranged from 0.33 mg dry weight (DW)/L at Lake Hartwell to 11.0 mg DW/L at Lake Thurmond (Table 5.1)(Figure 5.3). Approximately 99.7% of the AGPT dry weights were equal to or less than 5 mg/L (Fig. 5.3), an in-lake action level that will reasonably assure protection from nuisance algal blooms and fish kills in southeastern lakes (Raschke and Schultz, 1987). The 5 mg/L of dry weight translates to a potential chlorophyll standing crop of approximately 57 ug/L of chlorophyll a based on the following equation:

$\mathbf{Log_{10}\ chl\ a = 1.15\ Log_{10}(DW) + 0.95}$ (Raschke and Schultz,1987).

The sampled maximum chlorophyll a of 12 ug/L is much lower than the 57 ug/L of chlorophyll a derived from the 5mg DW/L AGPT action level suggesting that the present phytoplankton biomass does not pose a threat to the integrity of the lake system. Figure 5.4 depicts the AGPT cdf's for cycles one through three. The curve for Cycle 2 rises more gradually than that for cycles one and three suggesting the mean AGPT is not only higher in Cycle 2, but also shows greater variability within this cycle. The Cramer-von Mises test statistic confirms that the difference between the three cycles at the alpha .05 level is statistically significant (W=8.60; k=3).

Total phosphorus (TP), another indicator like AGPT of potential production, ranged from 3.0 ug/L in most lakes to 60 ug/L in Lake Russell (Table 5.1). Approximately 87.0% of the embayment acreage was equal to or less than 10 ug/L TP (Figure 5.5). If all of the phosphorus were available for algal growth, at high values of 40 to 60 ug/L one could expect severe bloom conditions, but this was not the case as seen by the relatively low chlorophyll a values. This is not surprising; besides needing optimum conditions for maximum growth, the phytoplankton need sufficient nutrients that are bioavailable to them. Generally, not all of the TP in lakes is available for phytoplankton growth. Peters (1981) estimated that bioavailable phosphorus (BP) is 83% of TP in natural lakes and 18 to 57% in rivers. Since our lakes are reservoirs and thus an extension of a river system one would expect bioavailability to be much less than that found in natural lakes. Previous work on the 18 Mile Creek embayment of Lake Hartwell showed that the average percent of BP to TP was 38% (Raschke et al., 1985). Sometimes the BP portion of TP can be as low at 3% (Raschke and Schultz, 1987). At the alpha .05 level there was a significant difference

(W=3.16; k=3) between Cycle 1 and the other two cycles, but higher values were observed in Cycle 1 (Figure 5.6). The significant differences between cycles for chlorophyll, AGPT, and TP suggests that other variables are influencing differences from one cycle to the other. We are not in a position with three years of data to focus on particular stress indicators at this time. Samples were collected from two to three weeks after rainfall events in the basin. Thus rainfall or unusually high stream flows would not seemingly cause the differences observed between cycles with respect to these three phytoplankton growth related indicators. Presumably the cyclic differences were caused by internal lake influences like internal nutrient cycling. Even these differences may be within the normal suite of variability experienced in a natural setting.

For water supply, a mean growing season average Secchi disc (SD) transparency of equal to or greater than 1.5 meters is desirable (Raschke, 1993). For non-water supply embayment situations a mean SD of greater than 1 meter is acceptable for fishing and swimming (Raschke, 1993). Secchi disc transparency ranged from 0.7 meters at Lake Russell to a high of 10 meters at Lake Hartwell (Table 5.1). An examination of Figure 5.7 shows that in about 2.6% of the embayment acreage, less than desirable conditions exist for recreational purposes, and only 5.3% of the acreage was less than the water supply criterion of equal to or greater than 1.5 meters. Where SD was less than one meter, measurements were located near shore or at the upper end of the tributary embayments.

The National Academy of Sciences (1973) has set TSS levels for different levels of stream protection. High protection can be maintained if the TSS is 25 mg/L or less, moderate protection is possible if the range is between 25 to 80 mg/L, low protection is from 80 to 400 mg/L, and there is very little protection from TSS at concentrations greater than 400 mg/L TSS. According to these criteria, our embayment population is highly protected in more than 95% of the embayment acreage and moderately protected in the remaining acreage (Fig. 5.8). Buck (1956) divided impoundments into 3 categories: clear with total suspended solids (TSS) less than 25 mg/L; intermediate with TSS 25-100 mg/L; and muddy with TSS greater than 100 mg/L. The mean harvest of game fish was 162 lbs/acre for clear lakes, 94 lbs/acre in intermediate lakes, and muddy lakes only yielded 30 lbs/acre. The TSS ranged from a low of 0.7 mg/L at Lakes Keowee and Thurmond to a high of 34 mg/L at Lake Jocassee, the uppermost lake in the Savannah Chain of lakes (Table 5.1). Again these high values were attributed to near shore stations receiving wind fetch at the time of sampling. Ninety-seven percent of the embayment acreage would fall into Buck's clean category, with only 3% being intermediate with respect to water clarity (Fig.

5.8). There were significant differences between the cycles ($W=2.84$, $k=3$)(Figure 5.9). Presumably, cycle three was significantly different from the other two cycles, because there were no significant differences at alpha .05 between cycles one and two ($W=0.15$; $k=2$).

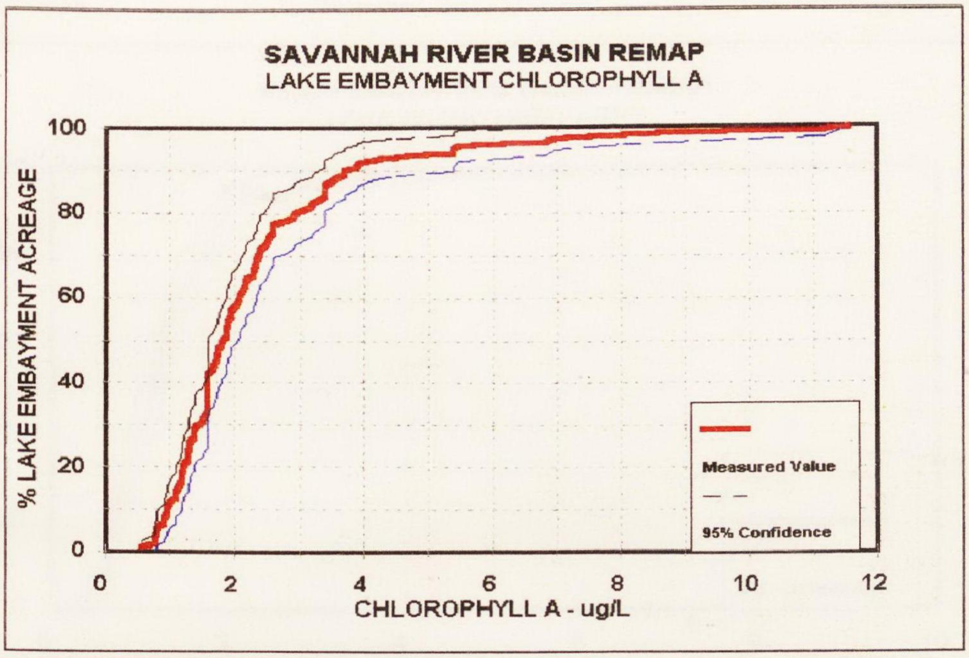


Figure 5.1. Cdf for Chlorophyll a.

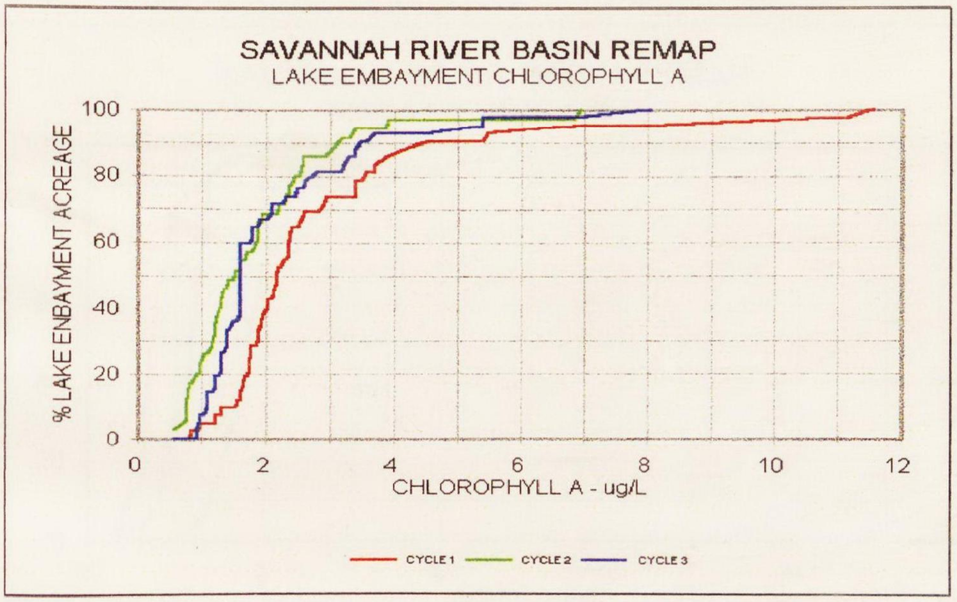


Figure 5.2. Cdf Curve Showing Differences Between Cycles.

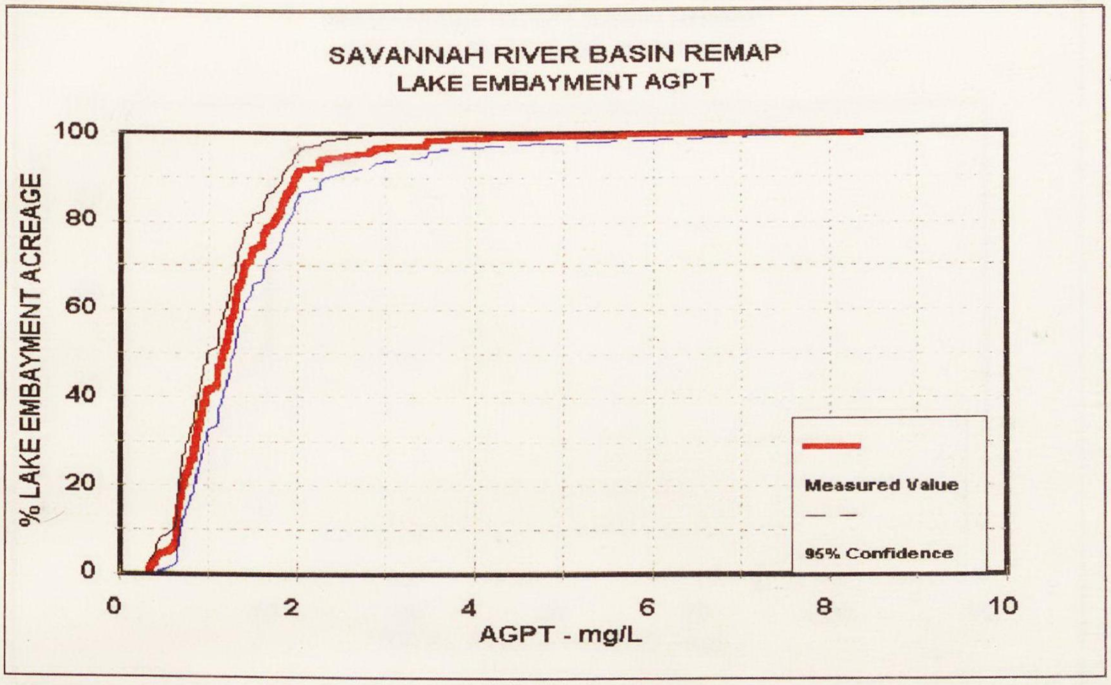


Figure 5.3. Cdf Curve for AGPT.

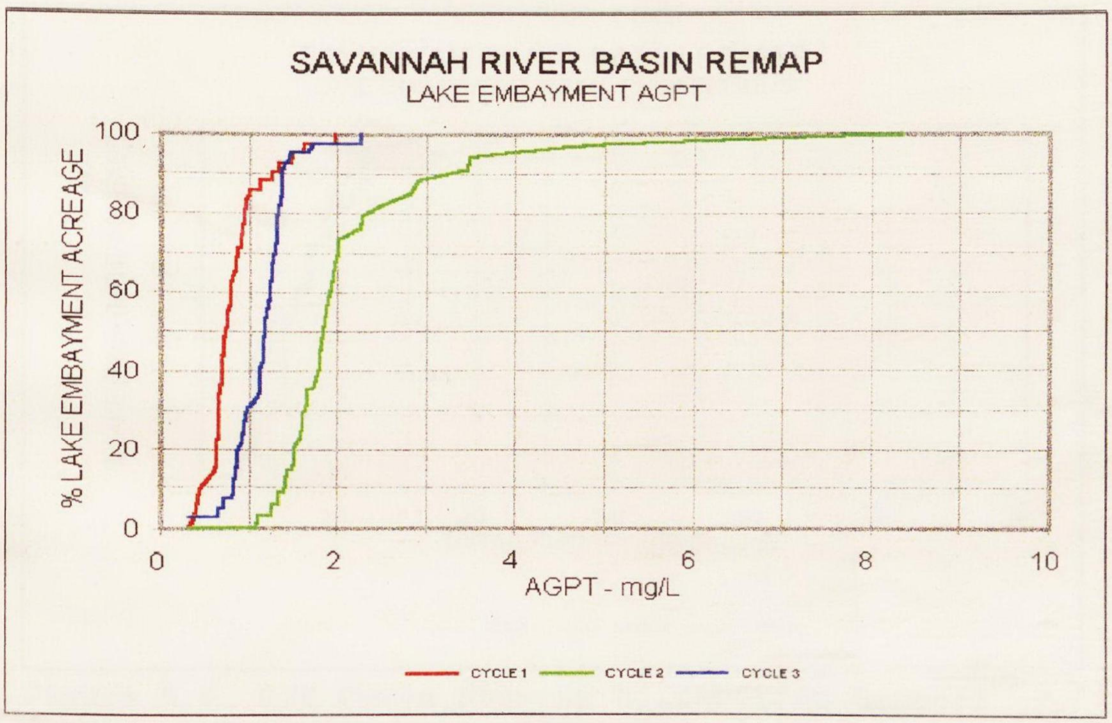


Figure 5.4. Cdf Curve Showing Differences Between Cycles.

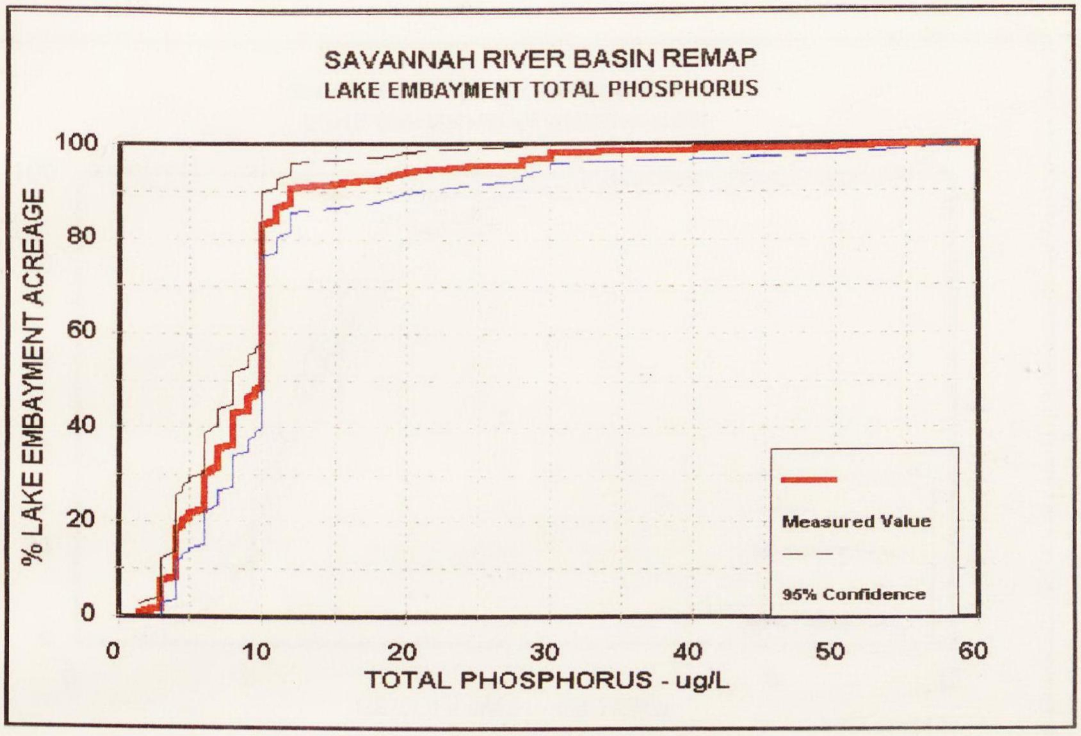


Figure 5.5. Cdf Curve for TP.

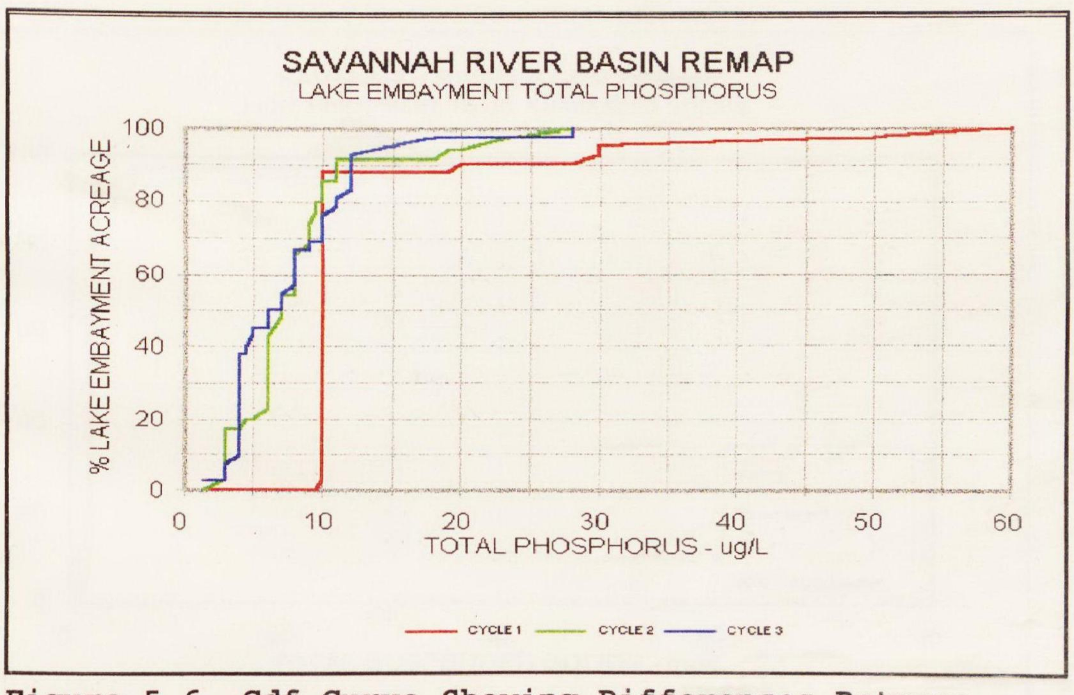


Figure 5.6. Cdf Curve Showing Differences Between Cycles.

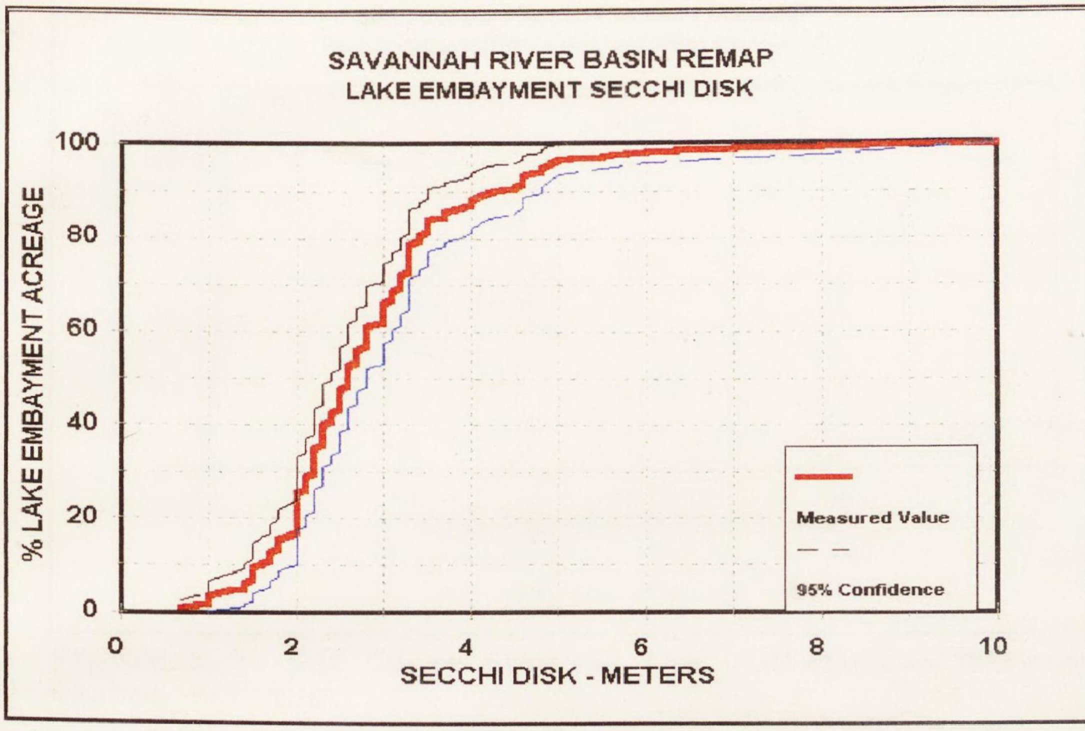


Figure 5.7. Cdf Curve for Transparency.

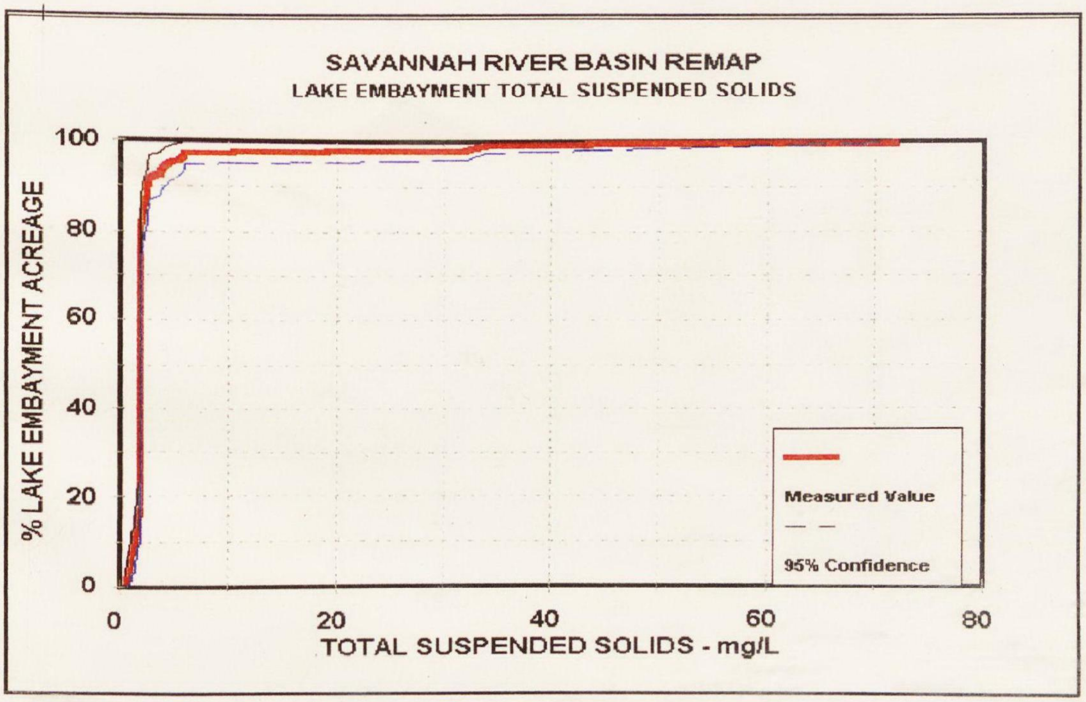


Figure 5.8. Cdf Curve for TSS.

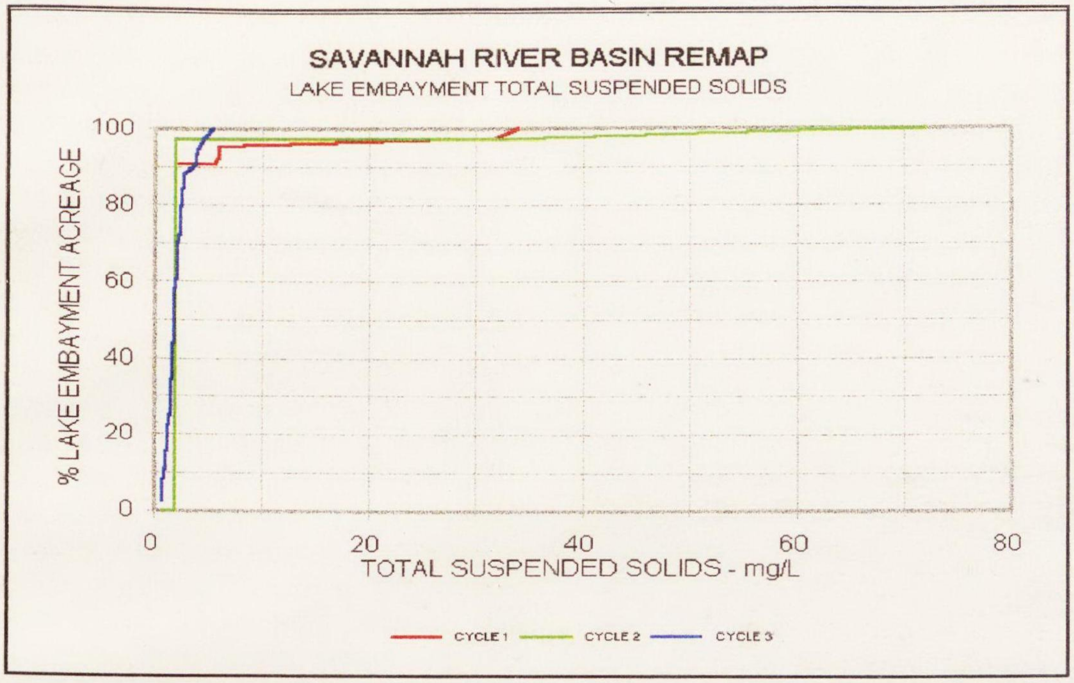


Figure 5.9. Cdf Curve Showing the Differences Between Cycles for TSS.

5.1.2 Streams

The report by Raschke et al. (1996) (Appendix H) demonstrated the applicability of the EMAP approach to stream monitoring in basins. The information in this section is a summary of four years of stream data. It is not an exhaustive analysis of basin response. Rather, we devoted our energies to demonstrating the applicability of the EMAP approach to an ecoregion and the application of modified indicators and a new index that incorporates macroinvertebrate and fish metrics (Section 5.2).

The family level EPT Index ranged from 1 - 20 across all six ecoregions (Appendix C). EPT Index scores exhibited a general decline southward along successive ecoregion belts (Fig. 5.10). However it should be pointed out that the small sample sizes within each ecoregion, with the exception of the Lower Piedmont, is inadequate to confirm this observation. The Blue Ridge Mountain Ecoregion had the highest EPT Index scores (range = 8 - 20; n = 11). Mean EPT Index value in the Blue Ridge was 15. Only 3 sampling stations were in the Upper Piedmont; EPT Index scores for the 3 Upper Piedmont stations were 9, 16, and 16. EPT Index scores in the Lower Piedmont (range = 1 - 18; n = 88) were lower than the Blue Ridge and Upper Piedmont and the mean EPT Index value of 7 was much lower than that of the Blue Ridge (15). Five stations were located in the Sand Hills where the EPT Index ranged from 3 to 11. Ten stations were located in the Southeastern Plains; the EPT Index ranged from 2 to 11 with a mean EPT Index value of 7. Only two stream stations were located in the Middle Atlantic Coastal Plain. An EPT Index value of 1 was recorded for the Middle Atlantic Coastal Plain stations.

Habitat evaluation scores for all sites ranged from 30 to 123 Figure 5.11). Habitat evaluation scores for each stream station are presented in Appendix C. Unlike the EPT Index results, habitat evaluation scores did not reveal any marked patterns from an ecoregional perspective (Figure 5.10). The Blue Ridge habitat evaluation scores(N = 11) ranged from 58 to 123 with a mean of 90. The Upper Piedmont (only 3 stations) had habitat evaluation scores of 82, 102, and 112. The Lower Piedmont's 88 stations had a wide range in habitat scores (30 to 119) with a mean score of 71. The Sand Hills ecoregion stations (N = 5) had a range in habitat evaluation scores of 92 to 108. Habitat evaluation scores for stations in the Southeastern Plains(N = 10)ranged from 73 to 120. The two stations in the Middle Atlantic Coastal Plain had habitat evaluation scores of 96 and 99.

Of the 118 sampling stations for the Savannah REMAP Project, 88 of them are in the Lower Piedmont ecoregion. Seventy-eight of these Lower Piedmont stations had data for all three indicators

(EPT, Fish IBI, and Habitat) utilized for ecological assessment. The other ecoregions within the project area did not have a sufficient number of sampling stations to adequately assess ecological condition. Statistical analysis was therefore restricted to the 78 station data set for the Lower Piedmont ecoregion.

During the four year study period, fish were collected from 108 stream stations. Over 10,000 fish, comprising 49 different species (Table 5.3), were collected. Appendix C list the species and the number collected at each stream station.

Stream fish were collected from six different ecoregions in the Savannah Basin (Table 5.3). Eighty eight (over 75%) of the stream stations were located in the Lower Piedmont ecoregion. The Lower Piedmont is the largest ecoregion in the Savannah River Basin. Only one stream station was located in the Mid-Atlantic Coastal Plain.

Table 5.3 Summary of the number of fish collected over the four year study.

Ecoregions	Stream Stations	Number Fish Species	Number of Fish Identified
Blue Ridge	11	17	318
Upper Piedmont	3	8	267
Lower Piedmont	82	43	9103
Sand Hills	3	9	48
Southern Plains	8	26	329
Mid-Atlantic Coastal Plain	1	2	9
Total	108	49*	10074

* - Number represents total number of different species collected during the study, not the column total.

Ranges of in-situ water quality measurements (pH, dissolved oxygen, conductivity, and temperature) are presented in Table 5.4. In regard to pH, no ecoregional pattern or characteristic emerged. Although the remaining water quality parameters are lacking in number of observations for the Upper Piedmont, Sand Hills, and Middle Atlantic Coastal Plain, there appears to be a gradient from the mountains to the coast (Figure 5.11). This occurs as a decrease in dissolved oxygen and an increase in the temperature regime from the Blue Ridge to the Middle Atlantic Coastal Plain. Although not as apparent as dissolved oxygen and temperature, conductivity, with the exception of the Sand Hills, also increased along this same ecoregional gradient. Again, more data points are necessary to validate this pattern.

Table 5.4 In-situ Water Quality Data

Ecoregion	pH	D.O. (mg/l)	Conductivity (μS/cm)	Temperature (°C)
Blue Ridge	6.6 - 7.6	7.9 - 9.5	16 - 29	16.5 - 23.7
Upper Piedmont	6.3 - 7.0	8.2 - 8.5	20 - 40	22.0 - 23.2
Lower Piedmont	5.1 - 9.1	3.6 - 11.3	15 - 3260	17.5 - 28.2
Sand Hills	5.2 - 6.9	6.7 - 7.9	18 - 914	20.9 - 25.6
South Eastern Plains	6.1 - 7.5	6.3 - 8.3	36 - 184	20.9 - 25.5
Mid-Atlantic Coastal Plain	4.1 - 6.0	5.1 - 6.9	58 - 60	25.6 - 25.8

Water quality violations were noted for dissolved oxygen and pH during the in-situ water quality measurements. Dissolved oxygen at Station 98, an unnamed tributary to Cliatt Creek, in Columbia County, Georgia was measured at 3.6 mg/L which is below the two state's water quality standards of 4.0 mg/L. This translates into about 2% of the stream miles being below the minimum standard dissolved oxygen in the basin (Figure 5.12). Likewise, about 8% of the stream miles were below both state's pH standard of 6.0 and approximately 2% were greater than the allowable level for streams in Georgia (8.5) and South Carolina (8.0) (Figure 5.13).

Algal growth potential tests were conducted for the first two years and analyzed from a basin perspective. The results of that effort and interpretation of the data are in a report by Raschke, et al. (1997) (Appendix H).

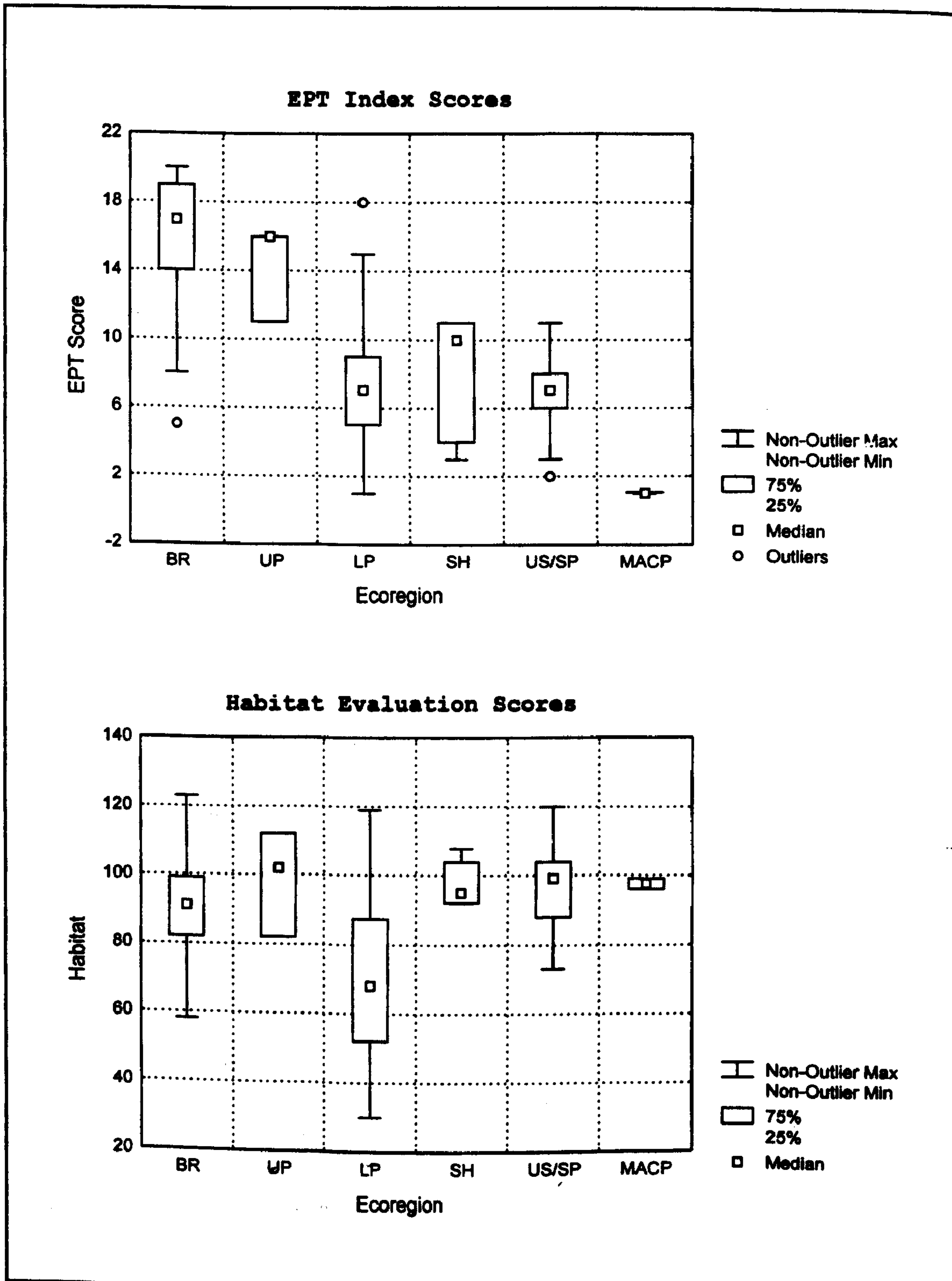


Figure 5.10 Box and Whisker Plots of Ecoregion EPT Index Scores and Habitat Scores.

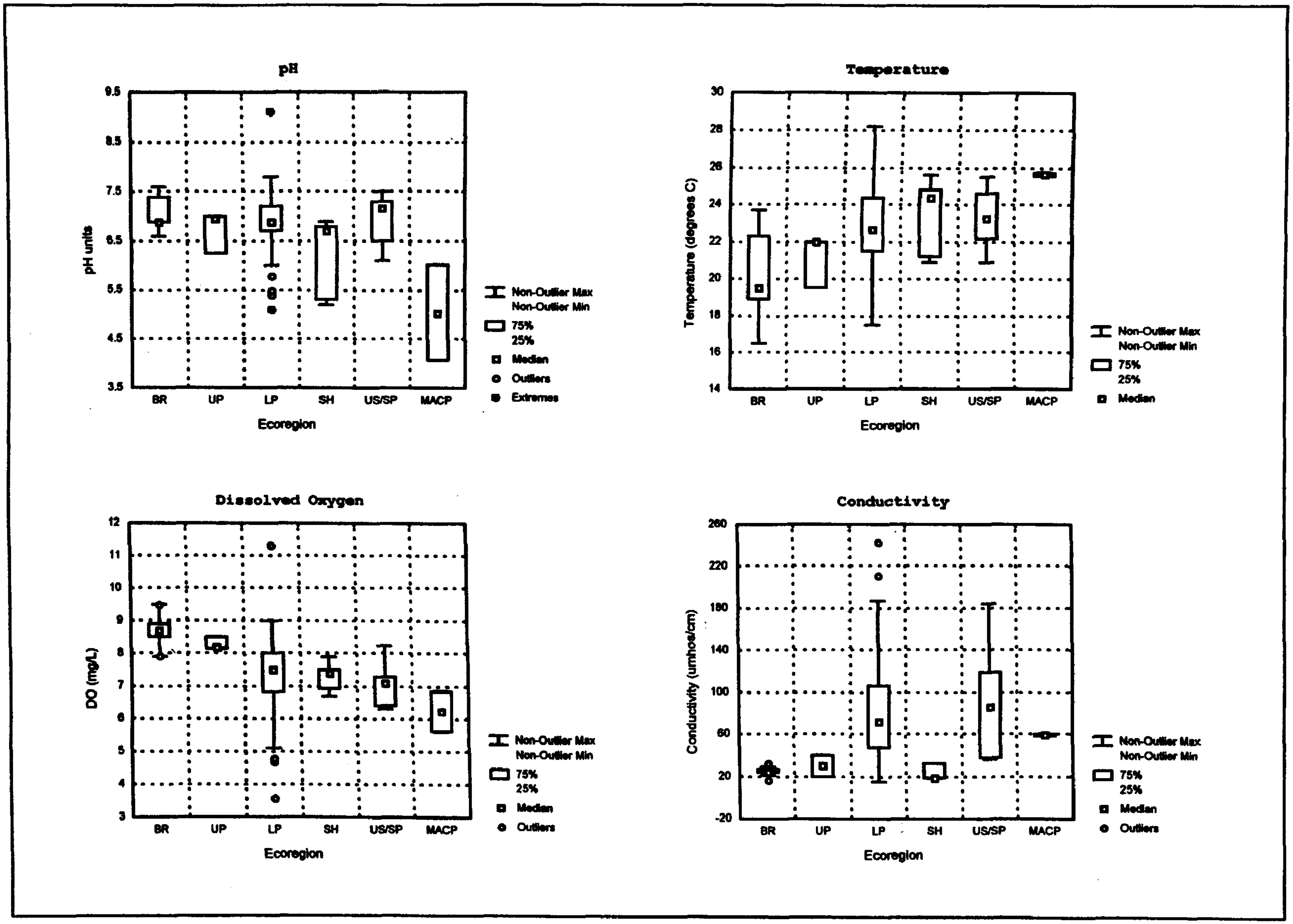


Figure 5.11. Box Whisker Plots of Ecoregion In-Situ Water Quality Parameters.

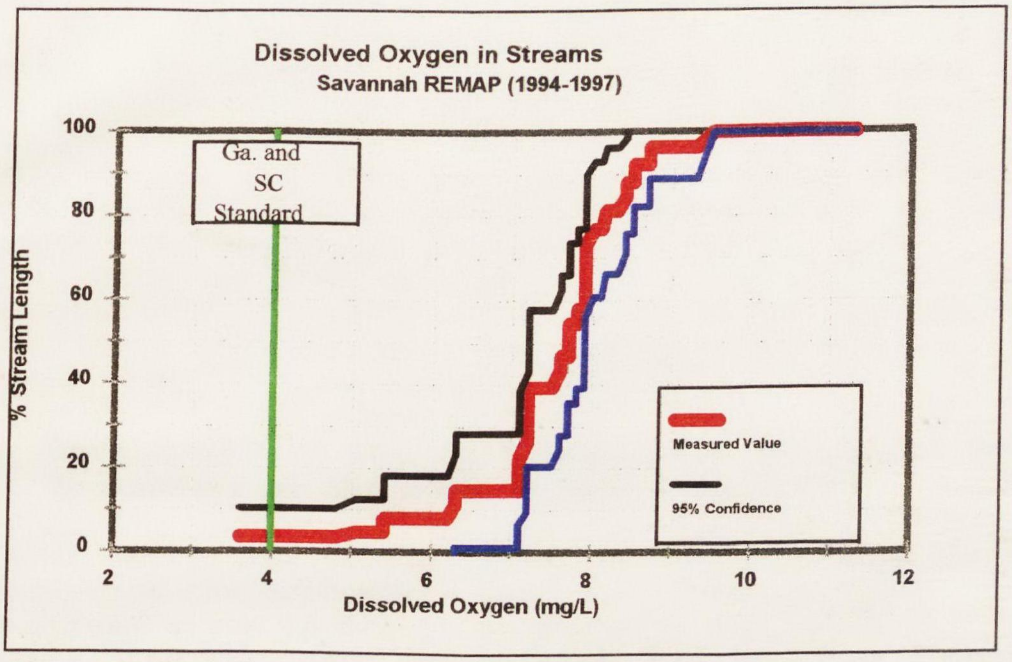


Figure 5.12 Cdf curve of Dissolved Oxygen Data.

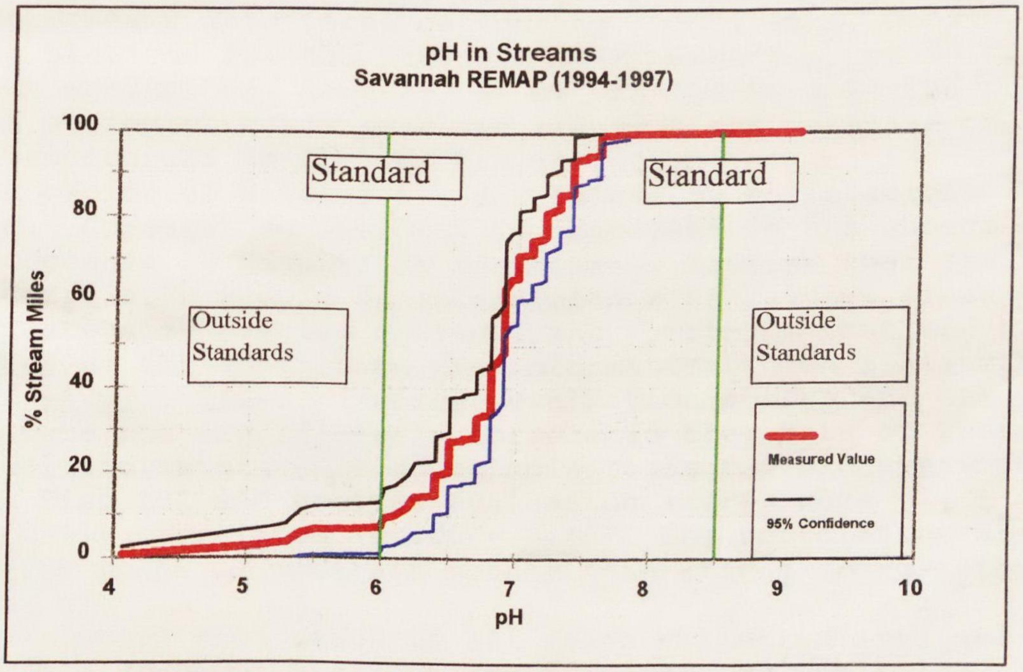


Figure 5.13 Cdf curve of pH data.

5.2 Ecoregion Perspective

Because of our original emphasis on Basin ecological condition, sampling locations were randomly selected over the whole Savannah River Basin, not by ecoregion. This skews the number of sampling locations in favor of the largest ecoregion, which was the Lower Piedmont. The Lower Piedmont ecoregion is a large geographical area that encompasses two states and many subwatersheds. There were not enough stream stations in all of the ecoregions to adequately develop an index for each ecoregion. Only the Lower Piedmont region had sufficient number of stream stations to produce enough data, in our opinion, to develop an index that realistically assesses ecological condition.

5.2.1 Development of Scoring Criteria for Ecological Health Assessment of the Lower Piedmont Ecoregion

Benthic macroinvertebrate, fish, and habitat were the basis for interpreting the ecological health of Savannah REMAP wadeable stream sites in the Lower Piedmont Ecoregion. Specifically, the EPT Index (macroinvertebrates), the fish IBI (Index of Biotic Integrity), and habitat evaluation scores were utilized to develop a scoring system for classifying Lower Piedmont streams into three categories (good, fair, poor). Sampling stations for the Savannah REMAP were located in six different ecoregions, however, 88 of the 119 were in the Lower Piedmont ecoregion which provided a sufficient database to examine ecological health in this ecoregion.

The choice of metrics was determined by correlation analysis. Correlation analysis is important in the choice of metrics because it identifies redundancy. Metrics that are very highly correlated should be interpreted with caution since they may indicate some overlap or redundancy; metrics that are highly correlated do not contribute new information to an assessment (Barbour et al., 1996). Habitat evaluation scores and EPT Index results were not significantly correlated thus both of these ecological indicators were acceptable tools for bioassessment. Although Fish IBI and habitat evaluation scores were significantly correlated ($p<.05 = 0.42$), the correlation was more on the order of moderate rather than strong correlation (Appendix C).

Descriptive statistics of all seven variables examined in all 88 Lower Piedmont stations are presented in Table 5.5. Box and whisker plots (Figure 5.15) were performed on the results for each indicator to define the boundaries for three categories (Good, Fair, and Poor). A scoring matrix based on boundaries defined by box and whisker plots was completed for the Lower Piedmont Ecoregion. The scoring matrix for the EPT Index, Fish IBI, and Habitat is provided in Table 5.6.

Table 5.5 Descriptive Statistics of the Stream Variables.

Descriptive Statistics					
Variables	# of Stations	Mean	Minimum	Maximum	Standard Deviation
Fish IBI Scores	82	26.00	13.00	43.00	6.26
Habitat Scores	84	70.99	30.00	119.00	21.68
EPT Scores	87	7.14	1.00	18.00	3.19
pH	84	6.91	5.10	9.10	0.53
Temperature (C)	83	23.02	17.5	28.2	2.05
Dissolved Oxygen (mg/l)	75	7.29	3.6	11.3	1.14
Conductivity (μS/cm)	76	80.58	15.00	243.00	43.92

Table 5.6 Scoring Matrix for Ecological Health of Lower Piedmont Streams

Indicator	5 points GOOD	3 points FAIR	1 point POOR
EPT Index	$\geq$ 9	6 - 8	$\leq$ 5
Fish IBI	$\geq$ 31	22 - 30	$\leq$ 21
Habitat	$\geq$ 87	53 - 86	$\leq$ 52

The next step was defining a final classification system based on the total score obtained from all three indicators for the 78 station Lower Piedmont data set. Again, box plots were utilized to define the boundaries for total scores in the "Good", "Fair", and "Poor" categories. This final classification system is termed the Savannah Basin-Lower Piedmont Ecological Index (SB-LPEI).

5.2.2 SB-LPEI and Ecological Condition of Lower Piedmont Streams

Final ecological health classification of Lower Piedmont streams, based on total points derived from the three ecological indicators (EPT Index, Fish IBI, and Habitat), was determined by the following scheme:

Classification	Total points
Good	$\geq$ 11
Fair	8 - 10
Poor	$\leq$7

(Note: a score of 1 in either of three ecological indicators does not warrant a "Good" ranking)

Based on this scoring scheme, 69% of the stream miles indicated some degree of impairment ("Fair" and "Poor" rankings) (Figure 5.14). A complete listing, by station, of the individual ecological indicator results and the final ecological health classification from the results of the SB-LPEI is provided in Appendix C. Habitat degradation, primarily from sedimentation, is apparently the leading cause affecting the aquatic life in Lower Piedmont streams. Habitat evaluation parameters such as bottom substrate/available cover, channel alteration, and bottom scouring and deposition specifically identify sedimentation concerns. Low scores in these three sediment-related parameters of the habitat evaluation worksheet translated into less than desirable benthic macroinvertebrate and fish populations. Conversely, ecoregional reference sites scored higher in these three sediment-related parameters and supported diverse fish and macroinvertebrate communities.

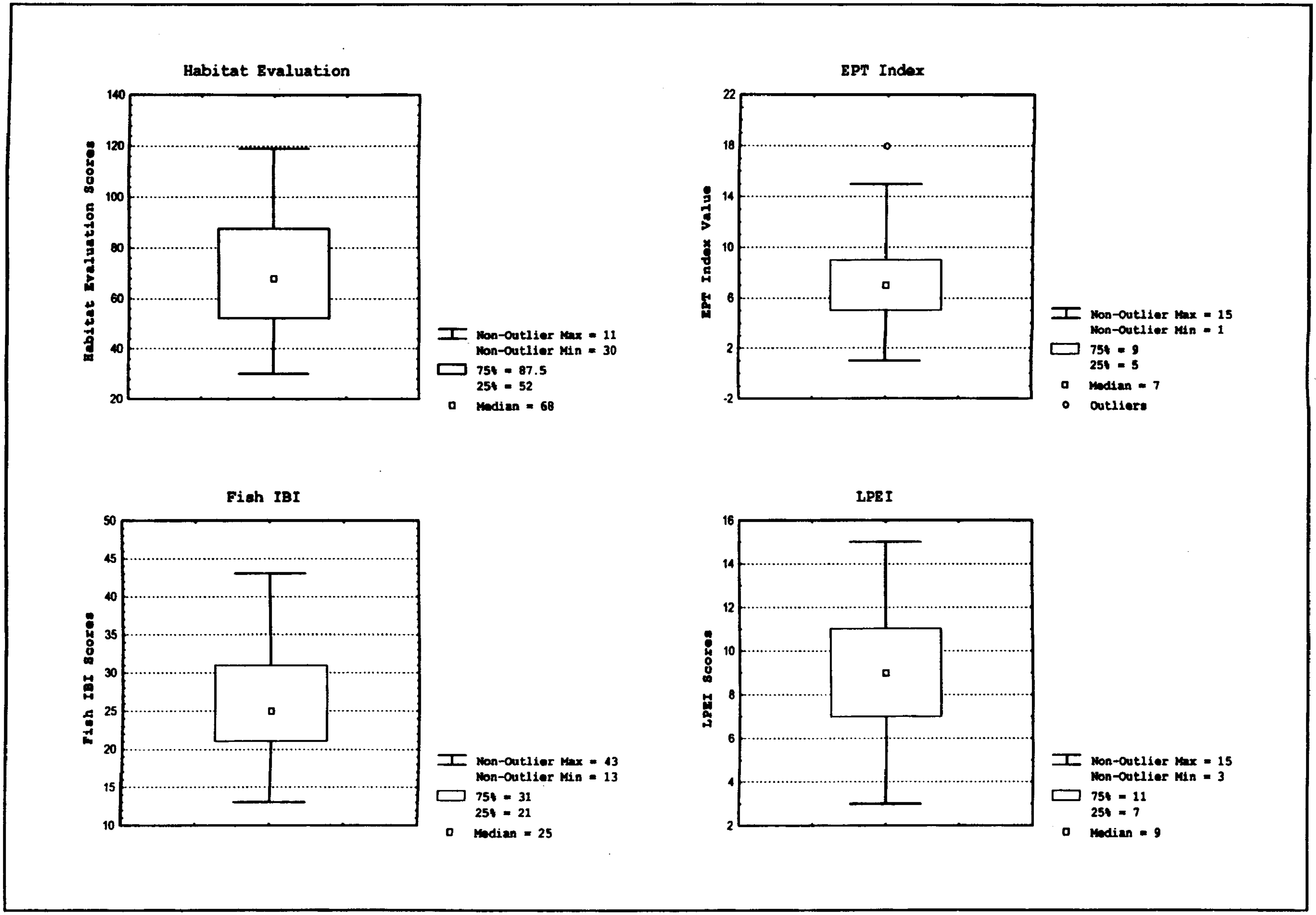


Figure 5.14. Box and Whisker Plots Used to Develop the Scoring Criteria for the Savanna Basin-Lower Piedmont Ecological Index.

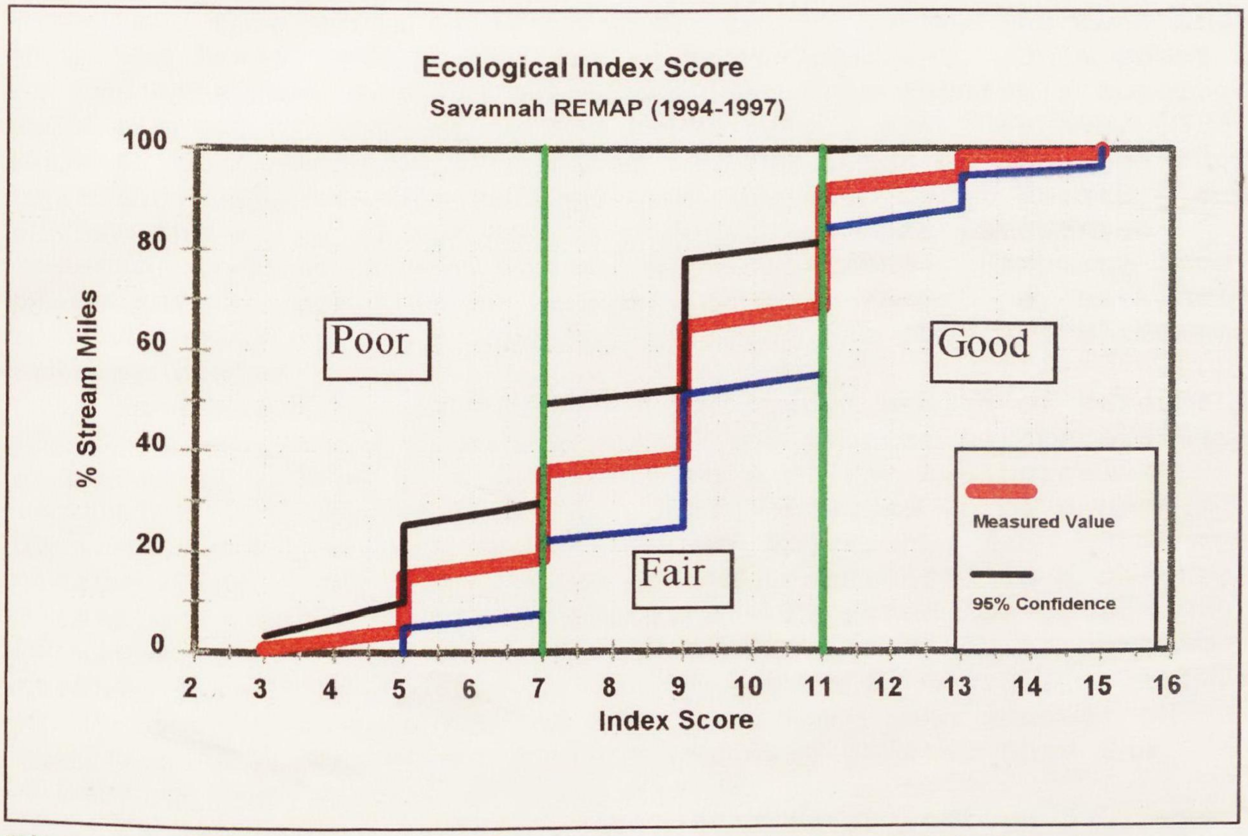


Figure 5.15. Cdf Curve of Savannah Basin-Lower Piedmont Ecological Index.

6.0 Discussion of Objectives

Estimate the status and change of the condition of water resources in the Savannah River Basin.

Based on three years of measuring trophic condition of the tributary embayments of large lakes in the basin, the data show that the lakes' embayments are in good condition. Only about 5% of the embayment acreage exhibited less than desirable conditions with respect to recreation and water supply use (Raschke, 1993). Much of that could be attributed to wind fetch at the near-shore stations. Significant changes from cycle to cycle possibly are within the realm of natural variability or some unmeasured stressor indicators within the lakes' environs. Sampling took place several weeks after rainfall events, therefore, external stream inputs were not expected to cause the observed differences between cycles.

In evaluating the status of ecological health of streams in the Savannah Basin, both biological and habitat parameters were examined to arrive at a final estimate of the ecological condition of wadeable streams. There appeared to be a general decline southward with respect to EPT Index, DO, and conductivity. The temperature gradient decreased in a northward direction. Water quality violations were noted for DO and pH. A DO violation of <4.0 was observed at Station 98 on an unnamed tributary to Cliatt Creek in Columbia County, Georgia. Likewise, about 8% of the stream miles were less than both states' pH standard of 6.0, and 2% of the miles were greater than the allowable South Carolina level of 8.0.

In-depth data analysis, as indicated in Section 5.0, was restricted to streams in the Lower Piedmont Ecoregion because there was not sufficient biological data for a thorough analysis of other ecoregions. Data analysis lead to the development of the Lower Piedmont Ecological Index (SB-LPEI). The components of the SB-LPEI were the fish IBI, macroinvertebrate EPT Index, and the RBP V habitat evaluation scores.

This SB-LPEI was successful in establishing ecological "status" of wadeable streams in the Lower Piedmont Ecoregion. Based on the SB-LPEI, sixty-nine percent of the streams were classified as "fair" or "poor" indicating ecological impairment. Impairment at these sites pointed to habitat degradation primarily from excessive sedimentation. The results of the SB-LPEI can be utilized to establish areas of concern for future evaluation.

Change in ecological condition was not established during this study. There was not enough data for all study years to confidently evaluate change over the four year study period.

Identify water quality spatial gradients that exist within the Savannah River Basin and associate current and changing condition with factors that may be contributing to this condition and spatial gradients.

Analysis of information by ORD, NERL-LasVegas (Appendix F) showed that landscape indicators like percent forest cover, forest edge, proportion of watershed area with agriculture or urban land cover(U-Index), agriculture edge, average patch, average forest patch, and agriculture on slopes >3% were significantly correlated with the stream indicators AGPT, EPT Index, Fish IBI, and Habitat Score (Appendix F). NERL-LasVegas showed that both the proportion and patterns of land use are useful in assessing potential causative effects of stream condition. Landscape indicators at the subbasin scale provided the best characterization of the basin.

In a previous Savannah REMAP report using two years of stream data, Raschke et al. (1996) identified one area that had an inordinate amount of bad sites clustered around Hart and Franklin Counties, Georgia near Interstate 85. Upon review of four years of data and taking a very conservative approach in developing criteria for poor ecological health, the information revealed that this area is much larger than expected. It has expanded into South Carolina (Figure 6.1). This area includes all or part of Hart and Elbert Counties, Georgia and Oconee, Pickens, and Anderson Counties, South Carolina. The designation of an area does not imply that every stream is in "poor" condition nor that the area has a certain confidence band. Our observations are qualitative, that is, there is an unusual number of poor areas clustered, in our professional opinion, along the Interstate 85 corridor. We believe streams in this area are most vulnerable to landscape perturbations and in need of further detailed investigation.

The landscape analysis showed that approximately 64% of this "poor" area is forest, 22.3% agriculture, 2.6% urban, and 3% barren. Two percent of the area is in agriculture on slopes >3%, there is approximately 21% agriculture on moderately erodible soils, and approximately 1% on highly erodible soils, and <0.1% agriculture on slopes >3% in highly erodible soils.

This area has been subjected to a considerable increase in population growth because of the large impoundments in the upper part of the Savannah River Basin. Furthermore, examination of GIS information shows that it has a high density of chicken production, extensive agriculture in large blocks, and the headwaters of streams in the subbasins have a high density of roads. In some subbasins of this "poor" area, the forest land is highly fragmented and the land has been opened up to industrial/urban/and agriculture development in the headwaters of some of the streams.

Demonstrate the utility of the REMAP approach for ecoregion and river basin monitoring and its applicability for state monitoring programs.

In the arena of state monitoring, the concept of probability sampling is like the "new kid on the block" - the one who dresses differently and acts differently. And we, the regions and states, mirroring real life, have been slow in warming up to this "kid," and rightfully so! For he embraces a new way of thinking that threatens stability, cultural traditions, and the past historical record. From the inception of this project, we were aware of the potential disruption that probability sampling could create among our state partners. So we diligently set a course of testing the EMAP approach and determined how we could best incorporate it into state monitoring schemes with as little disruption as possible. We sought out and found Dr. Steve Rathbun of the University of Georgia Statistical Department. He is a statistician who has experience in different types of probability sampling approaches and experience with the problems of incorporating the "new kid on the block" into traditional state monitoring programs. Rathbun addressed concerns regarding probability-based designs posed by the "Assessment Design Focus Group of the 305(b) Consistency Workgroup (Appendix G). His full report in Appendix G is an important first step in the integration of judgement and probability monitoring data without losing most of the historical data.

States and the federal government historically have established monitoring networks based on judgmental sampling. That is, stations were usually located where there were pollution problems or the area was vulnerable to pollution because of man's activities. Unfortunately, this type of site selection is biased and it is virtually impossible to relate to a whole population of streams/lakes, watersheds, basins, ecoregions etc. Sampling designs based on judgement sampling are not likely to yield representative samples.

With the need for preserving historical monitoring data and marrying it to a probability-based design, Rathbun (Appendix G) tested an approach using an interval overlap technique with historical judgement sites and probability-based sites located near judgement sites. The technique uses a back-prediction method that determines what the historical data should have been had a probability-based sample design been implemented from the very beginning of the program. If the above methods shows there is still some bias in the data, then a bias-corrective factor is calculated to best fit the data.

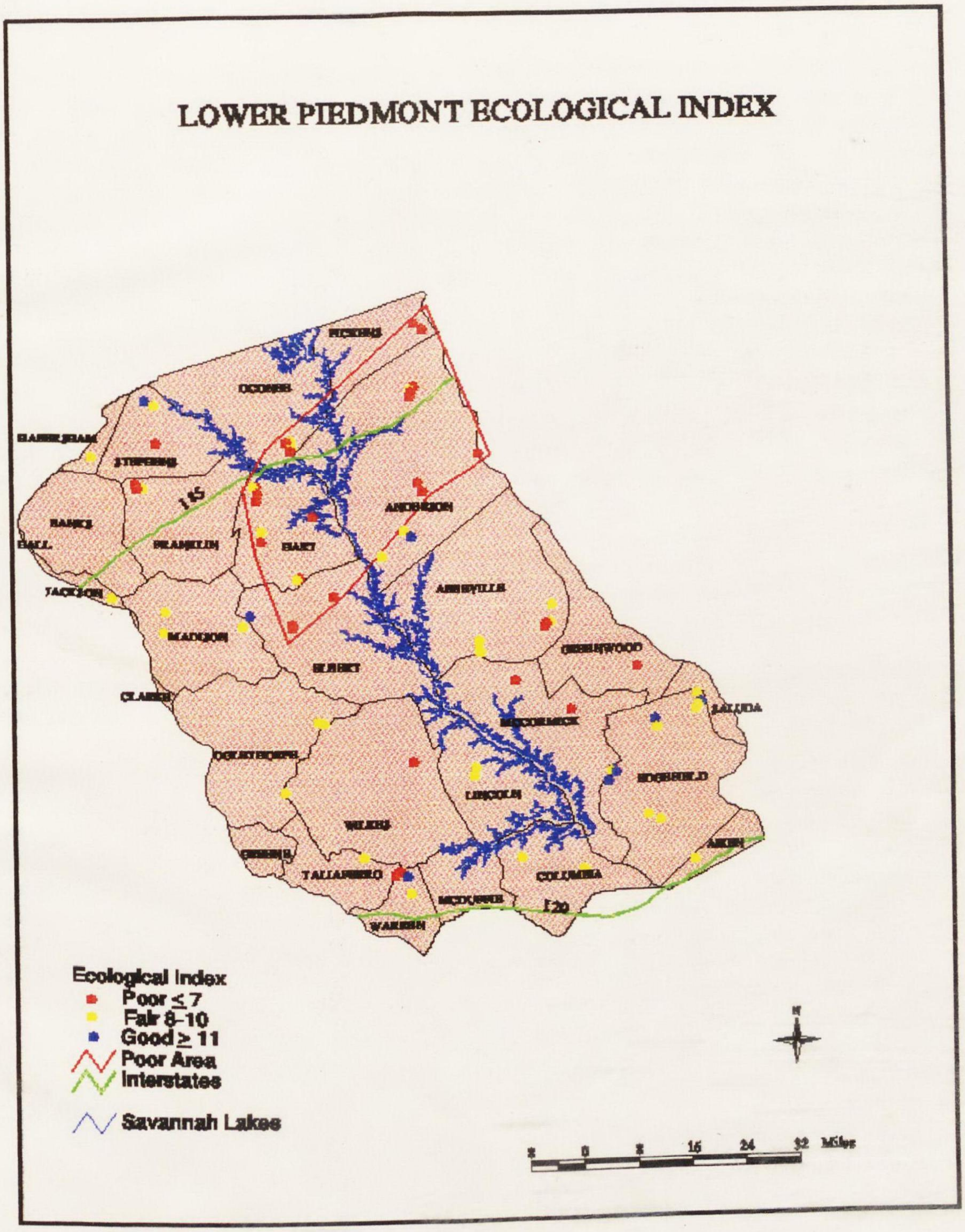


Figure 6.1. Area in the Lower Piedmont with an Unusual Amount of Poor Sites.

Incorporate the REMAP approach in the formulation and accomplishment of the state river basin management plans.

Most states are monitoring their basins on a cyclic schedule rather than doing state-wide monitoring every year. This report shows that it is possible to incorporate probabilistic sampling (the EMAP approach) into state monitoring programs at the basin level and even the ecoregional level. Rathbun (Appendix G) presents a method of incorporating historical judgement station data into a probabilistic design. This is important because the states can better estimate stream miles impacted etc. and have sufficient data for trend analysis. We can't predict to what degree each state will incorporate probability sampling into their monitoring programs. As of the distribution of this report, we have had a workshop on integration of judgement data with probability data. The workshop addressed state concerns and opened the door for joint discussions. Likewise, the Office of Water has directed the states to move toward probability sampling for purposes of including better estimates of ecological condition into the 305(b) reports. South Carolina is moving toward probability sampling, Alabama has partially incorporated it into their monitoring program and Kentucky is evaluating it presently.

Provide baseline information required to conduct comparative risk assessments in the Savannah River Basin.

REMAP is not a problem-specific program. It focuses on monitoring the condition or system response, and changes in the condition of the ecological resource; not specific physical alterations, chemical species or associated problems. Biological indicators are the focus of monitoring in REMAP, but selected abiotic indicators can be monitored to provide directional diagnostic ability if changes in condition are detected or existing condition of the resource is degraded. Additional and/or more intensive monitoring in a given region likely will be required to specifically determine problem causes and determine the existing or potential risk to the resource. A risk analysis consists of three phases: Problem Formulation, Analysis, and Risk Characterization (EPA, 1992b).

REMAP contributes primarily to problem formulation by providing comparable information on the condition of multiple resources in a region, basin, or ecoregion. As shown in the data analysis, it can highlight areas, stream miles, etc. that are affected. It can show areas in a basin or ecoregion that might be under man-induced assaults, thereby needing further investigation like the area along I-85 in Georgia and South Carolina (Figure 6.1; Appendix F).

7.0 REFERENCES

APHA. 1995. Standard Methods for the Examination of Water and Wastewater, 19th Edition. American Public Health Association, Washington, D.C.

Buck, D. H. 1956. Effects of turbidity on fishing. Transactions of the American Wildlife Conference 21:249-261.

Barbour, M. T., J. L. Plafkin, B. P. Bradley, C. G. Graves, and R. W. Wiseman. 1992. Evolution of EPA's Rapid Bioassessment Benthic Metrics: Metric Redundancy and Variability Among Reference Stream Sites. Environmental Toxicology and Chemistry 11:437-449.

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1998. Draft Revision to Rapid Bioassessment Protocols for Use in Streams and Rivers. EPA 841-D-97-002, Office of Water, Washington, D. C.

Bromberg, S. M. 1990. Identifying ecological indicators: An environmental monitoring and assessment program. Journal of the Air Pollution Control Association 40:976-978.

Callicott, J. B. 1991. Conservation ethics and fishery management. Fisheries 16:22-28.

Chow, Ven te (Editor). 1964. Handbook of Applied Hydrology: A Compendium of Water-Resources Technology. McGraw-Hill Book Co. New York, New York.

Doudoroff, P., and C. E. Warren. 1957. Biological indices of water pollution with special reference to fish populations. Pages 144-163. In: Biological Problems in Water Pollution. U. S. Public Health Service, Cincinnati, Ohio.

EPA. 1983. Methods for the Chemical Analysis of Water and Wastes. EPA-600/4-79-020. Revised March, 1983. U.S. Environmental Protection Agency, Cincinnati, OH.

EPA. 1991. The Watershed Protection Approach, an Overview. EPA 503/9-92/002. U. S. Environmental Protection Agency, Washington, D. C.

EPA. 1992a. Methods for the Determination of Chemical Substances in Marine and Estuarine Environmental Samples. U.S. Environmental Protection Agency, Cincinnati, OH.

EPA. 1992b. Framework of Ecological Risk Assessment: Risk Assessement Forum. EPA 630/ R-92/001. EPA, Washington, D. C.

EPA. 1996. Watershed Approach Framework. EPA 840/S-96/001. Office of Water. U. S. Environmental Protection Agency, Washington, D. C.

EPA, 1998. EPA Region 4, Science and Ecosystem Support Division, Ecological Assessment Branch SOP and QA Manual. Athens, GA.

Frey, D. G. 1977. Biological integrity of water - A historical approach. Pages 127-140. In: R. K. Ballentine and L. J. Guarria (Eds.). The Integrity of Water. U. S. Environmental Protection Agency, Washington, D. C.

FTN Associates, R. L. Raschke, H. Howard. 1994. Plan of Study for Characterizing Ecological Condition in the Savannah River Basin. U. S. Environmental Protection Agency, Region 4, Science and Ecosystem Support Division, Ecological Assessment Branch, Athens, Georgia.

Hughes, R. M., and R. F. Noss. 1992. Biological diversity and biological integrity: Current concerns for lakes and streams. Fisheries 17:11-19.

Hunsaker, C. T., D.E. Carpenter, and J. Messer. 1990. Ecological indicators for regional monitoring. Bulletin of the Ecological Society of America 71(2):165-172.

Hunsaker, C. T., and D. E. Carpenter. 1990. Environmental Monitoring and Assessment Program - Ecological Indicators. EPA 600/3-90/060. U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Karr, J. R., and D. R. Dudley. 1981. Ecological perspective on water quality goals. Environmental Management 5:55-68.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing Biological Integrity in Running Waters: A Method and its Rationale. Publication No. 15. Illinois Natural History Survey, Champaign, Illinois.

Klessig, L. 1988. Problem Identification. In: The Lake and Reservoir Restoration Guidance Manual. EPA 440/5-88-002. Criteria and Standards Division, Washington, D. C.

Larsen, D. P., D. J. Stevens Jr., A. R. Selle, and S. G. Paulsen. 1991. Environmental Monitoring and Assessment Program, EMAP - surface waters, a Northeast lakes pilot. Lake and Reservoir Management 7(1):1-11.

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

Management Committee. 1995. Savannah River Watershed Project: Initial Assessment and Prioritization Report for the Savannah River Basin. Volumes 1 and 2. Prepared by the Management Committee of the Savannah River Basin Watershed Project.

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

Miller, W., J. Green, and T. Shiroyama. 1978. The Selenastrum capricornutum Printz Algal Assay Bottle Test. EPA-600/9-78-018.

National Academy of Sciences and National Academy of Engineering. 1973. Water Quality Criteria 1972. Superintendent of Documents, U. S. Government Printing Office. Washington, D. C.

Noss, R. F. 1990. Indicators for monitoring biodiversity: A hierarchical approach. Conservation Biology 4:355-364.

Olem, H. and G. Flock (Eds.). 1990. Lake and Reservoir Restoration Guidance Manual. EPA 440/4-90/006. Second Edition Prepared by The North American Lake Management Society for the U. S. Environmental Protection Agency, Washington, D. C.

Overton, W. S., D. Stevens Jr., D. White 1990. Environmental Monitoring and Assessment Program - Design Report for EMAP. EPA 600/3-91/053. U. S. Office of Research and Development, Environmental Protection Agency, Washington, D. C.

Peters, H. 1981. Phosphorus availability in Lake Memphrenagog and its tributaries. Limnology and Oceanography 26:1150-?

Pister, E. P. 1987. A pilgrims progress from Group A to Group B. Pages 221-232. In: J. B. Callicott (Ed.). Companion to Sand County Almanac. University of Wisconsin Press, Madison, Wisconsin.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross, and R. M. Hughes. 1989. Rapid Bioassessment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrate and Fish. EPA 440/4-89-001. Office of Water, U. S. EPA.

Raschke, R. L. 1994. Phytoplankton bloom frequencies in a population of small southeastern impoundments. Lake and Reservoir Management 8(2):205-210.

Raschke, R. L. 1993. Guidelines for Assessing and Predicting Eutrophication Status of Small Southeastern Piedmont Impoundments. U. S. Environmental Protection Agency, Region 4, Science and Ecosystem Support Division, Ecological Assessment Branch, Athens, Georgia.

Raschke, R. L., and D. A. Schultz. 1987. The use of the algal growth potential test for data assessment. Journal of the Water Pollution Control Federation 59(4):222-227.

Raschke, R. L., B. Cosgrove, and M. Koenig. 1985. Phosphorus loading effects upon phytoplankton standing crop in a tributary embayment of a southern piedmont reservoir. Pages 139-148. In: Lake and Reservoir Management: Practical Applications. North American Lake Management Society, Madison, Wisconsin.

Raschke, R. L., H. S. Howard, J. R. Maudsley, and R. J. Lewis. 1996. The Ecological Condition of Small Streams in the Savannah River Basin: A REMAP Progress Report. EPA Region 4, Science and Ecosystem Support Division, Ecological Assessment Branch, Athens, GA.

Raschke, R. L., R. Quinn, T. Olsen, and D. Stevens Jr. 1997. Savannah River Basin REMAP Interim Report: Large Lake Embayments. EPA Region 4, Science and Ecosystem Support Division, Ecological Assessment Branch, Athens, GA.

Rathbun, S. C., R. Houghton, and C. Laurin. 1996. Cramer-von Mises Tests for Environmental Monitoring Data. A report for the

U. S. Environmental Protection Agency, Science and Ecosystem Support Division, Ecosystem Assessment Branch, Athens, Georgia.

Schultz, D., R. Raschke, and R. Jones. 1994. A shortened algal growth potential test. Environmental Monitoring and Assessment 32:201-205.

Snedecor, G. W., and W. G. Cochran. 1967. Statistical Methods. Sixth Edition. The Iowa State University Press, Ames, IA.

Stevens, D. L. Jr. 1994. Implementation of a national environmental monitoring program. Journal of Environmental Management 42:1-29.

Stevens, D. L. Jr. 1997. Variable density grid-based sampling designs for continuous spatial populations. Environmetrics 8:165-197.

Stevens, D. L. Jr., A. R. Olsen, and D. White. 1992. Integrated Sampling Design: Environmental Monitoring and Assessment Program. EPA 600/xxx/xxx. U. S. Environmental Protection Agency, Environmental Research Laboratory, Corvallis, Oregon.

Urquhart, M. S., W. S. Overton, and D. S. Birkes. 1993. Comparing sampling designs for ecological status and trends: Impact of temporal patterns. In: V. Barnett and K. F. Turkman (Eds.). Statistics for the Environment. Wiley, New York.

Wallace, J. B., J. W. Grubaugh, and M. R. Whiles. 1996. Biotic Indices and Stream Ecosystem Resources: Results from an Experimental Study. Ecological Applications 6(1):140-151.

Welch, P. S. 1948. Limnological Methods. McGraw-Hill Book Co., Inc., New York.

White, D., A. J. Kimerling, and W. S. Overton. 1992. Cartographic and geometric components of a global sampling design for environmental monitoring. Cartography and Geographic Information Systems 19:5-22.

Appendix A

QA Data

Table QC-1. SAVANNAH REMAP STUDY- FIELD DUPLICATES

AGPT in milligram Dry Weight per Liter

STA#	CYCLE	REP1	REP2	MEAN	SD	CV
5	1	1.01	0.86	0.94	0.134	14.3%
19	1	0.84	0.55	0.70	0.258	37.1%
39	1	0.66	0.67	0.67	0.009	1.3%
49	1	0.48	0.41	0.45	0.062	14.0%
59	1	4.46	5.08	4.77	0.552	11.6%
79	2	1.82	1.51	1.67	0.276	16.6%
97	2	2.96	1.60	2.28	1.210	53.1%
101	2	1.81	2.07	1.94	0.231	11.9%
105	2	1.99	1.71	1.85	0.249	13.5%
8	3	0.93	0.79	0.86	0.125	14.5%
16	3	1.41	1.35	1.38	0.053	3.9%
41	3	1.20	1.13	1.17	0.062	5.3%
43	3	1.08	1.17	1.13	0.080	7.1%
						15.7%

CHLOROPHYLL a in ug/L

STA#	CYCLE	REP1	REP2	MEAN	SD	CV
5	1	1.51	1.98	1.75	0.418	24.0%
19	1	2.44	2.13	2.29	0.276	12.1%
39	1	1.65	1.70	1.68	0.045	2.7%
49	1	2.98	2.83	2.91	0.134	4.6%
59	1	0.91	0.93	0.92	0.018	1.9%
79	2	0.83	1.55	1.19	0.641	53.8%
101	2	2.12	1.80	1.96	0.285	14.5%
8	3	1.50	0.90	1.20	0.534	44.5%
16	3	1.60	1.60	1.60	0.000	0.0%
41	3	7.80	8.30	8.05	0.445	5.5%
43	3	2.60	3.00	2.80	0.356	12.7%
						16.0%

TOTAL PHOSPHORUS in ug/L

STA#	CYCLE	REP1	REP2	MEAN	SD	CV
79	2	6	7	6.5	0.890	13.7%
97	2	12	7	9.5	4.450	46.8%
101	2	14	6	10.0	7.120	71.2%
105	2	9	10	9.5	0.890	9.4%
8	3	4	4	4.0	0.000	0.0%
16	3	5	4	4.5	0.890	19.8%
41	3	29	27	28.0	1.780	6.4%
43	3	12	12	12.0	0.000	0.0%
						20.9%

TSS in mg/L

STA#	CYCLE	REP1	REP2	MEAN	SD	CV
8	3	1U	2.0			
16	3	2.2	1.7	1.95	0.445	22.8%
41	3	4.7	3.6	4.15	0.979	23.6%
43	3	1.8	2.0	1.9	0.178	9.4%
						18.6%

U - material was analyzed for but not detected. The number is the minimum quantitation limit.

Table QC-2 SAVANNAH REMAP STUDY - FIELD BLANKS

SAMPLE#	STA#	CYCLE	TSS (mg/L)	TP (ug/L)	CHL a (ug/L)
9609	41C	3	1U	3U	0.5U
9683	8C	3	1U	3U	0.5U
9696	43C	3	1U	3	0.5U
9741	16C	3	1U	3U	0.5U

U = Material was analyzed for but not detected. The number is the minimum detection limit.

B

Appendix B

Lake Data

SAVANNAH RIVERBASIN LAKE DATA

CYCLE	LAKE	STATION	DATE	AGPT (MG/L)	LIMITING NUTRIENT	CHL a (UG/L)	TPHOS (UG/L)	TSS (MG/L)
1	JOCASSEE	1	07/20/95	0.68	N+P	2.54	20 U	34
1	KEOWEE	4	07/21/95	1.12	N+P	2.03	20 U	4.0 U
1	KEOWEE	5A	07/21/95	1.01	N+P	1.51	20 U	4.0 U
1	KEOWEE	5B	07/21/95	0.86	N+P	1.98	20 U	4.0 U
1	KEOWEE	7	07/21/95	0.64	N+P	1.77	20 U	4.0 U
1	KEOWEE	8	07/21/95	0.79	N+P	1.66	20 U	4.0 U
1	HARTWELL	10	07/20/95	1.49	N+P	3.55	20 U	6.0
1	HARTWELL	11	07/20/95	0.91	N+P	3.89	20 U	4.0 U
1	HARTWELL	13	07/20/95	1.97	N+P	2.35	20 U	4.0 U
1	HARTWELL	14	07/20/95	0.81	N+P	2.37	20 U	4.0 U
1	HARTWELL	15	07/20/95	0.86	N+P	2.39	20 U	4.0 U
1	HARTWELL	16	07/20/95	0.88	N+P	2.02	20 U	4.0 U
1	HARTWELL	17	07/20/95	0.80	N+P	1.22	20 U	4.0 U
1	HARTWELL	18	07/20/95	1.26	N+P	6.84	30	4.0 U
1	HARTWELL	19A	07/20/95	0.84	N+P	2.44	20 U	4.0 U
1	HARTWELL	19B	07/20/95	0.55	N+P	2.13	20 U	4.0 U
1	HARTWELL	20	07/20/95	1.63	N+P	1.55	20 U	4.0 U
1	HARTWELL	22	07/20/95	0.62	N+P	1.74	20 U	4.0 U
1	HARTWELL	24	07/19/95	0.71	N+P	0.84	20 U	4.0 U
1	HARTWELL	25	07/19/95	0.65	N+P	0.98	20 U	4.0 U
1	HARTWELL	26	07/19/95	0.55	N+P	1.77	20 U	4.0 U
1	HARTWELL	27	07/19/95	0.70	N+P	4.16	20 U	4.0 U
1	THURMOND	28	07/19/95	0.79	N+P	1.87	20 U	4.0 U
1	THURMOND	29	07/19/95	0.75	N+P	1.31	20 U	4.0 U
1	THURMOND	30	07/19/95	0.69	N+P	2.35	20 U	4.0 U
1	THURMOND	31	07/19/95	0.67	N+P	2.18	20 U	4.0 U
1	THURMOND	32	07/19/95	0.98	N+P	4.50	20 U	4.0 U
1	THURMOND	34	07/18/95	0.92	N+P	2.16	20 U	4.0 U
1	THURMOND	35	07/18/95	0.71	N+P	1.95	20 U	4.0 U
1	THURMOND	36	07/18/95	0.66	N+P	1.60	20 U	6.0
1	THURMOND	38	07/18/95	0.66	N+P	1.94	20 U	4.0 U
1	THURMOND	39A	07/18/95	0.67	N+P	1.65	20 U	4.0 U
1	THURMOND	39B	07/18/95	0.66	N+P	1.70		
1	THURMOND	40	07/18/95	1.01	N+P	3.40	20	4.0 U
1	THURMOND	41	07/18/95	1.34	N	11.56	50	4.0 U
1	THURMOND	42	07/18/95	0.74	N	11.17	30	4.0 U
1	RUSSELL	43	07/17/97	0.85	N+P	2.92	20 U	4.0 U
1	RUSSELL	44	07/17/95	0.95	N	3.70	20 U	4.0 U
1	RUSSELL	45	07/17/95	0.64	N	5.47	60	32
1	RUSSELL	46	07/17/95	0.75	N+P	2.59	20 U	4.0 U
1	RUSSELL	48	07/17/95	0.63	N+P	3.39	20 U	4.0 U
1	RUSSELL	49A	07/17/95	0.48	N+P	2.98	20 U	4.0 U
1	RUSSELL	49B	07/17/95	0.41	N+P	2.83		

SAVANNAH RIVERBASIN LAKE DATA

CYCLE	LAKE	STATION	DATE	AGPT (MG/L)	LIMITING NUTRIENT	CHL a (UG/L)	TPHOS (UG/L)	TSS (MG/L)
1	RUSSELL	50	07/17/95	0.42	N+P	2.17	20 U	4.0 U
1	RUSSELL	51	07/17/95	0.41	N+P	2.11	20 U	4.0 U
1	RUSSELL	52	07/17/95	0.39	N+P	1.88	20 U	4.0 U
2	JOCASSEE	53	07/05/96	1.67	N	2.59	6 U	4.0 U
2	JOCASSEE	55	07/05/96	1.84	N	1.89	6 U	4.0 U
2	JOCASSEE	56	07/05/96	1.95	N	1.35	6 U	4.0 U
2	BURTON	57	07/05/96	1.62	N	1.60	6	4.0 U
2	KEOWEE	59A	07/03/96	4.46	N	0.91	6 U	4.0 U
2	KEOWEE	59B	07/03/96	5.08	N	0.93	6	4.0 U
2	KEOWEE	60	07/03/96	NA		0.80	6 U	4.0 U
2	KEOWEE	62	07/03/96	2.51	N	1.04	6 U	4.0 U
2	KEOWEE	65	07/03/96	1.81	N	0.58	6 U	4.0 U
2	KEOWEE	66	07/03/96	1.11	N	0.77	6	4.0 U
2	KEOWEE	67	07/03/96	1.25	N	0.77	6	4.0 U
2	HARTWELL	70	07/02/96	1.33	N	1.28	7	4.0 U
2	HARTWELL	73	07/02/96	2.27	N	2.32	7	4.0 U
2	HARTWELL	74	07/02/96	1.62	N	1.54	6	4.0 U
2	HARTWELL	75	07/02/96	1.51	N	1.31	8	4.0 U
2	HARTWELL	77	07/02/96	1.41	N	1.85	6	4.0 U
2	HARTWELL	78	07/02/96	2.02	N	1.72	6	4.0 U
2	HARTWELL	79A	07/02/96	1.82	N	0.83	6	4.0 U
2	HARTWELL	79B	07/02/96	1.51	N	1.55	7	4.0 U
2	HARTWELL	80	07/02/96	1.87	N	0.77	8	4.0 U
2	HARTWELL	81	07/02/96	1.51	N	1.21	6	4.0 U
2	HARTWELL	84	07/01/96	1.44	N	1.17	6	4.0 U
2	THURMOND	87	06/28/96	8.35	N	0.98	11	4.0 U
2	THURMOND	88	06/26/96	3.47	N+P	6.95	28	4.0 U
2	THURMOND	89	06/26/96	1.60	N	3.19	19	4.0 U
2	THURMOND	93	06/25/96	1.97	P	1.87	7	4.0 U
2	THURMOND	95	06/25/96	1.83	P	1.42	8	4.0 U
2	THURMOND	96	06/24/96	2.82	P	2.56	10	4.0 U
2	THURMOND	97A	06/25/96	2.96	N+P	0	12	4.0 U
2	THURMOND	97B	06/25/96	1.60	P	1.62	7	4.0 U
2	THURMOND	98	06/25/96	3.49	N	3.90	23	72
2	THURMOND	99	06/24/96	1.76	P	2.20	9	4.0 U
2	THURMOND	100	06/24/96	2.90	N	3.39	9	4.0 U
2	THURMOND	101A	06/26/96	1.81	N	2.12	14	4.0 U
2	THURMOND	101B	06/26/96	2.07	P	1.80	6	4.0 U
2	THURMOND	103	06/24/96	1.80	N+P	2.47	9	4.0 U
2	RUSSELL	105A	07/01/96	1.99	N	1.2 U	9	4.0 U
2	RUSSELL	105B	07/01/96	1.71	N	1.85	10	4.0 U
2	RUSSELL	109	07/01/96	2.01	N	3.05	11	4.0 U
2	RUSSELL	110	07/01/96	1.90	N	2.38	8	4.0 U

SAVANNAH RIVERBASIN LAKE DATA

CYCLE	LAKE	STATION	DATE	AGPT (MG/L)	LIMITING NUTRIENT	CHL a (UG/L)	TPHOS (UG/L)	TSS (MG/L)
3	JOCASSEE	1	07/08/97	0.66	N+P	1.8	4	1.2
3	KEOWEE	4	07/09/97	0.87	N	1.4	11	5.5
3	KEOWEE	5	07/09/97	0.72	N	1.6	4	3.8
3	KEOWEE	7	07/09/97	0.88	N+P	0.93	3 U	0.8
3	KEOWEE	8A	07/09/97	0.93	N	1.5	4	0.9
3	KEOWEE	8B	07/09/97	0.79	N	0.90	4	2.0
3	KEOWEE	8C	07/09/97	0.49	N+P	0.5 U	3 U	0.0
3	KEOWEE	9	07/09/97	1.22	N	1.3	4	0.7
3	HARTWELL	10	07/09/97	1.25	N	1.6	8	2.1
3	HARTWELL	11	07/09/97	0.33	N+P	2.1	11	1.8
3	HARTWELL	13	07/10/97	1.18	N+P	1.6	8	2.4
3	HARTWELL	14	07/10/97	1.35	N+P	1.8	4	1.6
3	HARTWELL	15	07/10/97	1.73	N	1.6	4	1.6
3	HARTWELL	16A	07/10/97	1.41	N+P	1.6	5	2.2
3	HARTWELL	16B	07/10/97	1.35	N+P	1.6	4	1.7
3	HARTWELL	16C	07/10/97	0.42	N+P	0.5 U	3 U	0.1
3	HARTWELL	17	07/10/97	1.13	N+P	1.1	4	1.6
3	HARTWELL	18	07/10/97	2.27	N+P	5.4	18	4.6
3	HARTWELL	19	07/10/97	1.40	N+P	1.6	4	1.9
3	HARTWELL	20	07/10/97	1.37	N+P	1.4	4	1.5
3	HARTWELL	22	07/10/97	1.30	N+P	1.2	4	1.0
3	HARTWELL	24	07/09/97	0.93	N+P	0.99	4	1.0
3	HARTWELL	25	07/09/97	1.16	N	1.1	4	1.2
3	HARTWELL	26	07/09/97	0.98	N+P	0.94	4	0.9
3	HARTWELL	27	07/09/97	1.12	N+P	2.3	12	2.5
3	THURMOND	28	07/07/97	1.16	N	3.7	8	2.6
3	THURMOND	29	07/07/97	0.84	N+P	2.6	5	2.3
3	THURMOND	30	07/07/97	0.91	N	3.5	7	2.8
3	THURMOND	32	07/07/97	0.84	N+P	3.2	6	2.5
3	THURMOND	34	07/07/97	1.38	N	1.4	6	1.9
3	THURMOND	35	07/07/97	1.23	P	1.6	3	1.5
3	THURMOND	36	07/07/97	1.21	N+P	1.3	8	1.2
3	THURMOND	38	07/07/97	1.09	P	1.3	5	0.7
3	THURMOND	39	07/07/97	0.95	P	1.5	8	1.9
3	THURMOND	40	07/07/97	0.99	N+P	1.9	10	2.9 U
3	THURMOND	41A	07/07/97	1.20	N	7.8	29	4.7
3	THURMOND	41B	07/07/97	1.13	N	8.3	27	3.6
3	THURMOND	41C	07/07/97	0.40	N+P	0.5 U	3 U	0.8
3	THURMOND	42	07/07/97	1.48	N	5.4	12	2.6
3	RUSSELL	43A	07/09/97	1.08	N+P	2.6	12	1.8
3	RUSSELL	43B	07/09/97	1.17	N+P	3.0	12	2.0
3	RUSSELL	43C	07/09/97	0.44	N+P	0.5 U	3	0.0
3	RUSSELL	44	07/09/97	1.29	N	3.4	10	1.7

SAVANNAH RIVERBASIN LAKE DATA

CYCLE	LAKE	STATION	DATE	AGPT (MG/L)	LIMITING NUTRIENT	CHL a (UG/L)	TPHOS (UG/L)	TSS (MG/L)
3	RUSSELL	45	07/09/97	1.33	N	1.6	10	1.9
3	RUSSELL	46	07/09/97	1.32	N	1.6	9	2.1
3	RUSSELL	48	07/09/97	1.32	N+P	3.3	7	1.4
3	RUSSELL	49	07/09/97	1.19	N	2.5	12	2.1
3	RUSSELL	50	07/09/97	1.31	N+P	1.6	15	3.8
3	RUSSELL	51	07/09/97	1.23	N+P	2.1	12	2.2
3	RUSSELL	52	07/09/97	1.26	N	1.1	3	2.8

U = Material was analyzed for but not detected. The number is the minimum quantitation limit.

Chlorophyll a

Cycle 1 - USDA (HPLC)
Cycle 2 - USDA (HPLC)
Cycle 3 - EPA (HPLC)

Appendix C

Fish Protocol & Stream Data

Rapid Bioassessment Protocol V Metric Development

The EPA RBP V (Barbour et. al., 1998)is based primarily on the Index of Biotic Integrity (IBI)(Karr,1981; Fausch et al. 1984;Karr et al. 1986). The IBI incorporates up to twelve metrics which are scored to assess changes in the fish community compared to a reference stream, or a stream with minimal impact, of similar size and geographic area to that of the stream being sampled. Like stream insect communities, fish communities will respond to environmental change.

The EPA RBP V's (Barbour et. al., 1998) twelve metrics were originally developed for Midwestern streams. They are not intended to be used verbatim in other geographical areas. The metrics presented in the RBP document are prototypes to be used as guidance for developing metrics in other geographical areas. Barbour et. al. (1998) also present modifications to the IBI that other researches have made to make the IBI more applicable to their regions or study area. Metric development is based on reference fish community or a fish community with minimal impact.

After evaluating the fish data collected over the four year study, it was determined that selecting metrics that assessed the basic fish community structure was the most effective way to screen, or evaluate, streams in such a large geographical area, especially due to the nature of the study design. Ideally, when conducting any IBI study, metric development is based on a reference fish community in the area of study. As part of the study design a reference fish community would be established. However, the approach used in this study did not focus on establishing reference fish community data. This makes it difficult to develop metrics, with confidence, that are more discriminating of the subtle differences within a large watershed. The results of this IBI analysis should be used to identify problem areas in the Basin, at which point an IBI study, which involves the aqusition of the necessary reference data, can be implemented to address the problems areas.

Nine metrics were utilized to evaluate the data to assess the condition of stream fish assemblages (table 1). The metrics were selected from a pool of metrics listed in the EPA RBP document and other studies that have been conducted in Georgia (DeVivo 1996).

The first seven metrics assess the fish assemblage structure and the last two assess the fish assemblage function. The assemblage structure metrics will all decrease with increased stream degradation. Combined, these metrics assess impacts to the stream from physical and chemical degradation. Of the two assemblage function metrics, proportion of omnivores will increase with increased stream degradation, and proportion of benthic Invertivores will decrease with increased stream degradation.

No metrics that assess fish abundance and condition were utilized for this study. Metrics 11 and 12 listed in the RBP document (Barbour et. al.1998), "Proportion of disease/anomalies" and "Proportion of Hybrids" requires a certain level of training to properly assess these metrics. The skill level among the sampling crews varied. Therefore, consistent assessment of these metrics was not possible. The abundance metric was not incorporated because of too much variance in the data. It was too difficult to determine any patterns or trends and establish scoring criteria for the metric.

The scoring criteria for each of the nine metrics were based on the data collected from the 82 lower Piedmont streams. Reliable reference data was unavailable for this study due to the nature in which sampling locations were selected. Therefore, an alternative method for developing scoring criteria was utilized. The range of metric results were trisected to produce three different ranges of results. A good result is given a score of 5, a medium range result is given a score of 3 and a lower range result is given a score of 1. This is considered an acceptable method for developing scoring criteria for IBI metrics (Karr, 1996). Metric scoring criteria are presented in Table 2.

Table 1 List of IBI metrics utilized for the RBP analysis.

Index of Biotic Integrity Metrics	
	Fish Assemblage Structure Metrics
1	Number of Species
2	Proportion of Non-Native Species
3	Brillioun Diversity Index
4	Number of Native Suckers
5	Number of Native Sunfishes
6	Number of Minnow Species
7	Number of Darter Species
	Fish Assemblage Function Metrics
8	Proportion of Omnivores
9	Benthic Invertivores

Initially, scoring criteria were developed regardless of stream order. That is, the same criteria was applied to all three stream orders (1st, 2nd , and 3rd) that were sampled. Pearson Product Moment Correlation were calculated for all stream sampling parameters, which also included stream order and IBI score. A positive correlation was indicated for stream order and IBI score. This indicated that the IBI score increased with stream order designation, suggesting that the scoring criteria favored third order streams and that the other streams were scored unfairly. To resolve this issue the metric scores were recalibrated based on stream order. Separate scoring criteria were developed for each stream order, so that each stream order had its own set of scoring criteria for each metric. After the metrics were recalibrated to compensate for differences in stream order, Pearson Product Moment Correlation were recalculated and the results indicated there was no significant correlation between stream order and IBI score.

After all metrics were calculated and scored for a particular stream station, the metric scores were summed to give one final IBI score for that particular stream station. The condition of the fish community is then usually characterized as either "Good", "Fair", or "Poor", depending on how far the total score deviates from the total possible score. These characterizations were developed by applying box and wisker plots to the range of final scores. Scores that were in the upper 25 percentile (>29) were classified as being in Good condition. Scores that fell between the 25 and 75 percentile (22-29) were classified as being in Fair condition, and scores that were in the lower 25 percentile (<22) were classified as being Poor

Table 2 Scoring criteria for the metrics utilized for the RBP V (IBI) analysis.

Community Structure Metrics		Stream Order	Metric Score Criteria 1	3	5
1	Number of Species	1	<5	5-7	>7
		2	<7	7-13	>13
		3	<9	9-14	>14
2	Proprotion of Non-Native Species	1	>27	14-26	<14
		2	>16	8-16	<8
		3	>20	10-20	<10
3	Brillioun Diversity Index	1	<0.303	0.303-0.523	>0.523
		2	<0.34	0.34-0.60	>0.60
		3	<0.70	0.82-0.70	>0.82
4	Number of Native Suckers	1	0	0	1
		2	<2	2	>2
		3	<2	2	>2
5	Number Native Sunfish	1	<3	3-4	>5
		2	<3	3-4	>4
		3	<2	2-3	>3
6	Number of Minnow Species	1	<2	3-4	>4
		2	<3	3-4	>4
		3	<4	4-5	>5
7	Number of Darter Species	1	<1	1	>1
		2	<2	2	>2
		3	<1	1-2	>2
Community Function Metrics					
8	Proportion of Generalized Feeders	1	>62	31-62	<31
		2	>70	41-70	<41
		3	>43	22-43	<22

condition. An example of how a stream station is scored and

Table 3 Example of scoring IBI metric results.

Index of Biotic Integrity Metrics	Stream Station					
	93		122		37	
	Result	Score	Result	Score	Result	Score
1 Number of Species	19	5	2	1	13	3
2 Proportion of Non-Native Species	0.39	5	0.0	5	2.94	5
3 Brillioun Diversity Index	0.88	5	0.09	1	0.77	3
4 Number of Native Suckers	3	5	1	1	1	1
5 Number of Native Sunfishes	4	5	0	1	2	3
6 Number of Minnow Species	4	3	2	1	5	3
7 Number of Darter Species	4	5	0	1	3	5
8 Proportion of Omnivores	18.99	5	100	1	31.62	3
9 Benthic Invertivores	34.11	5	0.00	1	11.76	1
Total IBI Score		43		13		27
Classification		Good		Poor		Fair

categorized is presented in Table 3.

References

Barbour, M. T., J. Gerritsen, B. D. Snyder, and J. B. Stribling. 1998. Draft Revision to Rapid Bioassessment Protocols for Use in Streams and Rivers. EPA 841-D-97-002, Office of Water, Washington, D. C.

DeVivo, J. C., C.A Couch, B.J. Freeman. 1996. Preliminary Index of Biotic Integrity in Urban Streams Around Atlanta, Georgia. Proceedings of the 1997 Georgia Water Resources Conference, held at The University of Georgia.

Fausch, K.D., J.R. Karr, and P.R. Yant. 1984. Regional Application of an Index of Biotic Integrity based on Stream Fish Communities. Transactions of the American Fisheries Society 113:39-55.

Karr, J.R. 1981. Assessment of Biotic Integrity Using Fish Communities. Fisheries 6:21-27

Karr, J.R. 1986, K.D. Fausch, P.L. Angermeier, P.R. Yant, and I.J. Schlosser. Assessment of Biological Integrity in Running Water: a Method and its Rationale. Illinois Natural History Survey Special Publication No. 5. Champaign Illinois. 28 pages.

Karr, J.R. 1996. Rivers as Sentinels: Using the Biology of Rivers to Guide Landscape Management in R.J. Naiman amd R.E. Bilby (eds), Ecology and Management of Streams and Rivers in the Pacific Northwest Coastal Ecoregion. Springer-Verlag, New York, New York.

Summary of all Savannah REMAP Stream Data

Station ID	CYCLE	YEAR	Stream Order	Eco Region	Latitude (DMS)			Longitude (DMS)			State	AGPT	MI_HAB	MI RICH	MI_EPT	FISH_IBI	PH	TEMP (C)	DO (mg/l)	COND (uS/cm)
4	0	94	1	LP	34	36	53.7993	82	41	37.7301	SC	34.08	76	11	3		6.9	21.3	7.9	45.3
8	0	94	3	LP	34	28	48.5681	83	2	40.6523	Ga	28.19	64	11	3	19	6.4	17.5	0.0	•52.0
9	0	94	2	LP	34	28	47.0579	83	3	14.2553	Ga	36.17	94	12	4	25	6.5	23.0	0.0	60.0
10	0	94	3	LP	34	25	49.8077	83	2	28.9258	Ga	33.82	80	11	4	21	6.4	21.5	7.2	53.2
11	0	94	3	LP	34	27	22.0911	83	21	41.3815	Ga	11.84	47	12	3	25	6.8	21.0	0.0	55.0
12	0	94	3	LP	34	25	3.2074	83	2	40.6590	Ga	40.28	83	16	5	27	6.0	22.0	0.0	50.0
13	0	94	1	LP	34	26	46.6378	83	20	19.0241	Ga	18.63	61	9	3	33	6.5	19.5	0.0	50.0
14	0	94	3	LP	34	28	42.6550	83	21	27.4603	Ga	10.12	45	10	3	29	6.4	21.0	0.0	55.0
15	0	94	1	LP	34	19	42.0003	83	2	1.1208	Ga	41.33	87	7	2	23	5.4	21.0	0.0	45.0
19	0	94	3	LP	34	14	48.3613	82	56	14.6220	Ga	30.48	91	13	5	21	5.8	23.0	0.0	35.0
21	0	94	3	LP	34	14	28.5278	82	55	12.7644	Ga	38.43	67	14	5		6.9	21.0	7.9	33.8
22	0	94	2	LP	34	14	38.4142	82	56	46.4356	Ga	33.87	93	15	5	23	5.5	24.0	0.0	35.0
27	0	94	3	LP	34	2	9.1629	82	3	55.5302	SC	8.28	64	11	3	17	7.0	22.3	4.7	211.0
28	0	94	2	LP	33	58	4.9672	81	54	45.5302	SC	3.12	92	12	2	27	6.5	22.7	5.1	117.2
29	0	94	2	LP	33	57	2.2850	81	53	58.6145	SC	1.24	119	24	7	37	7.2	22.8	7.1	98.7
30	0	94	3	LP	33	58	35.4090	81	54	24.9065	SC	8.62	116	16	5	33	7.2	22.6	8.0	104.0
31	0	94	2	LP	34	2	42.8822	82	50	58.5292	Ga	8.16	87				7.0	20.9	8.1	78.0
32	0	94	3	LP	33	56	2.0761	81	54	52.4463	SC	4.23	112	14	5	35	7.4	22.7	7.9	101.6
33	0	94	1	LP	33	56	36.8479	82	14	21.7438	SC	8.42	75	15	4	21	7.2	22.3	7.2	243.0
34	0	94	3	LP	33	54	54.1530	82	1	9.6508	SC	10.18	113	20	6	39	7.0	22.4	5.9	93.2
37	0	94	3	LP	33	53	51.7284	82	1	9.4029	SC	23.09	111	21	7	27	7.1	22.9	7.5	91.4
38	0	94	3	LP	33	55	38.8549	82	52	49.6650	Ga	4.13	88	12	5	23	6.8	23.5	7.5	76.5
39	0	94	3	LP	33	55	42.2439	82	53	46.5835	Ga		82	16	2	31	6.7	23.2	6.1	78.7
41	0	94	2	LP	33	50	6.3653	82	39	2.7338	Ga	3.25	48	16	4	35	6.7	25.3	6.2	110.4
44	0	94	1	LP	33	43	58.8046	81	54	24.7820	SC	3.46	75	9	3		6.9	23.4	7.6	67.1
45	0	94	2	LP	33	41	49.2998	82	43	32.5372	Ga	59.51	45	7	1		7.5	27.4	5.9	3260.0
48	0	94	3	SH	33	15	21.9660	82	20	1.5107	Ga	30.38	108	8	3		6.7	24.8	6.9	914.0
49	0	94	3	SH	33	15	19.3664	82	19	59.9691	Ga	28.45	92	9	4		6.8	25.6	7.5	677.0
51	0	94	1	SP	33	0	5.8158	81	24	2.8706	SC	1.54	120	11	2		7.4	22.7	7.1	36.1
57	0	94	3	SP	32	56	4.1333	81	48	1.8028	Ga	8.74	111	17	3		7.5	21.2	6.4	184.0
61	1	95	2	BR	34	55	43.5823	83	7	50.1026	SC	1.41	99	23	14		6.9	16.6	8.9	16.0
64	1	95	2	LP	34	47	48.7826	82	37	51.8720	SC	1.71	45	15	7	19	6.8	20.0	8.1	
65	1	95	2	LP	34	47	3.6141	82	36	22.0255	SC	3.33	41	11	5	17	6.5	22.0	7.4	
68	1	95	2	LP REF	34	38	17.2747	83	19	55.0356	Ga	2.46	104	30	15	31	7.2	21.4	8.3	35.0
69	1	95	1	LP	34	37	42.2113	83	18	29.6489	Ga	1.40	84	23	12	19	7.4	21.2	8.4	37.0
71	1	95	1	LP	34	30	41.9353	82	27	58.4868	SC	17.96	76	19	8	19	6.4	22.5	7.6	
72	1	95	1	LP	34	30	57.1580	83	28	28.7501	Ga	3.79	82	27	18	21	7.3	20.4	8.4	36.0
74	1	95	2	LP	34	20	37.3651	82	39	48.5095	SC	2.06	41	20	9	35	6.8	21.1	7.8	50.6
75	1	95	2	LP	34	19	58.7644	82	38	39.8460	SC	11.23	54	23	9	31	6.9	21.4	7.7	49.6
77	1	95	1	LP REF	34	20	58.6465	83	2	3.1563	Ga	2.57	105	25	10	21	6.9	20.4	7.8	35.0
78	1	95	1	LP	34	16	45.7984	82	17	5.8484	SC	1.98	91	11	4	29	7.4	23.8	7.9	98.0

Summary of all Savannah REMAP Stream Data

Station ID	CYCLE	YEAR	Stream Order	Eco Region	Latitude (DMS)			Longitude (DMS)			State	AGPT	MI_HAB	MI RICH	MI_EPT	FISH_IBI	PH	TEMP (C)	DO (mg/l)	COND (uS/cm)
79	1	95	3	LP	34	8	27.1407	82	17	14.9103	SC		53	16	7	23	7.6		7.9	125.0
80	1	95	3	LP	34	8	11.3255	82	17	51.3748	SC	13.12	52	22	8	21	7.4	22.0	7.9	123.0
81	1	95	3	LP	34	7	43.1356	82	18	12.2656	SC	8.26	52	23	8	25	7.3	23.0	7.6	125.0
82	1	95	3	LP	34	6	6.7529	82	28	39.8788	SC	3.96	107	28	8	25	7.3	25.0	7.6	101.0
83	1	95	1	LP	34	12	15.4863	83	25	26.0343	Ga	0.55	66	23	11	19	6.7	20.3	7.7	44.0
85	1	95	1	LP	34	8	43.1670	82	57	23.1062	Ga	4.80	45	23	5	33	6.8	22.3	5.4	52.0
86	1	95	3	LP	34	5	6.1063	82	28	36.4788	SC	4.58	72	25	9	25	7.8	26.7	8.5	102.7
87	1	95	2	LP	34	8	18.6496	82	57	24.4861	Ga	7.03	30	19	7	21	6.7	22.2	5.9	51.0
88	1	95	3	LP	34	4	32.8378	82	28	27.1207	SC	5.97	68	27	11	27	7.3	22.8	6.5	106.0
89	1	95	2	LP	34	4	38.9293	82	30	16.6960	SC	4.55		18	4	27	7.2	25.1	4.8	150.0
93	1	95	3	LP REF	33	48	16.0738	82	7	57.4444	SC		104	26	10	43	7.6	26.0	8.2	146.9
94	1	95	3	LP REF	33	48	7.1642	82	8	13.9907	SC		59	17	8	39				
95	1	95	3	LP	33	47	59.2036	82	7	23.9716	SC	1.34	99	32	10	39	9.1	28.2	11.3	146.5
96	1	95	2	LP	33	46	51.1624	82	8	34.7175	SC	2.06	103	22	10	23	7.4	21.5	7.2	
97	1	95	1	LP	33	52	1.2324	83	9	52.4866	Ga	1.46	40	21	11		6.8	21.7	6.3	62.0
98	1	95	1	LP	33	37	11.4077	82	22	33.5640	Ga		69	30	12	25	6.4	22.2	3.6	76.0
99	1	95	2	LP	33	35	15.4617	82	12	46.2592	Ga	8.57	72	23	8	31	7.1	25.2	5.2	105.0
100	1	95	3	LP	33	35	31.1643	82	41	13.5212	Ga	4.97	49	21	8	29	7.2	26.4	6.2	107.0
101	1	95	3	LP	33	35	26.7033	82	41	51.5983	Ga	11.06	48	20	6	27	7.2	26.6	6.2	104.0
102	1	95	3	LP	33	35	7.9421	82	42	12.1633	Ga	0.82	49	15	6	27	7.3	26.3	5.9	103.0
103	1	95	1	LP	33	34	54.0984	82	40	22.8367	Ga	1.34	89	22	8	35	7.6	24.4	7.1	187.0
104	1	95	1	LP	33	32	41.1923	82	39	50.4017	Ga	3.07	70	12	6	37	7.0	22.7	7.9	91.1
113	1	95	2	SH	33	15	55.8165	81	57	21.6129	Ga	7.37	104	26	10		6.9	24.4	6.7	32.0
114	1	95	1	SP	33	7	22.9653	81	51	2.9919	Ga	6.66	78	14	6		6.5	20.9	6.3	37.0
121	2	96	2	BR	34	49	59.4444	83	35	29.5255	Ga		69	22	14		6.9	16.5	7.9	24.0
122	2	96	2	LP	34	32	45.2630	83	16	11.5174	Ga		53	16	5	13	7.2	24.2	8.2	691.0
123	2	96	2	LP	34	10	27.1961	83	17	2.2760	Ga		86	21	6	27	6.8	22.3	7.3	52.0
127	2	96	1	UP	34	49	25.9638	82	58	42.8228	SC		112	32	16		7.0	22.0	8.2	40.0
130	2	96	1	LP	34	32	36.8223	82	58	0.7518	SC		54	16	1	25	6.8	26.0	7.2	50.0
131	2	96	2	LP	34	32	45.4489	82	57	22.3898	SC		62	22	8	33	6.8	24.5	8.6	40.0
132	2	96	2	LP	34	31	28.9133	82	57	22.5618	SC		48	19	8	25	7.1	26.5	7.5	45.0
133	2	96	3	LP	34	31	46.5814	82	57	5.6190	SC		62	21	7	23	6.8	26.0	8.0	50.0
135	2	96	2	LP	33	46	25.7746	82	58	59.2704	Ga		75	22	7	31	6.9	24.1	5.4	142.0
136	2	96	1	UP	34	54	4.4831	82	46	53.2661	SC		102	19	11		6.3	19.5	8.5	20.0
138	2	96	2	UP	34	55	12.6118	82	45	38.6099	SC		82	26	16		7.0	22.0	8.2	30.0
143	2	96	3	LP	34	39	20.7134	82	38	31.9222	SC		93	22	10	21	7.1	24.0	8.0	50.0
144	2	96	3	LP	34	38	22.9730	82	38	37.6329	SC		51	24	10	19	7.1	25.0	7.5	50.0
145	2	96	3	LP	34	38	50.4543	82	38	25.1805	SC		47	26	11	19	6.9	22.0	8.0	45.0
147	2	96	3	LP	34	39	40.2584	82	37	47.9465	SC		48	19	10	21	7.1	25.5	8.1	45.0
148	2	96	1	LP	33	37	37.8644	82	47	12.3996	Ga		55	24	9	25	7.6	23.2	7.2	42.0
149	2	96	2	LP	34	26	55.3141	82	37	39.0583	SC		52	15	6	19	6.7	22.5	7.7	120.0

Summary of all Savannah REMAP Stream Data

Station ID	CYCLE	YEAR	Stream Order	Eco Region	Latitude (DMS)			Longitude (DMS)			State	AGPT	MI_HAB	MI RICH	MI_EPT	FISH_IBI	PH	TEMP (C)	DO (mg/l)	COND (uS/cm)
151	2	96	2	LP	34	25	50.3552	82	36	54.0540	SC		51	16	5	23	6.8	23.0	7.5	130.0
154	2	96	2	LP	33	48	13.7570	82	29	42.0120	Ga		68	27	11	27	6.9	23.6	7.0	82.0
155	2	96	2	LP	A	48	7.6710	82	29	41.8078	Ga		71	32	10	29	6.9	23.2	6.8	77.0
155.1	2	96	3	LP	B	49	17.0757	82	29	26.3316	Ga		72	20	8	27	7.1	24.7	7.4	112.0
162	2	96	1	LP	33	35	49.7966	81	55	35.6942	SC		95	16	9	15	5.1	20.0	7.7	15.0
163	2	96	2	SP REF	33	3	1.0792	81	54	53.5055	Ga		102	26	7		7.3	24.0	6.6	128.0
164	2	96	2	SP REF	33	4	3.3065	81	54	42.1236	Ga		98	24	7		7.3	23.3	7.1	119.0
166	2	96	2	SP REF	33	2	57.2257	81	54	21.7342	Ga		104	27	11		6.8	22.2	8.3	86.0
167	2	96	3	SP	33	3	40.2696	81	53	41.3649	Ga		99	22	7		7.2	24.6	8.2	100.0
176	2	96	2	MACP	32	35	19.1747	81	26	41.7662	Ga		99	14	1		6.0	25.8	6.9	60.0
177	2	96	1	MACP	32	31	57.0026	81	27	18.6763	Ga		96	14	1		4.1	25.6	5.1	58.0
186	3	97	2	BR	35	0	31.3210	82	49	14.8715	SC		123	25	16		7.0	23.7	8.0	24.0
187	3	97	3	BR	35	0	12.3063	82	49	37.5836	SC		91	36	20		7.4	22.3	8.5	29.0
189	3	97	3	BR	34	59	46.4191	82	49	40.7735	SC		87	35	19		7.6	22.4	8.5	27.0
191	3	97	2	BR	34	52	3.5187	83	8	10.8764	SC		92	32	20		7.4	19.3	8.7	24.0
192	3	97	2	BR	34	51	41.4279	83	9	58.7073	SC		91	29	17		6.8	19.4	8.7	26.0
193	3	97	1	BR	34	51	15.5872	83	9	0.6364	SC		60	13	5		6.9	18.9	9.5	32.0
194	3	97	2	BR	34	51	6.1637	83	8	40.5458	SC		82	31	17		7.1	19.8	9.3	24.0
195	3	97	3	BR	34	50	36.4122	83	9	12.9877	SC		110	30	18		6.6	21.9	8.7	24.0
196	3	97	1	BR	34	46	55.1103	83	14	10.5392	SC		58	20	8		6.9	19.5	8.7	21.0
197	3	97	1	LP	34	22	46.7383	82	54	2.9064	Ga		52	16	6	21	6.2	21.9	7.2	29.0
200	3	97	2	LP	34	17	23.0683	82	43	33.0550	SC		58	24	11	19	7.0	24.0	7.3	65.0
205	3	97	2	LP	34	12	17.6608	82	50	58.5352	Ga		63	17	5	23	6.7	21.5	8.2	32.0
207	3	97	1	LP	34	14	14.2111	83	22	43.9620	Ga			19	6	33				
210	3	97	3	LP	34	9	47.0635	83	3	59.4713	Ga		103	18	9	23	7.2	22.7	9.0	40.0
211	3	97	3	LP	34	9	23.5813	83	6	5.3189	Ga			19	7	25				
213	3	97	1	LP	34	8	23.9605	83	5	18.1675	Ga		57	24	12	21	6.9	21.0	7.9	541.0
214	3	97	1	LP	34	7	46.7219	83	17	19.2083	Ga		74	15	7	27	6.5	19.8	8.0	379.0
216	3	97	2	LP	34	0	42.8311	82	22	52.1874	SC		46	21	9	17	6.8	22.8	8.1	111.0
221	3	97	2	LP	33	42	0.1361	82	2	45.8011	SC		67	21	12	23	7.0	25.3	7.5	80.0
222	3	97	2	LP	33	41	22.6965	82	0	43.3443	SC		66	18	8	25	6.9	26.4	6.4	121.0
224	3	97	3	LP	33	40	40.1499	82	36	7.4409	Ga			24	11	35				
231	3	97	3	SH REF	33	26	43.6063	81	36	4.6469	SC		92	24	11		5.2	20.9	7.9	19.0
232	3	97	3	SH	33	25	57.1961	81	38	17.6399	SC		95	21	11		5.3	21.2	7.4	18.0
236	3	97	2	SP	33	5	35.6071	81	31	9.3483	SC		73	19	8		6.1	25.4	6.4	38.0
237	3	97	2	SP	33	4	44.9632	81	30	4.5215	SC		88	21	9		6.4	25.5	7.3	47.0
238	3	97	1	SP	33	6	15.7579	81	47	38.8681	Ga			16	7					

BR*= Bluridge, *LP*= Lower Piedmont, *UP*= Upper Piedmont, *SH*= Sand Hills, *SP*= Southeastern Plains, *MACP*= Midatlantic Coastal Plains**, ***REF**= Reference**

ORGANISMS	STATIONS 4	8	9	10	11	12	13	14	15	19	21	22	27	28	29	30	32	33	34	37
DIPTERA																				
Tabanidae											X				X	X		X		
Empididae																				
Tipulidae																				X
Tanyderidae																				
Rhagionidae																				
Culicidae															X					
Ptychopteridae																				
Dixidae																				
Ceratopogonidae															X					
Chironomidae	X	X	X	X	X	X	X	X	X	X	X	X		X	X			X	X	X
Blephariceridae																				
Simuliidae	X		X	X	X		X	X				X				X		X	X	X
TRICHOPTERA																				
Hydroptilidae																				
Helicopsychidae																				
Hydropsychidae			X	X		X	X	X				X			X	X	X	X	X	X
Rhyacophilidae																				
Limnephilidae										X										
Phryganeidae																				
Polycentropidae			X			X				X								X		X
Psychomiidae																				
Philopotamidae															X		X		X	X
Dipseudopsidae																				
Brachycentridae																				
Goeridae																				
Lepidostomatidae																				
Calamoceratidae																				
Uenoidae																				
Molannidae																				
Odontoceridae																				
Leptoceridae																				
Glossosomatidae																				
PLECOPTERA																				
Perlidae	X									X	X				X				X	
Perlodidae															X	X				X
Peltoperlidae																				
Capniidae																				
Leuctridae																				

ORGANISMS	STATIONS 4	8	9	10	11	12	13	14	15	19	21	22	27	28	29	30	32	33	34	37
Nemouridae																				
Pteronarcyidae																				
EPHEMEROPTERA																				
Baetidae	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X
Tricorythidae																				
Heptagenidae	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Oligoneuriidae							X				X		X		X	X	X		X	X
Leptophlebiidae					X															
Caenidae																				
Neoephemeridae																				
Ephemeridae		X				X					X	X								
Ephemerellidae				X								X								
ODONATA																				
Libellulidae																				
Corduliidae																				
Cordulegasteridae																				
Gomphidae	X	X	X	X	X	X		X	X	X	X	X	X		X		X	X	X	
Aeschnidae		X	X								X	X								
Macromiidae		X		X	X	X						X	X	X		X			X	
Calopterygidae					X		X		X	X	X	X	X	X			X	X	X	
Coenagrionidae								X						X	X					
MEGALOPTERA																				
Corydalidae				X	X	X	X		X	X		X		X	X	X	X	X	X	X
Sialidae	X					X			X	X			X		X					
NEUROPTERA																				
Sisyridae																				
HEMIPTERA																				
Corixidae																				
Veliidae																				
Belostomatidae															X					
Nepidae																X				
Gerridae	X	X	X			X		X		X	X			X	X	X	X	X		X
COLEOPTERA																				
Elmidae	X	X	X	X	X	X	X	X		X	X	X	X		X	X	X	X	X	X
Hydrophilidae		X		X	X	X												X	X	X
Gyrinidae	X					X							X			X				X
Dytiscidae											X									
Noteridae																				
Dryopidae																		X		

	STATIONS																			
ORGANISMS	4	8	9	10	11	12	13	14	15	19	21	22	27	28	29	30	32	33	34	37
Psephenidae															X				X	
Helodidae														X		X	X			
Haliplidae																				
Eubriidae																				
Ptilodactylidae																				
CRUSTACEA																				
Astacidae	X	X	X		X	X	X	X		X	X	X	X	X	X		X	X	X	
Isopoda																				
Amphipoda														X	X				X	
Palaemonidae																				
OLIGOCHAETA																				
Glossosophoniidae																				
Naididae																				
Tubificidae												X					X			X
Lumbriculidae			X																X	
HIRUDINEA						X														X
HYDRACARINA														X	X	X			X	X
MOLLUSCA																				
Bivalvia																				
Gastropoda undet. sp.																				
Unionidae																X	X			X
Corbiculidae													X		X					X
Lymnaeidae																				
Sphaeriidae																			X	
Physidae																				
Viviparidae																				
Planorbidae															X					X
Pleuroceridae																				
Ancylidae																				
TAXA RICHNESS	11	11	12	11	12	16	9	10	7	13	14	15	11	12	24	16	14	15	20	21
EPT INDEX	3	3	4	4	3	5	3	3	2	5	5	5	3	2	7	5	5	4	6	7
HABITAT SCORE	76	64	94	80	47	83	61	45	87	91	67	93	64	92	119	116	112	75	113	111

ORGANISMS	38	39	41	44	45	48	49	51	57	61	64	65	68	69	71	72	74	75	77	78
DIPTERA																				
Tabanidae																				
Empididae																				
Tipulidae										X			X	X	X	X	X	X	X	
Tanyderidae																				
Rhagionidae																				
Culicidae																	X			
Ptychopteridae																				
Dixidae													X	X	X	X			X	
Ceratopogonidae			X							X										
Chironomidae	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Blephariceridae																				
Simuliidae	X	X	X			X		X		X	X	X	X	X	X	X		X	X	
TRICHOPTERA																				
Hydroptilidae																				
Helicopsychidae																				
Hydropsychidae		X	X		X					X	X	X	X	X	X	X	X	X	X	X
Rhyacophilidae										X			X	X		X				
Limnephilidae											X	X	X	X	X	X		X	X	
Phryganeidae																				
Polycentropidae										X			X	X		X				
Psychomiidae													X			X			X	
Philopotamidae										X			X	X				X		
Dipseuopsidae																				
Brachycentridae										X			X							
Goeridae																				
Lepidostomatidae						X	X							X	X	X			X	
Calamoceratidae																				
Uenoidae																				
Molannidae																				
Odontoceridae																			X	
Leptoceridae	X						X					X			X		X			
Glossosomatidae													X			X				
PLECOPTERA																				
Perlidae	X					X	X			X	X	X	X	X	X	X	X	X	X	
Perlodidae										X						X			X	
Peltoperlidae										X			X	X		X				
Capniidae										X						X				
Leuctridae																X				

ORGANISMS	38	39	41	44	45	48	49	51	57	61	64	65	68	69	71	72	74	75	77	78
Nemouridae																				
Pteronarcyidae										X	X		X			X		X		
EPHEMEROPTERA																				
Baetidae	X	X	X	X		X	X	X	X	X	X		X	X	X	X	X	X	X	X
Tricorythidae																	X			
Heptagenidae	X		X	X					X	X	X	X	X	X	X	X	X	X	X	X
Oligoneuriidae	X		X							X	X		X	X	X	X	X	X	X	X
Leptophlebiidae														X			X			
Caenidae								X												
Neoephemeridae																				
Ephemeridae																X		X		
Ephemerellidae										X			X			X	X			
ODONATA																				
Libellulidae										X										
Corduliidae											X									
Cordulegasteridae																				
Gomphidae		X		X	X	X	X		X	X	X	X	X	X	X	X		X	X	X
Aeschnidae								X	X		X	X	X	X	X		X	X	X	
Macromiidae																				
Calopterygidae		X	X	X	X			X		X									X	
Coenagrionidae	X	X			X	X		X				X					X			
MEGALOPTERA																				
Corydalidae	X	X			X						X		X	X		X		X		X
Sialidae		X															X	X	X	X
NEUROPTERA																				
Sisyridae																				
HEMIPTERA																				
Corixidae																				
Veliidae										X	X		X		X		X	X		X
Belostomatidae		X																		
Nepidae																				
Gerridae	X		X	X	X				X											
COLEOPTERA																				
Elmidae	X	X	X	X		X	X	X	X				X		X	X	X	X	X	X
Hydrophilidae			X				X		X				X				X	X		
Gyrinidae		X	X				X										X	X		
Dytiscidae				X																
Noteridae																				
Dryopidae													X	X	X			X	X	X

ORGANISMS	38	39	41	44	45	48	49	51	57	61	64	65	68	69	71	72	74	75	77	78
Psephenidae													X	X		X			X	
Helodidae																				
Haliplidae								X	X											
Eubriidae																				
Ptilodactylidae																			X	
CRUSTACEA																				
Astacidae		X	X					X	X	X				X	X	X	X	X	X	
Isopoda									X										X	
Amphipoda		X	X					X												
Palaemonidae		X							X											
OLIGOCHAETA													X						X	
Glossesophoniidae																				
Naididae																				
Tubificidae	X		X																	
Lumbriculidae											X	X			X					
HIRUDINEA			X						X											
HYDRACARINA																				
MOLLUSCA																				
Bivalvia																				
Gastropoda undet. sp.													X	X						
Unionidae		X							X											
Corbiculidae													X							
Lymnaeidae																				
Sphaeriidae																				
Physidae																		X		
Viviparidae																				
Planorbidae									X											
Pleuroceridae									X											
Ancylidae																				
TAXA RICHNESS	12	16	16	9	7	8	9	11	17	23	15	11	30	23	19	27	20	23	25	11
EPT INDEX	5	2	4	3	1	3	4	2	3	14	7	5	15	12	8	18	9	9	10	4
HABITAT SCORE	88	82	48	75	45	108	92	120	111	99	45	41	104	84	76	82	41	54	105	91

ORGANISMS	79	80	81	82	83	85	86	87	88	89	93	94	95	96	97	98	99	100	101	102
DIPTERA																				
Tabanidae						X										X				
Empididae																				
Tipulidae				X	X	X	X	X	X		X		X		X	X				
Tanyderidae																	X			
Rhagionidae																				
Culicidae			X	X					X	X			X							
Ptychopteridae															X					
Dixidae				X	X	X	X		X											
Ceratopogonidae																		X	X	
Chironomidae	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Blephariceridae																				
Simuliidae	X		X			X	X	X				X	X			X	X			
TRICHOPTERA																				
Hydroptilidae											X									
Helicopsychidae																				
Hydropsychidae	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X
Rhyacophilidae																				
Limnephilidae	X	X	X		X		X	X	X		X			X	X					
Phryganeidae											X									
Polycentropidae																				
Psychomiidae															X					
Philopotamidae						X					X	X	X	X		X	X	X		
Dipseuopsidae																				
Brachycentridae															X			X		
Goeridae																				
Lepidostomatidae					X										X					
Calamoceratidae																				
Uenoidae																				
Molannidae															X					
Odontoceridae															X	X				
Leptoceridae		X	X	X			X	X	X		X		X	X			X		X	X
Glossosomatidae					X															
PLECOPTERA																				
Perlidae		X	X		X		X	X	X			X	X	X	X	X		X	X	
Perlodidae						X		X							X	X				
Peltoperlidae					X													X		
Capniidae															X					
Leuctridae															X					

ORGANISMS	79	80	81	82	83	85	86	87	88	89	93	94	95	96	97	98	99	100	101	102
Nemouridae																				
Pteronarcyidae	X	X	X	X																
EPHEMEROPTERA																				
Baetidae	X	X	X	X	X		X		X	X	X	X	X	X		X	X	X	X	
Tricorythidae									X		X	X	X	X			X			X
Heptagenidae	X	X	X	X	X	X	X	X	X		X	X	X	X		X	X	X	X	X
Oligoneuriidae	X	X	X	X	X		X		X	X	X	X	X	X		X	X			
Leptophlebiidae																X				
Caenidae				X			X		X	X	X	X	X	X		X	X			
Neoephemeridae																				
Ephemeridae	X			X	X	X	X	X	X	X						X		X	X	X
Ephemerellidae					X				X				X			X				X
ODONATA																				
Libellulidae						X				X			X	X			X			
Corduliidae				X		X				X	X		X	X		X	X			
Cordulegasteridae					X											X				
Gomphidae	X	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X		X	
Aeschnidae	X	X	X	X	X	X	X	X	X		X		X	X	X	X	X	X		X
Macromiidae		X	X	X			X		X	X	X		X						X	
Calopterygidae		X	X		X	X	X				X		X		X		X	X	X	X
Coenagrionidae			X	X			X	X	X	X	X		X			X	X	X		
MEGALOPTERA																				
Corydalidae	X	X	X	X	X	X	X		X		X	X	X		X	X	X	X	X	X
Sialidae	X	X	X	X		X	X	X	X	X	X		X			X		X	X	X
NEUROPTERA																				
Sisyridae															X	X				
HEMIPTERA																				
Corixidae										X										
Veliidae	X	X	X	X	X									X	X	X				
Belostomatidae																				
Nepidae																				
Gerridae				X																
COLEOPTERA																				
Elmidae	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X	X
Hydrophilidae		X	X			X			X		X									
Gyrinidae			X	X	X			X					X				X		X	
Dytiscidae		X	X	X		X		X		X										
Noteridae																			X	
Dryopidae		X				X	X		X		X		X	X	X	X		X		X

ORGANISMS	79	80	81	82	83	85	86	87	88	89	93	94	95	96	97	98	99	100	101	102
Psephenidae														X						
Helodidae																				
Haliplidae				X							X		X							
Eubriidae																				
Ptilodactylidae																				
CRUSTACEA																				
Astacidae		X		X	X	X	X		X	X						X	X	X	X	
Isopoda																X				
Amphipoda				X																
Palaemonidae																		X		X
OLIGOCHAETA								X									X		X	
Glossosophoniidae																				
Naididae																				
Tubificidae						X														
Lumbriculidae																				
HIRUDINEA																				
HYDRACARINA					X															
MOLLUSCA																				
Bivalvia																				
Gastropoda undet. sp.								X												
Unionidae									X			X								
Corbiculidae	X	X		X			X		X	X	X	X	X	X			X	X	X	X
Lymnaeidae																				
Sphaeriidae										X										
Physidae										X			X							
Viviparidae							X					X	X							
Planorbidae												X	X	X						
Pleuroceridae											X		X	X				X	X	
Ancylidae																X				
TAXA RICHNESS	16	22	23	28	23	23	25	19	27	18	26	17	32	22	21	30	23	21	20	15
EPT INDEX	7	8	8	8	11	5	9	7	11	4	10	8	10	10	11	12	8	6	6	6
HABITAT SCORE	53	52	52	107	66	45	72	30	68	N/A	104	59	99	103	40	69	72	49	48	49

ORGANISMS	103	104	113	114	121	122	123	127	130	131	132	133	135	136	138	143	144	145	147	148
DIPTERA																				
Tabanidae	X																			
Empididae																				
Tipulidae	X		X	X				X	X	X	X			X				X		
Tanyderidae																				
Rhagionidae																				
Culicidae		X																		X
Ptychopteridae																				
Dixidae							X													X
Ceratopogonidae					X								X			X				X
Chironomidae	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Blephariceridae					X															
Simuliidae	X		X	X			X	X			X					X		X		
TRICHOPTERA																				
Hydroptilidae																				
Helicopsychidae																				
Hydropsychidae	X	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Rhyacophilidae					X									X						
Limnephilidae		X	X	X				X			X			X	X					X
Phryganeidae																				
Polycentropidae										X			X							
Psychomiidae				X											X					
Philopotamidae	X	X			X		X	X						X	X	X				X
Dipseuopsidae																		X		
Brachycentridae					X										X		X	X	X	
Goeridae																				
Lepidostomatidae								X						X	X					
Calamoceratidae																				
Uenoidae																				
Molannidae																				
Odontoceridae														X						
Leptoceridae			X			X	X	X		X	X	X	X		X	X	X	X	X	X
Glossosomatidae					X			X												
PLECOPTERA																				
Perlidae			X	X	X		X	X		X	X	X		X	X	X	X	X	X	X
Perlodidae					X															
Peltoperlidae	X				X			X			X			X	X					
Capniidae				X																
Leuctridae			X		X	X		X						X						X

ORGANISMS	103	104	113	114	121	122	123	127	130	131	132	133	135	136	138	143	144	145	147	148
Nemouridae					X															
Pteronarcyidae					X										X	X	X	X	X	
EPHEMEROPTERA																				
Baetidae	X	X	X		X	X	X	X		X		X	X		X	X	X	X	X	
Tricorythidae								X												
Heptagenidae	X	X	X		X	X	X	X		X	X	X	X	X	X	X	X	X	X	X
Oligoneuriidae	X							X		X	X	X			X	X	X	X	X	X
Leptophlebiidae			X	X	X			X												
Caenidae		X	X										X							
Neoephemeridae															X					
Ephemeridae	X							X		X		X	X	X	X	X	X	X	X	
Ephemerellidae	X		X	X	X			X			X				X	X	X	X	X	X
ODONATA																				
Libellulidae	X		X	X									X							
Corduliidae			X						X								X			
Cordulegasteridae								X												
Gomphidae	X		X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Aeschnidae	X		X			X	X	X	X	X	X	X		X	X	X	X	X	X	
Macromiidae												X			X		X			X
Calopterygidae		X		X		X	X		X	X		X	X				X	X	X	X
Coenagrionidae							X	X	X				X							X
MEGALOPTERA																				
Corydalidae	X		X	X		X	X	X	X	X	X	X		X	X	X	X	X	X	X
Sialidae	X						X	X	X	X			X		X		X	X		
NEUROPTERA																				
Sisyridae																				
HEMIPTERA													X							X
Corixidae									X			X								
Veliidae	X	X										X			X		X	X		
Belostomatidae										X										
Nepidae																				
Gerridae		X								X		X			X	X	X	X		
COLEOPTERA																				
Elmidae			X		X		X	X		X	X	X	X	X	X	X	X	X	X	X
Hydrophilidae						X				X	X	X				X	X	X		
Gyrinidae			X			X			X	X	X	X				X	X	X	X	
Dytiscidae			X	X					X											
Noteridae																				
Dryopidae	X						X	X		X		X	X	X		X		X	X	

ORGANISMS	103	104	113	114	121	122	123	127	130	131	132	133	135	136	138	143	144	145	147	148
Psephenidae								X												
Helodidae																				
Haliplidae		X											X							
Eubriidae																				
Ptilodactylidae								X												
CRUSTACEA																				
Astacidae	X		X		X	X	X	X		X	X	X			X	X	X	X	X	X
Isopoda	X												X							
Amphipoda							X													
Palaemonidae			X																	
OLIGOCHAETA	X					X		X												
Glossosophoniidae																				
Naididae																				
Tubificidae																				
Lumbriculidae																				
HIRUDINEA			X																	
HYDRACARINA													X							
MOLLUSCA																				
Bivalvia			X																	
Gastropoda undet. sp.													X							
Unionidae																				
Corbiculidae				X							X									
Lymnaeidae																				
Sphaeriidae																				
Physidae									X											
Viviparidae																				
Planorbidae									X											
Pleuroceridae			X					X						X						
Ancylidae																				
TAXA RICHNESS	22	12	26	14	22	16	21	32	16	22	19	21	22	19	26	22	27	26	19	24
EPT INDEX	8	6	10	6	15	5	6	16	1	8	8	7	7	11	16	10	10	11	10	9
HABITAT SCORE	89	70	104	78	89	54	86	112	54	62	48	62	97	102	82	93	51	47	48	55

ORGANISMS	149	151	154	155	162	163	164	166	167	176	177	186	187	189	191	192	193	194	195	196
DIPTERA																				
Tabanidae			X						X	X		X						X		
Empididae																				
Tipulidae		X	X	X	X	X	X				X	X	X	X	X	X		X	X	X
Tanyderidae																				
Rhagionidae																				
Culicidae				X		X		X												X
Ptychopteridae																				
Dixidae			X	X		X														X
Ceratopogonidae			X	X		X		X	X	X	X									
Chironomidae	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Blephariceridae														X						
Simuliidae			X	X						X		X	X	X		X	X		X	X
TRICHOPTERA																				
Hydroptilidae																		X	X	
Helicopsychidae																			X	
Hydropsychidae	X	X	X	X	X			X		X	X	X	X	X	X	X	X	X	X	X
Rhyacophilidae												X	X	X					X	
Limnephilidae	X	X		X	X							X	X		X	X		X	X	X
Phryganeidae																				
Polycentropidae			X	X		X	X		X					X	X	X				
Psychomiidae						X		X		X		X	X		X	X		X	X	
Philopotamidae			X	X			X	X											X	
Dipseuopsidae																X				
Brachycentridae								X					X	X	X				X	
Goeridae													X	X						
Lepidostomatidae					X							X	X		X			X		
Calamoceratidae					X															
Uenoidae												X	X	X	X	X		X		
Molannidae					X			X												
Odontoceridae					X															
Leptoceridae		X	X	X		X	X	X	X	X		X	X	X	X			X		X
Glossosomatidae												X	X	X	X	X				
PLECOPTERA																				
Perlidae					X							X	X	X	X	X		X	X	X
Perlodidae												X								
Peltoperlidae					X							X	X	X	X	X		X	X	
Capniidae																				
Leuctridae								X				X	X	X	X	X	X	X	X	X

ORGANISMS	149	151	154	155	162	163	164	166	167	176	177	186	187	189	191	192	193	194	195	196
Nemouridae																				
Pteronarcyidae												X	X	X	X	X		X	X	
EPHEMEROPTERA																				
Baetidae	X		X	X		X	X	X	X	X		X	X	X	X	X	X	X	X	X
Tricorythidae			X	X					X				X					X		
Heptagenidae	X	X		X	X	X	X	X	X	X		X	X	X	X	X		X	X	X
Oligoneuriidae	X	X	X	X									X	X	X	X		X	X	
Leptophlebiidae			X			X	X	X	X	X						X	X			
Caenidae			X	X		X	X	X	X					X	X					
Neoephemeridae													X	X	X			X	X	
Ephemeridae	X		X											X	X	X		X	X	X
Ephemerellidae				X								X	X	X	X	X	X		X	
ODONATA																				
Libellulidae										X										
Corduliidae											X									
Cordulegasteridae				X	X					X			X	X				X		
Gomphidae	X	X	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X	X
Aeschnidae	X	X	X	X	X	X		X		X			X	X	X	X	X		X	X
Macromiidae		X	X			X	X		X										X	
Calopterygidae	X	X	X	X		X	X	X		X								X		X
Coenagrionidae		X	X	X		X	X		X						X					X
MEGALOPTERA																				
Corydalidae	X	X	X	X		X	X	X	X		X		X	X	X	X	X	X		
Sialidae				X		X	X		X				X	X		X	X			
NEUROPTERA																				
Sisyridae																				
HEMIPTERA				X					X	X										
Corixidae																				
Veliidae	X	X			X								X							X
Belostomatidae																				
Nepidae						X														
Gerridae	X	X																		
COLEOPTERA																				
Elmidae			X	X		X	X	X	X	X		X	X	X	X	X		X	X	
Hydrophilidae											X		X							
Gyrinidae	X	X	X				X			X				X	X	X		X	X	
Dytiscidae		X		X			X			X	X								X	
Noteridae																				
Dryopidae							X	X	X				X	X	X		X	X		

ORGANISMS	149	151	154	155	162	163	164	166	167	176	177	186	187	189	191	192	193	194	195	196
Psephenidae												X	X	X	X					
Helodidae			X	X																
Haliplidae																				
Eubriidae								X												
Ptilodactylidae								X									X			
CRUSTACEA																				
Astacidae	X	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Isopoda				X			X	X			X									
Amphipoda				X		X			X	X	X									
Palaemonidae																				
OLIGOCHAETA				X		X			X	X	X									
Glossosophoniidae																			X	
Naididae			X																	
Tubificidae																				
Lumbriculidae			X																	
HIRUDINEA																				
HYDRACARINA						X			X											
MOLLUSCA																				
Bivalvia							X		X	X										
Gastropoda undet. sp.						X		X												
Unionidae																				
Corbiculidae						X														
Lymnaeidae																				
Sphaeriidae																		X		
Physidae																				
Viviparidae																				
Planorbidae																				
Pleuroceridae												X	X	X	X	X	X			
Ancylidae				X						X			X	X				X	X	X
TAXA RICHNESS	15	18	28	32	16	27	22	27	23	24	14	25	36	35	32	29	13	31	30	20
EPT INDEX	6	5	11	11	9	7	7	11	7	6	1	16	20	19	20	17	7	17	18	8
HABITAT SCORE	52	51	68	71	95	102	98	104	99	99	96	123	91	87	92	91	60	82	110	58

ORGANISMS	197	200	205	207	210	211	213	214	216	221	222	224	231	232	236	237	238	155a
DIPTERA																		
Tabanidae																		
Empididae				X			X											
Tipulidae	X				X	X	X		X									
Tanyderidae																		
Rhagionidae																		
Culicidae																		
Ptychopteridae																		
Dixidae						X												
Ceratopogonidae	X																	
Chironomidae	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Blephariceridae																		
Simuliidae		X		X	X		X		X							X	X	
TRICHOPTERA	X																	
Hydroptilidae			X						X	X								
Helicopsychidae																		
Hydropsychidae	X	X		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Rhyacophilidae																		
Limnephilidae							X			X	X	X	X	X				
Phryganeidae																		
Polycentropidae	X						X						X					X
Psychomiidae																		
Philopotamidae		X			X	X	X			X		X			X	X		
Dipseudopsidae														X	X	X		
Brachycentridae					X								X	X	X		X	
Goeridae																		
Lepidostomatidae							X						X					
Calamoceratidae														X				
Uenoidae																		
Molannidae																		
Odontoceridae																		
Leptoceridae	X	X	X	X		X			X	X	X	X	X	X	X	X	X	X
Glossosomatidae																		
PLECOPTERA																		
Perlidae	X	X			X	X	X	X	X	X	X		X	X	X			
Perlodidae							X											
Peltoperlidae					X		X											
Capniidae																		
Leuctridae		X															X	

ORGANISMS	197	200	205	207	210	211	213	214	216	221	222	224	231	232	236	237	238	155a
Nemouridae																		
Pteronarcyidae		X																
EPHEMEROPTERA																		
Baetidae		X	X	X	X	X	X	X	X		X	X	X	X	X	X	X	X
Tricorythidae			X	X		X			X	X	X	X				X	X	X
Heptagenidae	X	X	X	X	X	X	X	X	X	X	X	X	X		X	X		X
Oligoneuriidae		X					X	X	X	X		X				X		
Leptophlebiidae								X					X	X				
Caenidae				X	X				X	X	X	X					X	X
Neoephemeridae																		
Ephemeridae		X						X		X		X						X
Ephemerellidae		X			X		X			X		X	X	X		X		
ODONATA																		
Libellulidae																		X
Corduliidae													X					
Cordulegasteridae						X	X		X									
Gomphidae	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Aeschnidae	X	X	X	X	X	X		X		X	X	X	X	X	X	X	X	X
Macromiidae			X	X		X						X	X	X				
Calopterygidae			X	X				X	X	X	X	X	X	X	X		X	X
Coenagrionidae	X		X						X		X							X
MEGALOPTERA																		
Corydalidae	X	X			X	X		X	X	X	X	X	X	X		X	X	X
Sialidae	X	X								X								X
NEUROPTERA																		
Sisyridae																		
HEMIPTERA												X						
Corixidae		X																
Veliidae		X											X	X				
Belostomatidae																		
Nepidae																		
Gerridae		X								X								
COLEOPTERA																		
Elmidae	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Hydrophilidae		X	X	X														X
Gyrinidae		X	X	X	X						X	X	X	X	X	X		
Dytiscidae												X						
Noteridae																		
Dryopidae						X	X		X	X		X						X

ORGANISMS	197	200	205	207	210	211	213	214	216	221	222	224	231	232	236	237	238	155a
Psephenidae							X										X	
Helodidae																		
Haliplidae				X														
Eubriidae																		
Ptilodactylidae			X				X	X										
CRUSTACEA																		
Astacidae	X		X	X		X	X	X			X	X	X	X	X	X	X	
Isopoda																		
Amphipoda																		
Palaemonidae													X		X	X		
OLIGOCHAETA																		
Glossosophoniidae																		
Naididae																		
Tubificidae																		
Lumbriculidae																		X
HIRUDINEA							X											
HYDRACARINA																		
MOLLUSCA																		
Bivalvia																		
Gastropoda undet. sp.			X	X														
Unionidae															X	X		
Corbiculidae		X			X	X					X	X	X		X	X		
Lymnaeidae									X									
Sphaeriidae																		
Physidae																		
Viviparidae															X	X		
Planorbidae																		
Pleuroceridae																		
Ancylidae																		
TAXA RICHNESS	16	24	17	19	18	19	24	15	21	21	18	24	24	21	19	21	16	20
EPT INDEX	6	11	5	6	9	7	12	7	9	12	8	11	11	11	8	9	7	8
HABITAT SCORE	52	56	63	N/A	103	N/A	57	74	46	67	66	N/A	92	95	73	88	N/A	71

Savannah River REMAP Fish Collection

Common Name	193	196	121	191	186	192	194	61
American Eel								
Pirate Perch								
Creek Chubsucker								
Northern hogsucker								
Spotted Sucker								
Striped jumprock								
Silver redhorse								
Flier								
Blue spotted sunfish								
Redbreast Sunfish						2		
Green Sunfish								
Warmouth								1
Bluegill								
Pumpkinseed								
Longear sunfish								
Redear sunfish								
spotted sunfish								
Redeye Bass								
largemouth Bass					1			
Black Crappie								
Mottled Sculpin			10		2			
Whitefin Shiner								
Silvery Minnow								
Rosyface Chub								
Bluehead Chub		1			4		1	
Golden Shiner								
Highfin Shiner						1		
Spottail Shiner								
Yellowfin Shiner	73	21		7	16	11	18	
Sandbar Shiner								
Creek Chub		22			1	2		6
Chain Pickerel								
Redfin Pickerel								
Yellow Bullhead								
Brown Bullhead								
Black Bullhead								
Snail Bullhead								
Margined Madtom								
Tadpole Madtom								
Speckled Madtom								
Flat Bullhead								
Savannah Darter								
Christmas darter								
turquiose darter				1				
Tessellated Darter				1				
Blackbanded darter								
Yellow perch							3	
Mosquitofish								
Rainbow trout			7					
Total	73	44	17	9	24	16	22	7
Stream Order	1	1	2	2	2	2	2	2
Number of Species	1	3	2	3	5	4	3	2
Ecoregion	BR	BR	BR	BR	BR	BR	BR	BR

Savannah River REMAP Fish Collection

Common Name	195	187	189	162	13	197	15	33
American Eel								
Pirate Perch				2				1
Creek Chubsucker					1			
Northern hogsucker		1	1		4			
Spotted Sucker								
Striped jumprock								
Silver redhorse								
Flier								
Blue spotted sunfish								
Redbreast Sunfish			2		3	9		
Green Sunfish						21		1
Warmouth								
Bluegill						5	9	
Pumpkinseed								
Longear sunfish								
Redear sunfish								
spotted sunfish								
Redeye Bass								
largemouth Bass			1				3	
Black Crappie								
Mottled Sculpin		25	6					
Whitefin Shiner			3		33			
Silvery Minnow								
Rosyface Chub		1			6			
Bluehead Chub		1			25	15	47	11
Golden Shiner								
Highfin Shiner								
Spottail Shiner								
Yellowfin Shiner	18	14	9		181		101	19
Sandbar Shiner								
Creek Chub	3	3	3	31	8	1	35	43
Chain Pickerel								
Redfin Pickerel								
Yellow Bullhead							1	
Brown Bullhead								
Black Bullhead								
Snail Bullhead								
Margined Madtom					6			
Tadpole Madtom								
Speckled Madtom			3					
Flat Bullhead								
Savannah Darter								
Christmas darter								
turquiose darter			2		1			
Tessellated Darter	1	1						
Blackbanded darter		4	1					
Yellow perch	2				4			
Mosquitofish								
Rainbow trout		1						
Total	24	51	31	33	272	51	196	75
Stream Order	3	3	3	1	1	1	1	1
Number of Species	4	9	10	2	11	5	6	5
Ecoregion	BR	BR	BR	LP	LP	LP	LP	LP

Savannah River REMAP Fish Collection

Common Name	207	72	213	214	103	85	83	130
American Eel								
Pirate Perch					1			
Creek Chubsucker								
Northern hogsucker	9				1			
Spotted Sucker								
Striped jumprock				3				
Silver redhorse								
Flier								
Blue spotted sunfish								
Redbreast Sunfish	3			2	7	6		10
Green Sunfish						2		4
Warmouth	2					2		
Bluegill	5					3		8
Pumpkinseed								
Longear sunfish								
Redear sunfish								1
spotted sunfish								
Redeye Bass								
largemouth Bass								1
Black Crappie								
Mottled Sculpin								
Whitefin Shiner								1
Silvery Minnow								
Rosyface Chub	1				23	2		
Bluehead Chub	35	15	3		81	14	24	
Golden Shiner								
Highfin Shiner								
Spottail Shiner				2				
Yellowfin Shiner	90	60	62	70	112	19	10	
Sandbar Shiner								
Creek Chub	8	26	39	9	17	10	15	
Chain Pickerel								
Redfin Pickerel								
Yellow Bullhead						1		1
Brown Bullhead								
Black Bullhead								
Snail Bullhead						1		
Margined Madtom					1			
Tadpole Madtom								
Speckled Madtom								
Flat Bullhead								
Savannah Darter								
Christmas darter					1			
turquiose darter								
Tessellated Darter					9			
Blackbanded darter						2		
Yellow perch								
Mosquitofish								
Rainbow trout								
Total	153	101	104	86	253	62	49	26
Stream Order	1	1	1	1	1	1	1	1
Number of Species	8	3	3	5	10	11	3	7
Ecoregion	LP	LP	LP	LP	LP	LP	LP	LP

Common Name *	148	78	69	71	104	98	77	68
American Eel								
Pirate Perch	1					2		
Creek Chubsucker					2			
Northern hogsucker								7
Spotted Sucker								
Striped jumprock								5
Silver redhorse								
Flier								
Blue spotted sunfish								
Redbreast Sunfish		6			3		3	
Green Sunfish		5						
Warmouth								
Bluegill				9	1	1		
Pumpkinseed								
Longear sunfish								
Redear sunfish		1						
spotted sunfish								
Redeye Bass								2
largemouth Bass				3				
Black Crappie								
Mottled Sculpin								
Whitefin Shiner								
Silvery Minnow								
Rosyface Chub		1				3		14
Bluehead Chub	35	19	8	1	9	41	35	35
Golden Shiner						1		
Highfin Shiner								
Spottail Shiner								
Yellowfin Shiner	31	66	18	17	29	72	99	22
Sandbar Shiner								
Creek Chub	8	19	23	125	20	20	34	1
Chain Pickerel								
Redfin Pickerel								
Yellow Bullhead		2						
Brown Bullhead				1				
Black Bullhead								
Snail Bullhead								
Margined Madtom								
Tadpole Madtom								
Speckled Madtom								
Flat Bullhead								
Savannah Darter								
Christmas darter					4			
turquiose darter								
Tessellated Darter	1				1			
Blackbanded darter								5
Yellow perch								1
Mosquitofish								
Rainbow trout								
Total	76	119	49	156	69	140	171	92
Stream Order	1	1	1	1	1	1	1	2
Number of Species	5	8	3	6	8	7	4	9
Ecoregion	LP	LP	LP	LP	LP	LP	LP	LP

Savannah River REMAP Fish Collection

Common Name *	65	64	29	9	22	28	41	74
American Eel								
Pirate Perch			4					
Creek Chubsucker			3		2	1	3	
Northern hogsucker			3	2	2			18
Spotted Sucker								
Striped jumprock								3
Silver redhorse								
Flier								
Blue spotted sunfish								
Redbreast Sunfish		1	28		3	10	38	10
Green Sunfish			2		21		1	3
Warmouth						1	6	
Bluegill			6	1		13	3	2
Pumpkinseed			7			2	4	
Longear sunfish								
Redear sunfish								
spotted sunfish								
Redeye Bass								2
largemouth Bass			3				2	1
Black Crappie								
Mottled Sculpin								
Whitefin Shiner								
Silvery Minnow								
Rosyface Chub			5	1	2		12	5
Bluehead Chub	16	27	61	1	19	6	34	34
Golden Shiner					1		7	
Highfin Shiner			18			12		
Spottail Shiner								
Yellowfin Shiner	9	43	46	5	18		54	66
Sandbar Shiner								
Creek Chub	25	36	3	2	18	7		8
Chain Pickerel						2		
Redfin Pickerel								
Yellow Bullhead			4	1	1	1	2	
Brown Bullhead								
Black Bullhead								
Snail Bullhead								
Margined Madtom			1					1
Tadpole Madtom								
Speckled Madtom								1
Flat Bullhead								
Savannah Darter								
Christmas darter			7			2	2	
turquiose darter								
Tessellated Darter			4				1	
Blackbanded darter					1		2	3
Yellow perch								
Mosquitofish			1		1		9	
Rainbow trout								
Total	50	107	206	13	89	57	180	157
Stream Order	2	2	2	2	2	2	2	2
Number of Species	3	4	18	7	12	11	16	14
Ecoregion	LP	LP	LP	LP	LP	LP	LP	LP

Savannah River REMAP Fish Collection

Common Name *	87	75	89	96	222	155	154	151
American Eel								
Pirate Perch			4		1			
Creek Chubsucker			1			1		
Northern hogsucker		18		2		3	1	5
Spotted Sucker							2	
Striped jumprock		4						
Silver redhorse		1						
Flier								
Blue spotted sunfish								
Redbreast Sunfish	12	9	1			28	7	9
Green Sunfish		4	8	10		4	7	
Warmouth	1				1	5	3	
Bluegill	6		9			21	7	
Pumpkinseed								
Longear sunfish								
Redear sunfish								
spotted sunfish								
Redeye Bass		2						
largemouth Bass					1	1	2	
Black Crappie								
Mottled Sculpin								
Whitefin Shiner					2			
Silvery Minnow								
Rosyface Chub			7			4	1	
Bluehead Chub	2	64	13	88		8	18	35
Golden Shiner								
Highfin Shiner								
Spottail Shiner						4	4	29
Yellowfin Shiner		67	47	117	18	16	9	157
Sandbar Shiner								
Creek Chub	1	19	5	16	14			20
Chain Pickerel								
Redfin Pickerel								
Yellow Bullhead							1	1
Brown Bullhead								
Black Bullhead								
Snail Bullhead			1					
Margined Madtom			3		4			
Tadpole Madtom								
Speckled Madtom								
Flat Bullhead								
Savannah Darter								
Christmas darter			3	4				
turquiose darter								
Tessellated Darter	1			7		2	1	
Blackbanded darter		4						2
Yellow perch								
Mosquitofish					7			
Rainbow trout								
Total	23	192	102	244	48	97	63	258
Stream Order	2	2	2	2	2	2	2	2
Number of Species	6	10	12	7	8	12	13	8
Ecoregion	LP	LP	LP	LP	LP	LP	LP	LP

Savannah River REMAP Fish Collection

Common Name	99	131	123	221	135	200	132	149
American Eel	2							
Pirate Perch				1				
Creek Chubsucker					5			
Northern hogsucker	6	1		2		2		
Spotted Sucker								
Striped jumprock								
Silver redhorse								
Flier								
Blue spotted sunfish								
Redbreast Sunfish	17	3	6	1	15	2	6	
Green Sunfish		1	3		2			1
Warmouth		1						
Bluegill		10	35			3	1	
Pumpkinseed								
Longear sunfish								
Redear sunfish		5					1	
spotted sunfish								
Redeye Bass		2						
largemouth Bass	1		1					
Black Crappie								
Mottled Sculpin								
Whitefin Shiner		1		1			1	
Silvery Minnow								
Rosyface Chub	21		5		6			
Bluehead Chub	14	10	24		1		25	72
Golden Shiner					28			
Highfin Shiner								
Spottail Shiner	11	3					1	
Yellowfin Shiner	4	2	60	51	31	8	2	42
Sandbar Shiner								
Creek Chub		2	6	22	5	13	3	19
Chain Pickerel								
Redfin Pickerel								
Yellow Bullhead	1	2			2			
Brown Bullhead		2						
Black Bullhead				1				
Snail Bullhead								
Margined Madtom		2					2	
Tadpole Madtom	1							
Speckled Madtom								
Flat Bullhead	1							
Savannah Darter								
Christmas darter					7			
turquiose darter					1			
Tessellated Darter	5							
Blackbanded darter	10					1		
Yellow perch								
Mosquitofish					24			
Rainbow trout								
Total	94	47	140	79	127	29	42	134
Stream Order	2	2	2	2	2	2	2	2
Number of Species	13	15	8	7	12	6	9	4
Ecoregion	LP	LP	LP	LP	LP	LP	LP	LP

Savannah River REMAP Fish Collection

Common Name	122	216	205	81	224	82	86	102
American Eel								
Pirate Perch					1	1		1
Creek Chubsucker								
Northern hogsucker			1	7	5	5	12	1
Spotted Sucker					4			
Striped jumprock						2		
Silver redhorse								
Flier								
Blue spotted sunfish								
Redbreast Sunfish				3	4	9	6	9
Green Sunfish			1	7		7	14	
Warmouth					1			
Bluegill			1	6	9			6
Pumpkinseed								
Longear sunfish								
Redear sunfish					4			
spotted sunfish								
Redeye Bass					1			
largemouth Bass			1	1	1			
Black Crappie					1			1
Mottled Sculpin								
Whitefin Shiner								
Silvery Minnow				3				
Rosyface Chub			5		10	6		3
Bluehead Chub	12		62	68	14	54	88	
Golden Shiner					1			
Highfin Shiner								
Spottail Shiner				1	6		5	
Yellowfin Shiner		11	82	53	16	66	98	12
Sandbar Shiner								
Creek Chub	1	3	1	6				1
Chain Pickerel								
Redfin Pickerel								
Yellow Bullhead				3				
Brown Bullhead								
Black Bullhead								
Snail Bullhead				4		1		
Margined Madtom				2	1	12	5	
Tadpole Madtom								
Speckled Madtom								
Flat Bullhead							1	
Savannah Darter								
Christmas darter						6	6	
turquiose darter								
Tessellated Darter			1	1			3	3
Blackbanded darter				4	3	6	14	3
Yellow perch					1			2
Mosquitofish								1
Rainbow trout								
Total	13	14	155	169	83	175	252	43
Stream Order	2	2	2	3	3	3	3	3
Number of Species	2	2	9	15	18	12	11	12
Ecoregion	LP	LP	LP	LP	LP	LP	LP	LP

Savannah River REMAP Fish Collection

Common Name	101	94	143	144	88	147	79	100
American Eel								
Pirate Perch	1							1
Creek Chubsucker								
Northern hogsucker	5	25	9		3	2	5	6
Spotted Sucker			2					
Striped jumprock		7			1			
Silver redhorse		2						
Flier				1				
Blue spotted sunfish								
Redbreast Sunfish	7	15		2	42	8	8	8
Green Sunfish	3	5			13		12	
Warmouth	1			3	1	1		
Bluegill		1			1	1	11	
Pumpkinseed							2	
Longear sunfish								
Redear sunfish				1				
spotted sunfish								
Redeye Bass		2						
largemouth Bass								
Black Crappie								
Mottled Sculpin								
Whitefin Shiner								
Silvery Minnow	3							1
Rosyface Chub	6	23						10
Bluehead Chub	38	53	24	12	34	33	89	11
Golden Shiner								
Highfin Shiner								
Spottail Shiner		9			22		4	
Yellowfin Shiner	46	51	28	10	34	26	78	19
Sandbar Shiner								
Creek Chub	4		2				19	
Chain Pickerel						1		
Redfin Pickerel								
Yellow Bullhead				2		3		
Brown Bullhead								
Black Bullhead								
Snail Bullhead		2			1			
Margined Madtom		2	4	1	1	3	1	
Tadpole Madtom								
Speckled Madtom						1	1	
Flat Bullhead	2				2		2	1
Savannah Darter								
Christmas darter		3						
turquiose darter		2						
Tessellated Darter	9				3			6
Blackbanded darter	10	30	4		5	8	4	8
Yellow perch								
Mosquitofish								
Rainbow trout								
Total	135	232	73	32	163	87	236	71
Stream Order	3	3	3	3	3	3	3	3
Number of Species	13	16	7	8	14	11	13	10
Ecoregion	LP	LP	LP	LP	LP	LP	LP	LP

Savannah River REMAP Fish Collection

Common Name	93	95	80	155A	39	38	211	210
American Eel								
Pirate Perch		3		1				
Creek Chubsucker								
Northern hogsucker	25	7	9	1			1	2
Spotted Sucker								
Striped jumprock	5	2					2	3
Silver redhorse	5	1						
Flier								
Blue spotted sunfish								
Redbreast Sunfish	5	16	1	16	17	11	9	3
Green Sunfish	1		7	3	7	4	3	1
Warmouth	1			3	2			
Bluegill	1		4	11	3	2	1	
Pumpkinseed								
Longear sunfish								
Redear sunfish								
spotted sunfish								
Redeye Bass	3	2			1		1	2
largemouth Bass	4							
Black Crappie								
Mottled Sculpin								
Whitefin Shiner						2		
Silvery Minnow								
Rosyface Chub	6	9			4	2	2	2
Bluehead Chub	47	11	76	7	20	3	18	42
Golden Shiner	1							
Highfin Shiner								
Spottail Shiner			35	6	2			
Yellowfin Shiner	94	13	73	4	5	3	11	9
Sandbar Shiner		3			3			
Creek Chub			12	1		5		
Chain Pickerel					11			
Redfin Pickerel								
Yellow Bullhead		1	1	7				
Brown Bullhead								
Black Bullhead								
Snail Bullhead								
Margined Madtom	6	1	5		2	1	2	1
Tadpole Madtom								
Speckled Madtom								
Flat Bullhead	1	1						
Savannah Darter								
Christmas darter		2						
turquiose darter	16						2	4
Tessellated Darter	9	1		1	2	1		
Blackbanded darter	17	5	4		20	1		
Yellow perch							2	5
Mosquitofish								
Rainbow trout								
Total	258	78	227	61	99	35	54	74
Stream Order	3	3	3	3	3	3	3	3
Number of Species	19	16	11	12	14	11	12	11
Ecoregion	LP	LP	LP	LP	LP	LP	LP	LP

Savannah River REMAP Fish Collection

Common Name *	133	37	34	32	12	27	19	14
American Eel								
Pirate Perch			2	1		1		
Creek Chubsucker			9	1				1
Northern hogsucker	1	5	7	1	7		1	2
Spotted Sucker			2					
Striped jumprock				3	2			1
Silver redhorse					1			
Flier								
Blue spotted sunfish								
Redbreast Sunfish	11	7	21	6	5	5	5	5
Green Sunfish	2	4	13	2	4		21	
Warmouth	1		2					2
Bluegill	7		5		1		1	1
Pumpkinseed								
Longear sunfish								
Redear sunfish								
spotted sunfish								
Redeye Bass								
largemouth Bass	1				1			
Black Crappie								
Mottled Sculpin								
Whitefin Shiner					2			
Silvery Minnow								
Rosyface Chub		13	4	16	1	8	1	5
Bluehead Chub	3	41	60	50	88	5	18	27
Golden Shiner					1		2	
Highfin Shiner		2	20	1				
Spottail Shiner								3
Yellowfin Shiner		43	66	66	31	13	11	23
Sandbar Shiner								
Creek Chub		2	3		14		4	1
Chain Pickerel								
Redfin Pickerel								
Yellow Bullhead					3		1	1
Brown Bullhead	1	2			1			
Black Bullhead								
Snail Bullhead			1	1		1		3
Margined Madtom	1	6	2	3				3
Tadpole Madtom								
Speckled Madtom								
Flat Bullhead								
Savannah Darter								
Christmas darter		7	6	5				
turquiose darter								
Tessellated Darter		1	3	9			1	
Blackbanded darter		3	2	3			3	
Yellow perch								5
Mosquitofish						3		1
Rainbow trout								
Total	28	136	228	168	162	36	69	84
Stream Order	3	3	3	3	3	3	3	3
Number of Species	9	13	18	15	15	7	12	16
Ecoregion	LP	LP	LP	LP	LP	LP	LP	LP

Savannah River REMAP Fish Collection

Common Name *	10	11	8	145	30	176	113	232
American Eel								
Pirate Perch		3			3	1		
Creek Chubsucker		2						
Northern hogsucker	11		4	3	2			
Spotted Sucker								
Striped jumprock								
Silver redhorse								
Flier								
Blue spotted sunfish								1
Redbreast Sunfish	2	7	9	5	10		1	
Green Sunfish	13	1	14		1			
Warmouth		5		1	1		3	1
Bluegill				3	5		6	
Pumpkinseed							4	
Longear sunfish								
Redear sunfish								
spotted sunfish								
Redeye Bass								
largemouth Bass								
Black Crappie								
Mottled Sculpin								
Whitefin Shiner	2		2					
Silvery Minnow			1					
Rosyface Chub		1	2		3			
Bluehead Chub	72	13	47	21	23			
Golden Shiner								
Highfin Shiner		2			1			
Spottail Shiner	1		1					
Yellowfin Shiner	50	17	26	6	16		6	8
Sandbar Shiner								
Creek Chub	11	4	2	2	2			
Chain Pickerel				2				
Redfin Pickerel								
Yellow Bullhead								
Brown Bullhead			2					
Black Bullhead								
Snail Bullhead					2			
Margined Madtom					1			1
Tadpole Madtom								
Speckled Madtom				1				
Flat Bullhead								
Savannah Darter							3	
Christmas darter		2			4			
turquiose darter								
Tessellated Darter					9			
Blackbanded darter	3			2				
Yellow perch								
Mosquitofish		1				8		
Rainbow trout								
Total	165	58	110	46	83	9	23	11
Stream Order	3	3	3	3		2	2	3
Number of Species	9	12	11	10	15	2	6	4
Ecoregion	LP	LP	LP	LP	LP	MACP	SH	SH

Savannah River REMAP Fish Collection

Common Name	231	127	136	138	238	51	163	166
American Eel					1			
Pirate Perch					5	2	2	2
Creek Chubsucker							4	1
Northern hogsucker				2				
Spotted Sucker								
Striped jumprock								
Silver redhorse								
Flier						1		
Blue spotted sunfish								
Redbreast Sunfish		1		1		3	1	2
Green Sunfish								
Warmouth						4	1	
Bluegill							1	
Pumpkinseed						2		
Longear sunfish								
Redear sunfish								
spotted sunfish						2	5	11
Redeye Bass								
largemouth Bass						1		
Black Crappie								
Mottled Sculpin			6	3				
Whitefin Shiner								
Silvery Minnow								
Rosyface Chub								
Bluehead Chub		27		7		1	2	20
Golden Shiner								
Highfin Shiner								
Spottail Shiner				3				
Yellowfin Shiner	11	119	13	10	2	2	38	15
Sandbar Shiner					2			
Creek Chub		35	34	5	11			3
Chain Pickerel							2	3
Redfin Pickerel					1			
Yellow Bullhead						2		
Brown Bullhead								
Black Bullhead								
Snail Bullhead								
Margined Madtom				1	1			
Tadpole Madtom								
Speckled Madtom	1				2			
Flat Bullhead								
Savannah Darter	2				2			
Christmas darter								
turquiose darter								
Tessellated Darter					3		1	1
Blackbanded darter							2	
Yellow perch					1			
Mosquitofish								
Rainbow trout								
Total	14	182	53	32	31	20	59	58
Stream Order	3	1	1	2	1	1	2	2
Number of Species	3	4	3	8	11	10	11	9
Ecoregion	SH	UP	UP	UP	US/SP	US/SP	US/SP	US/SP

Savannah River REMAP Fish Collection

Common Name	236	237	164	167	Total
American Eel					2
Pirate Perch	1	1	5	3	45
Creek Chubsucker					33
Northern hogsucker					237
Spotted Sucker					10
Striped jumprock					29
Silver redhorse					10
Flier					1
Blue spotted sunfish			1		1
Redbreast Sunfish		1	16	13	533
Green Sunfish			1		250
Warmouth			2	1	44
Bluegill				6	194
Pumpkinseed					19
Longear sunfish			3	8	11
Redear sunfish					12
spotted sunfish			3	3	22
Redeye Bass					15
largemouth Bass				2	29
Black Crappie					2
Mottled Sculpin					42
Whitefin Shiner					16
Silvery Minnow					8
Rosyface Chub					225
Bluehead Chub			10		2098
Golden Shiner					42
Highfin Shiner					54
Spottail Shiner					147
Yellowfin Shiner	4	6	42		2866
Sandbar Shiner					6
Creek Chub		3			816
Chain Pickerel			2	2	25
Redfin Pickerel	1				0
Yellow Bullhead					40
Brown Bullhead					8
Black Bullhead					1
Snail Bullhead					15
Margined Madtom		1	1		75
Tadpole Madtom					1
Speckled Madtom		2			7
Flat Bullhead					11
Savannah Darter					61
Christmas darter					40
turquiose darter					43
Tessellated Darter			11	1	70
Blackbanded darter			4		198
Yellow perch	1				3
Mosquitofish					46
Rainbow trout					1
Total	7	14	101	39	10074
Stream Order	2	2	2	3	
Number of Species	4	6	13	9	
Ecoregion	US/SP	US/SP	US/SP	US/SP	

Savannah River REMAP IBI Final Results

	1 Benthic Invertivores		2 Brillouin Diversity Index		3 Non-native species		4 Number of native suckers		5 Number of native sunfishes		6 Number of minnows species		7 Number of species		8 Number of darter species		9 Proportion of omnivores		
Stream Station	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Total IBI Score
13	2.21	1	0.51	3	1.47	5	2.00		1.00	1	5.00	5	11.00	5	1.00	3	12.13	5	33
15	0.00	1	0.51	3	0.00	5	0.00	3	1.00	1	3.00	3	6.00	3	0.00	1	41.84	3	23
33	0.00	1	0.43	3	1.33	5	0.00	3	1.00	1	3.00	3	5.00	3	0.00	1	72.00	1	21
69	0.00	1	0.41	3	0.00	5	0.00	3	0.00	1	3.00	3	3.00	1	0.00	1	63.27	1	19
71	0.00	1	0.29	1	0.00	5	0.00	3	1.00	1	3.00	3	6.00	3	0.00	1	80.77	1	19
72	0.00	1	0.39	3	0.00	5	0.00	3	0.00	1	3.00	3	3.00	1	0.00	1	40.59	3	21
77	0.00	1	0.43	3	0.00	5	0.00	3	1.00	1	3.00	3	4.00	1	0.00	1	40.35	3	21
78	0.00	1	0.54	5	4.20	5	0.00	3	3.00	3	4.00	3	8.00	5	0.00	1	31.93	3	29
83	0.00	1	0.41	3	0.00	5	0.00	3	0.00	1	3.00	3	3.00	1	0.00	1	79.59	1	19
85	3.23	3	0.74	5	3.23	5	0.00	3	4.00	3	4.00	3	11.00	5	1.00	3	38.71	3	33
98	0.00	1	0.49	3	0.00	5	0.00	3	1.00	1	5.00	5	7.00	3	0.00	1	44.29	3	25
103	4.35	3	0.59	5	0.00	5	1.00	5	1.00	1	4.00	3	10.00	5	2.00	5	38.74	3	35
104	10.14	5	0.59	5	0.00	5	1.00	5	2.00	1	3.00	3	8.00	5	2.00	5	42.03	3	37
130	0.00	1	0.54	5	15.38	3	0.00	3	4.00	3	1.00	1	7.00	3	0.00	1	0.00	5	25
148	1.32	1	0.43	3	0.00	5	0.00	3	0.00	1	3.00	3	5.00	3	1.00	3	56.58	3	25
162	0.00	1	0.08	1	0.00	5	0.00	3	0.00	1	1.00	1	2.00	1	0.00	1	93.94	1	15
197	0.00	1	0.52	5	41.18	1	0.00	3	3.00	3	2.00	1	5.00	3	0.00	1	31.37	3	21
207	5.88	3	0.51	3	0.00	5	1.00	5	3.00	3	4.00	3	8.00	5	0.00	1	28.10	5	33
213	0.00	1	0.32	3	0.00	5	0.00	3	0.00	1	3.00	3	3.00	1	0.00	1	40.38	3	21
214	3.49	3	0.27	1	0.00	5	1.00	5	1.00	1	3.00	3	5.00	3	0.00	1	12.79	5	27
9	15.38	5	0.00	1	0.00	5	1.00	1	1.00	1	4.00	3	7.00	3	0.00	1	23.08	5	25
22	5.62	1	0.74	5	23.60	1	2.00	3	2.00	1	5.00	5	12.00	3	1.00	1	42.70	3	23
28	5.26	1	0.77	5	3.51	5	1.00	1	4.00	3	3.00	3	11.00	3	1.00	1	22.81	5	27
29	8.25	3	0.88	5	4.37	5	2.00	3	4.00	3	5.00	5	18.00	5	2.00	3	31.07	5	37
41	4.44	1	0.83	5	2.78	5	1.00	1	5.00	5	4.00	3	16.00	5	3.00	5	22.78	5	35
64	0.00	1	0.48	3	0.00	5	0.00	1	1.00	1	3.00	3	4.00	1	0.00	1	58.88	3	19
65	0.00	1	0.41	3	0.00	5	0.00	1	0.00	1	3.00	3	3.00	1	0.00	1	82.00	1	17
68	18.48	5	0.67	5	1.09	5	2.00	3	0.00	1	4.00	3	9.00	3	1.00	1	39.13	5	31
74	15.29	5	0.73	5	1.91	5	2.00	3	3.00	3	4.00	3	14.00	5	1.00	1	26.75	5	35
75	14.08	5	0.68	5	2.08	5	3.00	5	2.00	1	3.00	3	10.00	3	1.00	1	43.23	3	31
87	4.35	1	0.48	3	0.00	5	0.00	1	3.00	3	2.00	1	6.00	1	1.00	1	13.04	5	21
89	3.92	1	0.72	5	7.84	5	1.00	1	3.00	3	4.00	3	12.00	3	1.00	1	17.65	5	27
96	5.33	1	0.52	3	4.10	5	1.00	1	1.00	1	3.00	3	7.00	3	2.00	3	42.62	3	23
99	22.34	5	0.85	5	0.00	5	1.00	1	1.00	1	4.00	3	13.00	3	2.00	3	29.79	5	31
122	0.00	1	0.09	1	0.00	5	1.00	1	0.00	1	2.00	1	2.00	1	0.00	1	100.00	1	13
123	0.00	1	0.62	5	2.14	5	1.00	1	3.00	3	4.00	3	8.00	3	0.00	1	21.43	5	27
131	2.13	1	0.87	5	2.13	5	1.00	1	5.00	5	5.00	5	15.00	5	0.00	1	31.91	5	33
132	0.00	1	0.52	3	0.00	5	0.00	1	3.00	3	5.00	5	9.00	3	0.00	1	69.05	3	25
135	10.24	3	0.81	5	1.57	5	1.00	1	2.00	1	5.00	5	12.00	3	2.00	3	26.77	5	31
149	0.00	1	0.42	3	0.75	5	0.00	1	1.00	1	3.00	3	4.00	1	0.00	1	67.91	3	19
151	2.71	1	0.53	3	0.00	5	1.00	1	1.00	1	4.00	3	8.00	3	1.00	1	32.56	5	23

Savannah River REMAP IBI Final Results

	1		2		3		4		5		6		7		8		9		
	Benthic Invertivores		Brillouin Diversity Index		Non-native species		Number of native suckers		Number of native sunfishes		Number of minnows species		Number of species		Number of darter species		Proportion of omnivores		
Stream Station	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Total IBI Score
154	6.35	1	0.83	5	11.11	3	2.00	3	4.00	3	4.00	3	13.00	3	1.00	1	34.92	5	27
155	6.19	1	0.80	5	4.12	5	2.00	3	4.00	3	4.00	3	12.00	3	1.00	1	12.37	5	29
200	10.34	3	0.52	3	0.00	5	1.00	1	2.00	1	2.00	1	6.00	1	1.00	1	44.83	3	19
205	1.29	1	0.41	3	0.65	5	1.00	1	2.00	1	4.00	3	9.00	3	1.00	1	40.65	5	23
216	0.00	1	0.18	1	0.00	5	0.00	1	0.00	1	2.00	1	2.00	1	0.00	1	21.43	5	17
221	2.53	1	0.37	3	1.27	5	1.00	1	1.00	1	3.00	3	7.00	3	0.00	1	27.85	5	23
222	0.00	1	0.80	5	0.00	5	0.00	1	1.00	1	3.00	3	8.00	3	0.00	1	29.17	5	25
8	3.64	1	0.66	1	12.73	3	1.00	1	2.00	3	7.00	5	11.00	3	0.00	1	45.45	1	19
10	8.48	1	0.61	1	7.88	5	1.00	1	2.00	3	5.00	3	9.00	3	1.00	3	50.91	1	21
11	6.90	1	0.78	3	1.72	5	1.00	1	3.00	3	5.00	3	12.00	3	1.00	3	29.31	3	25
12	6.17	1	0.63	1	2.47	5	3.00	5	3.00	3	6.00	5	15.00	5	0.00	1	63.58	1	27
14	4.76	1	0.80	3	5.95	5	3.00	5	3.00	3	5.00	3	16.00	5	0.00	1	36.90	3	29
19	7.25	1	0.73	3	30.43	1	1.00	1	3.00	3	5.00	3	12.00	3	2.00	3	34.78	3	21
27	0.00	1	0.62	1	0.00	5	0.00	1	1.00	1	3.00	1	7.00	1	0.00	1	13.89	5	17
32	13.10	3	0.71	3	1.19	5	3.00	5	2.00	3	4.00	3	15.00	5	3.00	5	29.76	3	35
34	12.72	3	0.87	5	5.70	5	3.00	5	4.00	5	5.00	3	18.00	5	3.00	5	27.63	3	39
37	11.76	1	0.77	3	2.94	5	1.00	1	2.00	3	5.00	3	13.00	3	3.00	5	31.62	3	27
38	5.71	1	0.76	3	11.43	3	0.00	1	3.00	3	5.00	3	11.00	3	2.00	3	22.86	3	23
39	22.22	3	0.88	5	7.07	5	0.00	1	4.00	5	5.00	3	14.00	3	2.00	3	22.22	3	31
79	3.61	1	0.70	1	5.93	5	1.00	1	4.00	5	4.00	3	13.00	3	1.00	3	46.31	1	23
80	5.73	1	0.70	1	3.08	5	1.00	1	3.00	3	4.00	3	11.00	3	1.00	3	54.19	1	21
81	7.10	1	0.71	3	4.14	5	1.00	1	3.00	3	5.00	3	15.00	5	2.00	3	44.38	1	25
82	10.86	1	0.71	3	4.00	5	2.00	3	2.00	3	3.00	1	12.00	3	2.00	3	30.86	3	25
86	13.89	3	0.67	1	5.56	5	1.00	1	2.00	3	3.00	1	11.00	3	3.00	5	37.30	3	25
88	7.36	1	0.79	3	7.98	5	2.00	3	4.00	5	3.00	1	14.00	3	2.00	3	35.58	3	27
93	34.11	5	0.88	5	0.39	5	3.00	5	4.00	5	4.00	3	19.00	5	4.00	5	18.99	5	43
94	29.74	5	0.90	5	2.16	5	3.00	5	3.00	3	4.00	3	16.00	5	3.00	5	26.72	3	39
95	23.08	5	0.90	5	0.00	5	3.00	5	1.00	1	4.00	3	16.00	5	3.00	5	15.38	5	39
100	26.17	5	0.79	3	0.00	5	1.00	1	1.00	1	4.00	3	10.00	3	2.00	3	16.90	5	29
101	17.78	3	0.77	3	2.22	5	1.00	1	3.00	3	5.00	3	13.00	3	2.00	3	32.59	3	27
102	16.28	3	0.77	3	4.65	5	1.00	1	2.00	3	3.00	1	12.00	3	2.00	3	2.33	5	27
133	3.57	1	0.61	1	7.14	5	1.00	1	4.00	5	1.00	1	9.00	3	0.00	1	10.71	5	23
143	20.55	3	0.60	1	0.00	5	2.00	3	0.00	1	3.00	1	7.00	1	1.00	3	35.62	3	21
144	0.00	1	0.59	1	0.00	5	0.00	1	4.00	5	2.00	1	8.00	1	0.00	1	37.50	3	19
145	10.87	1	0.67	1	0.00	5	1.00	1	3.00	3	3.00	1	10.00	3	1.00	3	50.00	1	19
147	11.49	1	0.66	1	0.00	5	1.00	1	3.00	3	2.00	1	11.00	3	1.00	3	37.93	3	21
155.1	3.28	1	0.81	3	4.92	5	1.00	1	4.00	5	4.00	3	12.00	3	1.00	3	22.95	3	27
210	12.16	3	0.61	1	8.11	5	2.00	3	2.00	3	3.00	1	11.00	3	1.00	3	56.76	1	23
211	9.26	1	0.74	3	9.26	5	2.00	3	3.00	3	3.00	1	12.00	3	1.00	3	33.33	3	25
224	14.46	3	0.94	5	1.20	5	2.00	3	4.00	5	5.00	3	18.00	5	1.00	3	25.30	3	35
30	18.07	3	0.85	5	1.20	5	1.00	1	4.00	5	5.00	3	15.00	5	2.00	3	30.12	3	33

Savannah River REMAP IBI Final Results

	1 Benthic Invertivores		2 Brillouin Diversity Index		3 Non-native species		4 Number of native suckers		5 Number of native sunfishes		6 Number of minnows species		7 Number of species		8 Number of darter species		9 Proportion of omnivores		
Stream Station	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Total IBI Score
13	2.21	1	0.51	3	1.47	5	2.00		1.00	1	5.00	5	11.00	5	1.00	3	12.13	5	33
15	0.00	1	0.51	3	0.00	5	0.00	3	1.00	1	3.00	3	6.00	3	0.00	1	41.84	3	23
33	0.00	1	0.43	3	1.33	5	0.00	3	1.00	1	3.00	3	5.00	3	0.00	1	72.00	1	21
69	0.00	1	0.41	3	0.00	5	0.00	3	0.00	1	3.00	3	3.00	1	0.00	1	63.27	1	19
71	0.00	1	0.29	1	0.00	5	0.00	3	1.00	1	3.00	3	6.00	3	0.00	1	80.77	1	19
72	0.00	1	0.39	3	0.00	5	0.00	3	0.00	1	3.00	3	3.00	1	0.00	1	40.59	3	21
77	0.00	1	0.43	3	0.00	5	0.00	3	1.00	1	3.00	3	4.00	1	0.00	1	40.35	3	21
78	0.00	1	0.54	5	4.20	5	0.00	3	3.00	3	4.00	3	8.00	5	0.00	1	31.93	3	29
83	0.00	1	0.41	3	0.00	5	0.00	3	0.00	1	3.00	3	3.00	1	0.00	1	79.59	1	19
85	3.23	3	0.74	5	3.23	5	0.00	3	4.00	3	4.00	3	11.00	5	1.00	3	38.71	3	33
98	0.00	1	0.49	3	0.00	5	0.00	3	1.00	1	5.00	5	7.00	3	0.00	1	44.29	3	25
103	4.35	3	0.59	5	0.00	5	1.00	5	1.00	1	4.00	3	10.00	5	2.00	5	38.74	3	35
104	10.14	5	0.59	5	0.00	5	1.00	5	2.00	1	3.00	3	8.00	5	2.00	5	42.03	3	37
130	0.00	1	0.54	5	15.38	3	0.00	3	4.00	3	1.00	1	7.00	3	0.00	1	0.00	5	25
148	1.32	1	0.43	3	0.00	5	0.00	3	0.00	1	3.00	3	5.00	3	1.00	3	56.58	3	25
162	0.00	1	0.08	1	0.00	5	0.00	3	0.00	1	1.00	1	2.00	1	0.00	1	93.94	1	15
197	0.00	1	0.52	5	41.18	1	0.00	3	3.00	3	2.00	1	5.00	3	0.00	1	31.37	3	21
207	5.88	3	0.51	3	0.00	5	1.00	5	3.00	3	4.00	3	8.00	5	0.00	1	28.10	5	33
213	0.00	1	0.32	3	0.00	5	0.00	3	0.00	1	3.00	3	3.00	1	0.00	1	40.38	3	21
214	3.49	3	0.27	1	0.00	5	1.00	5	1.00	1	3.00	3	5.00	3	0.00	1	12.79	5	27
9	15.38	5	0.00	1	0.00	5	1.00	1	1.00	1	4.00	3	7.00	3	0.00	1	23.08	5	25
22	5.62	1	0.74	5	23.60	1	2.00	3	2.00	1	5.00	5	12.00	3	1.00	1	42.70	3	23
28	5.26	1	0.77	5	3.51	5	1.00	1	4.00	3	3.00	3	11.00	3	1.00	1	22.81	5	27
29	8.25	3	0.88	5	4.37	5	2.00	3	4.00	3	5.00	5	18.00	5	2.00	3	31.07	5	37
41	4.44	1	0.83	5	2.78	5	1.00	1	5.00	5	4.00	3	16.00	5	3.00	5	22.78	5	35
64	0.00	1	0.48	3	0.00	5	0.00	1	1.00	1	3.00	3	4.00	1	0.00	1	58.88	3	19
65	0.00	1	0.41	3	0.00	5	0.00	1	0.00	1	3.00	3	3.00	1	0.00	1	82.00	1	17
68	18.48	5	0.67	5	1.09	5	2.00	3	0.00	1	4.00	3	9.00	3	1.00	1	39.13	5	31
74	15.29	5	0.73	5	1.91	5	2.00	3	3.00	3	4.00	3	14.00	5	1.00	1	26.75	5	35
75	14.06	5	0.68	5	2.08	5	3.00	5	2.00	1	3.00	3	10.00	3	1.00	1	43.23	3	31
87	4.35	1	0.48	3	0.00	5	0.00	1	3.00	3	2.00	1	6.00	1	1.00	1	13.04	5	21
89	3.92	1	0.72	5	7.84	5	1.00	1	3.00	3	4.00	3	12.00	3	1.00	1	17.65	5	27
96	5.33	1	0.52	3	4.10	5	1.00	1	1.00	1	3.00	3	7.00	3	2.00	3	42.62	3	23
99	22.34	5	0.85	5	0.00	5	1.00	1	1.00	1	4.00	3	13.00	3	2.00	3	29.79	5	31
122	0.00	1	0.09	1	0.00	5	1.00	1	0.00	1	2.00	1	2.00	1	0.00	1	100.00	1	13
123	0.00	1	0.62	5	2.14	5	1.00	1	3.00	3	4.00	3	8.00	3	0.00	1	21.43	5	27
131	2.13	1	0.87	5	2.13	5	1.00	1	5.00	5	5.00	5	15.00	5	0.00	1	31.91	5	33
132	0.00	1	0.52	3	0.00	5	0.00	1	3.00	3	5.00	5	9.00	3	0.00	1	69.05	3	25
135	10.24	3	0.81	5	1.57	5	1.00	1	2.00	1	5.00	5	12.00	3	2.00	3	26.77	5	31
149	0.00	1	0.42	3	0.75	5	0.00	1	1.00	1	3.00	3	4.00	1	0.00	1	67.91	3	19
151	2.71	1	0.53	3	0.00	5	1.00	1	1.00	1	4.00	3	8.00	3	1.00	1	32.56	5	23

Savannah River REMAP IBI Final Results

	1 Benthic Invertivores		2 Brillouin Diversity Index		3 Non-native species		4 Number of native suckers		5 Number of native sunfishes		6 Number of minnows species		7 Number of species		8 Number of darter species		9 Proportion of omnivores		
Stream Station	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Metric Result	IBI Score	Total IBI Score
154	6.35	1	0.83	5	11.11	3	2.00	3	4.00	3	4.00	3	13.00	3	1.00	1	34.92	5	27
155	6.19	1	0.80	5	4.12	5	2.00	3	4.00	3	4.00	3	12.00	3	1.00	1	12.37	5	29
200	10.34	3	0.52	3	0.00	5	1.00	1	2.00	1	2.00	1	6.00	1	1.00	1	44.83	3	19
205	1.29	1	0.41	3	0.65	5	1.00	1	2.00	1	4.00	3	9.00	3	1.00	1	40.65	5	23
216	0.00	1	0.18	1	0.00	5	0.00	1	0.00	1	2.00	1	2.00	1	0.00	1	21.43	5	17
221	2.53	1	0.37	3	1.27	5	1.00	1	1.00	1	3.00	3	7.00	3	0.00	1	27.85	5	23
222	0.00	1	0.80	5	0.00	5	0.00	1	1.00	1	3.00	3	8.00	3	0.00	1	29.17	5	25
8	3.64	1	0.66	1	12.73	3	1.00	1	2.00	3	7.00	5	11.00	3	0.00	1	45.45	1	19
10	8.48	1	0.61	1	7.88	5	1.00	1	2.00	3	5.00	3	9.00	3	1.00	3	50.91	1	21
11	6.90	1	0.78	3	1.72	5	1.00	1	3.00	3	5.00	3	12.00	3	1.00	3	29.31	3	25
12	6.17	1	0.63	1	2.47	5	3.00	5	3.00	3	6.00	5	15.00	5	0.00	1	63.58	1	27
14	4.76	1	0.80	3	5.95	5	3.00	5	3.00	3	5.00	3	16.00	5	0.00	1	36.90	3	29
19	7.25	1	0.73	3	30.43	1	1.00	1	3.00	3	5.00	3	12.00	3	2.00	3	34.78	3	21
27	0.00	1	0.62	1	0.00	5	0.00	1	1.00	1	3.00	1	7.00	1	0.00	1	13.89	5	17
32	13.10	3	0.71	3	1.19	5	3.00	5	2.00	3	4.00	3	15.00	5	3.00	5	29.76	3	35
34	12.72	3	0.87	5	5.70	5	3.00	5	4.00	5	5.00	3	16.00	5	3.00	5	27.63	3	39
37	11.76	1	0.77	3	2.94	5	1.00	1	2.00	3	5.00	3	13.00	3	3.00	5	31.62	3	27
38	5.71	1	0.76	3	11.43	3	0.00	1	3.00	3	5.00	3	11.00	3	2.00	3	22.86	3	23
39	22.22	3	0.86	5	7.07	5	0.00	1	4.00	5	5.00	3	14.00	3	2.00	3	22.22	3	31
79	3.81	1	0.70	1	5.93	5	1.00	1	4.00	5	4.00	3	13.00	3	1.00	3	48.31	1	23
80	5.73	1	0.70	1	3.08	5	1.00	1	3.00	3	4.00	3	11.00	3	1.00	3	54.19	1	21
81	7.10	1	0.71	3	4.14	5	1.00	1	3.00	3	5.00	3	15.00	5	2.00	3	44.38	1	25
82	10.86	1	0.71	3	4.00	5	2.00	3	2.00	3	3.00	1	12.00	3	2.00	3	30.86	3	25
86	13.89	3	0.67	1	5.58	5	1.00	1	2.00	3	3.00	1	11.00	3	3.00	5	37.30	3	25
88	7.36	1	0.79	3	7.98	5	2.00	3	4.00	5	3.00	1	14.00	3	2.00	3	35.58	3	27
93	34.11	5	0.86	5	0.39	5	3.00	5	4.00	5	4.00	3	19.00	5	4.00	5	18.99	5	43
94	29.74	5	0.90	5	2.16	5	3.00	5	3.00	3	4.00	3	16.00	5	3.00	5	26.72	3	39
95	23.06	5	0.90	5	0.00	5	3.00	5	1.00	1	4.00	3	16.00	5	3.00	5	15.38	5	39
100	26.17	5	0.79	3	0.00	5	1.00	1	1.00	1	4.00	3	10.00	3	2.00	3	16.90	5	29
101	17.78	3	0.77	3	2.22	5	1.00	1	3.00	3	5.00	3	13.00	3	2.00	3	32.59	3	27
102	16.28	3	0.77	3	4.65	5	1.00	1	2.00	3	3.00	1	12.00	3	2.00	3	2.33	5	27
133	3.57	1	0.61	1	7.14	5	1.00	1	4.00	5	1.00	1	9.00	3	0.00	1	10.71	5	23
143	20.55	3	0.60	1	0.00	5	2.00	3	0.00	1	3.00	1	7.00	1	1.00	3	35.62	3	21
144	0.00	1	0.59	1	0.00	5	0.00	1	4.00	5	2.00	1	8.00	1	0.00	1	37.50	3	19
145	10.87	1	0.67	1	0.00	5	1.00	1	3.00	3	3.00	1	10.00	3	1.00	3	50.00	1	19
147	11.40	1	0.66	1	0.00	5	1.00	1	3.00	3	2.00	1	11.00	3	1.00	3	37.93	3	21
155.1	3.26	1	0.81	3	4.92	5	1.00	1	4.00	5	4.00	3	12.00	3	1.00	3	22.95	3	27
210	12.16	3	0.61	1	8.11	5	2.00	3	2.00	3	3.00	1	11.00	3	1.00	3	56.76	1	23
211	9.26	1	0.74	3	9.28	5	2.00	3	3.00	3	3.00	1	12.00	3	1.00	3	33.33	3	25
224	14.46	3	0.94	5	1.20	5	2.00	3	4.00	5	5.00	3	18.00	5	1.00	3	25.30	3	35
30	16.07	3	0.86	5	1.20	5	1.00	1	4.00	5	5.00	3	15.00	5	2.00	3	30.12	3	33

Summary of biological indicator results and scores for the Lower Piedmont Ecological Index

Station	Order	Station	Order	Habitat	Score	Macroinvertebrate EPT Index	Macroinvertebrate EPT Score	Fish IBI Result	Fish IBI Score	LPEI Total Index Score	LPEI Stream Classification
8	3	8	3	64	3	3	1	19	1	5	Poor
9	2	9	2	94	5	4	1	25	3	9	Fair
10	3	10	3	80	3	4	1	21	1	5	Poor
11	3	11	3	47	1	3	1	25	3	5	Poor
12	3	12	3	83	3	5	1	27	3	7	Poor
13	1	13	1	61	3	3	1	33	5	9	Fair
14	3	14	3	45	1	3	1	29	3	5	Poor
15	1	15	1	87	3	2	1	23	3	7	Poor
19	3	19	3	91	5	5	1	21	1	7	Poor
22	2	22	2	93	5	5	1	23	3	9	Fair
27	3	27	3	64	3	3	1	17	1	5	Poor
28	2	28	2	92	5	2	1	27	3	9	Fair
29	2	29	2	119	5	7	3	37	5	13	Good
30	3	30	3	116	5	5	1	33	5	11	Good
32	3	32	3	112	5	5	1	35	5	11	Good
33	1	33	1	75	3	4	1	21	1	5	Poor
34	3	34	3	113	5	6	3	39	5	13	Good
37	3	37	3	111	5	7	3	27	3	11	Good
38	3	38	3	88	5	5	1	23	3	9	Fair
39	3	39	3	82	3	2	1	31	5	9	Fair
41	2	41	2	48	1	4	1	35	5	7	Poor
64	2	64	2	45	1	7	3	19	1	5	Poor
65	2	65	2	41	1	5	1	17	1	3	Poor
68	2	68	2	104	5	15	5	31	5	15	Good
69	1	69	1	84	3	12	5	19	1	9	Fair
71	1	71	1	76	3	8	3	19	1	7	Poor
72	1	72	1	82	3	18	5	21	1	9	Fair
74	2	74	2	41	1	9	5	35	5	11	Good
75	2	75	2	54	3	9	5	31	5	13	Good
77	1	77	1	105	5	10	5	21	1	11	Good
78	1	78	1	91	5	4	1	29	3	9	Fair
79	3	79	3	53	3	7	3	23	3	9	Fair
80	3	80	3	52	1	8	3	21	1	5	Poor
81	3	81	3	52	1	8	3	25	3	7	Poor
82	3	82	3	107	5	8	3	25	3	11	Good
83	1	83	1	66	3	11	5	19	1	9	Fair
85	1	85	1	45	1	5	1	33	5	7	Poor
86	3	86	3	72	3	9	5	25	3	11	Good
87	2	87	2	30	1	7	3	21	1	5	Poor
88	3	88	3	68	3	11	5	27	3	11	Good
93	3	93	3	104	5	10	5	43	5	15	Good
94	3	94	3	59	3	8	3	39	5	11	Good
95	3	95	3	99	5	10	5	39	5	15	Good
96	2	96	2	103	5	10	5	23	3	13	Good
98	1	98	1	69	3	12	5	25	3	11	Good
99	2	99	2	72	3	8	3	31	5	11	Good
100	3	100	3	49	1	8	3	29	3	7	Poor
101	3	101	3	48	1	6	3	27	3	7	Poor
102	3	102	3	49	1	6	3	27	3	7	Poor
103	1	103	1	89	5	8	3	35	5	13	Good
104	1	104	1	70	3	6	3	37	5	11	Good
122	2	122	2	53	3	5	1	13	1	5	Poor
123	2	123	2	86	3	6	3	27	3	9	Fair
130	1	130	1	54	3	1	1	25	3	7	Poor
131	2	131	2	62	3	8	3	33	5	11	Good
132	2	132	2	48	1	8	3	25	3	7	Poor
133	3	133	3	62	3	7	3	23	3	9	Fair

Summary of biological indicator results and scores for the
Lower Piedmont Ecological Index

Station	Order	Station	Order	Habitat	Score	Macroinvertebrate EPT Index	Score	Fish IBI Result	Score	LPEI Total Index Score	Stream Classification
135	2	135	2	75	3	7	3	31	5	11	Good
143	3	143	3	93	5	10	5	21	1	11	Good
144	3	144	3	51	1	10	5	19	1	7	Poor
145	3	145	3	47	1	11	5	19	1	7	Poor
147	3	147	3	48	1	10	5	21	1	7	Poor
148	1	148	1	55	3	9	5	25	3	11	Good
149	2	149	2	52	1	8	3	19	1	5	Poor
151	2	151	2	51	1	5	1	23	3	5	Poor
154	2	154	2	68	3	11	5	27	3	11	Good
155	2	155	2	71	3	10	5	29	3	11	Good
155.1	3	155.1	3	72	3	8	3	27	3	9	Fair
162	1	162	1	95	5	9	5	15	1	11	Good
197	1	197	1	52	1	6	3	21	1	5	Poor
200	2	200	2	58	3	11	5	19	1	9	Fair
205	2	205	2	83	3	5	1	23	3	7	Poor
210	3	210	3	103	5	9	5	23	3	13	Good
213	1	213	1	57	3	12	5	21	1	9	Fair
214	1	214	1	74	3	7	3	27	3	9	Fair
216	2	216	2	46	1	9	5	17	1	7	Poor
221	2	221	2	67	3	12	5	23	3	11	Good
222	2	222	2	66	3	8	3	25	3	9	Fair

Station	Order	Habitat	Score	Index	Score	Score	Score	Total Index Score	Stream Classification
4	1	76	3	3	1				
13	1	61	3	3	1	33	5	9	Fair
15	1	87	3	2	1	23	3	7	Poor
33	1	75	3	4	1	21	1	5	Poor
44	1	75	3	3	1		1		
69	1	84	3	12	5	19	1	9	Fair
71	1	76	3	8	3	19	1	7	Poor
72	1	82	3	18	5	21	1	9	Fair
77	1	105	5	10	5	21	1	11	Fair
78	1	91	5	4	1	29	3	9	Fair
83	1	66	3	11	5	19	1	9	Fair
85	1	45	1	5	1	33	5	7	Poor
97	1	40	1	11	5		1		
98	1	69	3	12	5	25	3	11	Fair
103	1	89	5	8	3	35	5	13	Good
104	1	70	3	6	3	37	5	11	Fair
130	1	54	3	1	1	25	3	7	Poor
148	1	55	3	9	5	25	3	11	Fair
162	1	95	5	9	5	15	1	11	Fair
197	1	52	1	6	3	21	1	5	Poor
207	1			6	3	33	5		
213	1	57	3	12	5	21	1	9	Fair
214	1	74	3	7	3	27	3	9	Fair
9	2	94	5	4	1	25	3	9	Fair
22	2	93	5	5	1	23	3	9	Fair
28	2	92	5	2	1	27	3	9	Fair
29	2	119	5	7	3	37	5	13	Good
31	2	87	3				1		
41	2	48	1	4	1	35	5	7	Poor
45	2	45	1	1	1		1		
64	2	45	1	7	3	19	1	5	Poor
65	2	41	1	5	1	17	1	3	Poor
68	2	104	5	15	5	31	5	15	Good
74	2	41	1	9	5	35	5	11	Fair
75	2	54	3	9	5	31	5	13	Good
87	2	30	1	7	3	21	1	5	Poor
89	2			4	1	27	3		
96	2	103	5	10	5	23	3	13	Good
99	2	72	3	8	3	31	5	11	Fair
122	2	53	3	5	1	13	1	5	Poor
123	2	86	3	6	3	27	3	9	Fair
131	2	82	3	8	3	33	5	11	Fair
132	2	48	1	8	3	25	3	7	Poor
135	2	75	3	7	3	31	5	11	Fair
149	2	52	1	6	3	19	1	5	Poor
151	2	51	1	5	1	23	3	5	Poor
154	2	68	3	11	5	27	3	11	Fair
155	2	71	3	10	5	29	3	11	Fair
200	2	58	3	11	5	19	1	9	Fair
205	2	63	3	5	1	2[illegible]	3	7	Poor
216	2	46	1	9	5	17	1	7	Poor
221	2	67	3	12	5	23	3	11	Fair
222	2	66	3	8	3	25	3	9	Fair
8	3	64	3	3	1	19	1	5	Poor
10	3	80	3	4	1	21	1	5	Poor
11	3	47	1	3	1	25	3	5	Poor
12	3	83	3	5	1	27	3	7	Poor
14	3	45	1	3	1	29	3	5	Poor
19	3	91	5	5	1	21	1	7	Poor
21	3	67	3	5	1		1		

Station	Order	Habitat	Score	EPT Index	Score	IBI Score	Score	Total Index Score	Stream Classification
27	3	64	3	3	1	17	1	5	Poor
30	3	116	5	5	1	33	5	11	Fair
32	3	112	5	5	1	35	5	11	Fair
34	3	113	5	6	3	39	5	13	Good
37	3	111	5	7	3	27	3	11	Fair
38	3	88	5	5	1	23	3	9	Fair
39	3	82	3	2	1	31	5	9	Fair
79	3	53	3	7	3	23	3	9	Fair
80	3	52	1	8	3	21	1	5	Poor
81	3	52	1	8	3	25	3	7	Poor
82	3	107	5	8	3	25	3	11	Fair
86	3	72	3	9	5	25	3	11	Fair
88	3	68	3	11	5	27	3	11	Fair
93	3	104	5	10	5	43	5	15	Good
94	3	59	3	8	3	39	5	11	Fair
95	3	99	5	10	5	39	5	15	Good
100	3	49	1	8	3	29	3	7	Poor
101	3	48	1	6	3	27	3	7	Poor
102	3	49	1	6	3	27	3	7	Poor
133	3	62	3	7	3	23	3	9	Fair
143	3	93	5	10	5	21	1	11	Fair
144	3	51	1	10	5	19	1	7	Poor
145	3	47	1	11	5	19	1	7	Poor
147	3	48	1	10	5	21	1	7	Poor
210	3	103	5	9	5	23	3	13	Good
211	3			7	3	25	3		
224	3			11	5	35	5		
155.1	3	72	3	8	3	27	3	9	Fair

	Scores	Percent
Good	9	11.5
Fair	38	48.7
Poor	31	39.7
Total	78	100

D

Appendix D

Guidelines for Locating and Accessing Sites on Wadable Streams in Watersheds of the Southeastern United States

By
James R. Maudsley
and
Robert J. Lewis

Guidelines for Locating and Accessing EMAP Sites on Wadeable Streams in Watersheds of the Southeastern United States

Prepared by:

James R. Maudsley Ph.D. And Robert J. Lewis
ManTech Environmental Technology Inc.
under contract to Integrated Laboratory Systems
Athens, Georgia

For:

U.S. Environmental Protection Agency
Region IV
Athens, Georgia

DCN #ESAT-4B-7004
1997

Table of Contents

List of Figures

Appendices

Introduction

The following guidelines summarize the steps employed by EPA Region IV to locate and access stream sampling sites selected by the Environmental Monitoring and Assessment Program (EMAP) approach. The strength of the EMAP approach is that it is a probability-based survey (Volstad et. al. 1995) Sampling locations are randomly selected. The net result is that fewer stream sampling locations are needed to characterize a watershed or river basin than would be required with a non-random selection process (ie. bridge crossings), thereby effectively reducing the effort needed to characterize large, regional river basins to a logistically and economically feasible level. While the EMAP approach reduces the effort required to conduct regional surveys, it presents some unique challenges. Because stream sampling locations are randomly selected, the stream sites may not be near identifiable physical land structures (bridge crossing, bend in the stream) and miles from the nearest road or highway. Through trial and error EPA Region IV has developed the following set of procedures for efficiently locating and accessing EMAP sites:

Acquiring Permission to Access Stream Sampling Sites

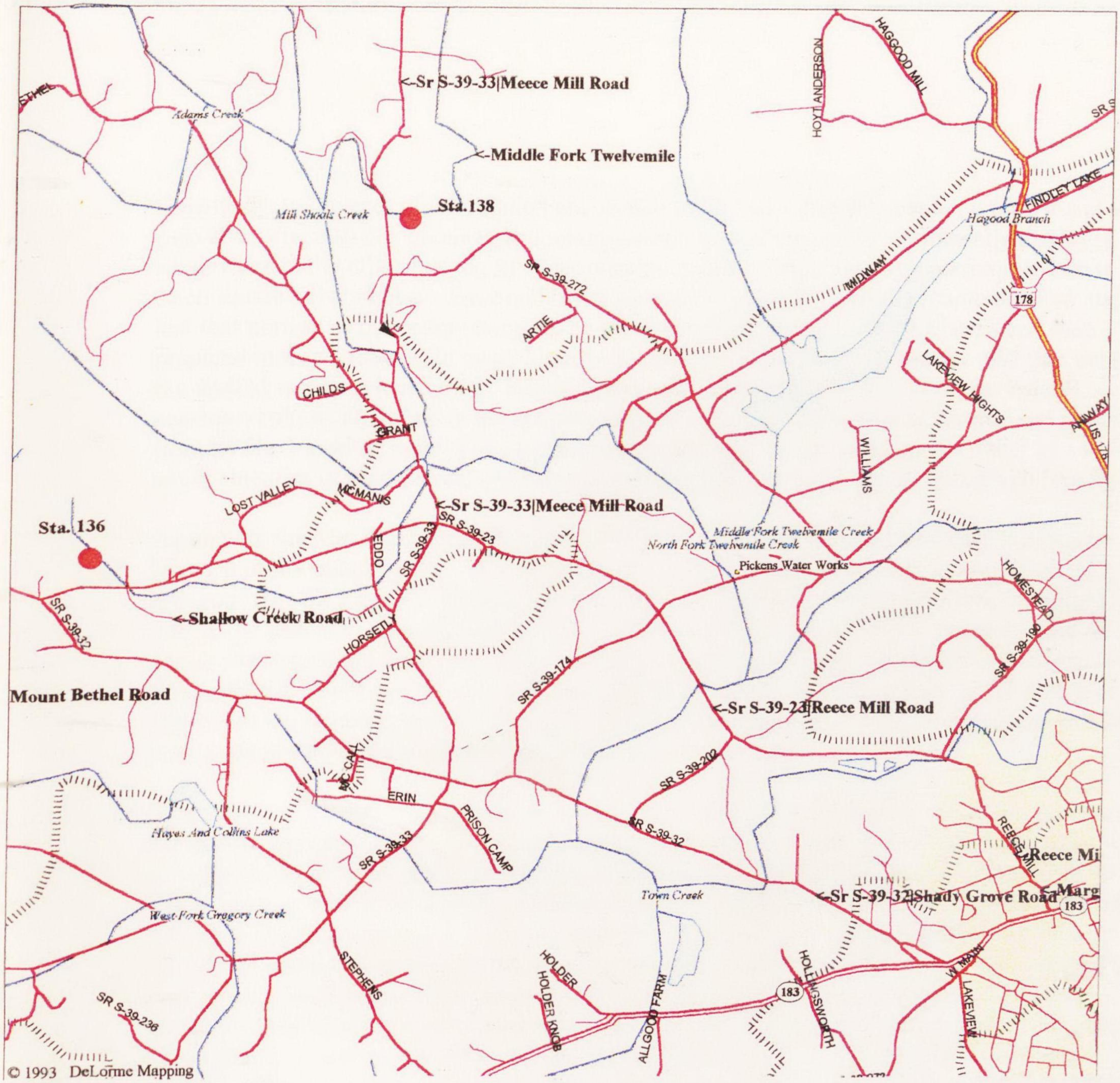
Step 1: Locating EMAP Coordinates on a Map

The first step in locating a randomly selected EMAP sampling location is to correctly pinpoint the site on a map. Accuracy is essential to prevent costly and time consuming mistakes. Each stream location generated by the EMAP approach is supplied as a pair of map coordinates, a latitude and a longitude. The method used to locate EMAP streams sites in EPA Region IV relies on computer software to mark the exact location of each set of coordinates on a computer generated map (see figure 1). Region IV uses a mapping program called MapeXpert® by Delorme but other mapping programs should work equally as well. The exact procedure for placing EMAP coordinates on a map using MapeXpert can be found in Appendix A of this document.

Step 2: Acquiring Names and Addresses of Landowners

Once a stream sampling location is accurately located on a map, the next step is to determine who owns the property adjacent to the stream and ask permission to access the site. This requires a visit to the Tax Assessor's Office in the county in which the site is located.

County tax assessor offices house records of property ownership. Most often the offices are located in the city or town serving as the county seat. Some offices have all records as a hard copy only, making the search for land ownership a rather slow process. However, more and more tax offices are converting to electronically stored information, easily accessible from computer terminals located at various stations throughout the office. Regardless of the information retrieval system certain basic information must be obtained to properly locate the owners of a given property. Personnel working in the tax offices are usually more than willing to guide you through the process.



LEGEND

- Geo Feature
- US Highway
- Population Center
- Street, Road
- Major Street/Road
- US Highway
- River
- Utility (powerline)
- Open Water
- Contours

Scale 1:37,500 (at center)

2000 Feet

1000 Meters

Mag 14.00
Thu Jul 25 15:57:52 1996

Fig. 1. Computer Generated Map

Essentially, two pieces of information are needed to lookup a property owner, a **map #** and a **parcel #**. The map # is obtained by looking a large map of the county prominently displayed somewhere in the office (see fig 2). This map has a grid of large squares superimposed on it. Each square has a number. Each numbered square corresponds to another more detailed map of just that portion of the county bounded by the sides of the square. With a MapeXpert map prepared in step 1 (above) in hand, identify the numbered square on the county map that contains the desired sampling location(s). Record the number of the square. Now (ask for help if necessary) physically locate in the tax office the map corresponding to that number and remove it from its file drawer or rack. The map is usually about 3 ft. square and appears either as a blue on white blueprint of roads, streams and other land structures or as an aerial photograph (see fig 3).

Again with the MapeXpert map from step 1 in hand, locate on the numbered map the exact location of the sampling site. The numbered map will have superimposed on it a mosaic of polygons (see fig 3 again). Each polygon is a parcel. Each parcel will have a whole number in it. This is the **parcel #.** Identify and record the number(s) of all parcels that must be crossed to access the stream site from the nearest road or pathway (power line cut, railroad track, trail etc). When the site is surrounded by multiple parcels, sketch a diagram of the position of each parcel in relation to the sampling site (see fig 4). If possible, sketch the parcels on the computer generated map made in step 1. This information will be important in Step 3.

Now, physically locate county records of land ownership. The records may be index cards in a file drawer, bound in books, or accessed through a computer terminal. Look up property owner(s) by **map #** and **parcel #**. Usually, file drawers and books are labeled by map number. Upon opening the drawer or book, individual records will be filed consecutively by parcel number starting with parcel number 1. Leaf through the records until the desired parcel # appears. Record the name and address of the property owner that appears on the record. As a cross check, look up each landowner by last name in the current year's tax records (another set of bound volumes listing all landowners alphabetically by last name) to obtain a current mailing address. For electronic retrieval, ask office personnel for assistance. Accessing records by computer usually requires that map and parcel numbers be entered into the computer in a specific format. This format often varies from tax office to tax office. With electronic retrieval, the names and addresses that appear on the computer screen will be current.

Step 3: <u>Requesting Permission to Access Stream Sampling Sites</u>

Send each property owner an envelope that includes a letter requesting permission to cross their property, a permission slip that grants permission for you to access their property, and an addressed, stampled envelop in which to return permission slips. An example of the letter used by Region IV to request permission to access a given property and an example of the permission slip that is to be returned by the landowner to EPA are included in this document in Appendix B. Note that the sampling location (e.g. site 175) must appear on the permission slip as well as on the letter requesting permission to access the property. This facilitates the matching of returned permission slips with the appropriate sample locations.

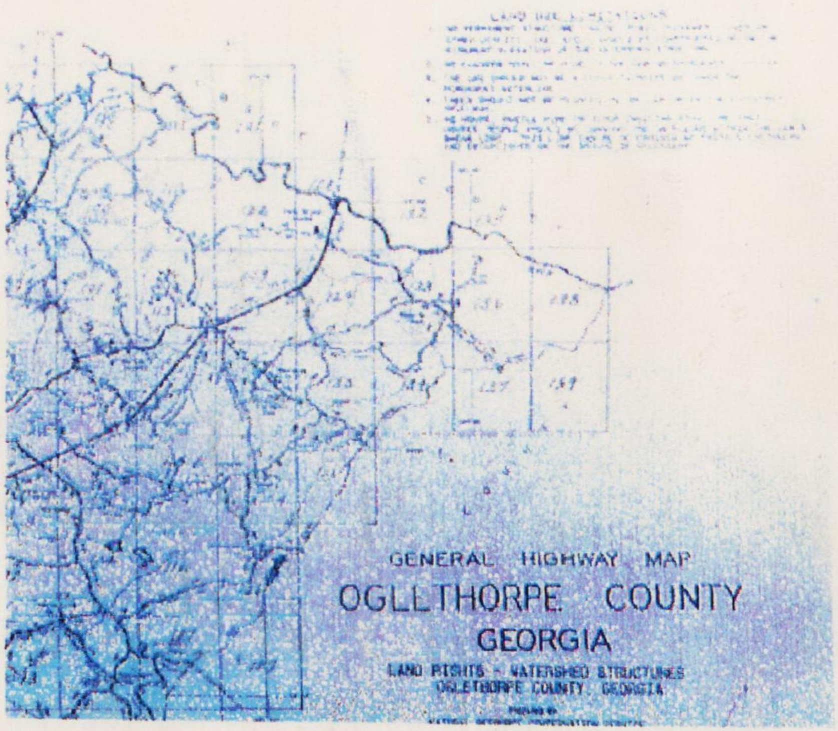


Fig. 2. Typical county map in tax assessor's office.

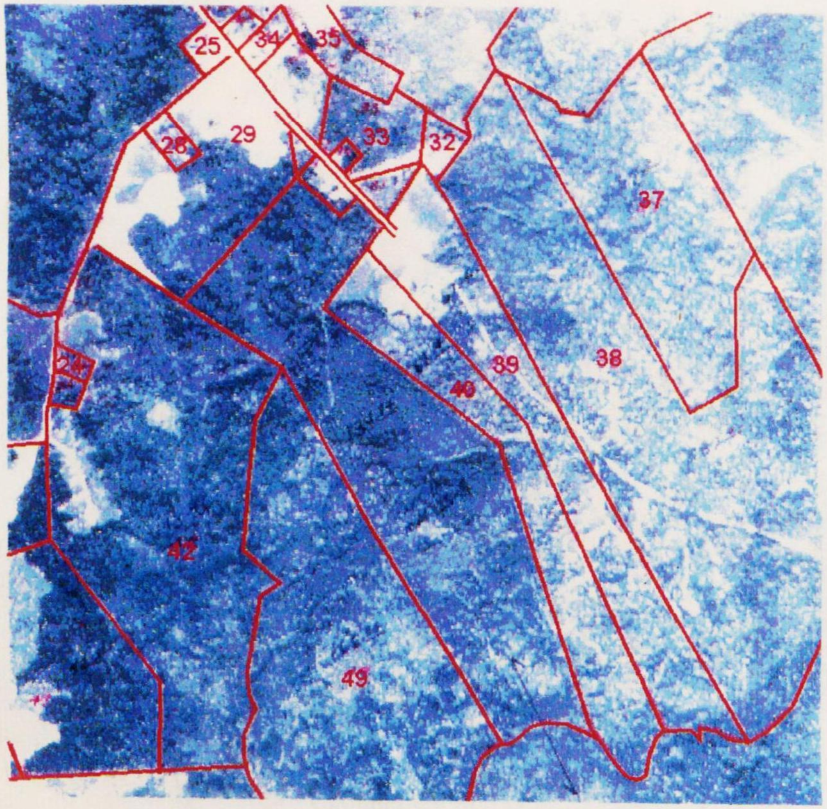
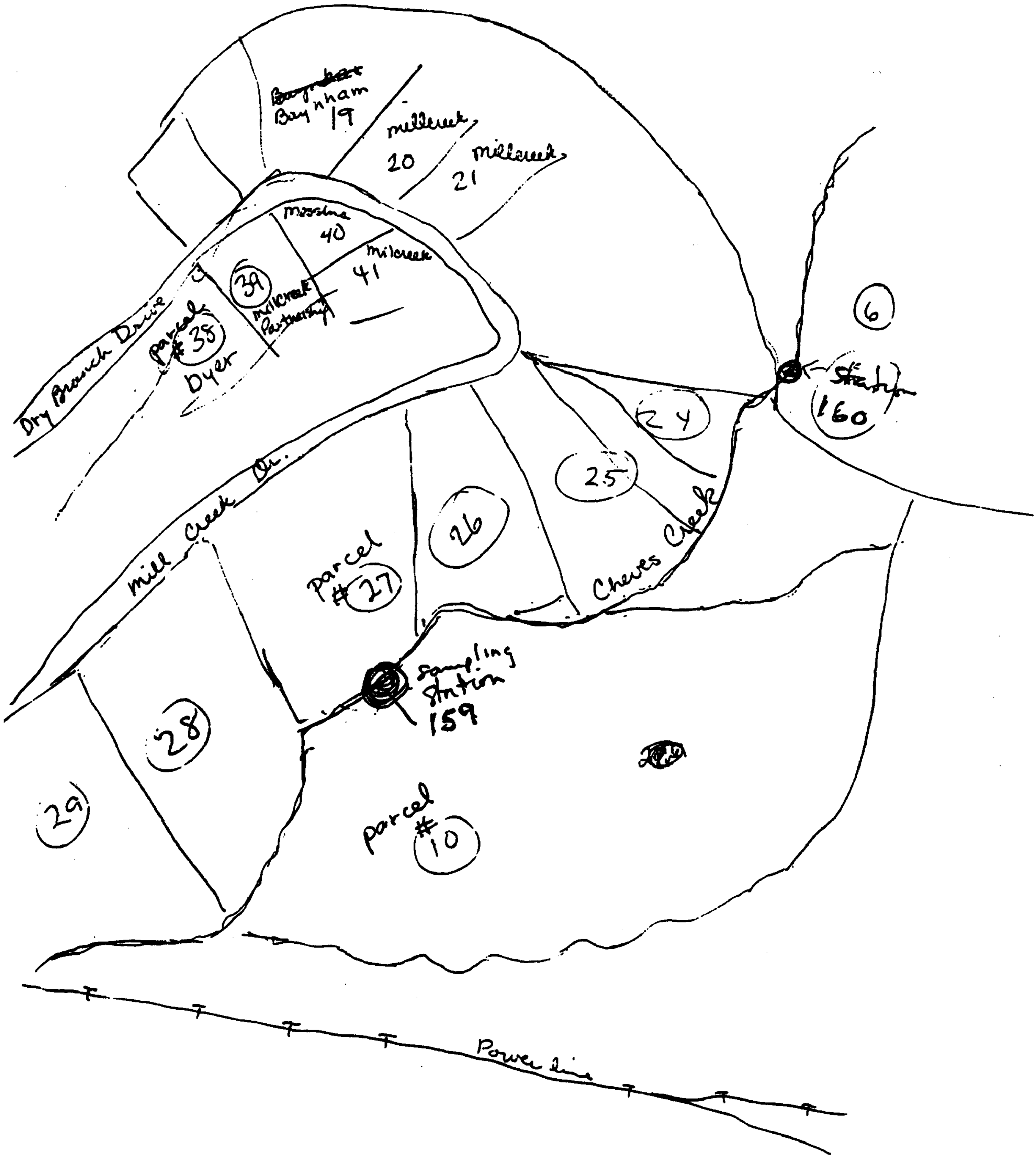


Fig 3. Parcel Map

Fig. 4. Sketch of parcels of land surrounding two sampling stations.



Locating EMAP Sites in the Field - Field Reconnaissance

Equipment and Supplies

Vehicle (4 wheel drive recommended) with a trip odometer
Hand held GPS unit
compass
flagging tape
machete
insect repellent
Maps accurately showing location of sampling sites.
Permission slips from property owners
Laptop computer loaded with software to link GPS signals to mapping program (optional)

Step 1: Drive as close to a stream sample site as possible

This is where a 4 wheel drive vehicle can make a difference. Once a vehicle can go no farther, the site must be within 30 minutes walking time from the vehicle. If not , the site is designated a "non-target" and will not be scheduled for sampling. So, the closer the better.

A factor that can significantly increase the success of finding a stream site is knowing where you are relative to the site. Following county road maps, computer generated maps, and topo maps work well as long as you can recognize landmarks along the route that pinpoint your location on a map. This is often difficult to do on large tracts of undeveloped forest or timberland where many of the roads, especially logging roads, are not marked or not shown on maps. This is where a GPS/Mapping System is employed. The basic system consists of a notebook (lap top) computer outfitted with a CD ROM drive and a GPS Receiver-PCMCIA card. A mapping program is run on the computer while the computer receives information from the GPS receiver. The result is a map on the computer screen which not only displays all previously marked stream locations but also displays the continuously updated position of the vehicle carrying the computer. In this way a vehicle's progress toward a predetermined location can be monitored to guide the vehicle as close the to site as possible, even on unmarked roads. The specifications of the GPS/Mapping system employed by EPA Region IV is as follows:

Notebook Computer

- 486-66mhz processor or greater
- CD ROM drive, 4X or greater
- Active matrix color screen (recommended for viewing screen in bright sunlight)
- 8MB memory minimum (recommend 16 MB or more)

MapeXpert with GPS Link (called Mapkit, $495.00)

available from: Delorme Mapping
2 Delorme Drive
P.O. Box 298
Yarmouth, ME. 04096
1-800-227-1656

GPS Receiver

PCMCIA Card, SATNAV LP NMEA Output Model PM50154	$550.00
with GPSpac software for Windows Model PM50154	$99.00

available from: Centennial Technologies Inc.
180 Cherokee Rd.
Ashville, NC 28804
(704)281-0044

Step 2: Locate the site on foot

From the parked vehicle navigate to the stream site with the aid of a hand-held GPS and a compass. To do this enter the coordinates of the stream site into the GPS unit as a WAYPOINT. Then, after allowing the GPS unit time to acquire your position at the parked vehicle, ask the unit to navigate to the WAYPOINT (refer to the GPS User's Manual for details on how to enter, retrieve, and navigate using way points[1]). The GPS unit will display a bearing (in degrees) and range (in meters or fractions of a mile). Using the compass, determine the direction of travel specified by the bearing on the GPS screen and start walking. Take advantage of deer trails, power line cuts, railroad beds and other pre-existing paths whenever possible. If necessary, hack through underbrush and briers with the machete. The GPS unit will update your progress as you go. Set new bearings with the compass as needed. Technically you have reached a site when the range displayed on the GPS unit reads 0.1 mile (500 feet) or less, but get as close to the stream site as physically possible. The "500 ft." rule was adopted for situations in which it is physically impossible to reach the exact WAYPOINT (e.g. site just over property line where permission to access was denied or stream site is impounded by a beaver dam). EMAP grid units are approximately 2000 feet across. The 500 ft limit should keep the sampling site a valid sampling point within the grid.

---

[1] EPA Region IV utilizes an ENSIGN XL GPS unit manufactured by Trimble Navigation. Copyrights prohibit the reproduction, translation, transformation or adaptation of the ENSIGN User's Manual in any form. Therefore, refer directly to the user's manual for details on the basic operation of the GPS unit and for specific procedures for entering (section 2-17) and navigating to (section 5-7) a waypoint.

Once a stream site is located, mark it with flagging. As you return to the vehicle mark the trail as needed and record walking time (remember, if walking time exceeds 30 minutes, eliminate the sampling site from the sampling schedule). Also eliminate streams that are impounded (e.g. beaver pond) or dry. On the trip back to the vehicle make notes (or take photographs) of any recognizable landmarks (big oak tree, rocks, ridges, gullies etc.) that would help guide someone back to the site should the flagging be lost or torn down. Immediately upon reaching the vehicle, sketch a map showing the trail and position of landmarks relative to the sampling site. As you drive away from the site continue to note landmarks on unimproved roads and use the trip odometer to measure distances between turn offs, forks etc. This information will be needed later to return to the site for the actual sampling.

Upon returning from the field, make necessary changes in maps (e.g. new roads, roads renamed or moved), write clearly written directions to each site, and redraw detailed site maps based on the notes and sketches made in the field.

It is best to conduct field reconnaissance after hunting season and not more than two months before sampling is scheduled. If not, hunters tend to remove flagging and paths hacked through underbrush and thickets tend to become overgrown again, making it difficult to follow paths and relocate the sampling sites.

Direction Packets - for Relocating EMAP Sites in the Field at a Later Date.

Because it may take considerable time to locate EMAP sites in the field, reconnaissance and sampling are treated as separate events. The reconnaissance is completed first. Then, at a later date, teams return to each site for the actual sampling. Because the personnel conducting the sampling may not be the same persons that conducted the field reconnaissance, a direction packet is prepared by the reconnaissance team that provides samplers with all the information needed to locate and safely access each stream site. A direction packet contains:

- Updated computer generated Map pinpointing the sampling site.
- Written directions to each site from the nearest town or main highway that contain highway routes, descriptions (or photos) of landmarks and the detailed site maps sketched during reconnaissance.
- Copies of the permission slips signed by property owners.

A sample direction packet is included at the end of this document as Appendix C.

References

Volstad, J. H., S. Weisberg, D. Heimbuch, H. Wilson, and J. Seibel. 1995. Answers to Commonly Asked Questions about REMAP Sampling Designs and Data Analyses. U. S. Environmental Protection Agency, Research Triangle Park, NC.

Appendix A

Accurately Locating EMAP Coordinates on a Computer Generated Map

Using MapeXpert™ Computerized Mapping Program to Pinpoint a Stream Sampling Site Given a Specific Latitude and Longitude.

MapeXpert ™ version 2.0 for Windows is available from:

DELORME MAPPING
2 Delorme Drive
P.O. Box 298
Yarmouth, Maine 04096
207-227-1656
estimated cost $495

Required Computer Hardware

IBM or 100% IBM-compatible microcomputer outfitted with an Intel® 80386 or higher or 100%-compatible processor.

Minimum of 2 MB RAM (4 MB RAM recommended).

3 MB of available hard drive space.

ISO 9660-compatible CD-ROM drive with Microsoft CD-ROM extensions.

Microsoft Windows ™-compatible mouse (recommended).

Microsoft Windows-compatible VGA card and monitor.

Microsoft Windows-compatible printer (recommended).

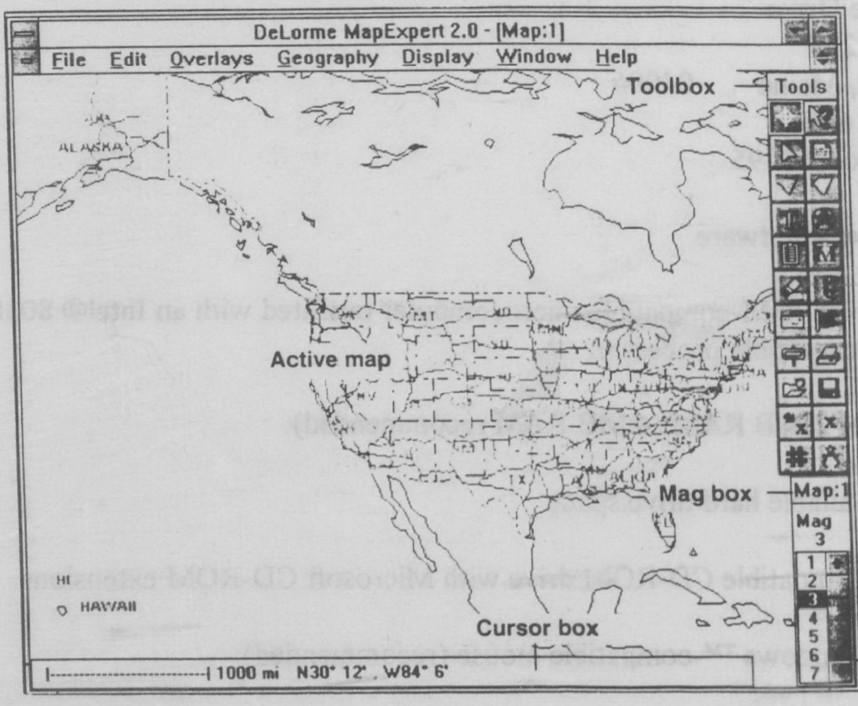
Microsoft Windows version 3.1

MS-DOS® 4.01 or higher.

Additional skills required to operate the MapeXpert program include a basic knowledge of personal computer operation, use of a "mouse," and familiarity with the features and techniques of Windows. The *Windows Tutorial* that accompanies Microsoft Windows Version 3.1 provides and excellent introduction to the skills needed to use MapeXpert.

To locate and label a stream sampling site given a set of map coordinates, use the **Point Box** feature of the MapeXpert program. The procedure is as follows:

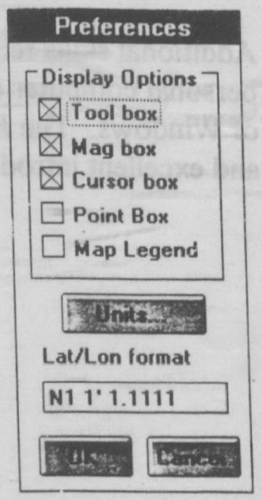
1. Start the MapeXpert program. You will see a map of the United States displayed at a magnification of 3 with the Toolbox, Mag(nification) box, and Cursor box displayed at the margins of the screen (see below).

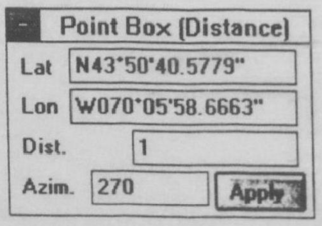


2. Using the mouse, **center** the **cross hairs** of the cursor **over** the region of the state where the **study stream** lies and **click** the left mouse button. This will place the area of the map that contains the study stream in the center of the viewing screen.

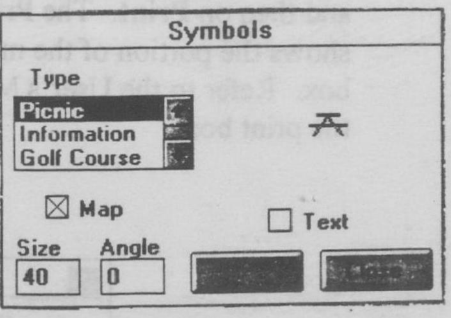
3. Slide the cursor to the Mag box and double **click** on **Mag 9**. Small streams will not be displayed at this magnification. Continue...

4. Move the cursor to the top of the screen . Choose **Display... Preferences...** and then **click** on the check box next to the **Point Box** option in the Preference dialog box (see right). Now **click** on **OK** at the bottom of the Preference box. The point box will appear in the lower right corner of the screen (see below). You will notice that the point box covers the Mag box. At this point **click** and **hold** the left mouse button anywhere on the words **Point Box [Distance]** and **drag** the entire **box** to the **left** until the Mag box is fully exposed.





5. Choose a symbol with which to mark the stream site. A red circle seems to stand out best. To choose a symbol move the cursor to the Toolbox and **click** on the Symbol tool . The Symbols dialogue box will appear (see right). Now **click** on the **Map** check box. This will add color to the symbols. Next use the **scroll** arrows on the **Type** list box to find **circle** and **highlight** it by clicking on it. Finally, **click** on **Close**.



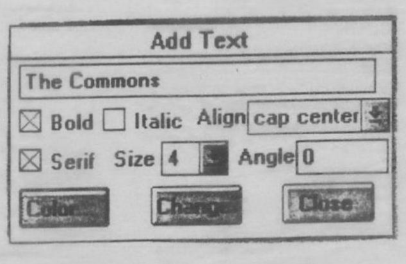
6. Now return to the Point Box to enter the given latitude and longitude. To do this move the cursor into the **Lat** box to the right of the letter **N** and **click**. Type the numbers representing latitude in the order: degrees, minutes, seconds (to the nearest tenth), leaving a space between each component. Next drop the cursor in the **Lon** box to the right of the letter **W, click**, and enter the numbers representing longitude

7. **Click** on the word **Apply** in the bottom right corner of the Point Box. A red circle should appear on the map displayed on the screen at the exact coordinates entered into the Point Box. Close the Point Box by clicking in the square in the upper right hand corner of the box. Now position the cursor (which now appears as a pencil) over the circle and **click** the **right** mouse button once. (If the red circle is not visible, it is probably just off-screen. Temporarily drop to Mag 8 to find the circle, then position the cursor over the circle and click the right mouse button once).

8. Now increase magnification so that even the smallest streams are displayed. This usually requires **Mag 13** or higher. The red circle should now be located over a stream even if no stream was visible under the circle at lower magnification.

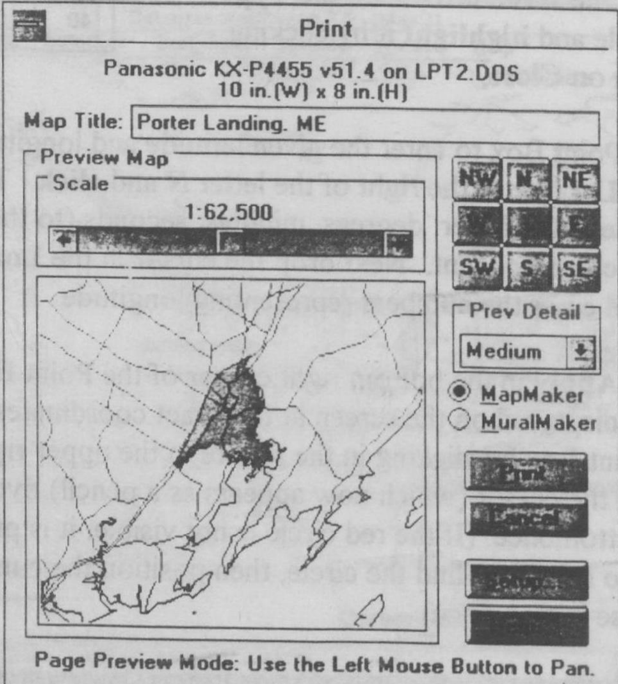
At this point additional sites can be located and marked on the map without repeating steps 1 through 5. by simply entering new numbers for latitude and longitude in the Point Box (step 6) and clicking **Apply** (step 7).

9. Finally, **after** all stream locations have been marked on the map, add text to the map to identify the stream site. To do this **click** on **Text tool** . The cursor should be flashing in the top box of the Text dialogue box (see right). If not, move the cursor into that box and type a label for the stream site. Now move the cursor (still a pencil) to the right, above, or below the red circle (wherever space permits) and **click** the **left** mouse button. The label should appear next to the circle. When finished adding text, **click** on **Close**.



10. To print a copy of the map **click** on the word **File** in the upper left corner of the screen

and then on **Print**. The Print screen will appear (see below). The Preview Map window shows the portion of the map that will be printed. To print the map **click** on the **Print** box. Refer to the User's Manual to use any of the other features (e.g. Scale) displayed in the print box.



11. Any red dot(s) and text added to the map is an **overlay** on the original base map. The overlay can be **saved** for future use by clicking on the word **Overlay** at the top of the screen and then following the "Save As" feature designed for Windows.

Appendix B

Form Letter to Landowners Requesting Permission to Access a Stream Site
and
Permission Slip

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

REGION 4

Science and Ecosystem Support Division
980 College Station Road
Athens, Georgia 30605-2720

Robert Sims
RR2 Box 198
Comer, Georgia 30629

Dear Mr. Sims:

The Science and Ecosystem Support Division of the United States Environmental Protection Agency, Region IV, in conjunction with the states of South Carolina and Georgia, will be conducting biological stream monitoring of the Savannah River Basin. The study, known as the Savannah River - Regional Environmental Monitoring and Assessment Program (SR-REMAP), is intended to gather data on condition of stream ecosystems in the Savannah River Basin, identify stressors to these systems, and provide a baseline of information for future trends. The data is not intended for uses pertaining to litigation, tax assessment, law enforcement, or similar purposes.

Randomly selected sites have been chosen in the Savannah River Basin and a site or sites are located on your property. The EPA and the state agencies assisting in this effort are seeking permission to access the following site(s):

211

Scull Shoal Creek

Enclosed you will find a self-addressed envelope and access permission form. Please sign the access permission form and mail by March 21, 1997. We will visit the sampling site on two occasions, once for fish sampling and once for macroinvertebrate (aquatic insects) sampling. On receiving permission, we will contact you by telephone or letter informing you of the scheduled sampling dates for the stream(s) on or near your property.

Mr. Hoke Howard of EPA, Region IV, Science and Ecosystems Support Division is the project coordinator for the SR-REMAP team. He can be reached at EPA in Athens, Georgia at (706)355-8721 if you have questions concerning this project. The entire team is grateful for your willingness to participate in this project.

Sincerely,

Hoke S Howard

Hoke S. Howard
Project Coordinator
SAVANNAH RIVER

SAVANNAH RIVER
REGIONAL ENVIRONMENTAL MONITORING AND ASSESSMENT PROGRAM

ACCESS TO STREAM MONITORING SITES

I,____, owner or representative of the owner of property adjacent to ____, grant permission to the staff of the U.S. EPA and state agencies assisting in the Savannah River Regional Environmental Monitoring and Assessment Program (SR-REMAP) access to said property for the purpose of stream monitoring as part of the SR-REMAP.

SIGNATURE ____

DATE ____

Site 138
Pickens Co. SC
M. Fork Twelve Mile Ck.

SAVANNAH RIVER
REGIONAL ENVIRONMENTAL MONITORING AND ASSESSMENT PROGRAM

ACCESS TO STREAM MONITORING SITES

I, John W Finley, owner or representative of the owner of property adjacent to __________ __________, grant permission to the staff of the U.S. EPA and state agencies assisting in the Savannah River Regional Environmental Monitoring and Assessment Program (SR-REMAP) access to said property for the purpose of stream monitoring as part of the SR-REMAP.

SIGNATURE John W Finley

DATE 2-23-96

138 - Middle fork twelve mile creek

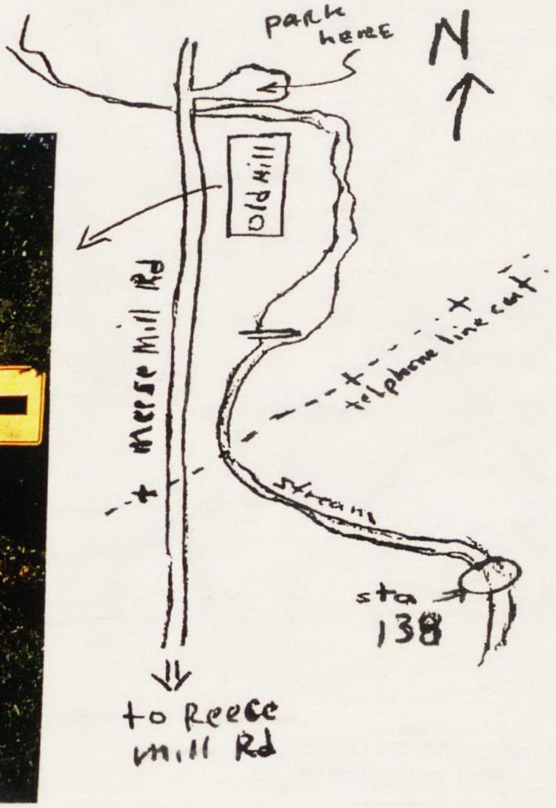
Appendix C

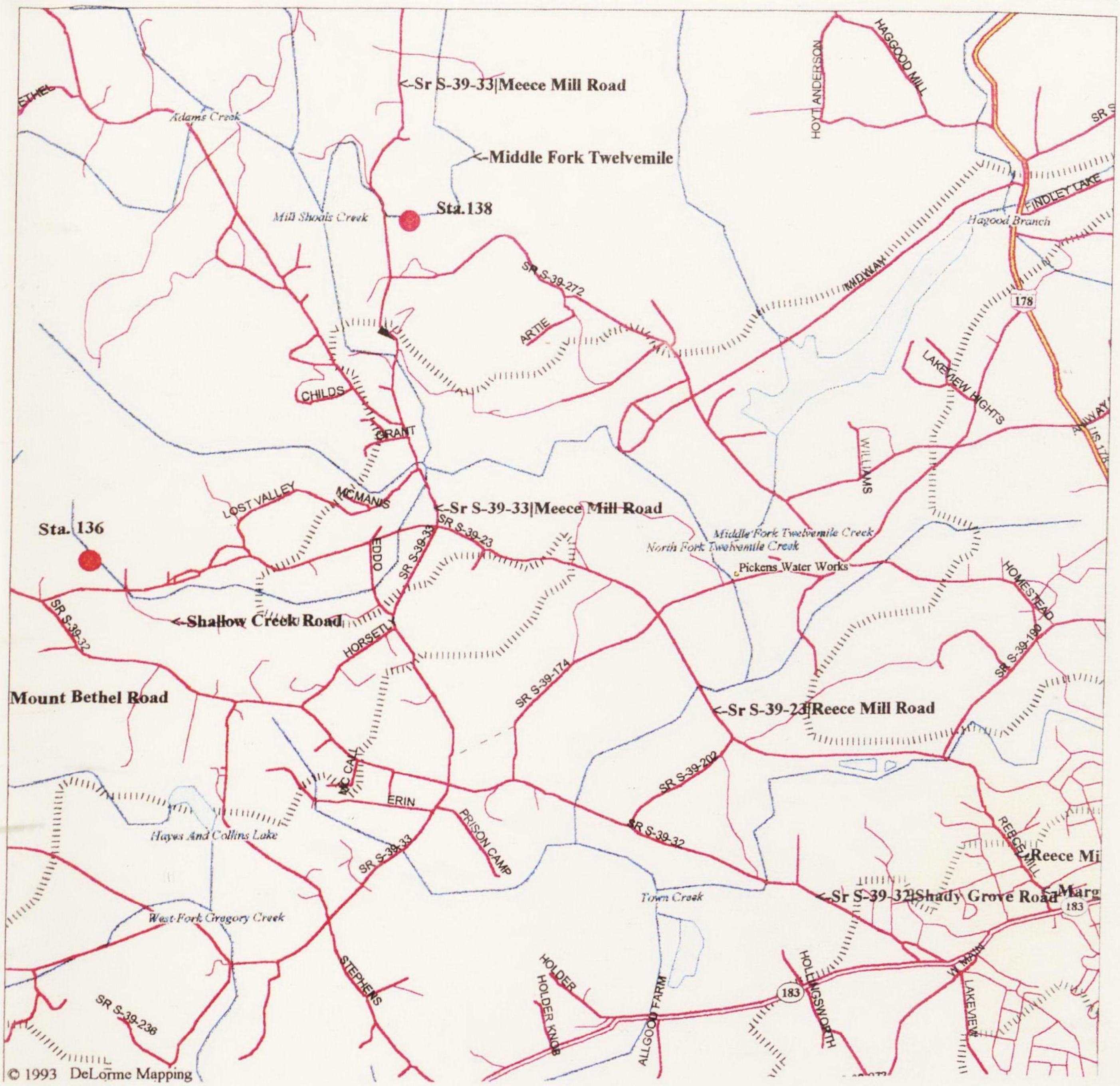
Sample Direction Packet

<u>Site 138</u> <u>Pickens Co. SC</u>

This site is located about 4 miles east of Pickens, SC off Meese Mill Rd. (not to be confused with Reece Mill Rd.)

To reach the site leave Pickens on Reese Mill Rd. as you did for Site 136. Follow Reece Mill Rd. for about to 3 mi. watching for the Meese Mill Rd. intersection. Turn right onto Meese Mill Rd. and drive approx. 1.4 mi. to an old mill (red building) on the right side of the road (see photo). There is a small parking area on the right Immediately after the mill (see inset map). Park . From the Mill walk back uphill to a telephone line cut. Enter the stream here and walk upstream about 100' to the site. The site is not flagged.





LEGEND

Geo Feature

US Highway

Population Center

Street, Road

Major Street/Road

US Highway

River

Utility (powerline)

Open Water

Contours

Scale 1:37,500 (at center)

2000 Feet

1000 Meters

Mag 14.00

Thu Jul 25 15:57:52 1996

E

Appendix E

Cramer Von Mises Test for Environmental Data

By

Stephen L. Rathbun

CRAMER-VON MISES TESTS FOR ENVIRONMENTAL MONITORING DATA

by

Stephen L. Rathbun, Richard Houghton

Department of Statistics

University of Georgia

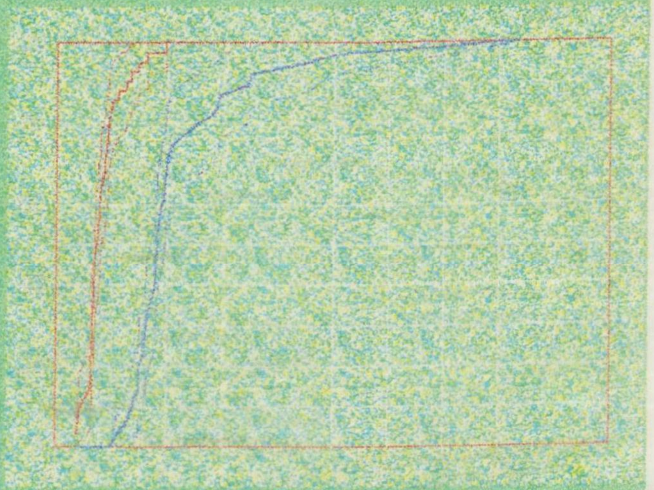
Athens, Georgia

Christina Laurin

FTN Associates, LTD

3 Innwood Circle, Suite 220

Little Rock, Arkansas 72211



1 Introduction

One of the objectives of both the Savannah River Initiative and the South Florida Initiative is to detect trends in important environmental variables (*e.g.*, algal growth potential test, mercury) over both time and space. Both of these environmental monitoring initiatives are conducted under the auspices of the Regional Environmental Monitoring and Assessment Program (REMAP), and employ the probability sampling design of the Environmental Monitoring and Assessment Program (Overton, White, and Stevens 1991) on a regional scale. Observations are collected over time and space in a serially alternating design (Urquhart, Overton, and Birkes 1993) that is spatially interpenetrating (Overton et al. 1991). The region of interest is partitioned into a grid of contiguous hexagonal quadrats, and quadrats are systematically partitioned into four cycles, corresponding to annual (Savannah River Initiative) or biannual (wet and dry seasons in the South Florida Initiative) sampling times. The interpenetrating component of the design comes from assigning two each of the six neighboring hexagons to the three remaining cycles. Thus, a hexagon assigned to cycle 0, will have a pair of neighboring hexagons each assigned to cycles 1, 2, and 3. Following the assignment of hexagons to cycles, sample points are randomly located within hexagons belonging to each cycle according to some probability sampling design. Then, in each of the first four sampling intervals, sites assigned to a successive cycles are sampled. This sampling pattern is then repeated in subsequent groups of four successive sample intervals.

The following considers statistical methods for testing the null hypothesis that the data obtained from two or more cycles are identically distributed. Although this discussion is couched in terms of comparing observations over time, these methods can also be used to test the null hypothesis that observations from different subregions are identically distributed. For example, Section 5.2 considers a test of the null hypothesis that data from different stream orders are identically distributed.

This paper is restricted to design-based methods of statistical inference. For each cycle, say t, the sampling units are locations, and the population is comprised of the collection of locations $\mathbf{s}$ in the region A_t, the set of hexagons assigned to that cycle. For design-based inference, the value of the variable of interest $Z_t(\mathbf{s})$ at a location $\mathbf{s} \in A_t$ is assumed to be fixed, not random. For each cycle, the data $Z_t(\mathbf{s}_{t1}), \cdots, Z_t(\mathbf{s}_{tn_t})$ are obtained from a probability sampling design, under which the locations $\mathbf{s}_{t1}, \cdots, \mathbf{s}_{tn_t}$ are sampled with known probabilities. The simplest example of such a design is the simple random sampling, where $\mathbf{s}_{t1}, \cdots, \mathbf{s}_{tn_t}$ are independently sampled from a uniform distribution on A. For design-based statistical inference, the source of random variation is the random selection of sample sites. This is in contrast to model-based statistical inference, where the source of random variation is in the assumed statistical model (*e.g.*, a regression model). Thus, design-based statistical inference has the advantage that no model assumptions are required. Design-based statistical inference for spatial sampling designs, such as employed by REMAP, is introduced by Cordy (1993), who considers Horvitz-Thompson estimation of population parameters.

There are several approaches that may be taken to comparing 2 or more cycles. Assuming homogeneity of variance, and that independent simple random samples are obtained from normally distributed populations with identical variances, a one-way analysis of variance may be used to test the null hypothesis that the population means are identical against the general alternative that at least one population mean is different. For large sample sizes, the central limit theorem says that sample means are approximately normally distributed even if the original data are not; so, the normality assumption can be relaxed. If a nonparametric procedure is desired, the Kruskal-Wallis test (Hollander and Wolfe 1973) can be used to test for identical population means.

More generally, we may wish to avoid making any assumptions concerning the forms of the distributions of the populations we wish to compare. Thus, we may wish to test the null hypothesis that the populations are identically distributed. Since the distribution of a population can be characterized through its cumulative distribution function (*cdf*), this is equivalent to testing the null hypothesis that the *cdf*'s are identical. There are two general classes of test statistics for comparing *cdf*'s. Kolmogorov-Smirnov test statistics are based on largest absolute differences between *cdf*'s, while Cramér-von Mises test statistics integrate squared differences between *cdf*'s over the possible values of the variable of interest. Since the latter looks at differences between *cdf*'s at more than one point, and not just the point where absolute differences are largest, Cramér-von Mises tests should be more powerful than Kolmogorov-Smirnov tests. The large-sample distributions of both classes of test statistics under simple random sampling were tabulated by Kiefer (1959).

This paper shall consider the application of Cramér-von Mises tests to the Savannah River and South Florida data. A Quattro-Pro Template for computing Cramér-von Mises test statistics shall be described. The South Florida Initiative involves data collection in four "cycles" corresponding to two different seasons (wet and dry) over a two year period. An "Analysis of Variance" analog shall be developed for partitioning variation in *cdf*'s between cycles into sources of variation due to year, season, and year by season interaction. Kiefer's large-sample results assume that simple random samples are obtained from each population, but neither the Savannah River nor the South Florida initiatives use simple random sampling designs. Results of a simulation study will be presented to investigate the distribution of the test statistic under sampling designs used by the two initiatives.

2 Cumulative Distribution Function

The distribution of data obtained from cycle t can be characterized through its cumulative distribution function (*cdf*). Since cycles are systematically assigned to hexagons, different regions are sampled in each cycle; let A_t denote the region sampled in cycle t. For south Florida marshes, A_t corresponds to that portion of the hexagons assigned to cycle t that are in marshlands, while for South Florida canals, and Savannah River Basin streams, A_t denotes the portion of the hexagons in these waterways. Let $Z_t(\mathbf{s})$ denote the variable of interest at location $\mathbf{s}$ in region A_t. Then the cumulative distribution function for $Z_t(\cdot)$ is defined to be

$$F_t(z) = \frac{1}{|A_t|} \int_{A_t} I\{Z_t(\mathbf{s}) \leq z\} d\mathbf{s} \,, \tag{1}$$

where the indicator function $I\{Z_t(\mathbf{s}) \le z\}$ is equal to one if $Z_t(\mathbf{s}) \le z$ and is equal to zero if otherwise, and $|A_t|$ is the area of region A_t. For rivers, streams, and canals, the integral is over the lengths of these waterways, and $|A_t|$ becomes the total length of waterways in region (cycle) t. The function $F_t(z)$ can be interpreted as the portion of the area (length) of region A_t for which the variable of interest takes values less than or equal to z.

Since it is not possible with a finite budget to observe $Z_t(\mathbf{s})$ at all locations $\mathbf{s} \in A_t$, the population *cdf* $F_t(z)$ is unknown and must be estimated from a sample. Let $z_{t1}, \cdots, z_{tn_t}$ denote the values of the variable of interest at the n_t sites sampled at time t, and let $\pi_1(t), \cdots, \pi_{n_t}(t)$ denote the corresponding inclusion probability densities. The inclusion probability $\pi_i(t)$ is defined to be equal to the likelihood that the i-th site is included in the sample at time t. Then the cumulative distribution function at time t may be estimated using the Horvitz-Thompson estimator

$$\hat{F}_t(z) = \frac{1}{|A_t|} \sum_{i=1}^{n_t} \frac{I\{z_{ti} \le z\}}{\pi_i(t)} \tag{2}$$

(Cordy 1993). If sampling probabilities are equal (within cycles), as in the case of South Florida canals, (2) reduces to

$$\hat{F}_t(z) = \frac{1}{n_t} \sum_{i=1}^{n_t} I\{z_{ti} \le z\}. \tag{3}$$

The variance of $\hat{F}_t(z)$ may be estimated by

$$\hat{V}_1(t) = \frac{1}{|A_t|^2} \left\{ \sum_{i=1}^{n_t} \frac{I\{z_{ti} \le z\}}{(\pi_i(t))^2} + \sum_{i \ne j} \frac{\pi_{ij}(t) - \pi_i(t)\pi_j(t)}{\pi_i(t)\pi_j(t)\pi_{ij}(t)} I\{z_{ti} \le z\} I\{z_{tj} \le z\} \right\} \tag{4}$$

where $\pi_{ij}(t)$ is the pairwise inclusion probability density that both sites i and j are included in the sample at time t. If sample sites are selected according to a simple random sampling design, the variance of $\hat{F}_t(z)$ is given by

$$\hat{V}_2(t) = \frac{\hat{F}_t(z)\{1 - \hat{F}_t(z)\}}{n_t - 1} \tag{5}$$

More generally, if sample sites are selected independently with inclusion probabilities $\pi_i(t)$, then the variance of $\hat{F}_t(z)$ may be estimated by

$$\hat{V}_3(t) = \frac{\hat{F}_t(z)\{1 - \hat{F}_t(z)\}}{|A_t|^2} \sum_{i=1}^{n_t} \left(\frac{1}{\pi_i(t)}\right)^2 \tag{6}$$

The (emprical) *cdf* $\hat{F}_t(z)$ is most readily interpreted by plotting $\hat{F}_t(z)$ against z. Figure 1 shows a plot of the *cdf* for total mercury in water in cycle 0 of the south Florida canals. Notice that the *cdf* is a step function (solid line); each step occurs at the location of a data point so that all of the information contained in the data is retained. The curve is steepest at low levels of total mercury indicating that about 40% of the values lie below 1.2 ng/ℓ, and an additoinal 45% of the values lie between 1.2 and 4.5 ng/ℓ. The curce is very shallow for large values of total mercury indicating that only 15% of the values lie above 4.5 ng/ℓ. The dashed lines give 95% confidence bands for $\hat{F}_t(z)$. Notice that these bands are widest for intemediate values of total mercury and that the width of these bands converges to zero as total mercury decreases towards the smallest observed value, or increases towards its largest observed value.

Although $\hat{V}_1(t)$ is an unbiased estimator for the variance of $\hat{F}_t(z)$, it is unstable and can sometimes take negative values. Moreover, $\hat{V}_1(t)$ requires values for the pairwise inclusion probability densities $\pi_{ij}(t)$, which cannot be easily obtained for either South Florida canals or Savannah River Basin streams. Thus, two *ad hoc* procedures may be considered:

Procedure 1: If inclusion probability densities are identical, then we might treat the data as if it came from a simple random sampling design, and the variance of $\hat{F}_t(z)$ might be estimated using $\hat{V}_2(t)$. For South Florida canals, this might be justified as follows: Here canals are partitioned and placed in random order. Then sites are placed along the randomized canal segments according to a systematic design. Now consider partitioning canals into larger and larger numbers of smaller and smaller segments. If we assume that canal segments are placed in completely random order, then as the length of the smallest canal segment converges to zero, the distribution of the sample sites converges to that of a simple random sampling design. Thus, assuming that canal segments are partitioned finely enough, the sample sites can be treated as if they came from a simple random sampling design. Note, however, that canal segments are not placed in completely random order, but according to a clustered sampling design in which clusters of canals are placed in random order, and then locations of canals are randomized within clusters. This was done to achieve better spatial coverage of sample sites. It also has the consequence that $\hat{V}_2(t)$ should over-estimate the variance of $\hat{F}_t(z)$, as confirmed by results of simulations to be described later. A better estimate of the variance of $\hat{F}_t(z)$ might be achieved using the post-stratification estimator given by Procedure 2.

Procedure 2: In the Savannah River Basin, one might expect differences in environmental variables between different orders of streams. Likewise, there may be differences among and between the various water conservation areas, Big Cypress National Preserve, and the Everglades National Park in South Florida. Similar differences may also exist between different canal reaches in South Florida. In response to this heterogeneity in environmental conditions, the respective sample regions can be partitioned into strata corresponding to orders of streams in the Savannah River Basin, water conservation areas and parks for south Florida marshes, and canal reaches in South Florida. The sample designs used in both environmental monitoring initiatives further lend themselves to this post-stratification, since, within each cycle, sampling probability densities are constant within these strata.

Let Z_{thi} denote the data from sample site i in stratum h in cycle t, let n_h denote the number of observations from stratum h, and let L denote the number of strata. Then the cumulative distribution function in stratum h and t is estimated by

$$\hat{F}_{th}(z) = \frac{1}{n_{th}} \sum_{i=1}^{n_{th}} I\{z_{thi} \leq z\}, \tag{7}$$

and the variance of $\hat{F}_{th}(z)$ is approximately

$$\tilde{V}_{th} = \frac{\hat{F}_{th}(z)\{1 - \hat{F}_{th}(z)\}}{n_{th} - 1}. \tag{8}$$

The cumulative distribution function for the population in cycle t is then given by expression (2), but an approximate expression for its variance is given by

$$\hat{V}_3(t) = \frac{1}{|A_t|^2} \sum_{h=1}^{L} \left(\frac{n_{th}}{\pi_{th}}\right)^2 \tilde{V}_{th} \ , \tag{9}$$

where π_{th} is the inclusion probability density for sites in stratum h in cyle t. This approximation is based on treating sample sites as if there generated by a simple random sampling design within strata.

3 Cramér-von Misés Test

Consider testing the null hypothesis that the cumulative distribution functions (*cdf* 's) of two or more populations are identical against the general alternative that at least one of the *cdf* 's is different. A Cramér-von Mises test statistic may be defined as

$$\begin{aligned} W &= \int_{-\infty}^{\infty} \sum_{t=1}^{k} n_t |\hat{F}_t(z) - \bar{F}(z)|^2 d\bar{F}(z) \\ &= \frac{1}{n} \sum_{t=1}^{k} n_t \sum_{i=1}^{n_t} |\hat{F}_t(z_{ti}) - \bar{F}(z_{ti})|^2 \end{aligned} \tag{10}$$

where

$$\bar{F}(z) = \frac{1}{n} \sum_{t=1}^{k} n_t \hat{F}_t(z), \tag{11}$$

and $n = n_1 + \cdots + n_k$. The function $\bar{F}(z)$ can be interpreted to be the average *cdf* over all cycles. The right-hand side of (10) involves integration of the squared difference between the *cdf* 's $\hat{F}_t(z)$ and the average *cdf* $\bar{F}(z)$ over all possible values of the variable z. Thus, *cdf* 's are compared over the whole range of the variable, not just at a single point as in the k-sample Kolmogorov-Smirnov test, yielding improved power when compared to the latter testing procedure.

The large-sample distribution of W was derived by Kiefer (1959), under the assumption of simple random sampling from all populations. Table 1 gives the critical values for α-level tests of $H_o : F_1 = F_2 = \cdots = F_k$. Thus, an $\alpha = 0.05$ level test for equality of *cdf* 's over $k = 4$ cycles would reject H_o if $W \geq 1.2373$.

As indicated above, the critical values in Table 1 are obtained under the assumption that simple random samples are obtained from each of the k-populations. However, none of the data considered here are obtained from simple random samples. Results of Monte Carlo simulations (Section 6) suggest that the critical values in Table 1 are conservative in the sense that we may fail to reject the null hypothesis when it is false less often than we should. Likewise, under the null hypothesis, the true α level is less than what is tabulated in Table 1.

The South Florida Initiative involves the collection of data over two seasons (wet and dry) over a two-year period. If we reject the null hypothesis that the *cdf* 's over the four cycles, then the next step would be to ask if there are significant differences between years, between seasons, or if there is an interaction between years and seasons. Since the integrand on the right-hand side of expression (10) takes the form of a sum of squares over all k-cycles, we can partition W into terms for testing differences between years, seasons, and interactions between years and seasons. Assume that the data are balanced; that is, $n_1 = n_2 = \cdots = n_k$; let n denote the common sample size for all seasons. Suppose observations are collected over a years and b seasons. Let z_{ijk} denote the observation from sample k in season j in year i, let $\hat{F}_{ij}(z)$ denote the *cdf* for season j in year i, let

$$\bar{F}_{i\cdot}(z) = \frac{1}{b}\sum_{j=1}^{b}\hat{F}_{ij}(z) \tag{12}$$

denote the average *cdf* for year i, let

$$\bar{F}_{\cdot j}(z) = \frac{1}{a}\sum_{i=1}^{a}\hat{F}_{ij}(z) \tag{13}$$

denote the average *cdf* for season j, and let

$$\bar{F}_{\cdot\cdot}(z) = \frac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\hat{F}_{ij}(z) \tag{14}$$

denote the average *cdf* over all seasons and years. Then, we may partition the Cramér-von Mises test statistic as follows:

$$W = W_y + W_s + W_{y\times s} \ , \tag{15}$$

where

$$\begin{aligned} W &= n\int_{-\infty}^{\infty}\sum_{i=1}^{a}\sum_{j=1}^{b}|\hat{F}_{ij}(z) - \bar{F}_{\cdot\cdot}(z)|^2 d\bar{F}_{\cdot\cdot}(z) \\ &= \tfrac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}|\hat{F}_{ij}(z_{ijk}) - \bar{F}_{\cdot\cdot}(z_{ijk})|^2 \end{aligned} \tag{16}$$

tests for variation between cycles,

$$\begin{aligned} W_y &= nb\int_{-\infty}^{\infty}\sum_{i=1}^{a}|\bar{F}_{i\cdot}(z) - \bar{F}_{\cdot\cdot}(z)|^2 d\bar{F}_{\cdot\cdot}(z) \\ &= \tfrac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}|\bar{F}_{i\cdot}(z_{ijk}) - \bar{F}_{\cdot\cdot}(z_{ijk})|^2 \end{aligned} \tag{17}$$

tests for variation between years,

$$\begin{aligned} W_s &= na\int_{-\infty}^{\infty}\sum_{j=1}^{b}|\bar{F}_{\cdot j}(z) - \bar{F}_{\cdot\cdot}(z)|^2 d\bar{F}_{\cdot\cdot}(z) \\ &= \tfrac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}|\bar{F}_{\cdot j}(z_{ijk}) - \bar{F}_{\cdot\cdot}(z_{ijk})|^2 \end{aligned} \tag{18}$$

tests for variation between seasons, and

$$\begin{aligned} W_{s\times y} &= n\int_{-\infty}^{\infty}\sum_{i=1}^{a}\sum_{j=1}^{b}|\hat{F}_{ij}(z) - \bar{F}_{i\cdot}(z) - \bar{F}_{\cdot j}(z) + \bar{F}_{\cdot\cdot}(z)|^2 d\bar{F}_{\cdot\cdot}(z) \\ &= \tfrac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}|\hat{F}_{ij}(z_{ijk}) - \bar{F}_{i\cdot}(z_{ijk}) - \bar{F}_{\cdot j}(z_{ijk}) + \bar{F}_{\cdot\cdot}(z_{ijk})|^2 \end{aligned} \tag{19}$$

tests for interaction between years and seasons. The interaction between years and seasons is more easily computed by subtraction:

$$W_{s\times y} = W - W_y - W_s.$$

If $a = b = 2$, then we may also compute

$$\begin{aligned} W_{s\times y} &= n\int_{-\infty}^{\infty}\left\{\left|\frac{\widehat{F}_{11}(z)+\widehat{F}_{22}(z)}{2} - \mathcal{F}_{\cdot\cdot}(z)\right|^2 - \left|\frac{\widehat{F}_{12}(z)+\widehat{F}_{21}(z)}{2} - \mathcal{F}_{\cdot\cdot}(z)\right|^2\right\} d\mathcal{F}_{\cdot\cdot}(z) \\ &= \frac{1}{4}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{n}\left\{\left|\frac{\widehat{F}_{11}(z_{ijk})+\widehat{F}_{22}(z_{ijk})}{2} - \mathcal{F}_{\cdot\cdot}(z_{ijk})\right|^2 - \left|\frac{\widehat{F}_{12}(z_{ijk})+\widehat{F}_{21}(z_{ijk})}{2} - \mathcal{F}_{\cdot\cdot}(z_{ijk})\right|^2\right\} \end{aligned} \tag{20}$$

The α-level tests for the various null hypotheses may then be tested as follows:

- The null hypothesis that there are no differences between cycles is rejected if $W \geq W_{\alpha,ab}$.
- The null hypothesis that there are no differences between years is rejected if $W_y \geq W_{\alpha,a}$.
- The null hypothesis that there are no differences between seasons is rejected if $W_s \geq W_{\alpha,b}$.

4 Analysis of *cdf*'s using Quattro Pro

The QuattroPro template CDFTEST2.WB1 can be used to calculate cumulative distribution functions for environmental monitoring data and 95% confidence intervals assuming simple random sampling. The template is set up to use three elements of the data; the station name, a grouping variable (*e.g.*, cycle, year, season, stratum), and the variable to be analyzed. The grouping variable must be a numerical variable with values between 0 and 9 yielding a possible 10 groups. Currently the template is limited to 500 total observations. Before using it, a working copy of the template CDFTEST2.WB1 should be created, and a backup copy should be kept in case the working copy is damaged. The template file is rather large, therefore we advise that a new working copy be created for each new variable analyzed instead of putting multiple analyses in a single file on multiple pages. Large Quattro Pro files become very slow to work with and are prone to causing memory errors.

A step by step procedure for using the Quattro Pro template is outlined as follows:

Step 1: Start Quattro Pro and open the Quattro Pro data file that contains the data for the variable of interest.

Step 2: Open the working copy of the Quattro Pro template file CDFTEST2.WB1.

Step 3: Go into the data file, copy the required columns (station ID, grouping variable, variable of interest), and paste them into page *B* of the template file. If sampling probabilities are not equal, they also must be copied and pasted into page *B* of the template file. Grouping variables may include, *e.g.*, cycles, years, seasons, or stream orders, and must take numerical values between 0 and 9. On page *B* of the template file, any extra data or lines containing missing values for the variable of interest must be deleted using block delete.

Step 4: Make sure the data in the grouping variable column, variable column, and sampling probability column are numerical data instead of alphanumeric. To convert alphanumeric data to numeric data, mark the columns to be changed, and perform a Search and Replace on the label indicator (', ", or ^), replacing these symbols with a plus sign (+). Search and Replace is either under EDIT in the main menu, or Block (Quattro Pro 6.0).

Step 5: Sort the data based on the variable of interest. Use the mouse to mark all of the data in the columns for station ID, grouping variable, variable of interest, and if necessary, sampling probability. Do not mark the column titles. The Sort option is under DATA in Quattro Pro 5.0 and under Block in Quattro Pro 6.0. Click in the first box in the Sort Keys section of the Sort box. Then on the notebook page, click on the top of the column with the variable of interest to select the entire column as the sort key. Click OK and then the data will be sorted.

Step 6: Copy the sorted data into the corresponding columns of page A using Copy and Paste. Do **not** select the entire column to copy to page A, select **only** the block of data. On page A, the data starts in the third row; *i.e.*, A3. Once the data is copied, the calculations and graphs are complete. The estimated *cdf* 's for each group are found in columns P to Y, and corresponding 95% confidence intervals appear in columns AK to BD. The average *cdf* appears in column Z. The Cramér-von Mises statistic for testing the null hypothesis that groups have identical *cdf* 's can be found at AL505.

Step 7: Save the completed template file under a new name.

Step 8: Graphs of the first four *cdf* 's and their corresponding 95% confidence bands are created in the completed file. Before these are printed, titles need to be modified to reflect the parameter being analyzed, and how they are grouped. Pull down Graph on the menu and select Edit Graph. Select the graph you wish to edit and a graph box will appear on the screen. Once the graph is open, click on the Graph menu again and select Titles. The variable, media, and source are specified in the Main Title. The group is specified as the Subtitle. Make sure the correct variable and units are specified in the Main Title and the axes titles. Once the necessary corrections are complete, click OK.

Step 9: To print a graph, open it (using Edit Graph or some other method) and then select *File/Print* in the main menu.

Step 10: If changes to the graphs are made, be sure to save the file again after they are complete.

Step 11: Close the file, reopen the template file and start again with the next variable of interest.

5 Examples

The following illustrates the use of the Cramér-von Mises test on data from both the Savannah River Initiative and the South Florida Initiative.

5.1 South Florida Canals

Data on a wide variety of physical variables were obtained from water and sediment samples collected from canals in southern Florida during the wet and dry seasons over a two-year period (fall 1993 to spring 1995). Fifty observations were available from each sampling cycle. The following considers temporal variation in total mercury and methyl mercury in water samples. To test for interaction between seasons and years, a grouping variable was defined to be equal to 0 if the observation was from cyles 0 or 3 (wet season year one or dry season year two, respectively), and equal to 1 if the observation was from cycles 1 or 2 (dry season year one or wet season year two, respectively).

Figure 2 depicts the *cdf*'s (solid lines) and corresponding 95% confidence bands (dashed lines) for methyl mercury in cycles 0, 1, 2, and 3. Note that the curves rise more gradually in cycles 0 and 3 than in cycles 1 and 2. This indicates that water samples from the former cycles not only tend to contain higher concentrations of methyl mercury, but also that methyl mercury readings show greater variability in these two cycles. Moreover, since cycle 0 was carried out in the wet season of the first year, while cycle 3 was carried out in the dry season of the second year, this pattern to temporal variation in *cdf* 's suggests a strong interaction between year and season. This result is confirmed by the Cramér-von Mises tests (Table 2). First the null hypothesis that the four *cdf* 's are identical is firmly rejected ($W = 4.69$; $k = 4$; $p < 0.0001$) with $W = 4.69$. The partitioning of W into terms for years, seasons, and year by season interaction indicates that while season accounts for most of the variation between cycles ($W_s = 3.11$; $k = 2$; $p < 0.0001$), there is a significant interaction between year and season ($W_{y \times s} = 1.31$; $k = 2$; $p = 0.003$). However, there is no evidence for a significant difference between years ($W_y = 0.28$; $k = 2$; $p = 0.494$).

Figure 3 depicts the *cdf* 's (solid lines) and corresponding 95% confidence bands (dashed lines) for total mercury. The curves for the wet season (cycles 0 and 2) rise more gradually than those for the dry season (cycles 1 and 3). This suggest the total mercury concentrations are not only higher in the wet season, but they also show greater variability. The Cramér-von Mises test W gives strong evidence that four *cdf* 's are not identical ($W = 7.24$; $k = 4$; $p < 0.0001$). The partitioning of W suggests that the year by season interaction accounts for most of the variation between cycles ($W_{y \times s} = 4.93$; $k = 2$; $p < 0.0001$). This interaction appears since total mercury concentrations are higher in the wet season of the first year than in the wet season of the second year, while the two dry seasons appear to be similar to one another. Moreover, there are significant differences between years and between seasons.

5.2 Savannah River Basin Streams

Data on a number of physical and biotic variables were obtained from the streams of the Savannah River Basin over two sampling cycles. The following considers temporal variation and variation between stream orders for the variables Algal Growth Potential Test (AGPT), a habitat code (Habitat), and a condition code (Condition).

Figure 4 depicts the *cdf* 's (solid lines) and corresponding 95% confidence bands (dashed lines) for AGPT for the two cycles. The curve for cycle 0 rises more gradually than that for cycle 1 suggesting that mean AGPT is not only higher in cycle 0, but also shows greater variability within this cycle. The Cramér-von Mises test W (Table 3) confirms that the difference between the two cycles is statistically significant ($W = 1.78; k = 2; p < 0.0001$). A comparison of *cdf* 's between cycles for Habitat and Condition reveals no significant difference between cycles (Table 3), as can be seen from Figures 5 and 6, respectively.

Comparing *cdf* 's for AGPT across stream orders reveals third order streams tend to have higher and more variable values of this variable than first or second order streams (Figure 7). The Cramér-von Mises test indicates that this difference between stream orders is statistically significant ($W = 1.31$; $k = 3$; $p = 0.012$). Likewise, first order streams tend to have lower condition codes than higher order streams (Figure 9); this difference is also statistically significant ($W = 1.71$; $k = 3$; $p = 0.002$). In contrast there is no significant difference between *cdf* 's for Habitat for different orders of streams ($W = 0.58$; $k = 3$;

$p = 0.294$); see Figure 8 for the *cdf* plots.

6 Monte Carlo Simulations

The α-level critical values W_α^* for the Cramér-von Mises test appearing in Table 1 are computed under the assumption of simple random sampling from the respective populations. However, neither the Savannah River Initiative nor the South Florida Initiative employ simple random sampling designs. Moreover, the tabulated critical values assume that we have a large sample. This section employs Monte Carlo simulations to investigate the distribution of the Cramér-von Mises test statistic under the designs employed by these initiatives. The general approach to carrying out the Monte Carlo simulation consists of simulating one realization of a Gaussian random field (defined below) with a given range of spatial correlation. Then 1000 independent samples of n sites per cycle are obtained from the realization according to a remap sampling design. For each sample, the 4-population Cramér-von Mises test statistic is computed; let W_i^* denote the value of the Cramér-von Mises test statistic for the i-th sample. Finally, the W_i^*'s are ranked from smallest to largest. The simulated critical value for an α-level test is then given by $(1-\alpha)\cdot 1000$th ranked value of W_i^*. The proportion $\hat{\alpha}$ of the W_i^*'s falling above the tabulated α-level critical value W_α was also computed. The tabulated values are conservative if the simulated critical values fall below W_α, or, equivalently if $\hat{\alpha} \leq \alpha$.

A stationary Gaussian random field is one of the simplest models for generating random functions $Z(\mathbf{s})$ of spatial locations $\mathbf{s}$ in a region A. Assume that

$$Z(\mathbf{s}) = \mu + \epsilon(\mathbf{s}), \tag{21}$$

where μ is the population mean, and $\epsilon(\mathbf{s})$ is a zero-mean normally-distributed error. The spatial dependence between data at locations $\mathbf{s}$ and $\mathbf{u}$ is modeled through the covariance function

$$C(\|\mathbf{s} - \mathbf{u}\|) \equiv cov\{Z(\mathbf{s}) - Z(\mathbf{u})\}, \tag{22}$$

which we shall assume to be a function of only the distance $\|\mathbf{s} - \mathbf{u}\|$ between the two sites. Simulations are carried out under an exponential covariance function model

$$C(r) = \sigma^2 e^{-\gamma r} \tag{23}$$

using the spectral method described by Shinozuka (1971) and Mejia and Rodriguez-Iturbe (1974). For the exponential covariance, the range of spatial correlation is defined to be equal to $3/\gamma$. The results of simulations for the Cramér-von Mises test statistic do not depend on μ and σ^2. So, without loss of generality, all simulations shall be carried out under $\mu = 0$ and $\sigma^2 = 1$.

The following gives the results of simulations of the sampling procedures used for the canals and marshes of south Florida.

6.1 South Florida Canals

Simulation of the design used to sample the south Florida canals was carried out conditional on the fixed location on the EMAP grid of contiguous hexagons. These hexagons partition the canals into short segments, that were placed into random order as follows: Hexagons were systematically partitioned into cycles, corresponding to the biannual sampling times used in south Florida, in such a manner that for each hexagon assigned to a given cycle, two of its six neighboring hexagons are assigned to each of the three remaining cycles. To ensure uniform spatial coverage of the study region, hexagons were also partitioned into 22 clusters. Then for each simulated realization, the clusters were randomly ordered, and canal segments were randomly ordered within clusters. Finally, the ordered canal segments were strung together, and a systematic random sample of 50 sites was located along the strung segments. This was accomplished by picking a random starting point between 0 and $\ell/50$ km and sampling every $\ell/50$ km thereafter, where ℓ is the total length of the strung segments.

The results show that the sample proportions of simulated Cramér-von Mises test statistics greater than the tabulated α-level critical values fall well below corresponding values of α (Table 4). Moreover, simulated estimates of α-level critical values all fall well below the tabulated critical values. These results indicate that the tabulated critical values are conservative in the sense that we will reject null hypothesis of equality of *cdf* 's less often than we should. Moreover, there is some suggestion that the test becomes more conservative with increasing range of spatial correlation.

6.2 South Florida Marsh

Simulation of the design used to sample the south Florida marshes was carried out on a 22 x 22 grid of contiguous hexagons whose centers are one unit apart. Hexagons were systematically partitioned into cycles, as described above, yielding a total of 121 hexagons per cycle. This is comparable to the sampling effort used in the south Florida marsh. One random sample sites was then located in each of the hexagons.

The results show that the sample proportions of simulated Cramér-von Mises test statistics greater than the tabulated α-level critical values fall well below corresponding values of α, and that simulated estimates of α-level critical values fall well below the tabulated critical values (Table 5). These results indicate that the tabulated critical values are conservative in the sense that we will fail the null hypothesis of equality of *cdf* 's less often than we should. Moreover, the test becomes more conservative with increasing range of spatial correlation, as indicated by decreasing simulated α-levels with increasing correlation.

7 Conclusions

The Cramér-von Mises test is a powerful procedure for testing the null hypothesis that two or more populations are identically distributed; that is, have identical cumulative distribution functions. This null hypothesis can be rejected if any feature of the distributions, including the means, variances, or even their shapes, varies significantly between the populations. Thus, the alternative hypothesis under a Cramér-von Mises test is very general. The Cramér-von Mises test is a nonparametric test; so, no distributional assumptions (e.g., normality, homogeneous variances) are required. However, this comes at the cost of loss of power in comparison to tests designed to detect specific alternatives. For example, the F-test in an analysis of variance is more powerful for detecting differences between population means, assuming identical population variances.

The results of the Monte Carlo simulations suggest that Keifer's tabulated critical values are conservative in the sense that this test will reject the null hypothesis of equality of *cdf*'s less often as it should. However, application of this test indicates that it has sufficient power to detect trends over time in the canals of southern Florida and the streams of the Savannah River Basin, and sufficient power to detect differences between stream orders in the Savannah River Basin.

The partitioning of the Cramer-von Mises test statistic into terms for year effects, season effects and year by season interaction can be extended to any factorial arrangement of factors (e.g., strata, habitat types, etc.) provided that the data are balanced; that is, there are identical numbers of observations in each of the factorial combinations. If the data are not balanced, then the results of hypothesis tests can become ambiguous with regards to their interpretations. For example, if one year has more wet season sample stations than another year, a test based on W_y for differences between years might be rejected if their are signficant between seasons, regardless of whether or not there are significant differences between years. Further research is required to test for differences between years after differences between seasons are taken into account.

REFERENCES

Cordy, C.B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statistics and Probability Letters*, **18**, 353-362.

Hollander, M., and Wolf, D.A. (1973). *Nonparametric Statistical Methods*. Wiley, New York.

Kiefer, J. (1959). K-sample analogues of the Kolmogorov-Smirnov and Cramér-V. Mises tests. *Annals of Mathematical Statistics*, **30**, 420-447.

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.

Overton, W.S., White, D., and Stevens, D.L. (1991). *Design Report for EMAP, the Environmental Monitoring and Assessment Program*. Washington, D.C., U.S. Environmental Protection Agency (EPA/600/3 - 91/053).

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

Urquhart, N.S., Overton, W.S., and Birkes, D.S. (1993). Comparing sampling designs for ecological status and trends: impact of temporal patterns. in V. Barnett and K.F. Turkman (eds.), *Statistics for the Environment*, Wiley, New York.

Table 1. Critical values of α-level Cramer–von Mises tests of H_0: $F_1 = F_2 = \cdots F_k$ against the general alternative that at least one cdf is different. This table is extracted from Kiefer (1959).

	k			
α	2	3	4	5
0.75	0.18545	0.31472	0.45103	0.59161
0.50	0.27757	0.44138	0.60668	0.77253
0.25	0.42098	0.62227	0.81775	1.00947
0.20	0.46640	0.67691	0.87980	1.07785
0.15	0.52481	0.74592	0.95734	1.16268
0.10	0.60704	0.84116	1.06311	1.27748
0.05	0.74752	1.00018	1.23730	1.46466
0.02	0.93320	1.20561	1.45913	1.70028
0.01	1.07366	1.35861	1.62263	1.87215
0.005	1.21412	1.51010	1.78345	2.03935
0.001	1.54027	1.85773	2.14949	2.40774
0.0001	2.00691	2.34950	2.66130	2.82500

Table 2. Cramer von Mises tests for equality of cumulative distribution functions for South Florida canals.

Total Mercury in Water

Source	W	P
Years	1.01	0.014
Seasons	1.31	0.003
Years x Seasons	4.93	<0.0001
Total	7.24	<0.0001

Methyl Mercury in Water

Source	W	P
Years	0.28	0.494
Seasons	3.11	<0.0001
Years x Seasons	1.31	0.003
Total	4.69	<0.0001

Table 3. Cramer–von Mises tests for equality of cumulative distribution functions for Savannah River Basin streams. Equality of cumulative distribution functions between cycles and between stream orders is tested.

	Cycles		Stream Orders	
Variable	W	P	W	P
AGPT	1.78	<0.0001	1.31	0.012
Habitat	0.27	0.518	0.58	0.294
Condition	0.12	0.929	1.71	0.002

Table 4. Simulated critical values W^*_α for 4–sample α–level Cramer–von Mises tests under different ranges of spatial correlation in the south Florida canals, and corresponding proportion $\hat{\alpha}$ of simulated Cramer–von Mises test statistics greater than α–level critical values. Tabulated critical values from Kiefer (1959) are given at the bottom.

	α					
	0.1		0.05		0.01	
Range (km)	W^*_α	$\hat{\alpha}$	W^*_α	$\hat{\alpha}$	W^*_α	$\hat{\alpha}$
0	0.8341	0.048	0.9903	0.017	1.3393	0.001
10	0.6887	0.013	0.8085	0.006	1.1559	0.001
20	0.8037	0.032	0.9467	0.011	1.2953	0.001
30	0.5913	0.002	0.6731	0.001	0.8573	0.000
60	0.6911	0.009	0.7773	0.002	1.0023	0.000
120	0.7959	0.019	0.9117	0.005	1.1885	0.000
tabulated	1.0631		1.2373		1.6226	

Table 5. Simulated critical W^*_α values for 4–sample α–level Cramer–von Mises tests under different ranges of spatial correlation in the south Florida marsh, and corresponding proportion $\hat{\alpha}$ of simulated Cramer–von Mises test statistics greater than α–level critical values. Tabulated critical values from Kiefer (1959) are given at the bottom.

	α					
	0.1		0.05		0.01	
Range (km)	W^*_α	$\hat{\alpha}$	W^*_α	$\hat{\alpha}$	W^*_α	$\hat{\alpha}$
0.5	0.8094	0.029	0.9564	0.014	1.2906	0.002
1.0	0.8279	0.040	0.9885	0.013	1.3666	0.002
2.0	0.5819	0.002	0.6981	0.002	0.9355	0.000
4.0	0.4557	0.000	0.5171	0.000	0.6233	0.000
8.0	0.3096	0.000	0.3489	0.000	0.5604	0.000
tabulated	1.0631		1.2373		1.6226	

FIGURE LEGENDS

Figure 1. Cumulative distribution function (solid line) for total mercury concentration in water samples collected from south Florida canals in cycle 0. The corresponding 95% confidence bands are given by the dashed lines.

Figure 2. Cumulative distribution functions (solid lines) for methyl mercury concentration in water samples collected from south Florida canals in cycles 0, 1, 2, and 3. The corresponding 95% confidence bands are given by the dashed lines.

Figure 3. Cumulative distribution functions (solid lines) for total mercury concentration in water samples collected from south Florida canals in cycles 0, 1, 2, and 3. The corresponding 95% confidence bands are given by the dashed lines.

Figure 4. Cumulative distribution functions (solid lines) for Algal Growth Potential Test in samples collected from Savannah River Basin streams in cycles 0 and 1. The corresponding 95% confidence bands are given by the dashed lines.

Figure 5. Cumulative distribution functions (solid lines) for Habitat Score in samples collected from Savannah River Basin streams in cycles 0 and 1. The corresponding 95% confidence bands are given by the dashed lines.

Figure 6. Cumulative distribution functions (solid lines) for Condition in samples collected from Savannah River Basin streams in cycles 0 and 1. The corresponding 95% confidence bands are given by the dashed lines.

Figure 7. Cumulative distribution functions (solid lines) for Algal Growth Potential Test in samples collected from Savannah River Basin streams of orders 1, 2, and 3 (top to bottom). The corresponding 95% confidence bands are given by the dashed lines.

Figure 8. Cumulative distribution functions (solid lines) for Habitat Score in samples collected from Savannah River Basin streams of orders 1, 2, and 3 (top to bottom). The corresponding 95% confidence bands are given by the dashed lines.

Figure 9. Cumulative distribution functions (solid lines) for Condition in samples collected from Savannah River Basin streams of orders 1, 2, and 3 (top to bottom). The corresponding 95% confidence bands are given by the dashed lines.

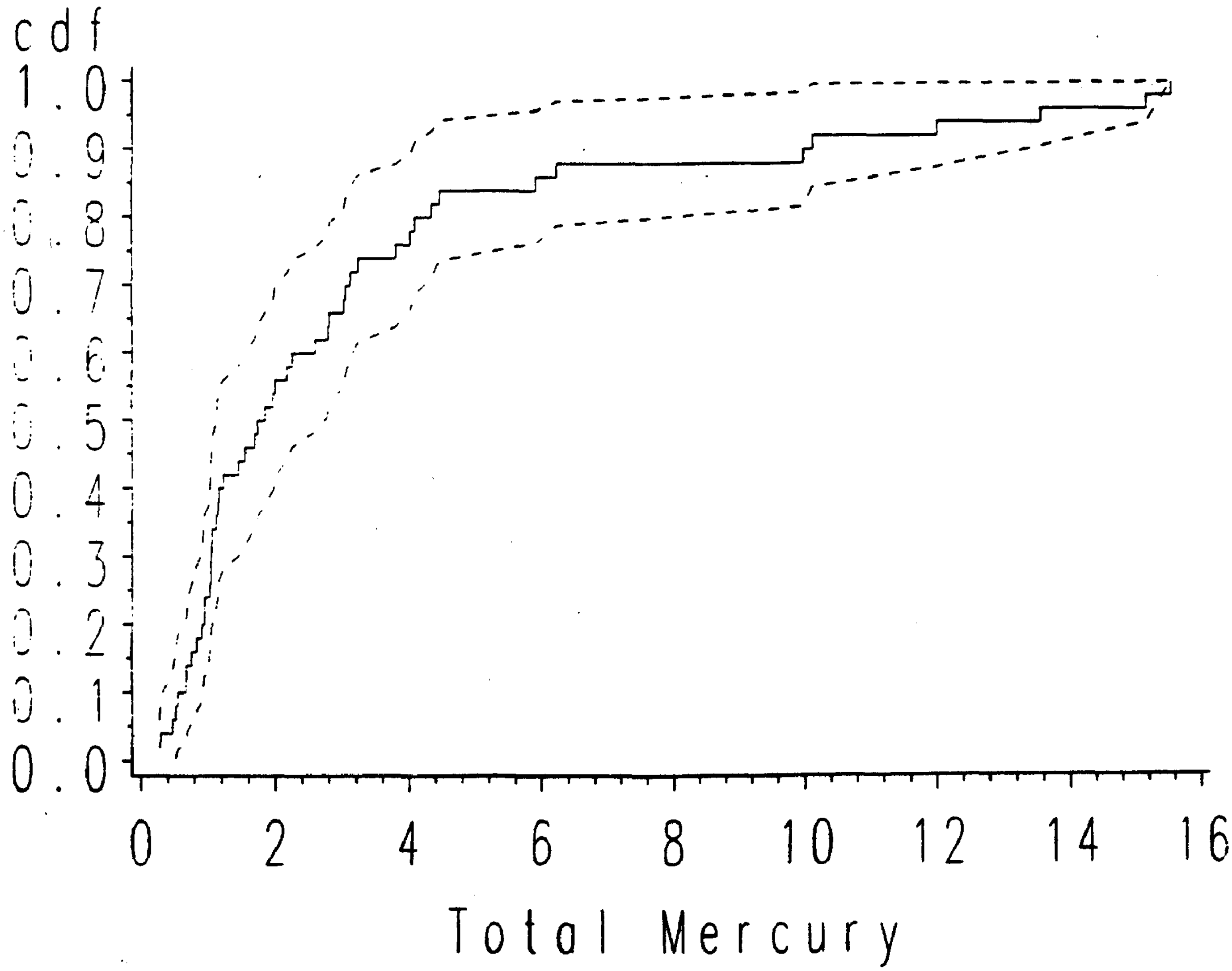

cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
12
14
16
Total Mercury

Figure 1

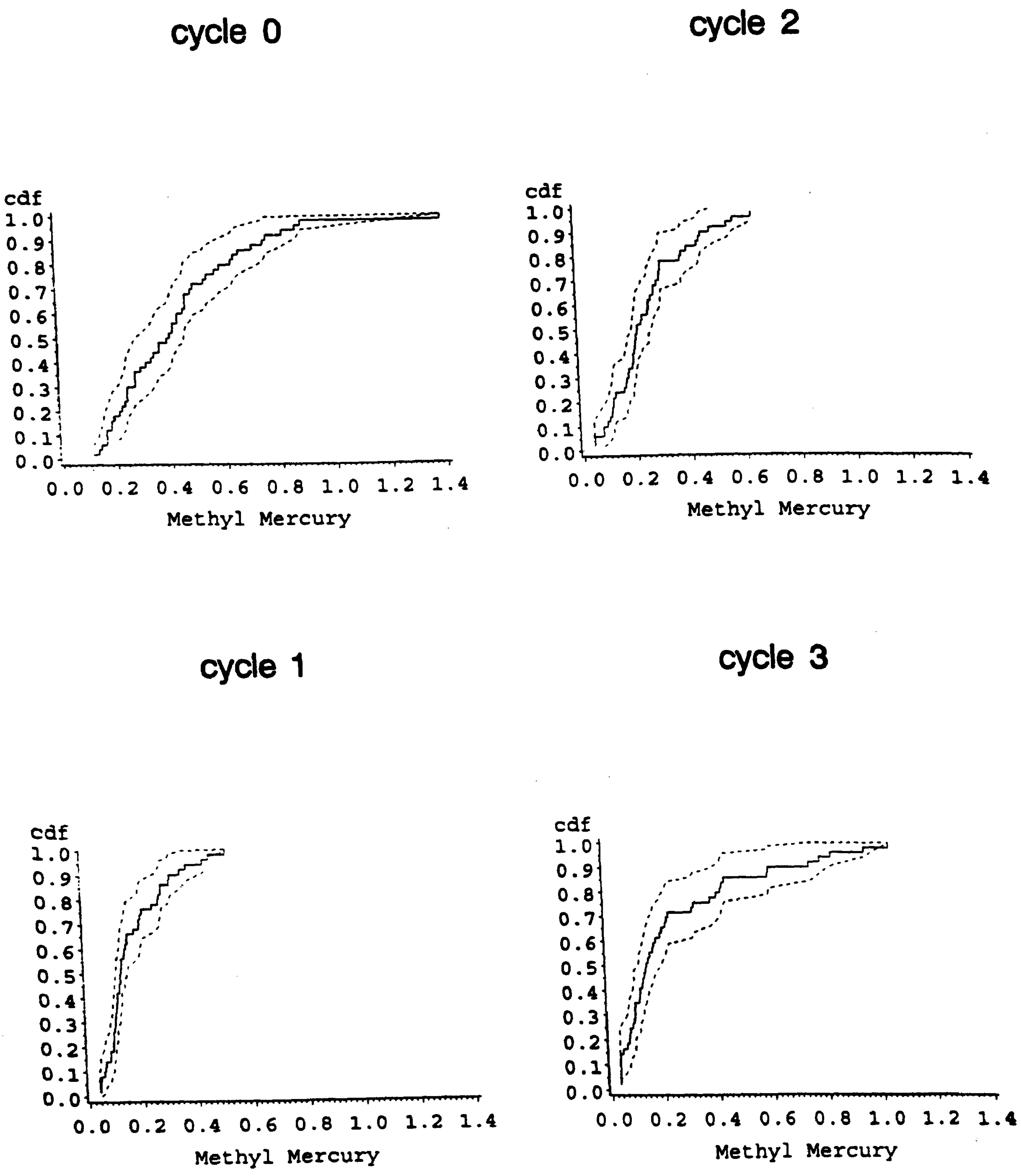

cycle 0
cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4
Methyl Mercury
cycle 2
cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4
Methyl Mercury
cycle 1
cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4
Methyl Mercury
cycle 3
cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4
Methyl Mercury

Figure 2

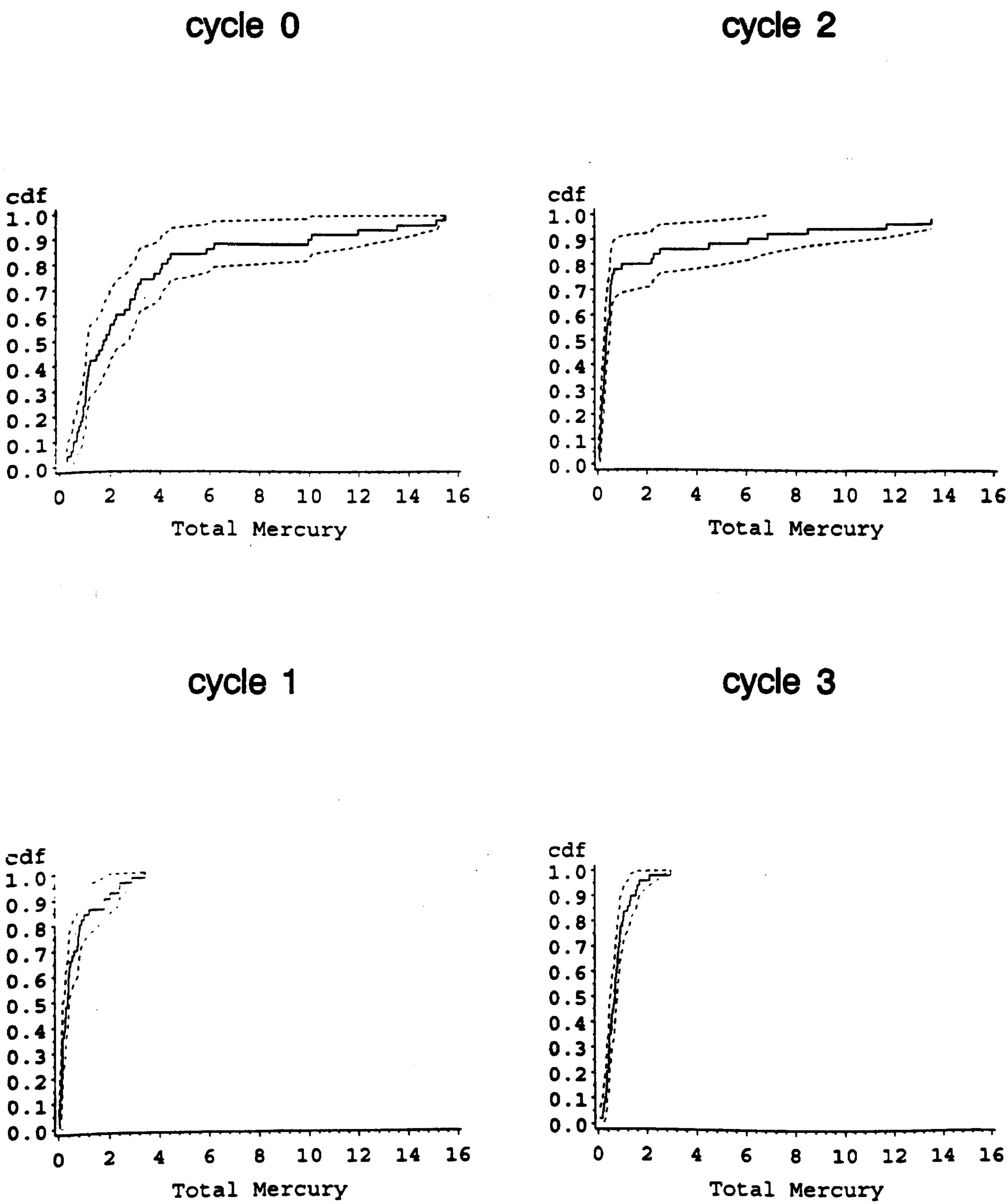
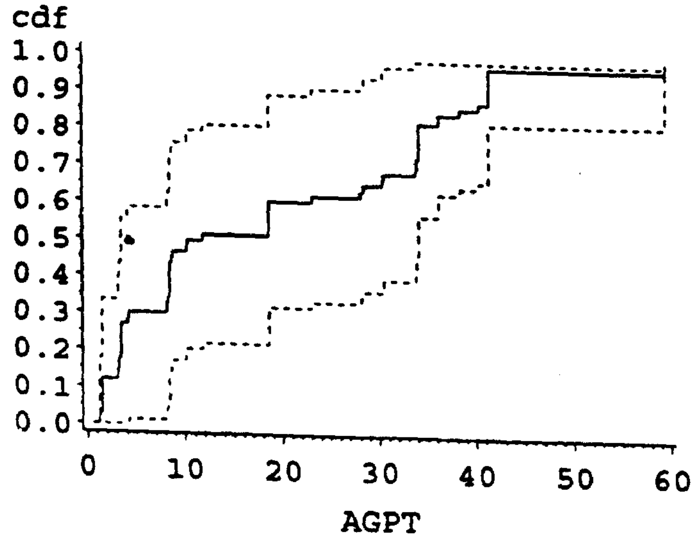

cycle 0
cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0 2 4 6 8 10 12 14 16
Total Mercury
cycle 2
cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0 2 4 6 8 10 12 14 16
Total Mercury
cycle 1
cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0 2 4 6 8 10 12 14 16
Total Mercury
cycle 3
cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0 2 4 6 8 10 12 14 16
Total Mercury

Figure 3

cycle 0



cycle 1

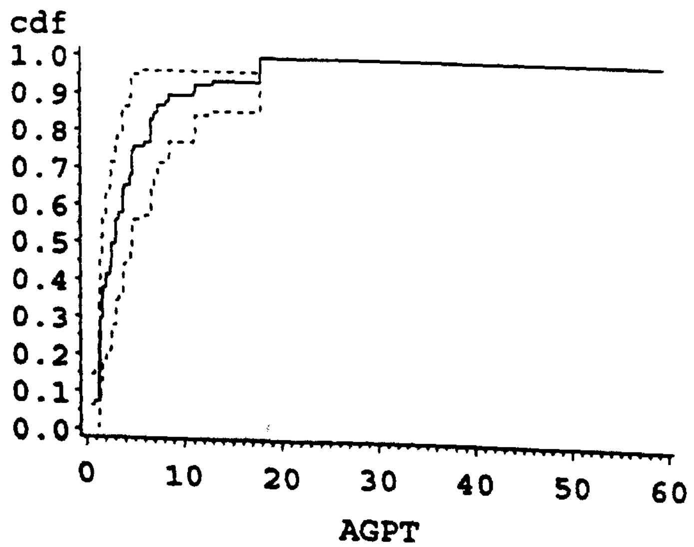
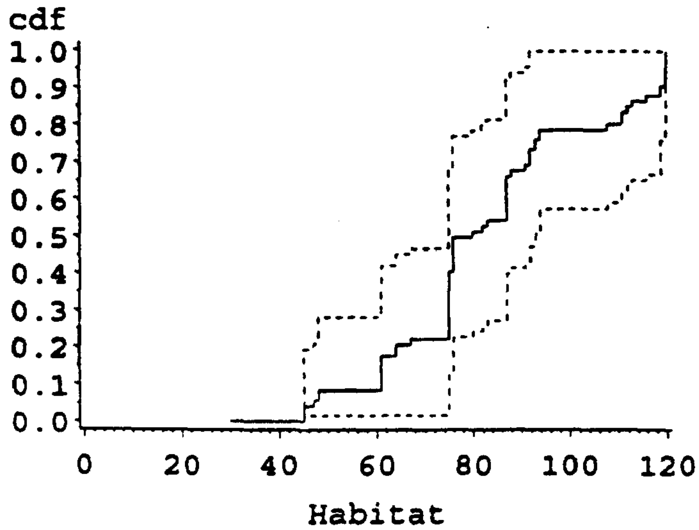


Figure 4

cycle 0



cycle 1

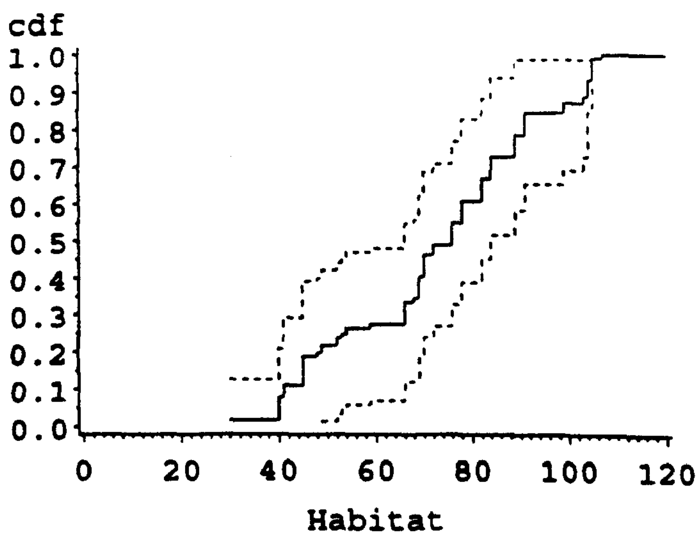
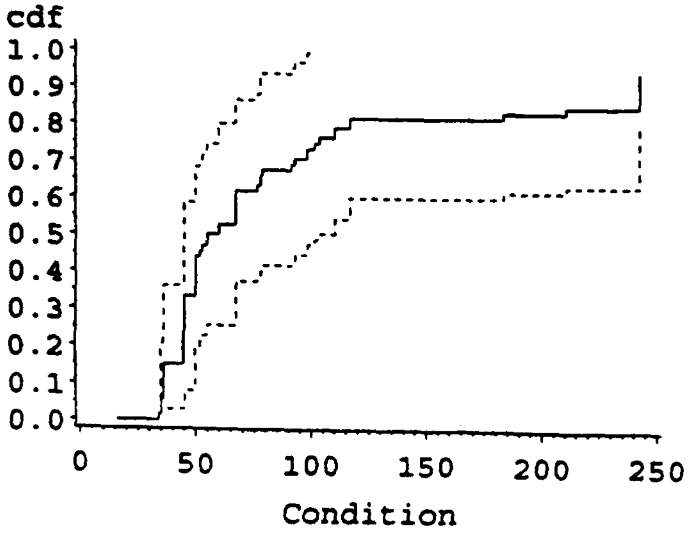


Figure 5

cycle 0



cycle 1

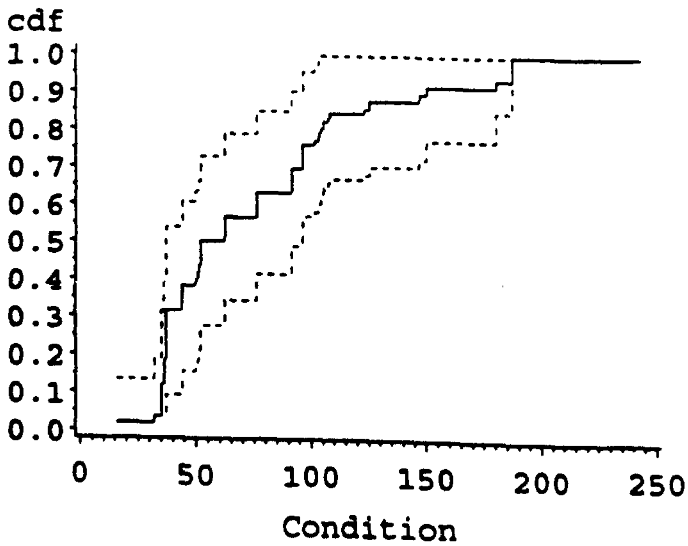


Figure 6

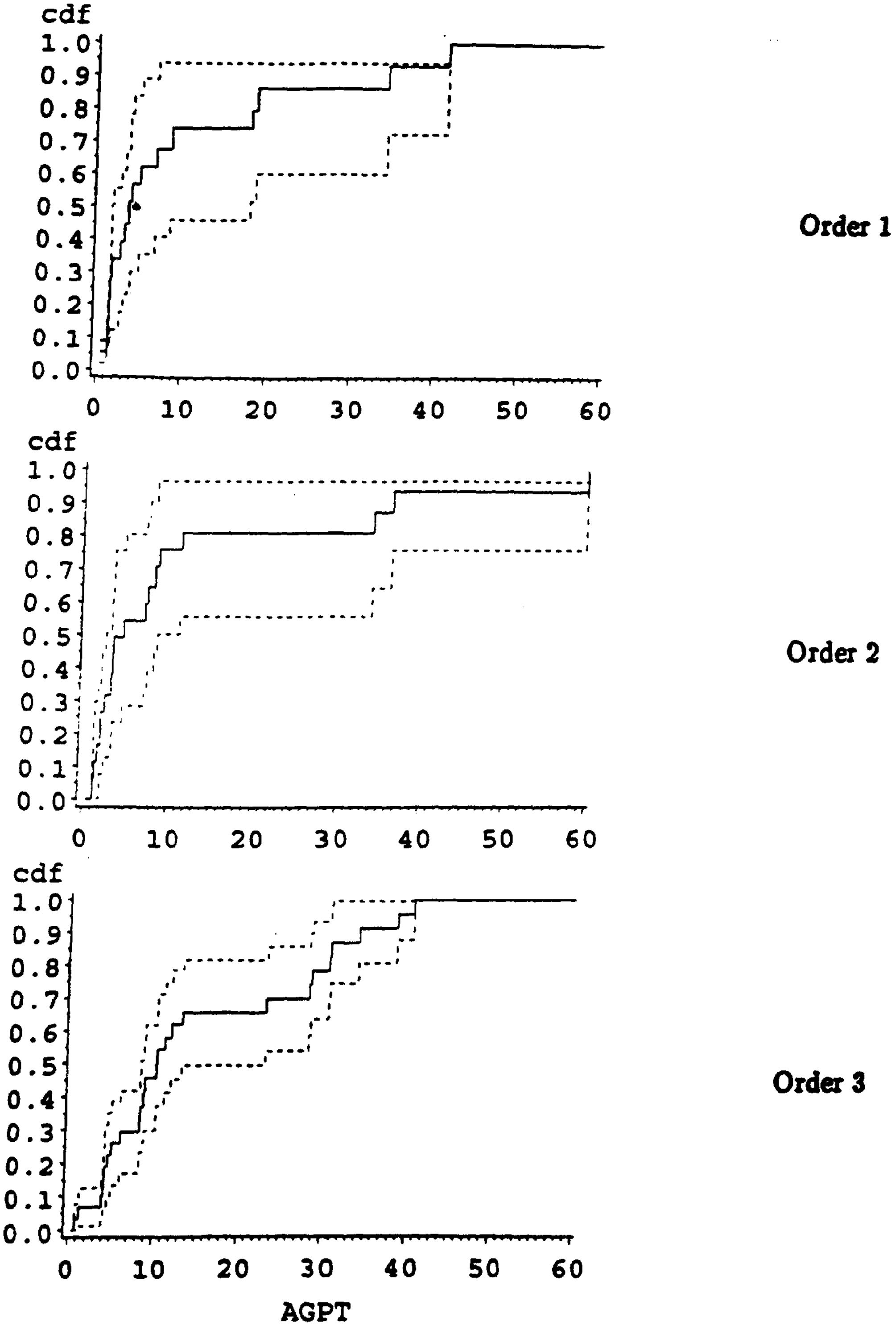
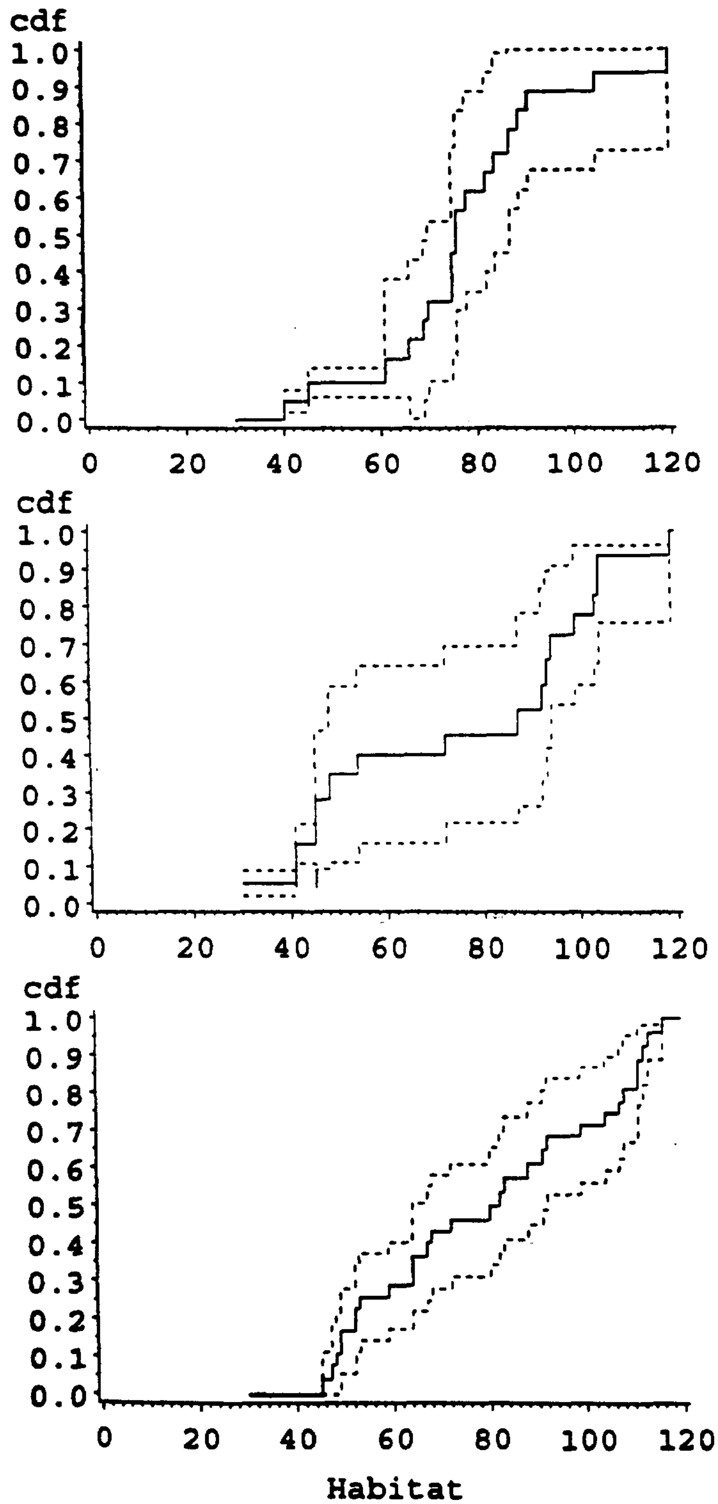

cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
10
20
30
40
50
60
Order 1
Order 2
Order 3
AGPT

Figure 7



Order 1

Order 2

Order 3

Figure 8

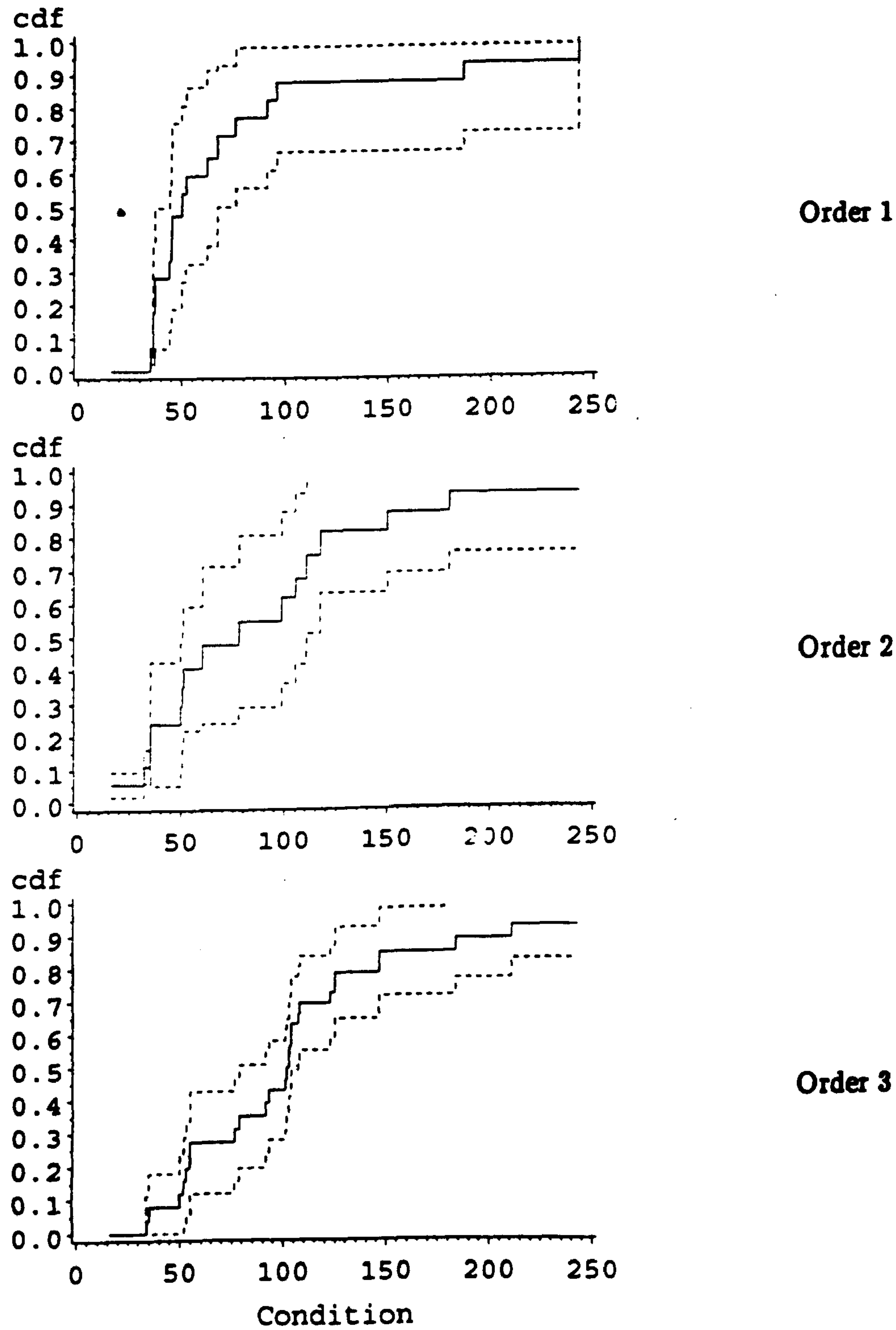

cdf
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
50
100
150
200
250
Order 1
Order 2
Order 3
Condition

Figure 9

F

Appendix F

Savannah River Basin Landscape Analysis
By
Chaloud et al.

Savannah River Basin
Landscape Analysis

by

Deborah J. Chaloud, Curtis M. Edmonds, and Daniel T. heggem

U.S. Environmental Protection Agency
Office of Research and Development
National Exposure Research Laboratory
Environmental Sciences Division
Landscape Ecology Branch
Las Vegas, Nevada

Savannah River Basin
Landscape Analysis

Deborah J. Chaloud, Curtis M. Edmonds, and Daniel T. Heggem

U.S. Environmental Protection Agency,
Office of Research and Development, National Exposure Research Laboratory, Environmental Sciences Division, Landscape Ecology Branch, Las Vegas, Nevada

INTRODUCTION

Scientists from the U.S. Environmental Protection Agency (EPA), Region 4, Science and Ecosystem Support Division, enlisted the assistance of the landscape ecology group of U.S. EPA, Office of Research and Development (ORD), National Exposure Research Laboratory, Environmental Sciences Division (ESD), in conducting a landscape assessment of the Savannah River Basin (Figure 1) as part of their ongoing Regional Environmental Monitoring and Assessment Program (REMAP) demonstration project. In the Scope of Work provided by Region 4, the goal was stated as "provide technical/scientific assistance ...to EPA Region 4 in assessing current wadeable stream conditions in the Savannah River Basin with landscape factors that may be contributing to these conditions or gradients." Three specific objectives were presented in the form of questions. These were:

> Are both the proportions of land uses and the spatial pattern of land uses important for characterizing and modeling stream condition in watersheds/ecoregions of different areas?
>
> Can land uses near the streams better account for the variability in ecological condition than land use for the entire watershed/ecoregion?
>
> Does the size of the watershed/ecoregion influence statistical relationships between landscape characteristics and ecological condition?

In addition, an assessment of landscape change was to be conducted as part of continuing ESD research in application of change detection techniques.

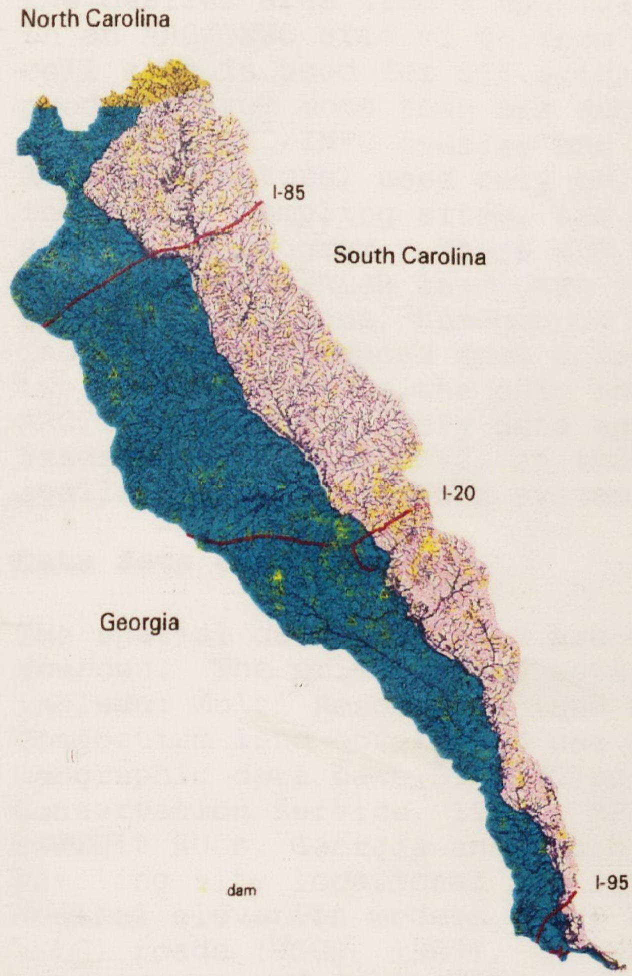


Figure 1. Savannah River Basin

The data analysis plan developed to address the objectives given above called for calculation of a specific suite of landscape indicators for all nine United States Geological Survey (USGS) 8-digit hydrological unit codes (HUC; USGS, 1982), a selected subset of the 94 Georgia and South Carolina subbasins, and the riparian corridors in the HUCs and selected subbasins. The subbasins are generally equivalent in area to USGS 11-digit HUCs. The riparian corridor was defined as 100 meters on either side of stream arcs; this size was selected from a review of state laws and literature available on the Internet (e.g., Santa Cruz County, 1998; U.S. EPA, 1998; South Carolina Department of Natural Resources, 1998). The suite of indicators included: landcover types, u-index, agriculture on slopes greater than 3 percent, agriculture on highly erodible soils, agriculture on moderately erodible soils, agriculture on highly erodible soils with slopes greater than 3 percent, number of occurrences of roads crossing streams, and number of impoundments. Landscape indicator statistics were also computed for the drainage areas and associated riparian corridors of a selected set of sites sampled by Region 4 using REMAP protocols. Region 4 provided an ARC/INFO coverage of the sampling locations and Quattro Pro spreadsheets of the water quality and biotic measurements.

METHODS

The selected landscape indicators are identical to, or based on, indicators used in the mid-Atlantic atlas (Jones et al, 1997). In the atlas, the indicators were calculated only for 8-digit HUCs; in this study, indicators are additionally calculated for smaller spatial units. The basic methodology is the same, however. In general, calculation of the landscape indicators involves ARC/INFO techniques of extracting or "cookie cutting" the desired area from a spatial data set. The data are formatted in an ARC/INFO grid of uniform cell size. In this study, a 30-m cell size is used for all grids. For indicators which are produced from more than one data set (e.g., roads crossing streams), ARC/INFO overlay and intersection techniques are used. A few indicators, used only on the drainage areas of the individual sampling sites, are produced from an in-house custom statistics program. These are indicators of fragmentation, i.e., the degree to which landcover types are present in patches rather than in continuous, homogenous blocks. The landscape change indicator is produced from comparison of satellite imagery from two dates. This is the only indicator which does not use ARC/INFO as the primary data analysis software. Landscape change assessment employs ENVI, an image processing software package available for PC or Unix systems.

Data Sets Used

The spatial data sets used are obtained from a variety of sources. The primary data sets used in this landscape assessment include: Multi Resolution Land Characteristics (MRLC) Interagency Consortium land cover/land use (Bara, 1994), State Soil Geographic data base(STATSGO)soils (Natural Resources Conservation Service, 1996), RF3 streams (U.S. EPA, 1997), USGS 8-digit HUCs, Georgia and South Carolina subbasins, Region 4 sampling site locational and sampling data, 30-m and 100-m digital elevation models (DEM; USGS, 1990), digital line graph (DLG) roads (USGS, 1989), and National Inventory of Dams impoundments (U.S. Army Corps of Engineers, 1997). Landscape change assessment used North American Land Characterization (NALC) imagery from the 1970s and 1990s (U.S. EPA, 1993). Data sets were subset to the area of interest using the basin boundary coverage.

Sampling Site Ranking, Selection, and Drainage Area Creation

A simple, unweighted scoring system was used to rank the sampling sites, shown in Figure 2, by their results. Water quality variables (pH, dissolved oxygen, conductivity) and biota [algal growth potential test (AGPT); Ephemeroptera, Plecoptera, and Trichoptera index (EPT); fish index of biological integrity (fish_ibi), macroinvertebrate habitat, and macroinvertebrate richness] were scored separately. The frequency distributions for each variable was examined. Most indicated a bimodal distribution, with reduced frequencies near the lower and upper ends of the variable's range. Measurement values corresponding to the inflection points of the curve were selected to divide the range into three classes. A score value was ascribed to the

Figure 2. Sampling Site Locations

measurement value, 1 for bad, 2 for fair, 3 for good, and 0 for missing data. Although these are labeled as good, fair, and bad, these terms apply to the measurement value compared to the range of measurement values, not to any applicable water quality standards or other measurement system. The scores were summed and recorded. The number of measurements used in the summation was also recorded; this was necessary because of the large number of sites missing results for one or more variables. The measurement data and scoring data were then associated with the site location coverage. Map compositions were prepared for each HUC, presented here as figures 3 through 11. The figures were useful in characterizing relative conditions across the basin and making preliminary decisions about areas for further investigation.

The sampling locations had been selected by the Region using the EMAP site selection protocol. Several discussions and correspondences were conducted with a lead EMAP Statistician, Dr. Tony Olson, about the spatial area represented by the sampling sites. It was determined that it would be necessary to develop the specific drainage area of each sampling location and to treat the water quality and biota information as point data. Accurate drainage area computation requires DEMs of 30-m intervals or better; at the time of analysis, these were available for only portions of the Savannah River Basin, primarily the north end and part of the central area.

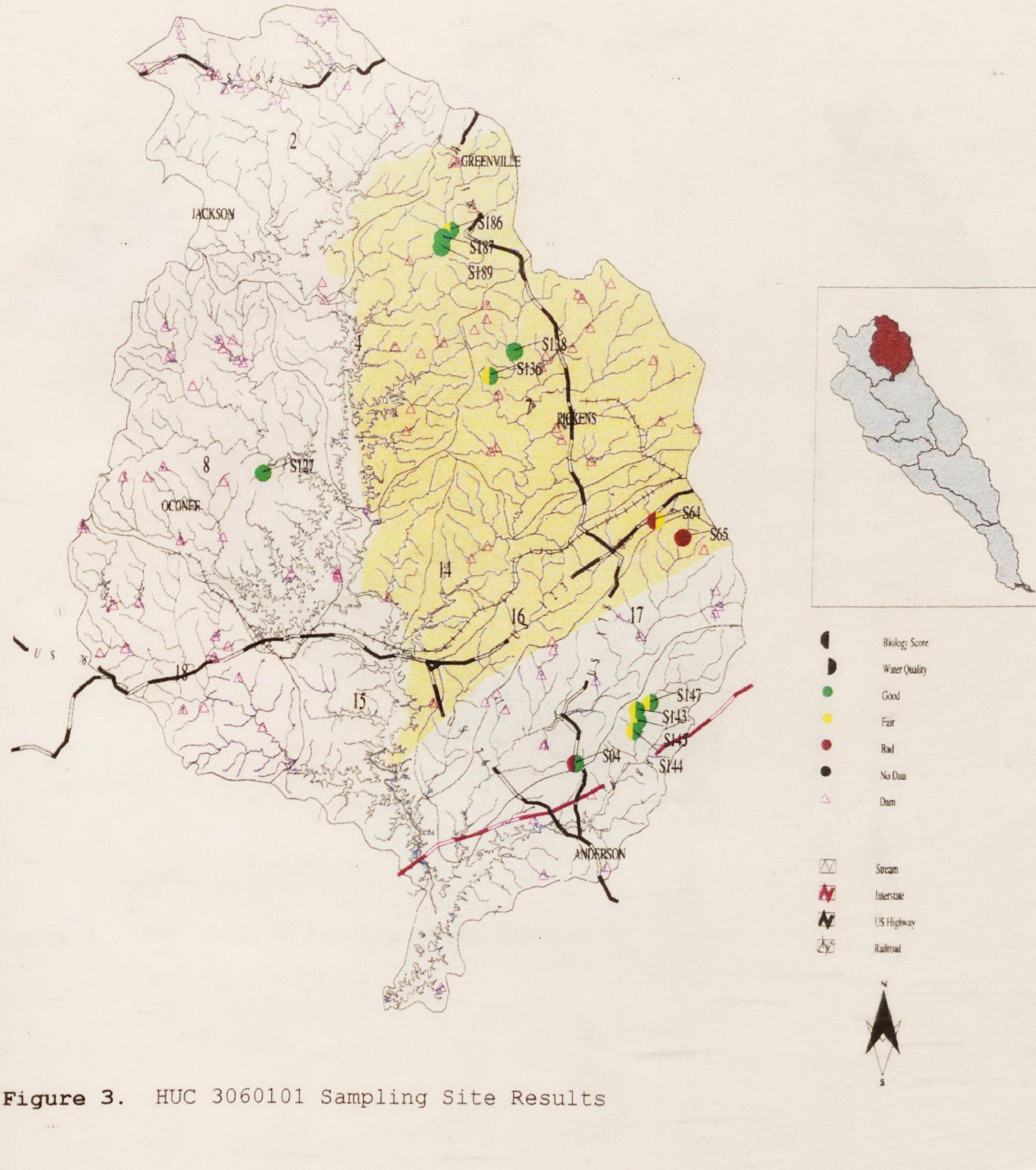


Figure 3. HUC 3060101 Sampling Site Results

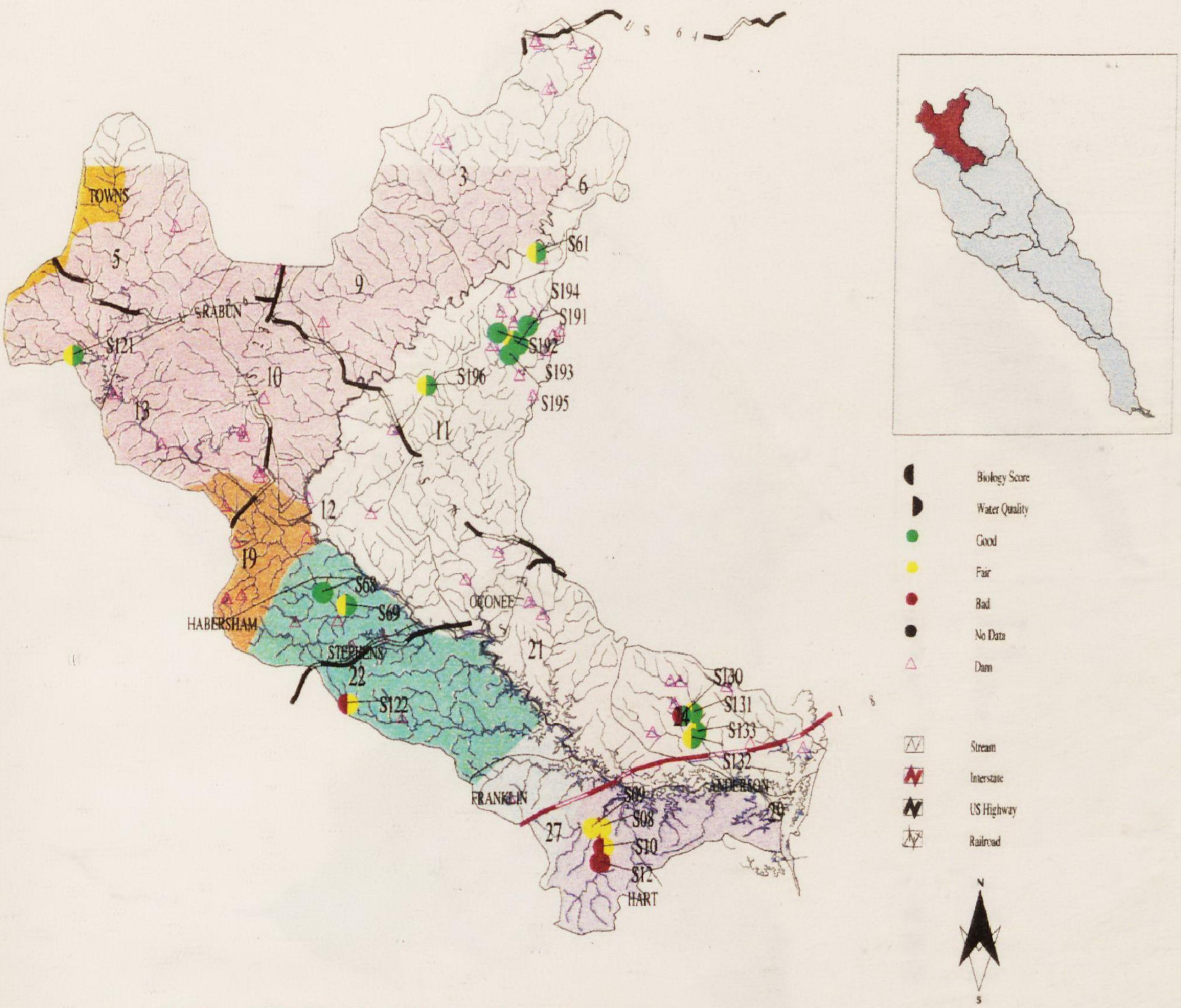


Figure 4. HUC 3060102 Sampling Site Results

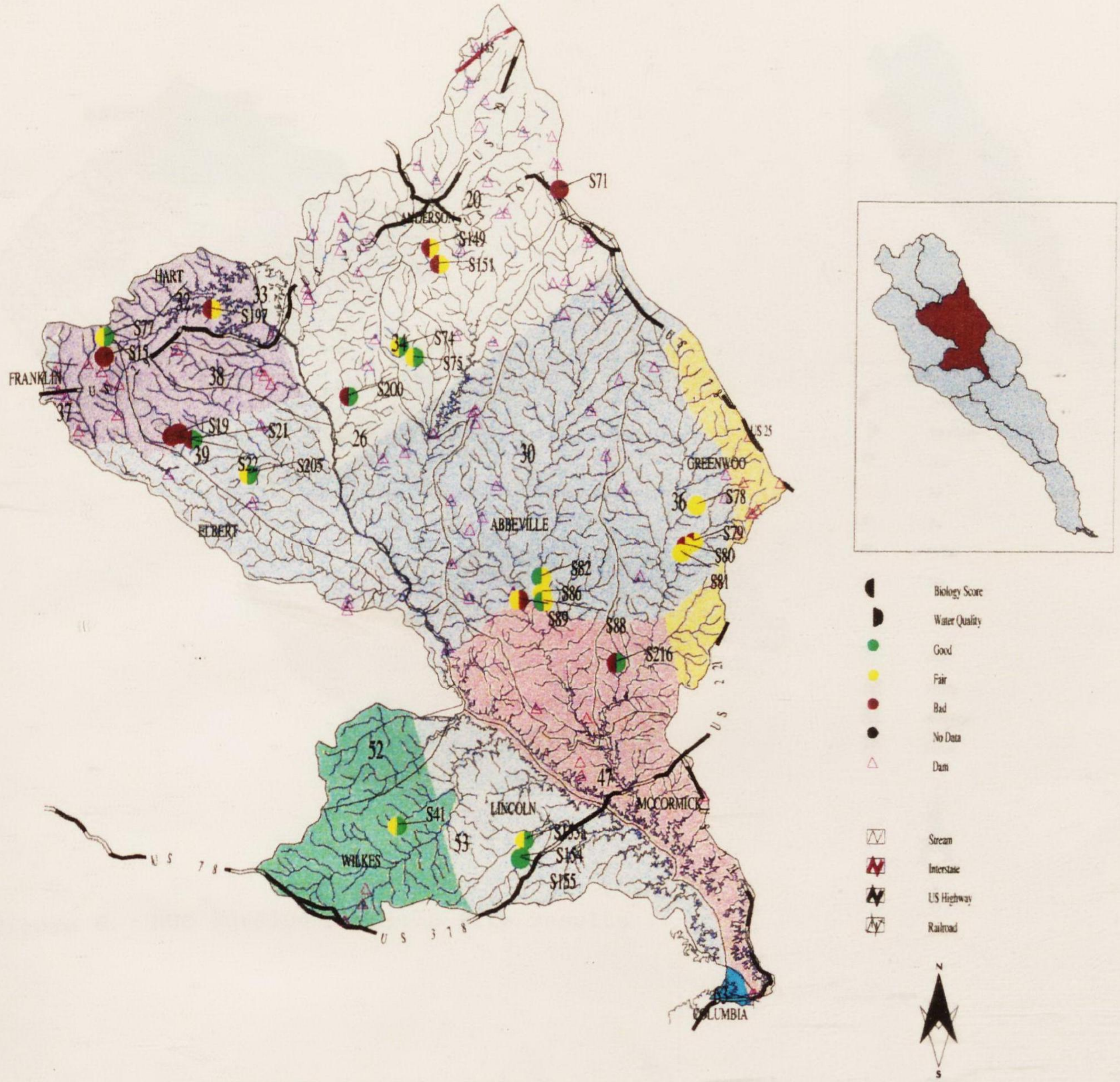


Figure 5. HUC 3060103 Sampling Site Results

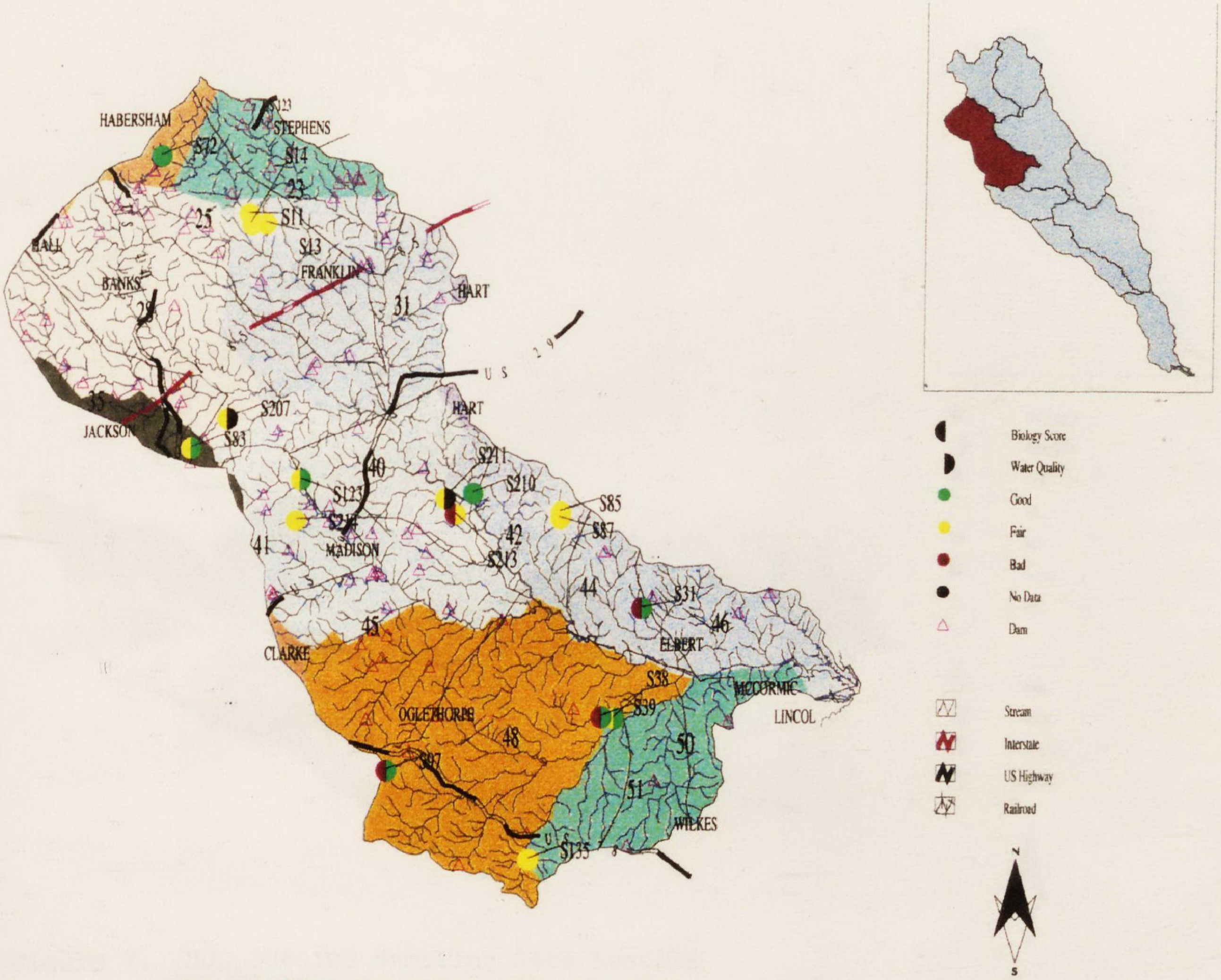


Figure 6. HUC 3060104 Sampling Site Results

Figure 7. HUC 3060105 Sampling Site Results

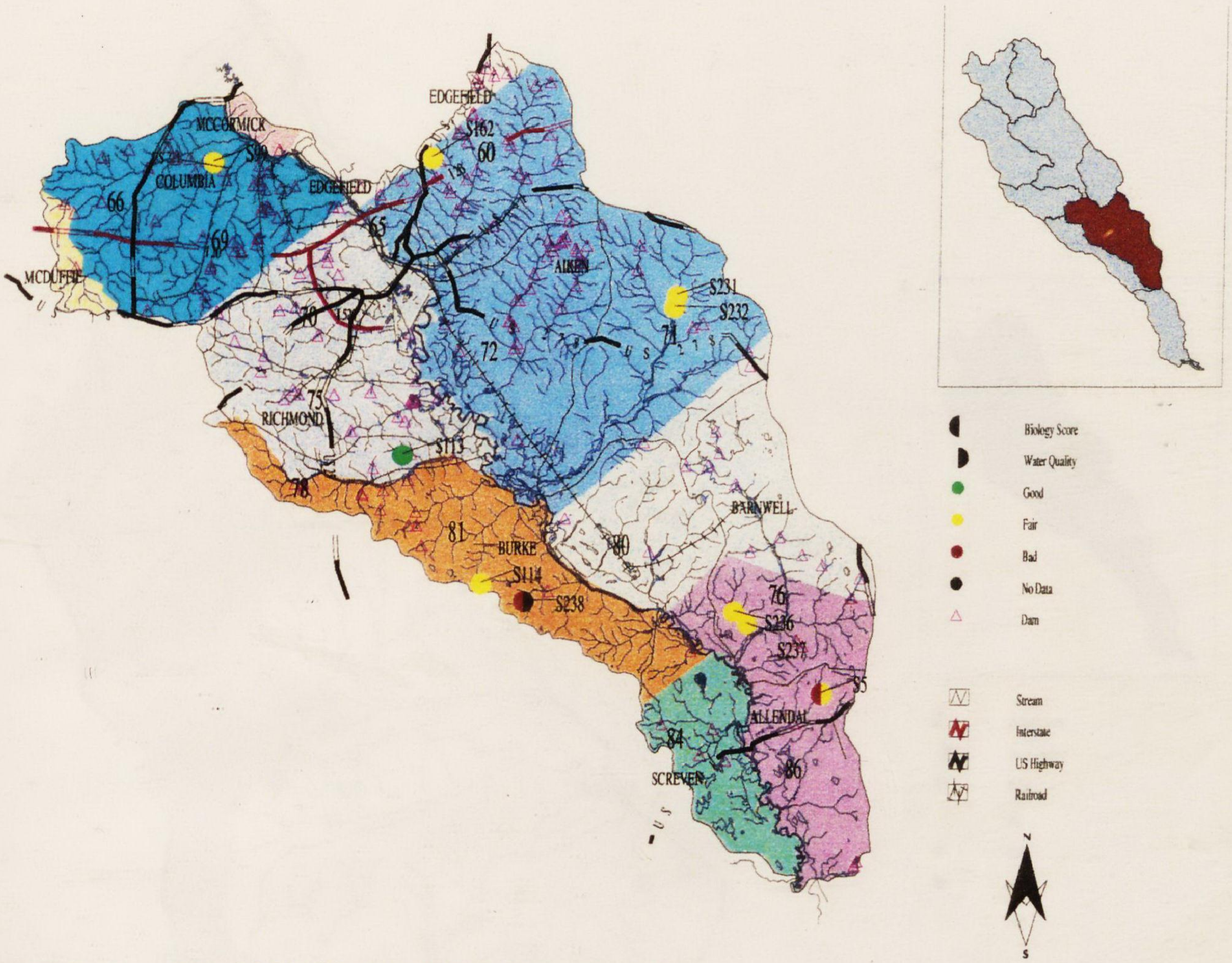


Figure 8. HUC 3060106 Sampling Site Results

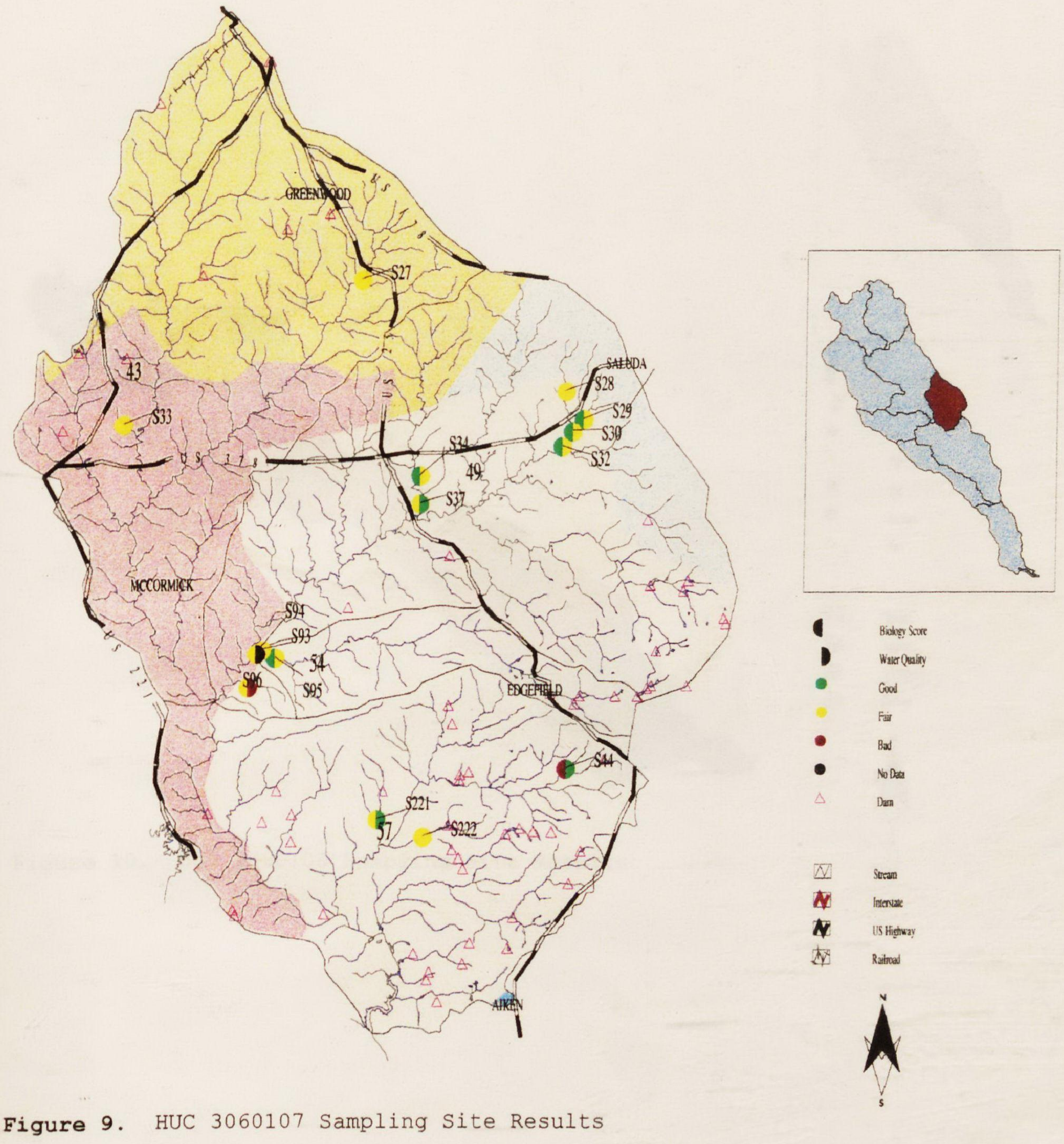


Figure 9. HUC 3060107 Sampling Site Results

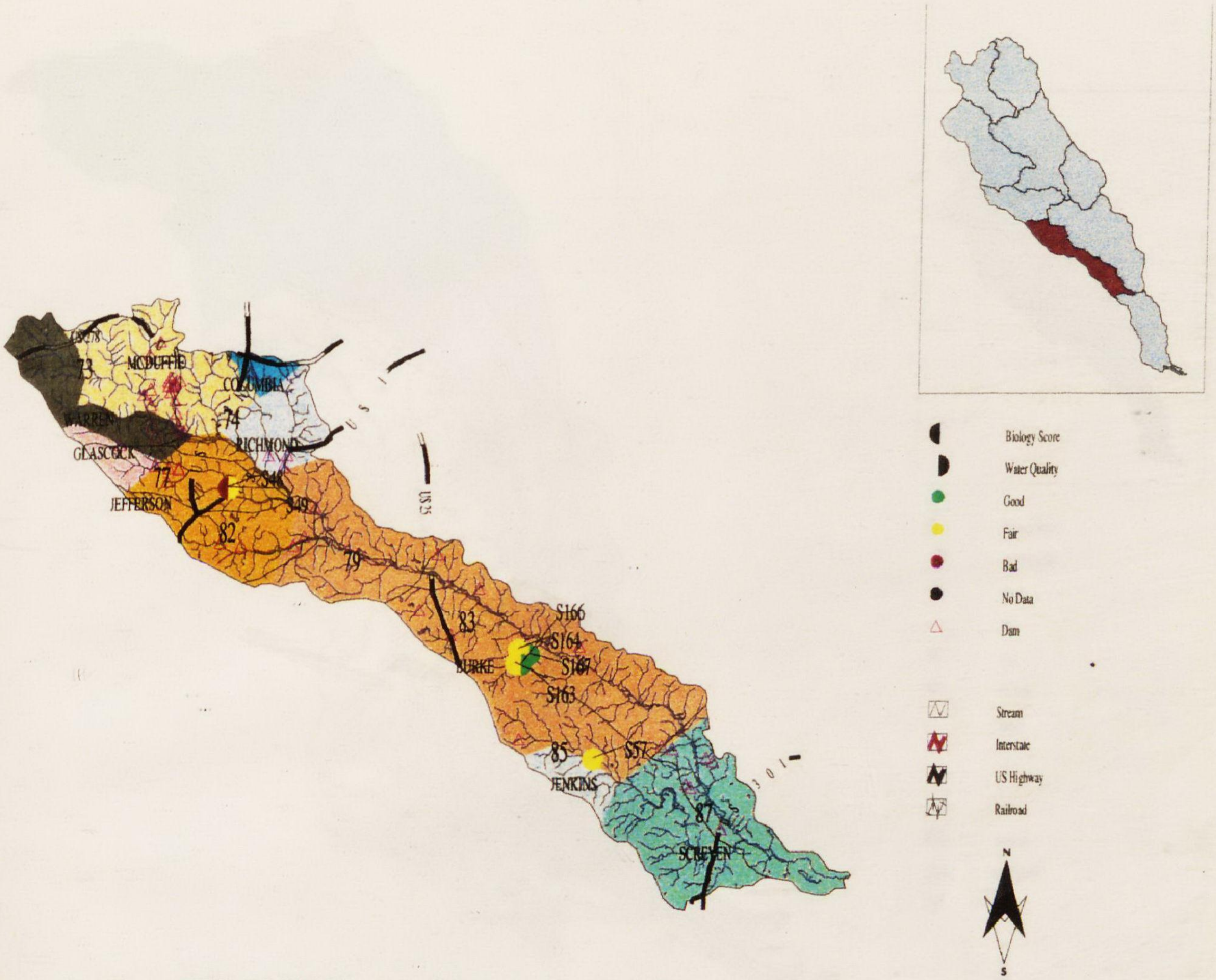


Figure 10. HUC 3060108 Sampling Site Results

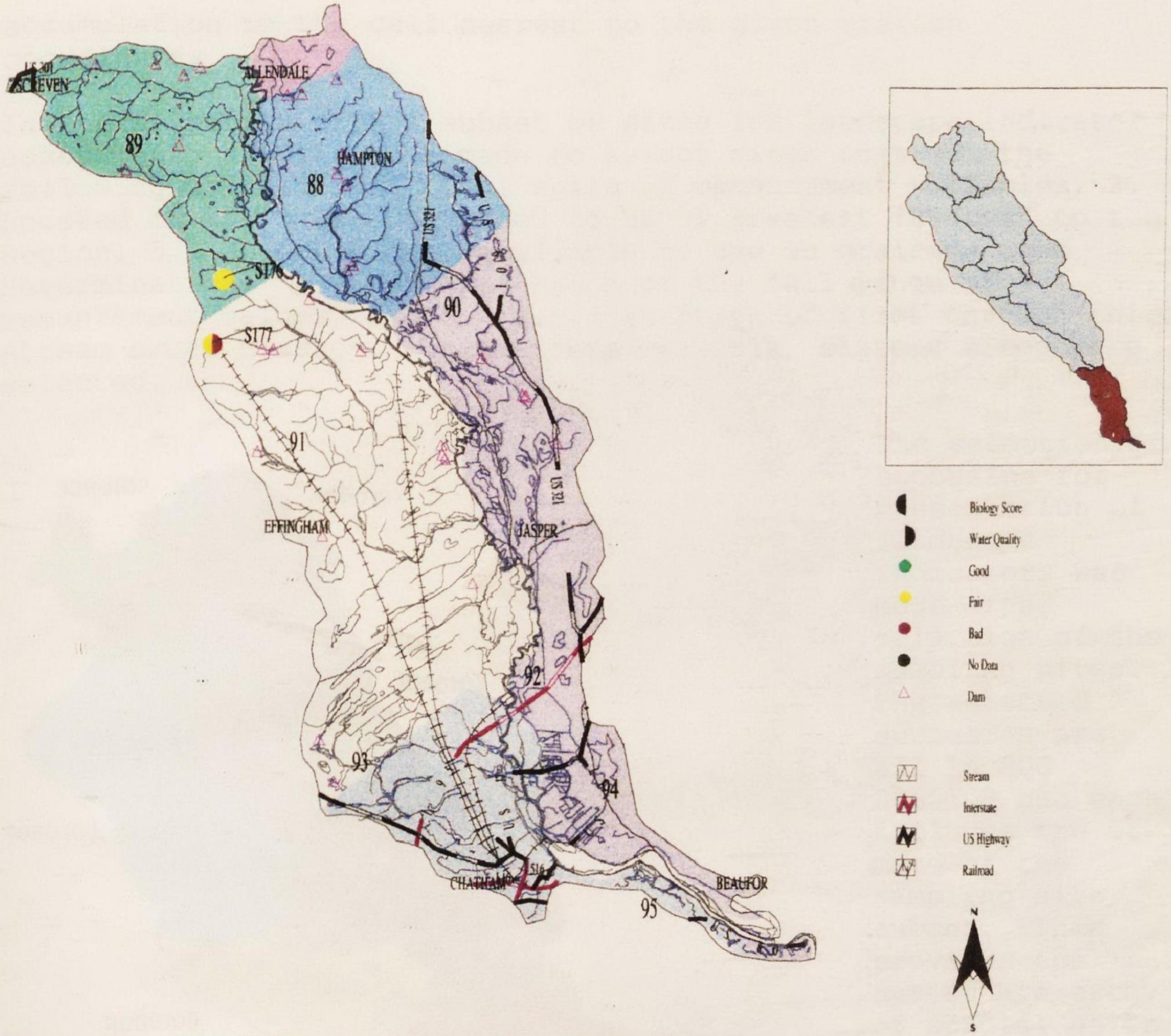


Figure 11. HUC 3060109 Sampling Site Results

The process used to delineate the drainage areas employs hydrological analyses tools contained in the Grid module of ARC/INFO. First sinks in the DEMs are identified and filled. Flow direction is computed as the direction from each 30-m cell towards its steepest downslope neighbor. From the flow direction grid, a flow accumulation grid is created by calculation of the number of cells which flow into each downslope cell; this grid resembles the existing stream network. The sampling station locations are input as pour points. In some cases, the sampling point coordinates did not fall directly on a flow accumulation path; in these instances, the pour point was placed on the flow accumulation in the cell nearest to the given station coordinates.

In the selection of the subset of sites for landscape indicator assessment, efforts were made to select sites that met the following criteria: 1. Full suite of measurement variables, 2. Located in the areas indicated to be of greatest interest to the Region, 3. 30-m DEM data available to use in drainage area determination, 4. Representation of the full range of measurement values, and 5. Representation of first through third stream order classes. Using these criteria, sixteen sites were selected.

The selection of subbasins for presentation of landscape indicators was made after selection of the sampling sites. The selected subbasins are all in HUC 3060103 and each includes one or more of the sampling site subset. This provides the nested hierarchy of spatial units in the assessment. An arbitrary number was assigned to each subbasin after merging the separate Georgia and South Carolina coverages. The subbasins are shown in Figure 12.

Figure 12. HUCs and Subbasins

LANDSCAPE ASSESSMENT

HUC Indicators

The Savannah River Basin is arrowhead-shaped, trending generally northwest to southeast. The basin is comprised of nine USGS 8-digit HUCs (numbered 3060101 through 3060109, hereafter referred simply by the last digit), spanning three ecoregions: Blue Ridge, Piedmont and Coastal Plains. As shown in Figure 13, HUCs 1 and 2 are primarily in the Blue Ridge ecoregion, HUCs 3, 4, 5, and 7 lie in the Piedmont, and the majority of HUCs 6, 8, and 9 are in the Coastal Plain. As shown in Table 1, the size of the HUCs varies from 200,987.55 ha (HUC 7) to 488,842.20 ha (HUC 6). Associated riparian areas vary from 31,324.14 ha (HUC 7) to 88,651.85 ha (HUC 3), based on a 100-m corridor on either side of all RF3 stream arcs.

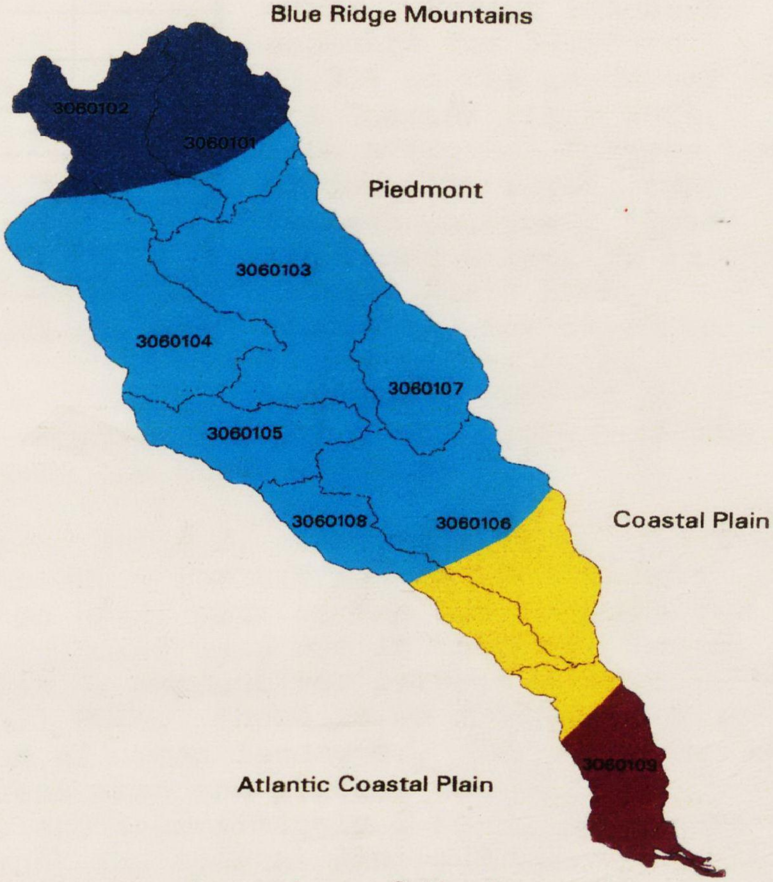


Figure 13. Ecoregions of the Savannah River Basin

Table 1. Physical Dimension Statistics for 8-digit HUCs

HUC	Total Area (ha)	Riparian Corridor (ha)	Stream Length (km)	Stream Density (m/ha)
3060101	272,812.23	55,585.95	3066.68	11.24
3060102	258,218.91	54,114.18	2994.61	11.60
3060103	483,189.03	88,651.85	4803.99	9.94
3060104	398,298.06	65,842.94	3463.28	8.70
3060105	204,446.97	32,453.26	1636.33	8.00
3060106	488,842.20	83,668.92	1765.63	3.61
3060107	200,987.55	31,324.14	4771.76	23.74
3060108	220,108.41	37,124.59	2044.32	9.29
3060109	248,158.71	47,316.49	2679.38	10.80

Landcover types are derived from MRLC data, nominal base year 1992. Differences among the three ecoregions are evident in the forest landcover statistics for the HUCs, Table 2. Deciduous and evergreen forests predominant in HUCs 1 through 5 and 7,

Table 2. Forest Cover Types, Percent Cover by HUC

HUC	Evergreen	Mixed	Deciduous	Woody Wetlands
3060101	23.36	10.45	37.92	0.62
3060102	25.66	12.15	45.68	0.27
3060103	28.35	11.07	25.28	0.55
3060104	23.72	9.65	38.66	0.36
3060105	39.95	8.85	28.57	0.69
3060106	33.39	7.22	12.74	11.54
3060107	50.21	9.72	18.69	0.74
3060108	24.17	7.50	15.38	10.86
3060109	25.24	4.63	7.33	31.46

the HUCs comprising the Blue Ridge and Piedmont ecoregions; all forest types account for 64.7 to 83.49% of the total land cover. Forest landcover accounts for 37.20 to 53.35% of the landcover in the Coastal Plain HUCs, with evergreen forests the predominant forest type. Wetland landcover types are found primarily in the Coastal Plain HUCs, accounting for 11.10 to 35.93% of the total landcover, most of it in woody wetlands. Wetlands comprise less than one percent of the landcover in the HUCs outside the Coastal Plains.

Agricultural landcover types, Table 3, comprise 9.91 to 32.47% of the total landcover in each HUC. Pasture/hay is the dominant agricultural land use in the upper part of the basin, while row crops are the largest agricultural land use in the lower basin. Urban landcover types, Table 4, account for between 0.85 to 5.33% of the total land use in all HUCs. There is no ecoregion-related pattern to the distribution of urban landcover. Barren landcover types, Table 5, comprise less than one percent of the total landcover in HUCs 1 and 2, and approximately 2 to 10 percent of the landcover of the Piedmont and Coastal Plains HUCs.

Table 3. Agricultural Land Cover Types, Percent Cover by HUC

HUC	Pasture /Hay	Row Crops	Other Grasses
3060101	10.18	5.05	0.76
3060102	6.76	3.46	0.28
3060103	13.21	9.08	0.48
3060104	15.51	7.59	0.35
3060105	4.01	6.90	0.12
3060106	1.60	14.60	0.46
3060107	3.32	6.59	0.09
3060108	2.76	29.71	0.06
3060109	1.78	14.15	0.38

Table 4. Urban Land Cover Types, Percent Cover by HUC

HUC	Low Intensity Residential	High Intensity Residential	High Intensity Commercial/ Industrial
3060101	3.35	0.29	1.16
3060102	1.03	0.06	0.42
3060103	1.74	0.25	0.60
3060104	0.88	0.06	0.49
3060105	0.60	0.08	0.32
3060106	2.72	1.01	1.60
3060107	0.71	0.10	0.30
3060108	0.49	0.10	0.26
3060109	1.03	0.64	1.18

Table 5. Other Land Cover Types, Percent Cover by HUC

HUC	Water	Emergent Wetlands	Barren: Quarries/ Strip Mines	Barren: Bare Rock/Sand	Barren: Transitional
3060101	6.41	0.04	0.19	<0.01	0.22
3060102	3.61	0.07	0.04	<0.01	0.49
3060103	6.60	0.03	0.14	<0.01	2.63
3060104	0.57	0.02	0.14	<0.01	1.98
3060105	4.87	0.03	0.19	<0.01	4.81
3060106	1.46	0.48	0.60	0.01	10.57
3060107	0.39	0.03	0.07	<0.01	9.04
3060108	0.47	0.24	0.54	<0.01	7.45
3060109	3.10	4.47	0.14	0.05	4.40

Table 6. Riparian Corridor Land Cover Types, Percent by HUC

HUC	Forest	Agriculture	Urban	Wetland	Barren	Water
3060101	71.96	8.16	3.19	1.92	0.44	14.51
3060102	79.06	7.03	1.31	1.09	0.33	11.20
3060103	70.63	10.10	1.38	1.77	1.26	14.86
3060104	83.20	12.16	0.76	1.21	0.64	2.04
3060105	76.83	4.84	0.57	2.23	4.17	11.33
3060106	48.79	6.54	3.51	28.90	6.23	6.06
3060107	86.00	4.63	0.40	1.99	5.51	1.50
3060108	48.53	12.78	0.33	32.43	4.20	1.77
3060109	24.13	6.38	1.34	57.23	2.48	8.44

The patterns of landcover/land use within the riparian corridors, Table 6, are not substantially different than those for the HUCs overall, with the exception that water is an appreciable percentage of the landcover within riparian corridors in most HUCs. Agricultural land use within the riparian corridor ranges from 4.63% to 12.78% and urban land use ranges from 0.33% to 3.51%. Barren landcover ranges from less than 1% to a little more than 6%. The predominant landcover types in the HUC riparian corridors are forest and wetlands in the Coastal Plains and forest in the other ecoregions.

While there is some variation in landcover types among the three ecoregions, overall the HUCs are relatively homogeneous in landcover/land use pattern. In all HUCs, natural landcover types comprise greater than 50% of the total landcover. Urban land uses account for only a small percent of the total landcover and agricultural uses account for 1/10 to approximately 1/3 of the total land cover/land use. These results contrast greatly with the results obtained for 8-digit HUCs in the mid-Atlantic region (Jones et al, 1997), where large differences were evident at this scale. The broad-scale patterns evident in the mid-Atlantic (e.g., intensive urbanization of the Coastal Plains, concentrated agricultural land uses in valleys, and isolation of forests to highland areas) are not in evidence in the Savannah River Basin.

Agriculture on slopes greater than 3% grade has been developed as a landscape indicator because the potential for erosion increases significantly at this grade. Similarly, agriculture practiced on highly or moderately erodible soils has a higher potential for erosion. These indicators are developed from overlays of DEMs, MRLC land cover/land use, and erodibility factors contained in the STATSGO soils data base. Results for all of these indicators are generally low, as shown in Table 7. Only HUCs 3 and 4 showed greater than 20% total land area for any of the agriculture-soil-slope indicators, that being agriculture on moderately erodible soils, most of it in pasture/hay. Results for these indicators within the riparian corridors are lower, ranging from nonexistent to less than 12% agriculture on moderately erodible soil in HUC 4, as shown in Table 8.

Table 7. Agriculture on Slopes and Erodible Soils, Percent by HUC

HUC	Pasture/ Hay on Slopes > 3%	Row Crops on Slopes >3%	Pasture/ Hay on Moderately Erodible Soils	Row Crops on Moderately Erodible Soils	Pasture/ Hay on Highly Erodible Soils	Row Crops on Highly Erodible Soils
3060101	2.24	0.98	10.03	4.96	--	--
3060102	1.33	0.67	6.54	3.36	--	--
3060103	1.25	0.75	12.51	8.41	0.61	0.58
3060104	2.25	0.95	15.24	7.45	0.26	0.14
3060105	0.22	0.42	2.22	3.84	1.76	2.71
3060106	0.13	0.72	0.22	1.67	<0.01	0.02
3060107	0.10	0.26	4.20	3.06	--	--
3060108	0.03	0.47	0.48	3.23	--	--
3060109	<0.01	0.02	0.41	2.51	--	--

Table 8. Agriculture-Related Indicators in Riparian Corridors, Percent Area

HUC	Pasture/ Hay on Slopes > 3%	Row Crops on Slopes >3%	Pasture/ Hay on Moderately Erodible Soils	Row Crops on Moderately Erodible Soils	Pasture/ Hay on Highly Erodible Soils	Row Crops on Highly Erodible Soils
3060101	1.04	0.55	4.66	2.50	--	--
3060102	0.95	0.48	4.08	1.94	--	--
3060103	0.51	0.33	5.68	3.53	0.20	0.21
3060104	1.00	0.47	7.98	3.89	0.08	0.07
3060105	0.11	0.16	0.73	1.46	0.98	1.56
3060106	0.05	0.23	0.21	1.17	<0.01	0.01
3060107	0.03	0.12	1.88	1.78	--	--
3060108	0.01	0.10	0.25	1.35	--	--
3060109	<0.01	<0.01	0.21	1.30	--	--

Table 9. Roads Crossing Streams and Impoundments

HUC	Road Crossings	No. Crossings/ Stream km	Dams	No. Dams/ Stream km
3060101	1235	0.40	117	0.038
3060102	964	0.32	58	0.019
3060103	1487	0.31	98	0.020
3060104	1227	0.35	102	0.029
3060105	362	0.22	35	0.021
3060106	1914	1.08	191	0.108
3060107	637	0.13	60	0.013
3060108	842	0.41	52	0.025
3060109	723	0.27	31	0.012

Roads frequently cause increased runoff to streams and contribute pollutants washed off the road surfaces. This phenomenon is represented by the roads-crossing-streams indicator, computed from intersecting digital line graph roads with RF3 stream arcs. As shown in Table 9, values for this indicator range from 362 in HUC 5 to 1,914 in HUC 6. Normalizing these values to the number of road crossings per stream kilometer, also shown in Table 9, shows the greatest frequency of roads crossings per stream kilometer is in HUC 6 with more than one road crossing per kilometer of stream length. The lowest frequency is in HUC 7 with approximately one road crossing for every 8 kilometers of stream length. The remaining HUCs have frequencies in the range of one road crossing for every 2.5 to 5 kilometers of stream length.

Information for dams was obtained from the National Inventory of Dams which tracks all dams greater than 6 feet in height for inspection purposes. As shown in Table 9, the fewest number of dams in any HUC is 31 in HUC 9 while the greatest number is 191 in HUC 6. Normalizing by the total stream length within each HUC shows the greatest frequency of dams is also in HUC 6, with one dam for every 9 kilometers of stream length. The lowest frequencies of dams are in HUC 9 and HUC 7, with roughly one dam for every 80 kilometers of stream length. The locations of dams are depicted in Figure 1.

Subbasin Indicators

As discussed above, the landscape indicators at the HUC level show some variation among HUCs attributable to natural landcover variation at the ecoregion level. However, the patterns of land use are generally consistent across ecoregions and among HUCs. This section focuses on the next scale, the subbasin. Landscape indicators are presented for several subbasins of HUC 3. These particular subbasins were selected because they each contain one or more of the sampling sites selected for analysis. The landscape indicators produced for the subbasins are the same as those produced for the HUCs.

Table 10. Physical Dimension Statistics for Selected Subbasins

Subbasin	Total Area (ha)	Riparian Corridor (ha)	Stream Length (km)	Stream Density (m/ha)
20	55,797.39	9,089.19	486.79	8.72
26	53,225.73	10,089.00	528.63	9.93
32	17,195.76	4,311.00	255.27	14.84
36	61,462.62	9,704.07	499.79	8.13
53	68,295.33	12,800.34	695.35	10.18

Physical dimensions of the selected subbasins are shown in Table 10. The total land area in each subbasin ranges from 17,195.76 ha in #32 to 68,295.33 ha in #53. The associated riparian corridors range from 4,311.00 to 12,800.34 ha.

Table 11. Land Cover Types for Selected Subbasins, Percent Area

Land Cover Type	Subbasin				
	20	26	32	36	53
Water	1.30	8.11	20.59	0.57	9.36
Low Intensity Residential	5.60	2.87	3.34	2.10	0.45
High Intensity Residential	0.83	0.57	0.43	0.28	0.03
High Intensity Commercial/Industrial	1.62	1.23	1.04	0.67	0.15
Pasture/ Hay	20.23	16.33	17.54	8.43	5.58
Row Crops	13.84	7.32	14.91	6.73	6.00
Other Grasses	1.53	0.95	0.75	0.36	0.14
Evergreen Forest	16.09	25.32	10.53	37.08	36.70
Mixed Forest	9.55	11.00	5.96	13.12	10.98
Deciduous Forest	28.30	21.12	23.77	24.61	25.98
Woody Wetlands	0.57	0.75	0.82	0.30	0.63
Emergent Wetlands	0.03	0.02	0.11	0.02	0.03
Barren: Quarries/ Strip Mines	0.26	0.14	0.21	0.07	0.09
Barren: Transitional	0.24	4.27	<0.01	5.66	3.88

The landcover statistics for HUC 3 overall are 64.70% forest (approximately 28% evergreen, 25% deciduous and 11% mixed forest), 22.29% agriculture (approximately 13% pasture/hay and 9% row crops), 2.59% urban, 6.60% water, approximately 3% barren, and less than one percent wetlands. Among the subbasins, the forest landcover classes vary from 40.26% in #32 to 73.66% in #53. As shown in Table 11, evergreen forests are the largest forest class in #26, #36, and #53; deciduous is the largest class in #20 and #32. Agricultural land use in #26 is about the same as in the HUC overall (23.65% of which approximately 16% is in pasture/hay). Greater agricultural land use is evident in #20 (34.07% with about 20% in pasture/hay) and #32 (32.45% of which almost 18% is pasture/hay). Less landcover is in agricultural land uses in #36 (15.16%, with more than 8% pasture/hay) and #53 (11.58%, with row crops slightly exceeding pasture/hay). Urban land use is lowest in #53 at less than one percent and highest in #20 at 8.05%. The remaining three subbasins have urban land use in slightly higher percentages than for the HUC overall, ranging 3.05% in #36 to 4.81% in #32.

Table 12. Land Cover Types for Selected Subbasin Riparian Corridors, Percent Area

Land Cover Type	Subbasin				
	20	26	32	36	53
Water	5.89	22.38	25.31	2.25	18.58
Low Intensity Residential	3.97	1.68	1.82	1.36	0.31
High Intensity Residential	0.26	0.19	0.07	0.08	<0.01
High Intensity Commercial/Industrial	0.76	0.47	0.39	0.20	0.08
Pasture/ Hay	8.84	8.68	7.60	3.84	1.94
Row Crops	5.54	3.40	8.32	2.84	2.47
Other Grasses	0.36	0.24	0.21	0.05	0.02
Evergreen Forest	15.78	18.59	14.30	25.59	29.40
Mixed Forest	12.76	12.30	8.49	15.37	10.54
Deciduous Forest	43.15	28.94	30.23	44.56	33.20
Woody Wetlands	2.11	1.69	2.73	0.79	2.27
Emergent Wetlands	0.09	0.05	0.37	0.04	0.10
Barren: Quarries/ Strip Mines	0.34	0.06	0.15	0.03	0.02
Barren: Transitional	0.16	1.32	<0.01	2.99	1.07

In the riparian corridors, forest comprises 53.02 to 85.52% of the total cover, with deciduous the most dominant forest cover type, as shown in Table 12. Agricultural land use within the riparian corridor ranges from 4.41% in #53 to 15.92% in #32 and urban land use comprises from 0.39% to 4.99% of the total riparian land cover. Wetlands account for approximately 3% or less of the riparian land cover types.

Table 13. Agriculture-Related Indicators for Selected Subbasins and Riparian Corridors

	Subbasin				
	20	26	32	36	53
Percent of Subbasin Total Area					
Agriculture on Slopes >3%	3.54	2.71	4.29	1.17	0.60
Agriculture on Moderately Erodible Soils	34.03	23.53	31.64	15.15	4.39
Agriculture on Highly Erodible Soils	--	--	--	--	7.11
Percent of Subbasin Riparian Corridor					
Agriculture on Slopes >3%	1.31	1.34	1.24	0.39	0.27
Agriculture on Moderately Erodible Soils	14.13	11.99	13.02	6.67	1.43
Agriculture on Highly Erodible Soils	--	--	--	--	2.61

The agriculture-soil-slope indicator results for HUC 3 are 2% agriculture on slopes greater than 3%, approximately 21% agriculture on moderately erodible soils, approximately 1% agriculture on highly erodible soils, and less than 0.1% agriculture on slopes greater than 3% in highly erodible soils. Among the subbasins, #20, #26, and #32 have more agriculture on slopes greater than 3% and more agriculture on moderately erodible soil than for the HUC overall; the remaining two subbasins are substantially lower than the HUC overall for both these indicators, as shown in Table 13. Only #53 has any agriculture on highly erodible soil (about 7%) and agriculture on slopes greater than 3% and highly erodible soils (0.35%). Results for these indicators are lower for the riparian corridors, with only subbasins #20, #26, and #32 having more than 10% riparian land cover in agriculture on moderately erodible soils.

Table 14. Roads Crossing Streams and Impoundments for Selected Subbasins

Subbasin	Road Crossings	No. Crossings/ Stream km	Dams	No. Dams/ Stream km
20	299	0.61	19	0.039
26	227	0.43	15	0.028
32	56	0.22	0	--
36	170	0.24	13	0.026
53	82	0.11	4	0.006

Table 14 provides results for the number and frequency of roads crossing streams and dams; these indicators are depicted in Figure 14. Roads crossing streams ranges from 56 in #32 to 299 in #20. There are no dams in #32, but 19 dams in #20. The frequency of roads crossing streams is highest in #20 with approximately one road crossing for every 1.6 kilometers of stream length; the lowest frequency is in #53 with one road crossing per approximately 9 kilometers of stream length. The frequency of roads crossing streams for the HUC overall is approximately one crossing per 3 kilometers of stream length. The frequency of impoundments for the HUC overall is approximately one dam for every 50 stream kilometers. The frequency of dams is lower than for the HUC overall in #32 with no dams and in #53 with approximately one dam for every 167 kilometers of stream length. The greatest frequency of dams among the subbasins is in #20 with one dam per approximately 25 stream kilometers.

At this scale, patterns which may impact water quality begin to be evident. In Figure 5, the sampling stations in #53 are indicated as fair to good (as compared to the overall data range). This subbasin has the highest proportion of landcover in forest among the subbasins, the lowest proportion of agriculture and urban land uses, and a low proportion of agriculture on slopes greater than 3%. Among the selected subbasins, it has the lowest frequency of roads crossing streams. Although is the largest of the subbasins in total area, this subbasin has only 4 dams. However, #53 is the only subbasin among those examined

Figure 14. Roads crossing Streams and Dams in Selected Subbasins

with agriculture on highly erodible soils and agriculture on slopes greater than 3% and highly erodible soils.

In contrast, the sampling sites in #32 and #20 rank as fair to bad compared to the overall data ranges. These two subbasins contain the greatest proportion of agriculture among the subbasins, 28 to 33% agriculture on moderately erodible soils, and 3 to 4% agriculture on slopes greater than 3%. In addition, #20 has the highest proportion of urban land use, the highest normalized roads crossing streams value, and the greatest frequency of dams among the selected subbasins.

Sampling Site Drainage Landscape Indicators

As described above, landscape analysis at the subbasin scale may be adequate to provide a generalized characterization of the Savannah River Basin. One of the objectives of this project, however, is to try to establish relationships among landscape indicators and water quality/aquatic biota indicators. The water quality data were collected at specific sampling sites. To investigate relationships with landscape indicators, it is necessary to delineate the drainage area to the individual sampling site. This was done for a subset of 16 sampling sites. The selection process was described earlier, as was the methodology for delineating the drainage areas.

The drainage areas for the sampling sites range from 122.58 to 10,665.18 ha, as shown in Table 15. In delineating the drainage areas, the locations for the sampling sites frequently did not lie on a stream arc, necessitating a best guess, based on the indicated stream order and proximity to stream arc, as to the point on the arc to use as the pour point. In addition to the landscape indicators calculated for the HUCs and subbasins, indicators of fragmentation were

Table 15. Physical Dimension Statistics for Sampling Site Drainages, Percent Area

Site	Total Area (ha)	Riparian Corridor (ha)	Stream Length (km)	Stream Density (m/ha)
S22	973.98	149.85	7.34	7.54
S27	4,950.90	939.78	47.57	9.61
S68	468.09	88.65	4.34	9.28
S80	6,499.71	884.16	44.90	6.91
S81	6,612.21	908.01	45.85	6.93
S95	10,665.18	1,727.73	89.30	8.38
S103	572.76	69.30	3.32	5.80
S113	747.00	83.79	4.51	6.03
S130	1,169.73	163.80	8.68	7.42
S149	776.52	139.50	6.87	8.84
S151	1,076.22	191.88	9.53	8.85
S155	2,556.72	381.06	18.98	7.42
S195	4,279.41	860.94	46.05	10.76
S197	122.58	43.56	2.06	16.78
S200	1,798.47	377.37	19.11	10.62
S216	551.16	116.19	5.69	10.33

Table 16. Aggregated Land Cover Types for Sampling Site Drainages, Percent Area

Site	Water	Urban	Agriculture	Forests	Wetlands	Barren
S22	0.11	<0.01	59.40	40.21	0.24	0.04
S27	0.34	0.34	18.76	80.06	0.37	0.11
S68	<0.01	<0.01	0.27	96.01	<0.01	3.73
S80	0.56	7.75	18.28	72.71	0.45	0.26
S81	0.55	7.62	17.97	73.17	0.44	0.25
S95	0.62	2.84	8.40	81.52	0.87	5.75
S103	<0.01	<0.01	0.04	89.92	0.08	9.96
S113	0.59	2.80	23.70	61.78	4.56	6.55
S130	1.02	<0.01	62.27	36.28	0.40	0.03
S149	0.03	16.02	44.33	39.38	0.10	0.13
S151	0.09	13.01	40.59	46.00	0.19	0.10
S155	0.11	3.21	17.77	73.21	0.56	5.14
S195	0.85	0.44	3.45	94.49	0.04	0.73
S197	<0.01	5.50	42.59	51.62	0.29	<0.01
S200	0.19	6.77	47.64	42.37	0.23	0.11
S216	0.10	0.07	4.90	87.62	0.13	7.18

generated using a custom, in-house software program. For the fragmentation indicators, the 15 landcover/land use classes of the MRLC data were aggregated to six classes: water, urban, forest, agriculture, wetlands, and barren, as shown in Table 16 for the overall drainage area and in Table 17 for the riparian corridors. In these aggregated land cover types, other grasses are included in agriculture and woody wetlands are included in the wetlands cover type.

Results for agriculture-related indicators over the entire drainage area and the riparian corridor are presented in Table 18. The number of road crossing streams and dams are shown in Table 19. Ten of the 16 sampling site drainages contain no dams; however, where dams are present, they are generally greater in frequency than in the HUC or subbasins overall. The frequency of roads crossing streams ranges from approximately one road crossing per 5.5 kilometers of stream length to a maximum of one road crossing for every stream kilometer.

Results for each indicator were encoded into ARC/INFO Grids. A Grid stack was generated and used to develop a correlation matrix. A separate Grid stack was generated for the riparian corridors contained in the drainage areas for the sixteen sampling sites. With an n of 16, the correlation coefficients are significant at values greater than 0.666 for $\alpha = 0.005$, at values greater than 0.601 for $\alpha = 0.01$, at values greater than 0.507 for $\alpha = 0.025$, and at values greater than 0.425 for $\alpha =$

Table 17. Aggregated Land Cover Types for Sampling Site Riparian Corridors, Percent Area

Site	Water	Urban	Agriculture	Forests	Wetlands	Barren
S22	<0.01	<0.01	29.91	69.37	0.72	<0.01
S27	0.48	0.05	8.19	90.41	0.82	0.06
S68	<0.01	<0.01	0.41	99.19	<0.01	0.41
S80	2.28	6.17	11.81	78.70	0.92	0.11
S81	2.21	6.02	11.63	79.14	0.88	0.11
S95	2.97	2.06	2.83	88.22	1.79	2.14
S103	<0.01	<0.01	<0.01	95.84	0.13	4.03
S113	4.19	<0.01	4.72	60.36	29.53	1.18
S130	6.21	<0.01	25.60	66.60	1.59	<0.01
S149	<0.01	11.35	20.70	67.93	<0.01	<0.01
S151	<0.01	9.52	17.82	72.47	0.19	<0.01
S155	0.07	1.49	6.55	89.14	0.64	2.13
S195	4.21	0.08	7.07	87.39	0.09	1.15
S197	<0.01	9.09	10.75	80.17	<0.01	<0.01
S200	0.45	6.13	26.18	66.71	0.45	0.07
S216	0.08	0.15	2.47	92.56	0.23	4.49

0.05. Using these values, a number of significant correlations between water quality/aquatic biology indicators and landscape indicators were indicated, as shown in Table 20. In general, correlations were the same or less for the riparian corridor than for landscape indicators over the whole drainage area. The primary exception is dissolved oxygen, which exhibited significant correlation only with total anthropogenic cover (U-index, comprised of an aggregation of urban and agriculture land cover types) in the riparian corridor. It should be noted that this analysis is preliminary and is based only on the nonrandomly selected subset of sixteen sampling locations. The data set size was insufficient to perform a cluster analysis. The strongest correlations were between landscape indicators; this is not surprising as several of the landscape indicators contain similar information. The redundancy is needed at this point in the research until the strongest and most sensitive relationships with aquatic indicators can be established.

Figures 15 through 20 depict six of the sampling station drainage areas. Sites S68, S113, and S195 are ranked as good data sites, based on the relative rankings of the data measurements. Site S68 is a small forested drainage located in HUC 2. Site 113 is also relatively small and is located in HUC 6; although agriculture and urban areas are evident within the drainage, they are fragmented as compared to the forest landcover; much of the riparian corridor is wetlands. Site S195 is a larger drainage and higher order stream located in HUC 2. All of the landcover

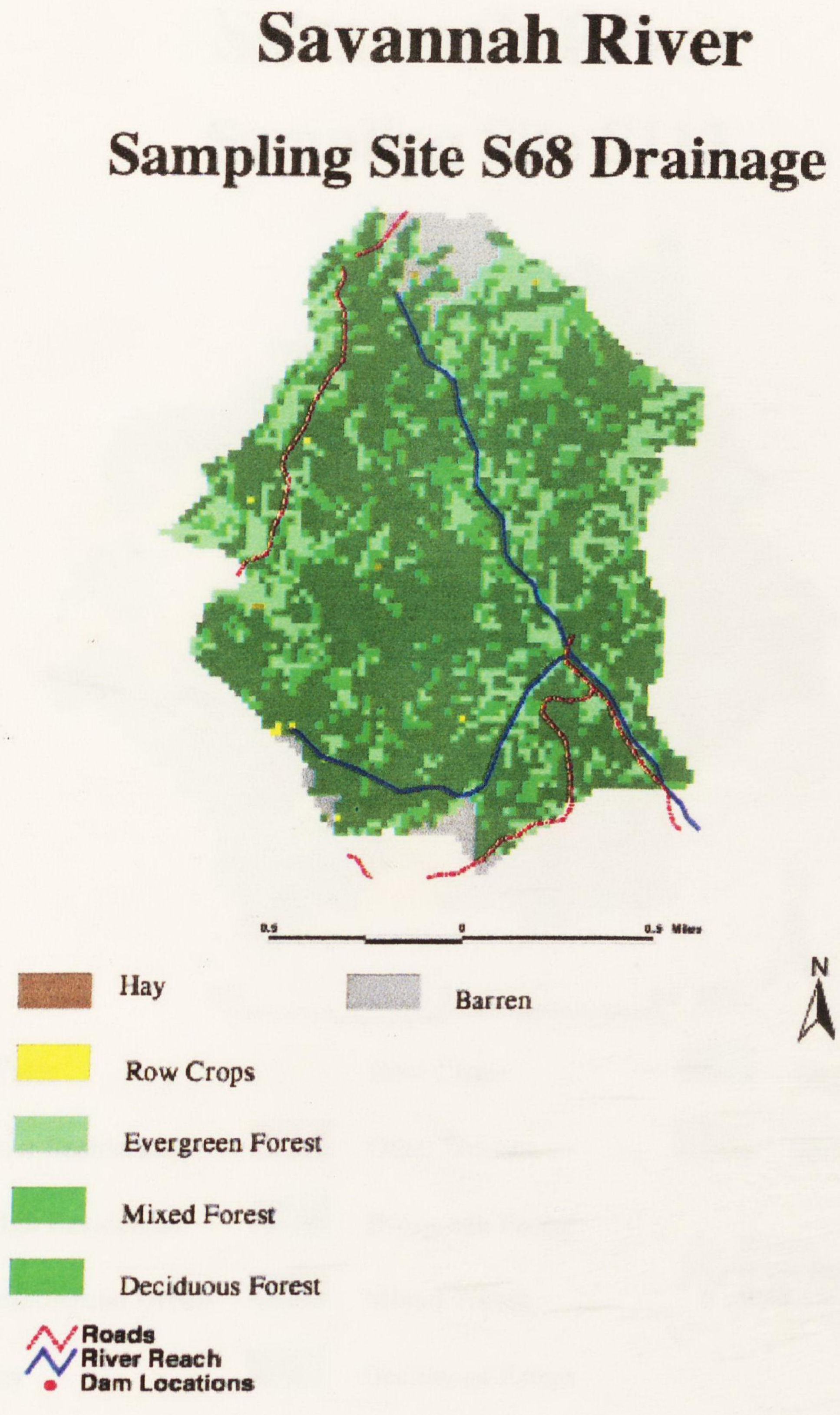


Figure 15. Sampling Site S68 Drainage.

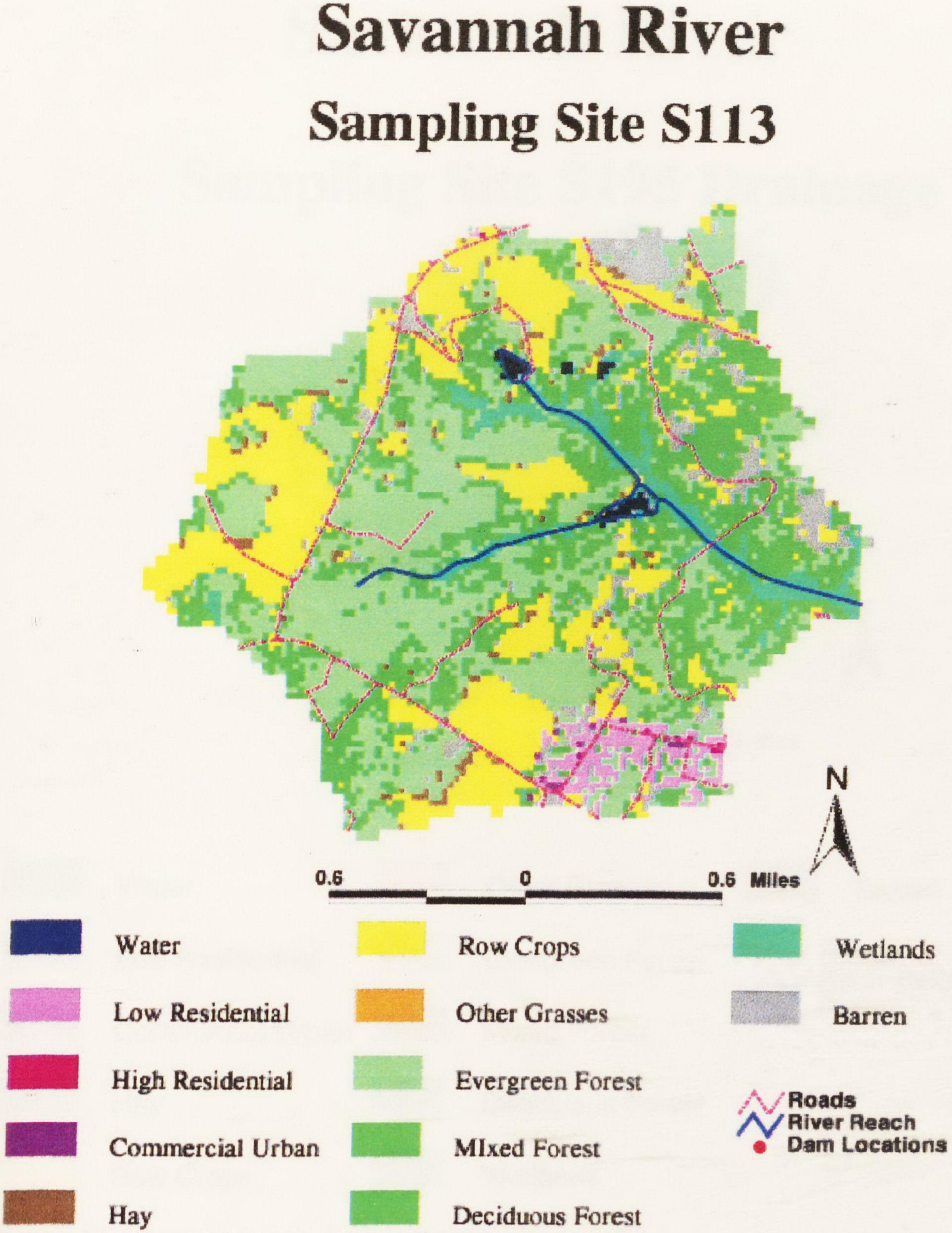


Figure 16. Sampling Site S113 Drainage

Savannah River

Sampling Site S195 Drainage

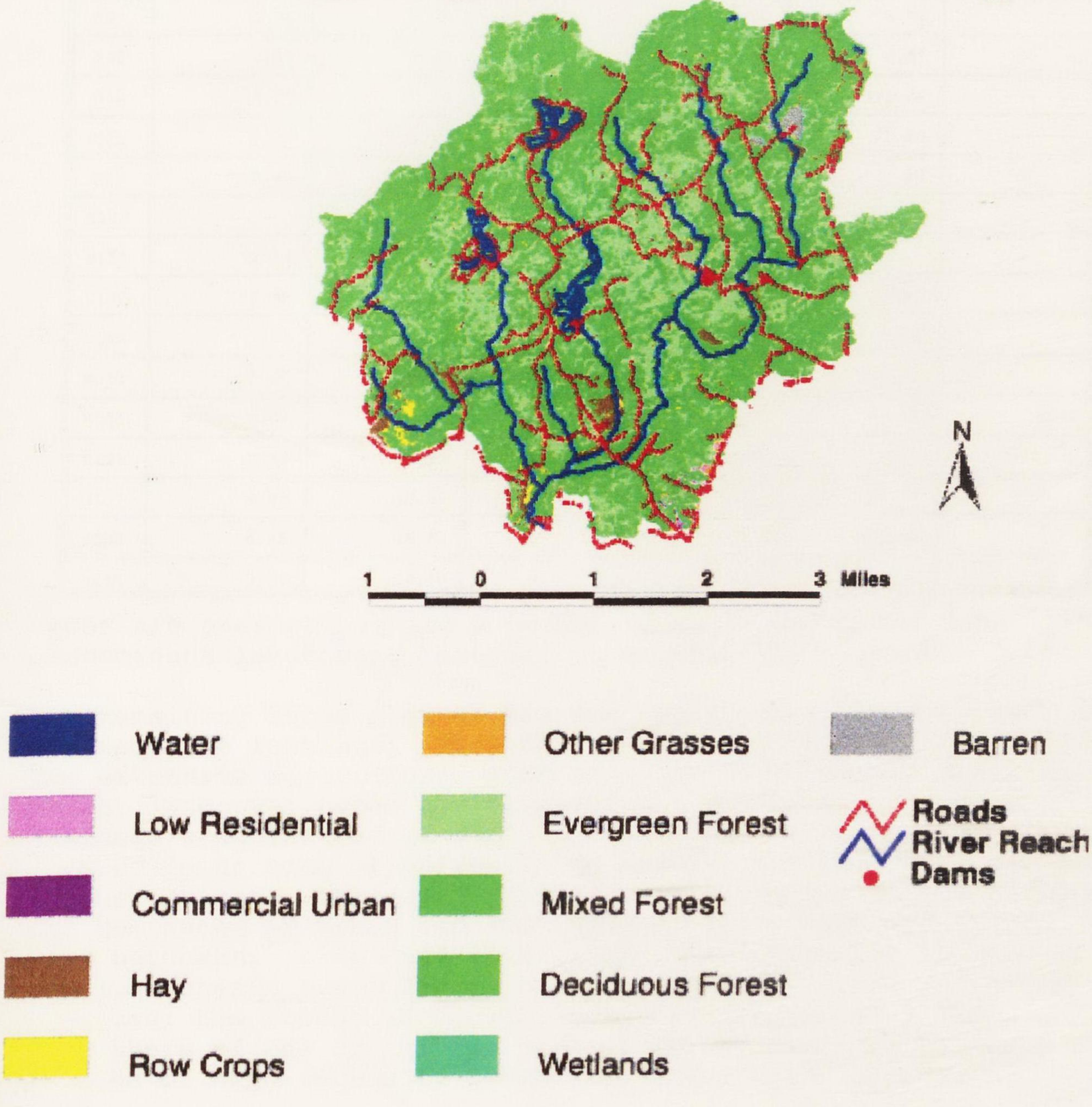


Figure 17. Sampling Site S195 Drainage

Table 18. Agriculture-Related Indicators for Sampling Site Drainages and Riparian Corridors, Percent Area

	Total Area			Riparian Corridor		
Site	Agriculture on Slopes > 3%	Agriculture on Moderately Erodible Soil	Agriculture on Highly Erodible Soils	Agriculture on Slopes >3%	Agriculture on Moderately Erodible Soils	Agriculture on Highly Erodible Soils
S22	3.91	59.40	--	2.22	29.91	--
S27	0.26	18.69	--	0.03	8.19	--
S68	0.19	0.27	--	0.30	0.41	--
S80	1.57	17.72	--	0.66	11.64	--
S81	1.54	17.42	--	0.65	11.43	--
S95	0.28	5.87	--	0.02	2.30	--
S103	--	0.04	0.04	--	--	--
S113	2.14	--	--	2.04	--	--
S130	9.59	62.27	--	1.70	25.60	--
S149	1.85	41.40	--	0.32	18.83	--
S151	2.35	38.36	--	0.28	16.46	--
S155	1.07	17.33	17.33	0.02	6.38	6.38
S195	0.64	3.37	--	0.96	7.07	--
S197	9.92	42.59	--	0.21	10.75	--
S200	4.25	45.51	--	2.43	25.42	--
S216	0.09	4.90	--	0.15	2.47	--

types are present, as are a number of roads and a few dams. The predominant landcover, however, is unfragmented forest.

The remaining three figures are indicative of sites with fair to bad relative rankings. Site S22, located in HUC 3, subbasin #39 has extensive agriculture, much of it in large blocks while the forest landcover types are fragmented. Site S80 is a large drainage area located in HUC 3 subbasin #36; the sampling site is located in an area of unfragmented forest, but the upper reaches of the drainage, including the headwaters of most of the streams are dominated by urban and agricultural landcovers and extensive road networks. Site S149 is a fairly small drainage located in HUC 3, subbasin #20. There is extensive agriculture and urban land use; the forest landcover is highly fragmented. The headwaters of one of the two streams in the drainage is found in an area of high intensity commercial/industrial land use.

Savannah River

Sampling Site S22 Drainage

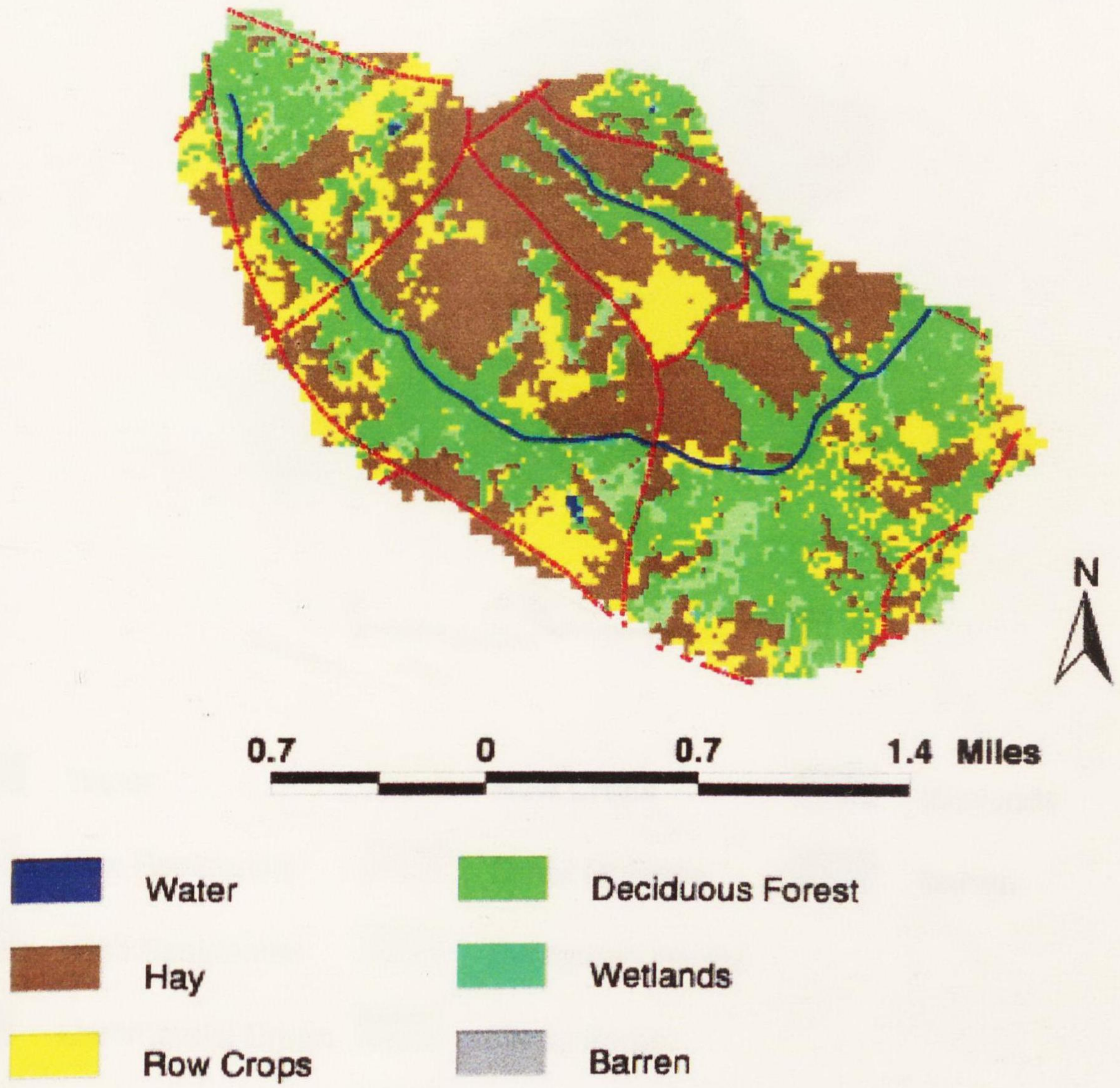


Figure 18. Sampling Site S22 Drainage

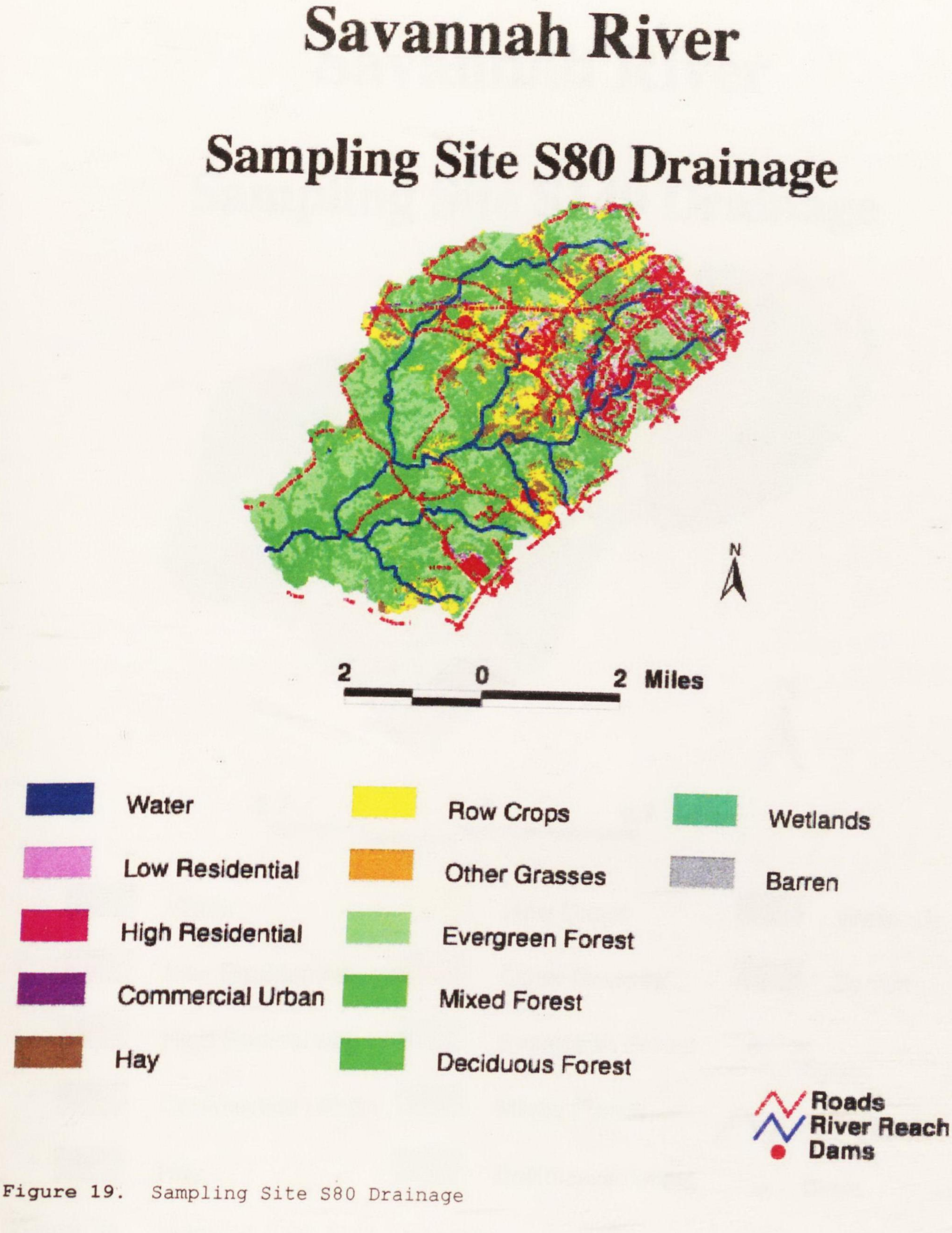


Figure 19. Sampling Site S80 Drainage

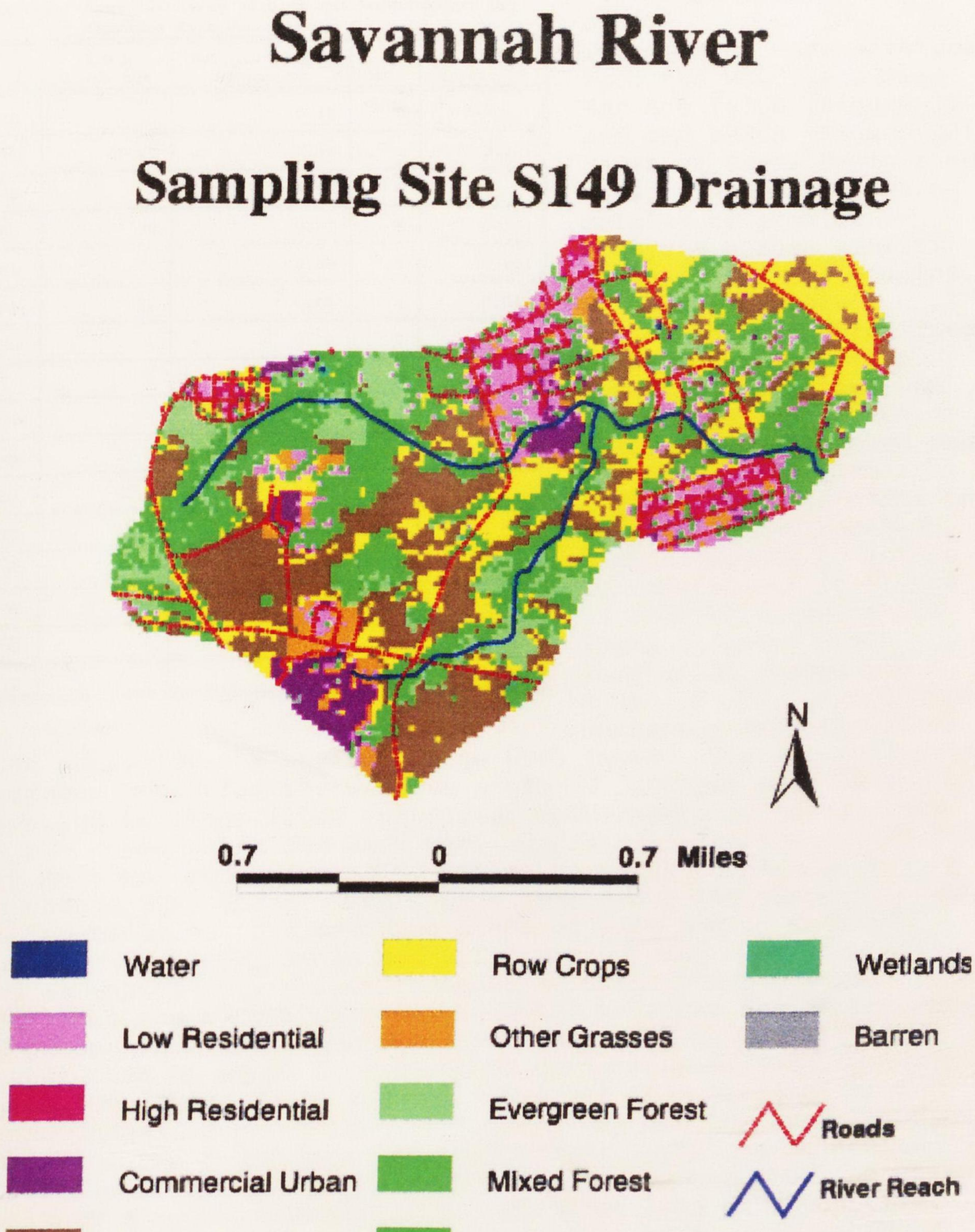


Figure 20. Sampling Site S149 Drainage

Table 19. Roads Crossing Streams and Impoundments for Sampling Site Drainages

Site	Road Crossings	No. Crossings/ Stream km	Dams	No. Dams/ Stream km
S22	3	0.41	0	--
S27	21	0.44	1	0.021
S68	2	0.46	0	--
S80	30	0.67	6	0.134
S81	30	0.65	6	0.131
S95	37	0.41	4	0.045
S103	1	0.30	0	--
S113	1	0.22	0	--
S130	4	0.46	1	0.115
S149	5	0.73	0	--
S151	8	0.84	0	--
S155	4	0.21	0	--
S195	27	0.59	4	0.087
S197	1	0.49	0	--
S200	19	0.99	0	--
S216	1	0.18	0	--

Landscape Change

Two mosaics were developed from the NALC data base for the 1970s (Figure 21) and the 1990s (Figure 22) Savannah River Basin study area. The mosaics were matched to provide analysis across similar areas of the two mosaics. Both mosaics were processed into normalized difference vegetation index (NDVI) images and the values in the 70s mosaic was subtracted from the 90s. Positive numbers indicate gains in vegetation and negative numbers equate to losses in vegetation; in Figure 23 vegetation gains are shown in green while vegetation losses are shown in red. A standard deviation was calculated using n-1, for the entire change NDVI image. The Arc/Info grid coverages depicting the various areas of interest were then converted to image files (hereafter referred to as masks), and the UTM coordinates for each were recorded. The resolution for each mask was converted to 60 meters to match the resolution of the change NDVI image. The change NDVI image was repeatedly sub-sampled to select the matching areas of each mask. Each sub-sampled change NDVI image and its corresponding mask were then used as inputs to a custom in-house software program which calculates the amount of cells (pixels) that are inside the mask and groups them into 4 categories. They are: cells which are greater than or equal to 4 standard deviations of loss in vegetation, those cells which are greater than or equal to 2 standard deviations of loss in vegetation, and the corresponding numbers of cells for gains in vegetation. In the following tables the losses and gains have been grouped together and shown as either a negative number for percent of loss or a positive number for percent of gain.

An additional column is used to represent the cells removed from the study area, which contain negative NDVI indices in either the

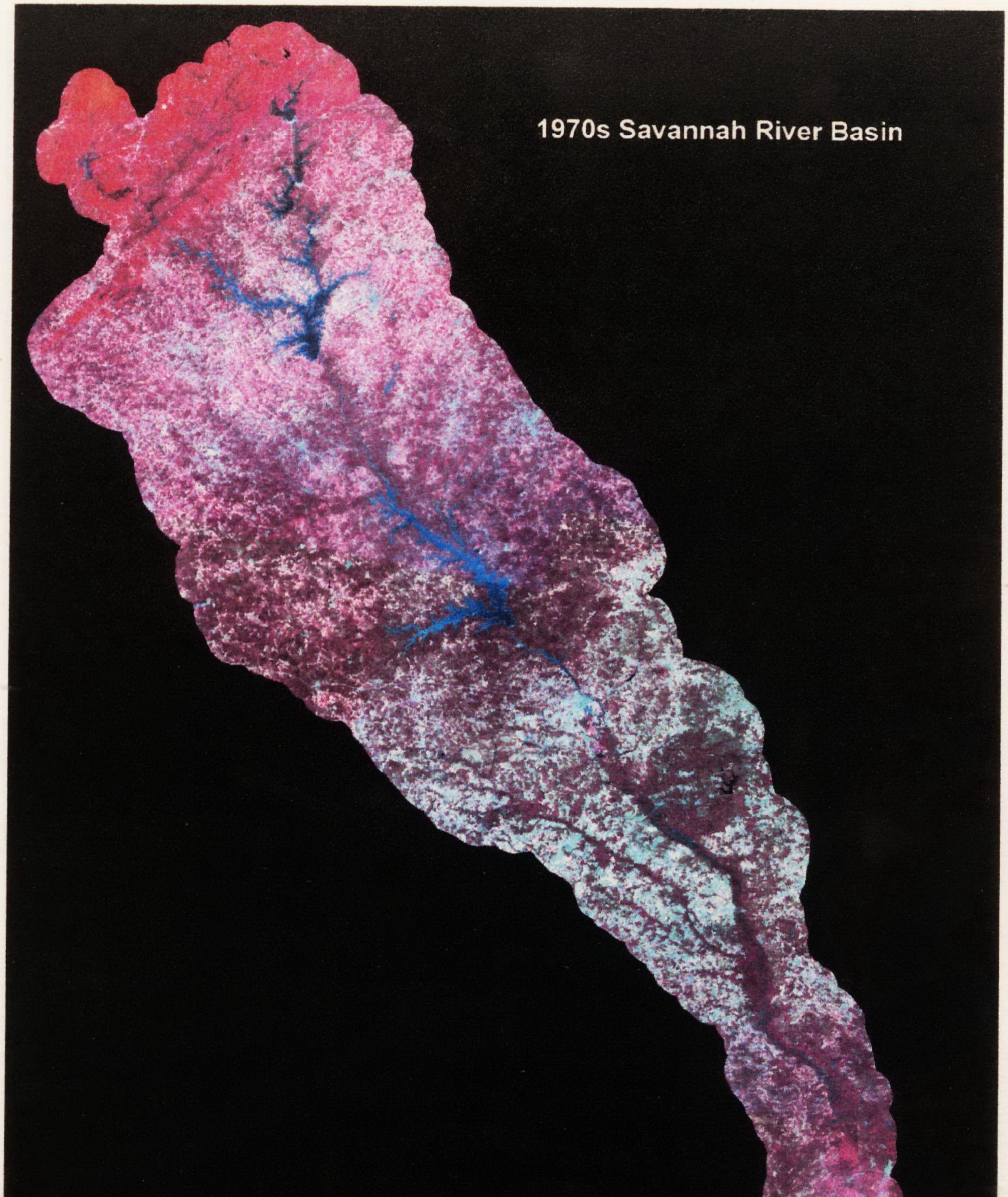


Figure 21. Mosaic of circa 1970 NALC images.

Figure 22. Mosaic of circa 1992 NALC images.

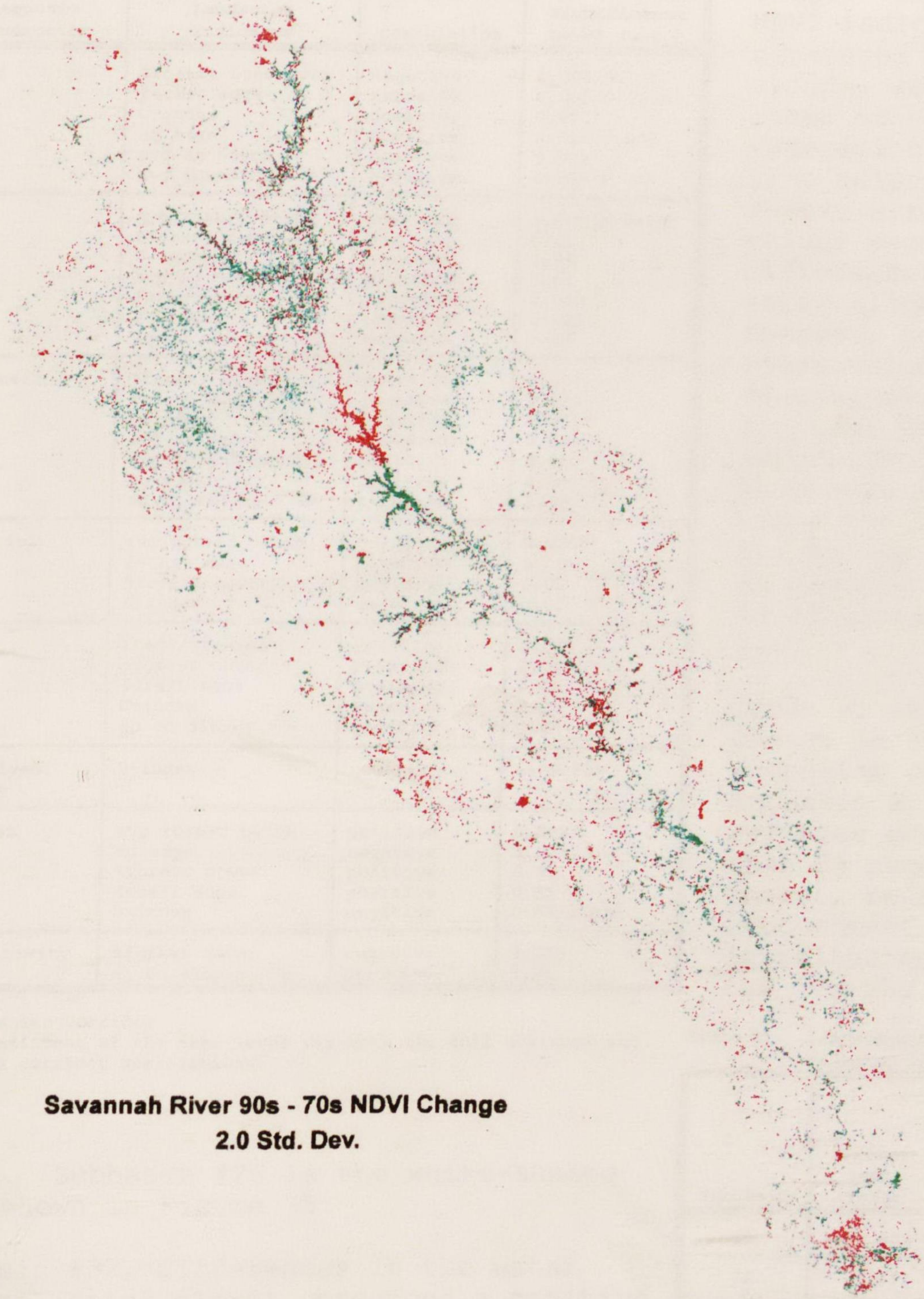


Figure 23. Gains (green) and losses (red) in vegetation, 1970s to 1990s

Table 20. Correlation of Aquatic and Landscape Indicators for Sample Site Drainages and Riparian Corridors

Aquatic Indicator	Landscape Indicator	Correlation	Significance Level (α =)
AGPT	%Forest Landcover	negative	0.005/0.01R
	forest edge	negative	0.005/0.025R
	U-index	positive	0.005*
	ag_edge	positive	0.025/0.05R
	avg ag patch	positive	0.025
	avg forest patch	negative	0.025/0.01R
EPT	avg forest patch	positive	0.005/0.025R
	%forest cover	positive	0.01
	U-index	negative	0.01
	avg ag patch	negative	0.01
	forest edge	positive	0.025
	ag edge	negative	0.025
	ag on slopes > 3%	negative	0.05
Richness	%forest cover	positive	0.025
	forest edge	positive	0.025
	U-index	negative	0.025*
	avg ag patch	negative	0.05
	avg forest patch	positive	0.05
	ag edge	negative	0.05
	ag on slopes > 3%	negative	0.05
Fish_ibi	avg forest patch	positive	0.025*
	forest edge	positive	0.05*
	%forest cover	positive	0.05
	U-index	negative	0.05
pH	roads/streams	positive	0.025
	%forest cover	positive	0.05
	forest edge	positive	0.05
	U-index	negative	0.05
	ag on slopes >3%	negative	0.05
Dissolved Oxygen	U-index	negative	--/0.05R
Habitat	avg forest patch	positive	0.025*
	ag edge	negative	0.025
	%forest cover	positive	0.05
	forest edge	positive	0.05
	U-index	negative	0.05/0.01R
Conductivity	%forest cover	negative	0.005
	ag on slopes >3%	positive	0.05

R = riparian corridor
* = significant at the same level for both the full drainage and riparian corridor correlations.

70s or 90s NDVI image. Negative NDVI indices are generated by clouds, water and other non-vegetation. This also helps to remove erratic NDVI values caused by differences in solar illumination. However, sometimes these values are meaningful, as in the case where an impoundment may have been installed after the 70s image and before the 90s. An example of this is shown in Figure 24.

Table 21 depicts change in selected subbasins of HUC 3. Subbasin #26 reflects greater than 3% negative change, because an impoundment was installed between the 70s and the 90s image. Subbasin #26 is the white-shaded area shown in Figure 25.

Subbasin #32, differences in the water surface (solar glare) produced a positive change in vegetation which offset the loss in that area. When the water areas

Table 21. Landscape Change for Selected Subbasins

Subbasin	Percent NDVI Change	Percent NDVI Change, Negative numbers removed
20	-9.803	-9.670
26	-9.993	-6.838
32	0.661	-6.384
36	-2.695	-2.826
52	-0.366	-1.873

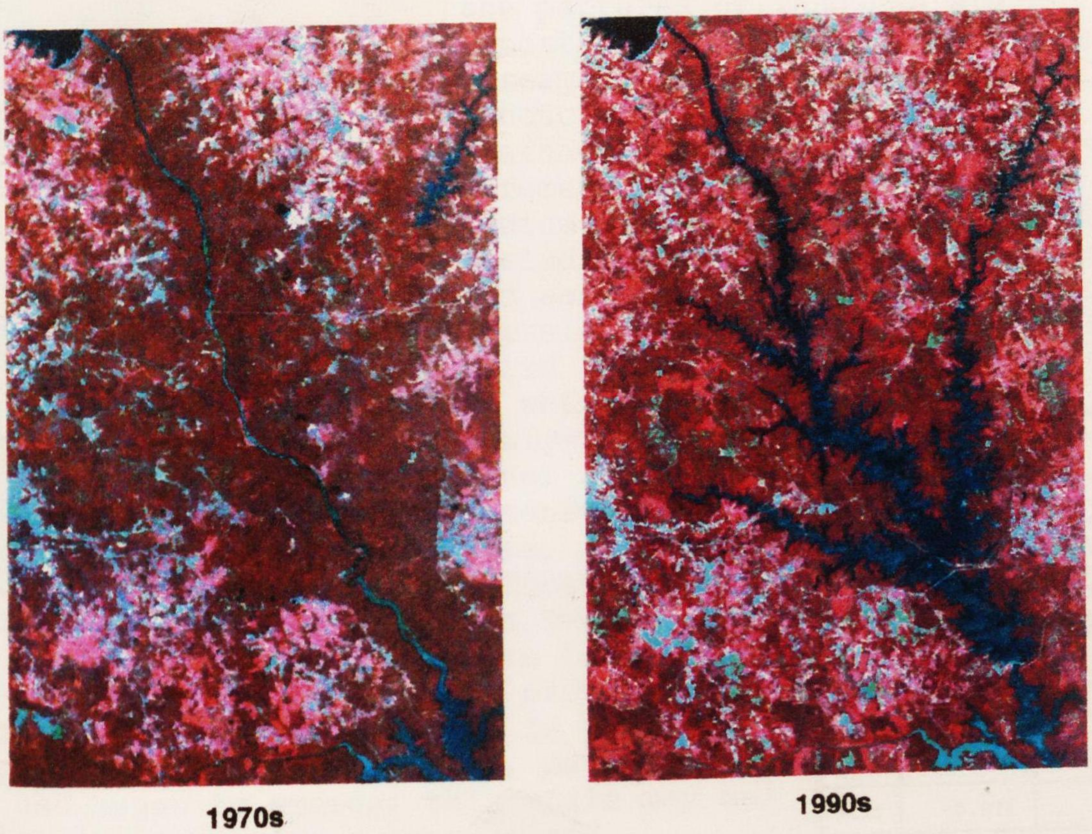


Figure 24. Lake Russell was created between the 1970s and 1990s.

(negative NDVI numbers) were removed the overall sub-watershed had a loss of greater than 6%.

Table 22 shows the NDVI change in the drainage areas of the selected sixteen sampling sites.

SUMMARY

The three questions posed as objectives by the Region can now be addressed:

Are both the proportions of land uses and the spatial pattern of land uses important for characterizing and modeling stream condition in watersheds/ecoregions of different areas?

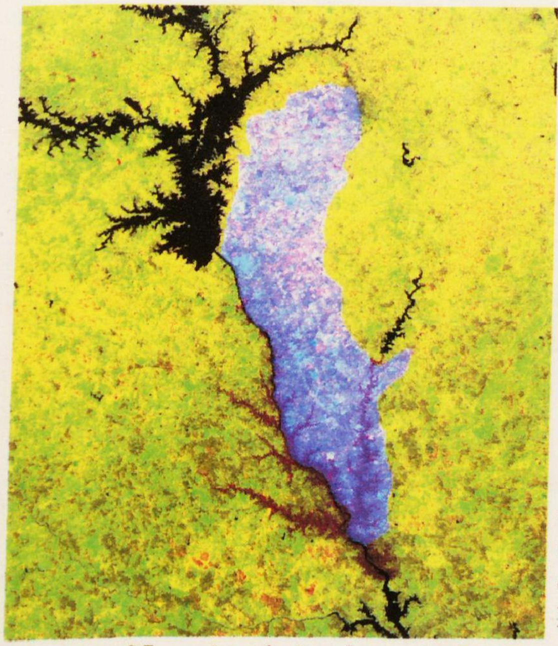


Figure 25. Land Surface Loss to Lake Russell in Subbasin #26.

As shown in this landscape assessment, both the proportion and the patterns of land use are important in assessing impacts on streams. In the correlation analysis conducted on the sampling site drainages, both total landcover types (%forest, U-index) and pattern indicators (fragmentation indicators including average patch size, forest and agriculture edges) were found to correlate with aquatic indicators. A third important element is the scale at which analysis is done. As demonstrated here, landscape indicators at the HUC level were too coarse to provide any indications of water quality. In the analysis of selected subbasins, patterns of land use began to emerge; this scale may be sufficient to provide a generalized characterization of the basin.

Table 22. Landscape Change for Sampling Site Drainages

Site	Percent NDVI Change	Percent NDVI Change, Negative numbers removed
Class "Good"		
S155	-0.873	-0.859
S68	-0.613	-0.613
S195	-0.976	-1.245
S113	-4.193	-4.000
Class "Bad"		
S80	-2.975	-2.748
S197	-8.235	-8.235
S149	-8.994	-7.325
S22	-18.404	-17.591
Class "Fair"		
S81	-2.930	-2.707
S216	-1.235	-1.235
S103	-0.626	-1.627
S27	-4.717	-4.674
Class "Other"		
S151	-6.992	-5.721
S200	-4.528	-4.328
S130	-13.372	-11.149
S95	-2.932	-2.871

Can land uses near the streams better account for the variability in ecological condition than land use for the entire watershed/ecoregion?

In this particular assessment, landscape indicators for the riparian corridors did not provide stronger correlation with aquatic indicators, with the exception of dissolved oxygen. It should be remembered, though, that this is one analysis of a small spatial area in one region with a particular suite of indicators. In other situations the riparian corridor may be of greater importance than the overall watershed. Even in this region, the southern portion of the basin has riparian corridors dominated by wetlands. Only one site from this area was used in the analysis and the entire sampling data set contains only a few sites in this ecoregion. A separate analysis of wetlands-dominated systems is probably worthwhile.

Does the size of the watershed/ecoregion influence statistical relationships between landscape characteristics and ecological condition?

There was no indication in this analysis of any relationship with the spatial extent of the drainage areas. This includes the landscape indicators developed for the HUCs and subbasins. In the sampling site analysis, one of the selection criteria was to include streams of varying order; by doing so, both small and large drainage areas were included. Drainage area was included in the correlation analysis; no correlation was shown with any of the aquatic indicators.

REFERENCES

Bara, T.J., ed. 1994. *Multi-Resolution Land Characteristics Consortium-Documentation Notebook*. U.S. Environmental Protection Agency, Environmental Monitoring and Assessment Program, Research Triangle Park, NC.

Jones, K.B., K.H. Riitters, J.D. Wickham, R.D. Tankersley, Jr., R.V. O'Neill, D.J. Chaloud, E.R. Smith, and A.C. Neale. 1997. *An Ecological Assessment of the United States Mid-Atlantic Region: A Landscape Atlas*. U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.

Natural Resource Conservation Service. 1996. *State Soil Survey Geographic Data Base (STATSGO) Metadata.* U.S. Department of Agriculture, Natural Resource Conservation Service, Washington, D.C.

Santa Cruz County. 1998. *Riparian Corridors* http://gate.cruzio.com/~countysc/pln/riparian.htm

South Carolina Department of Natural Resources. 1998. *South Carolina Rivers Program: Recommended Best Management Practices for River-bordering Lands* http://www.dnr.state.sc.us/water/envaff/river/bmps.html

U.S. Army Corps of Engineers. 1997. *The National Inventory of Dams* http://corpsgeo1.usace.army.mil/

U.S. Environmental Protection Agency. 1998. *Ecological Restoration* http://www.epa.gov/owowwtr1/NPS/Ecology/chap1.html

U.S. Environmental Protection Agency. 1997. *History of the U.S. EPA's River Reach File*. U.S. Environmental Protection Agency, Office of Water, Washington, D.C.

U.S. Environmental Protection Agency. 1993. *North American Landscape Characterization (NALC) Research Plan.* U.S. Environmental Protection Agency, Office of Research and Development, Washington, D.C.

U.S. Geological Survey. 1990. *Digital Elevation Models, National Mapping Program Technical Instructions, Data Users Guide 5*. U.S. Geological Survey, Second Printing (Revised), Reston, Virginia, 1990.

U.S. Geological Survey. 1989. *Digital Line Graphs from 1:100,000-Scale Maps-Data Users Guide 2*. U.S. Geological Survey, Reston, Virginia.

U.S. Geological Survey. 1982. *Codes for identification of hydrologic units in the United States and the Caribbean outlying areas.* U.S. Geological Survey, Circular 878-A, Reston, VA.

Acknowledgments

Data processing and analysis support was provided by Lockheed-Martin under Contract 68-C5-0065 to the U.S. Environmental Protection Agency. In particular, the contributions of Lee Bice, Karen Lee, and Dick Dulaney of Lockheed-Martin were essential in the preparation of analyses used in this report.

G

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

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 considered 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.

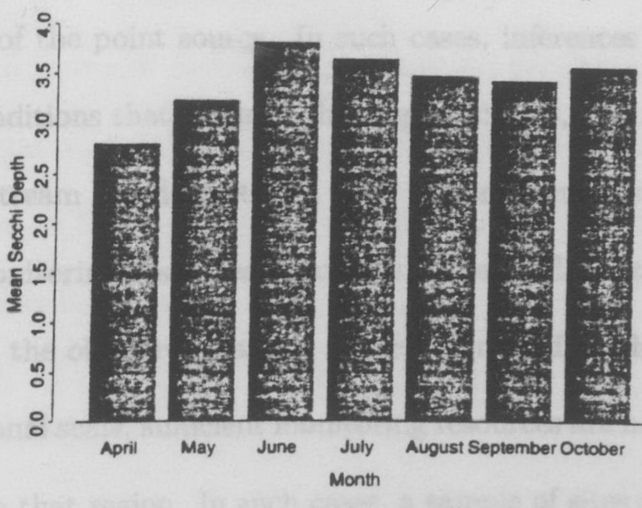


Figure 1: Mean secchi depth by month.

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 investigator. 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

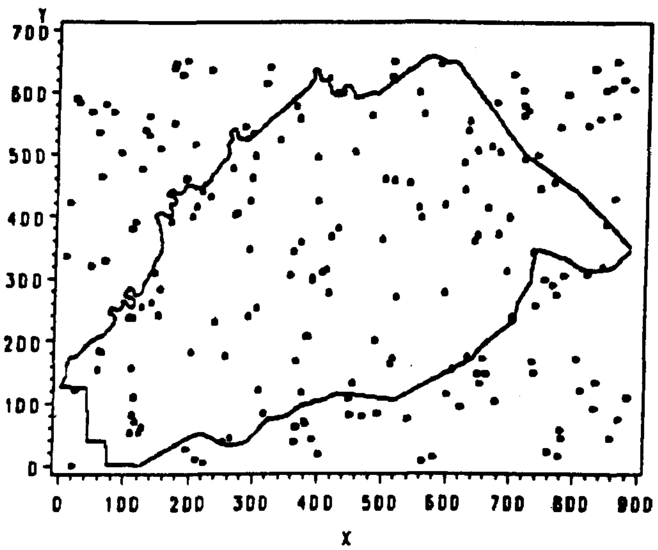


Figure 2: Simple random sample of 100 sites in Ebenezer Aquifer (closed circles). Sites falling outside the study region (open circles) are excluded from the sample.

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

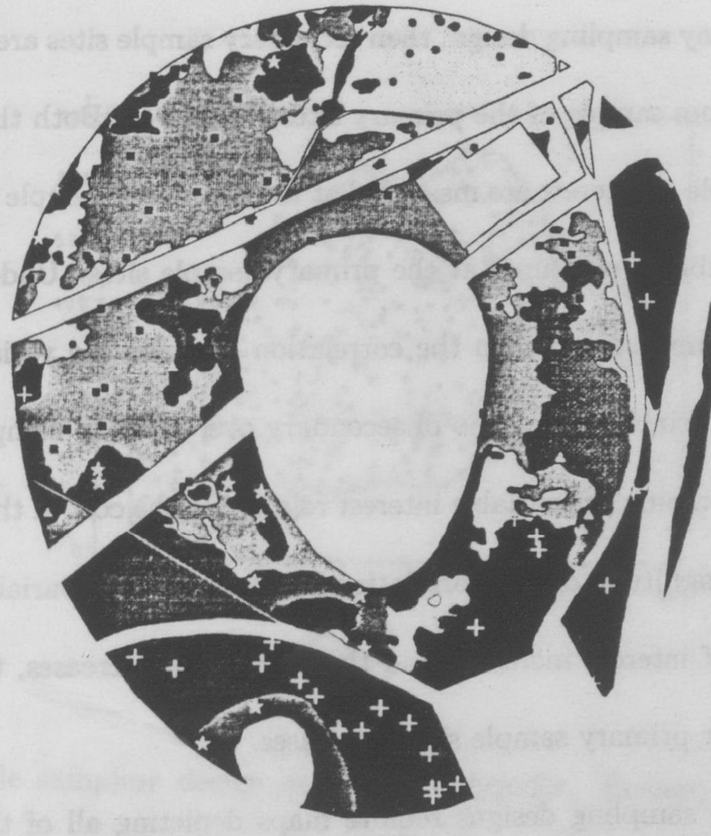


Figure 3: Stratified random sampling design in the wetlands surrounding a Carolina Bay. Circles are in grasslands, squares are in briars and shrubs, triangles are in vines and small trees, stars are in hardwoods and pines, and crosses are in pines.

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

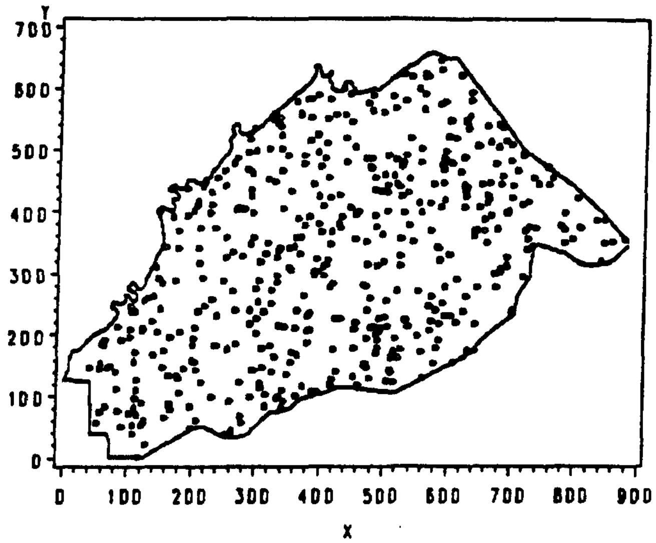


Figure 4: Double samplinr design in Ebenezer Aquifer. Primary sample sites are designated by open circles. and secondary sample sites are designated by closed circles.

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

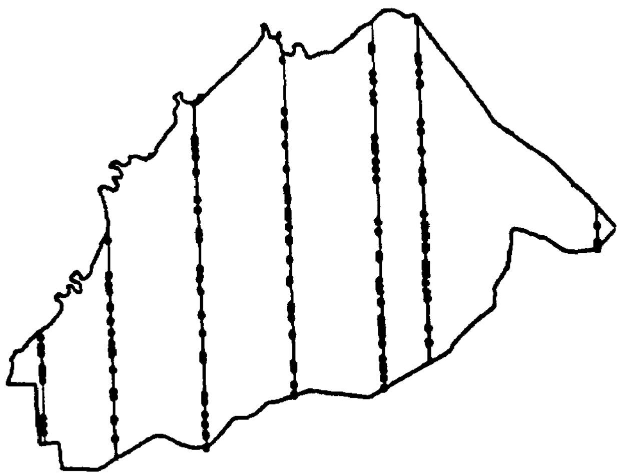


Figure 5: Line transect design in Ebenezer Aquifer.

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.

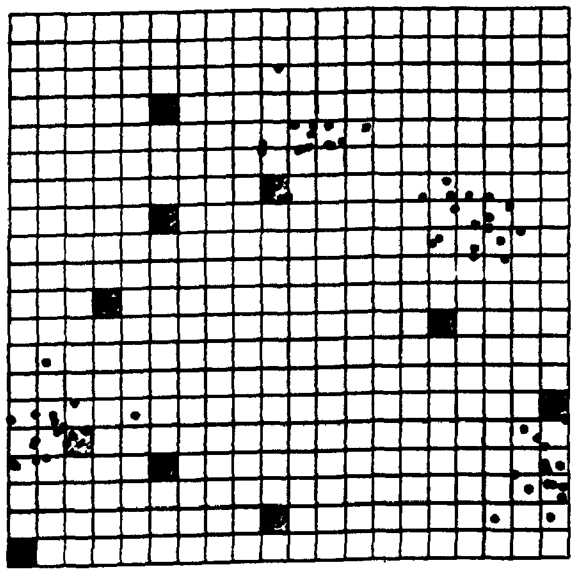


Figure 6: Adaptive cluster sampling desi㎍n. First 10 sample units are selected at random (dark shaded squares). Then adjacent unit are added to the sample whenever one or more points are observed in the selected unit (light shaded squares).

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

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

where $\hat{\alpha}_i$ and $\hat{\tau}_j$ 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

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

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

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

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

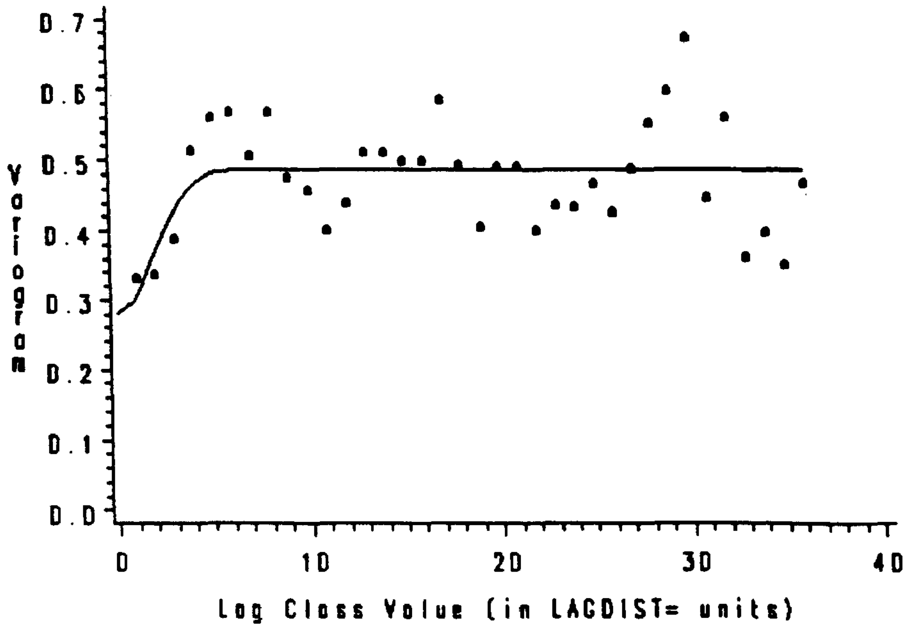


Figure 7: Temporal variogram for Clean Lakes Program data. The closed circles give the nonparametric estimates, while the curved line gives the fitted variogram model.

a range of temporal autocorrelation of $\sqrt{3/\hat{\alpha}} = 4.6$ 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

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

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

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

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

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

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 $\hat{c}_0 = 0.726$, $\hat{c}_e = 1.2937$, and $\hat{\alpha} = 0.0797$. The large nugget effect of $\hat{c}_0 = 0.726$ 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 $3/\hat{\alpha} = 37.7$ 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

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

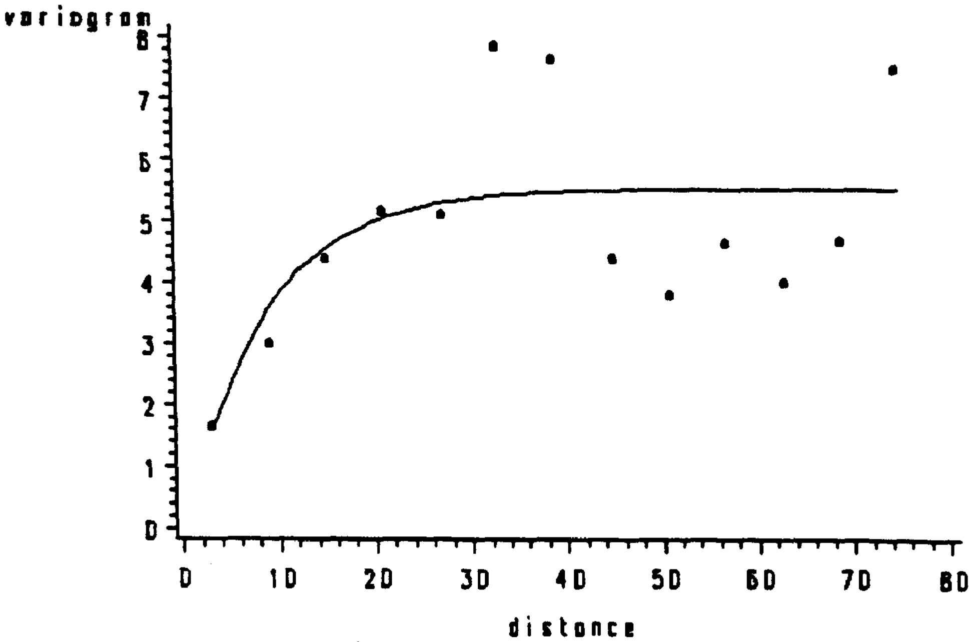


Figure 8: Spatial Variogram for REMAP data. The closed circles give the nonparametric estimates, while the curved line gives the fitted variogram model.

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

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

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

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 environmental 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 correlation, 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

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

where $\hat{\mu}_i$ 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

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

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

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

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

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

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

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

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,

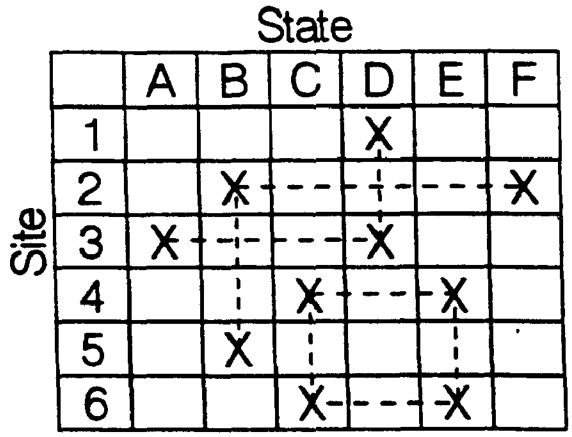


Figure 9: Connected subsets of states.

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

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

with corresponding variance estimate

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

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

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

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

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

with corresponding variance estimate

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

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

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

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

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

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

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

with corresponding variance estimate

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

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.

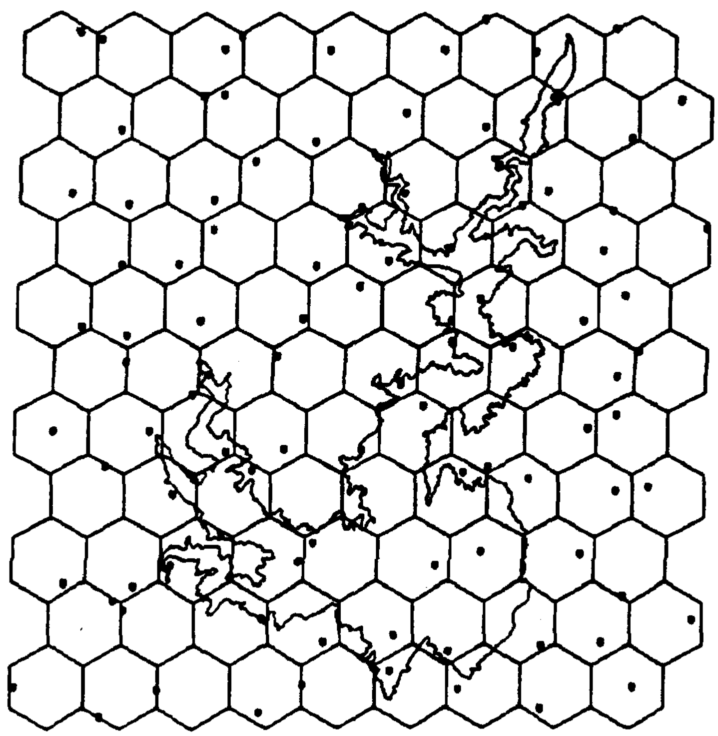


Figure 10: Randomized-tessellation stratified design for Lake Jocassee.

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 $\bar{y}$, with corresponding variance estimate

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

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

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

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

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

and

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

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

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

with corresponding variance estimate

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

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

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

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

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

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

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

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

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

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

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

and

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

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

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

and

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

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

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

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

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

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

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

from primary unit i, where M_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

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

where M 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

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

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

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

Although the total number of lakes M 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

$$\widehat{M} = \frac{N}{n} \sum_{i=1}^{n} M_i.$$

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

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

whose variance may be estimated by

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

where

$$\tilde{s}_u^2 = \frac{\sum_{i=1}^n M_i^2(\bar{y}_i - \hat{\mu}_R)^2}{n-1}.$$

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

$$\hat{p}_{II} = \frac{N}{nM}\sum_{i=1}^{n} M_i\hat{p}_i$$

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

$$\hat{p}_R = \frac{\sum_{i=1}^n M_i\hat{p}_i}{\sum_{i=1}^n M_i}$$

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

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

where

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

and

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

where

$$\tilde{s}_p^2 = \frac{\sum_{i=1}^n M_i^2(\hat{p}_i - \hat{p}_R)^2}{n-1}.$$

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

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

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

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

where

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

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

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

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

$$\widehat{\text{var}}(\hat{\tau}) = \frac{N(N-n)}{n(n-1)} \sum_{i=1}^{n} \left(\frac{M_i}{m_i} \sum_{j=1}^{m_i} |A_{ij}| \cdot d_{ij} \cdot y_{ij} - \frac{1}{n} \sum_{k=1}^{n} \frac{M_k}{m_k} \sum_{j=1}^{m_i} |A_{kj}| \cdot d_{kj} \cdot y_{kj} \right)^2 + \frac{N}{n} \sum_{i=1}^{n} \frac{M_i(M_i - m_i)}{m_i(m_i - 1)} \sum_{j=1}^{m_i} \left(|A_{ij}| \cdot d_{ij} \cdot y_{ij} - \frac{1}{m_i} \sum_{k=1}^{m_i} |A_{ik}| \cdot d_{ik} \cdot y_{ik} \right)^2$$

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 $\bar{y}$, with corresponding variance estimate

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

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

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

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 $\hat{\tau}$ may then be estimated by

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

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

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

where

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

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

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

where

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

is the sample variance of the observations from stratum h.

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

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

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

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

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

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

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

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

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

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

where

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

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

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

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

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

where

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

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

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

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

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

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:

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

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

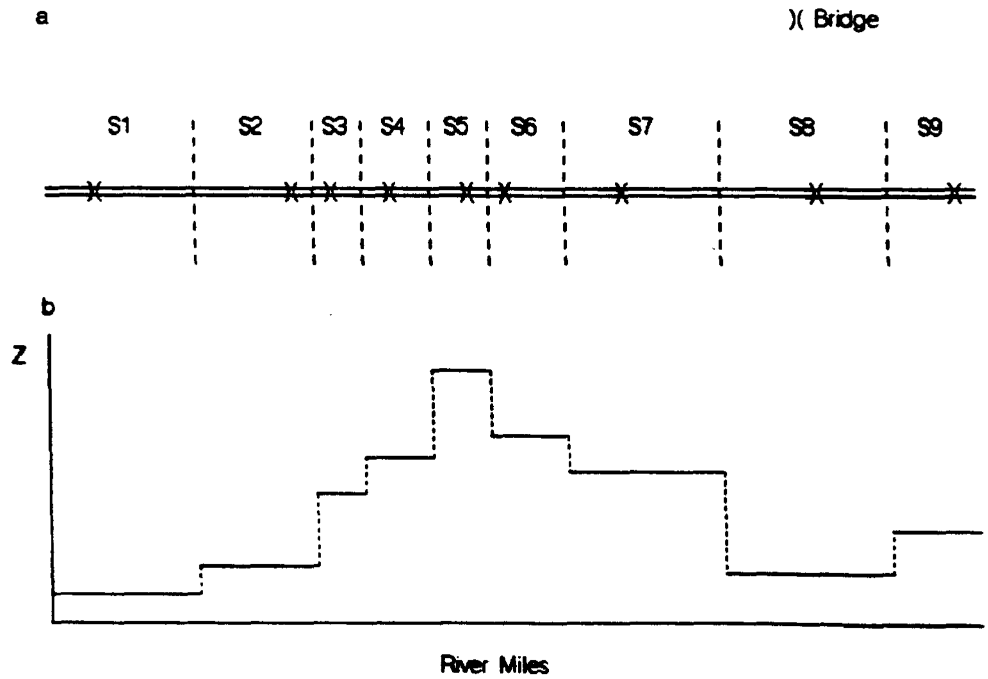


Figure 11: Sampling Bridges. (a) The locations of nine bridges along the length of a river. The river is partitioned into segments S1 to S9 as shown. A bridge will be sampled if a random point falls in that bridge's segment. (b) Assumed relationship between the variable of interest and location along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

variance estimate $\widehat{\text{var}}(\bar{y}) = s^2/n$, 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

$$\hat{\mu}_{st} = \frac{1}{L} \sum_{i=1}^{m} L_i \bar{y}_i,$$

where $\bar{y}_i$ 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

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

where s_i^2 is the sample variance of selected bridges in stratum i. Again, the estimator $\hat{\mu}_{st}$ 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 $\mathrm{var}(\widehat{\mu}_b)$ and $\mathrm{var}(\widehat{\mu}_a)$ 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

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

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 approach 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

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

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

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

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

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 $\hat{\varepsilon}(\mathbf{s}_i, t)$ 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 $\hat{\boldsymbol{\theta}}$ that minimizes

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

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

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

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 $t = M < T$, a probability-based sampling design is implemented, selecting sites $\mathbf{u}_1, \cdots, \mathbf{u}_m$. Data are then collected at times $t = M, M + 1, \cdots$, so that the probability sample data are $\{Z(\mathbf{u}_i, t) : i = 1, \cdots, m;\ t = M, \cdots, T\}$. 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 $\{Z(\mathbf{u}_i, t) : i = 1, \cdots, m; t = 1, \cdots, M-1\}$. 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 $nT + m(T - M + 1) + p + 1$ 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 M to predict the unobserved values of the data from the probability sample at time t. Then the universal kriging predictor is

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

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

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

$$\sum_{i=1}^{n}\lambda_{1i}+\sum_{i=1}^{m}\lambda_{2i}=1,$$

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

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

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

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

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

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 $a_{st} = 0$,

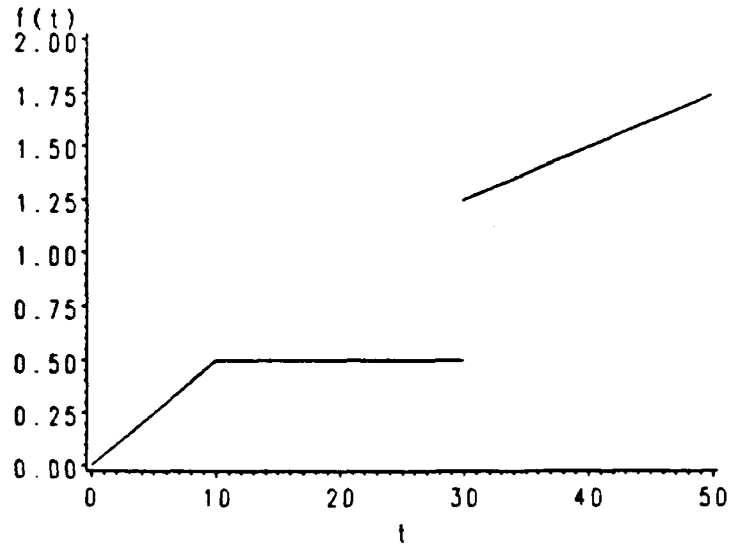


Figure 12: Temporal Trend.

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).

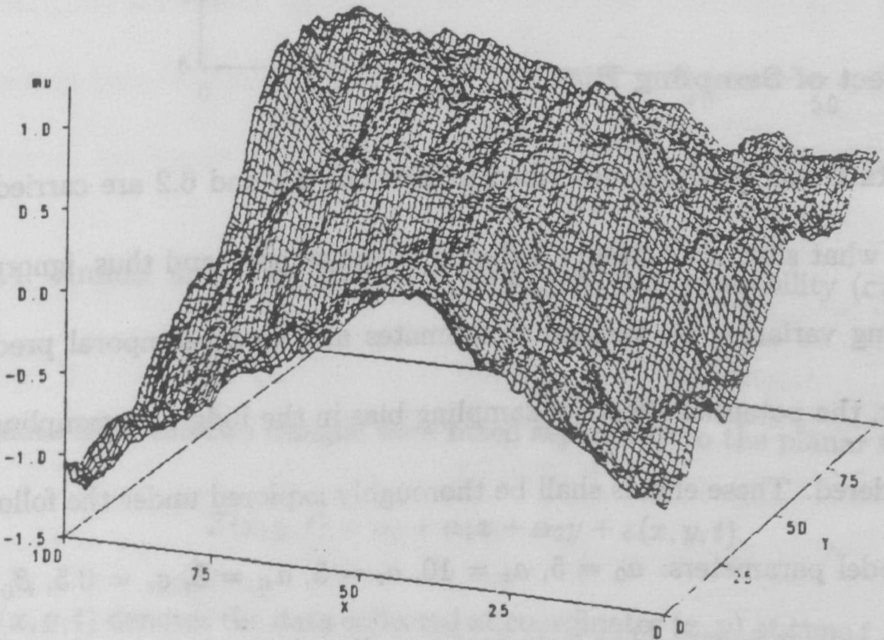


Figure 13: Spatial Trend.

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: $a_0 = 5$, $a_s = 10$, $a_t = 3$, $a_{st} = 3$, $a_\epsilon = 0.5$, $\beta_0 = -1$, 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.

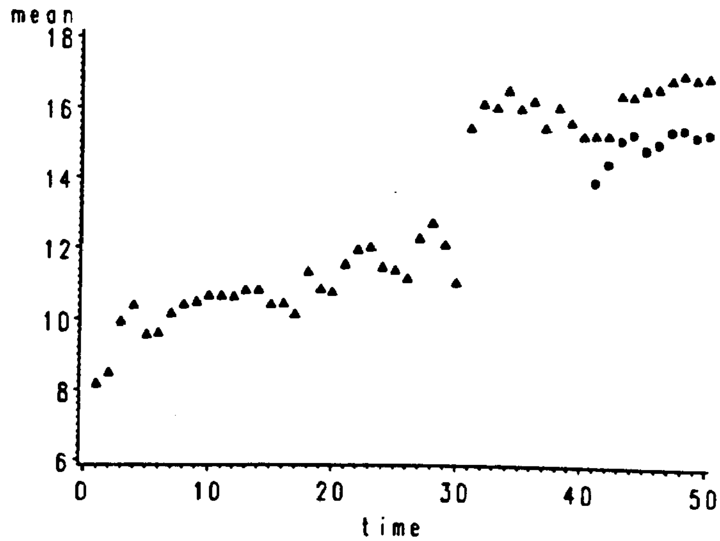


Figure 14: Annual means for judgment (triangles) and probability (circles) sample sites.

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 $2\hat{\gamma}_s(h)$ 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, $2\hat{\gamma}_s(h)$ 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 $2\hat{\gamma}_s(h)$ 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; $\hat{\kappa}_s = 31.7$ km for the probability-based sites, and $\hat{\kappa}_s = 30.6$ km for the judgment sites. However, the judgment sites show a higher variance ($\hat{\sigma}^2 = 12.02$) than the probability sites ($\hat{\sigma}^2 = 10.45$).

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 $2\hat{\gamma}_t(r)$ increases to an asymptote with increasing time lag r (Figure 16). Weighted least squares estimation was used to fit the exponential variogram model

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

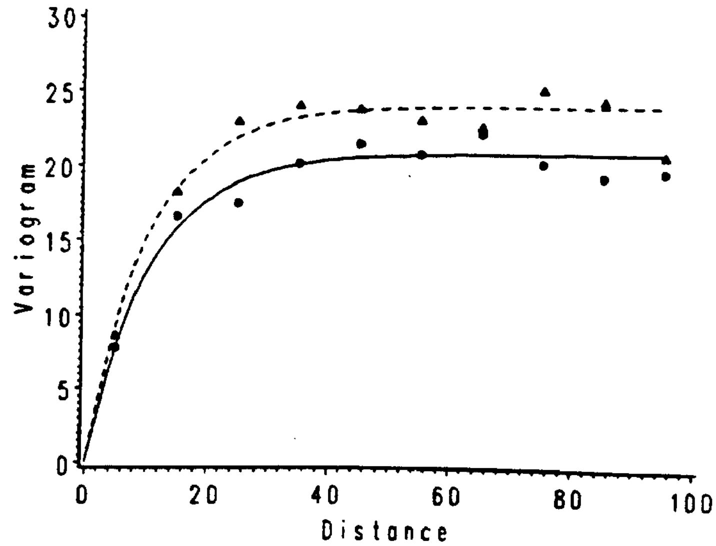


Figure 15: Fitted spatial variogram models for probability-based (solid line fit to the circles), and judgment (dashed line fit to the triangles) sample sites.

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

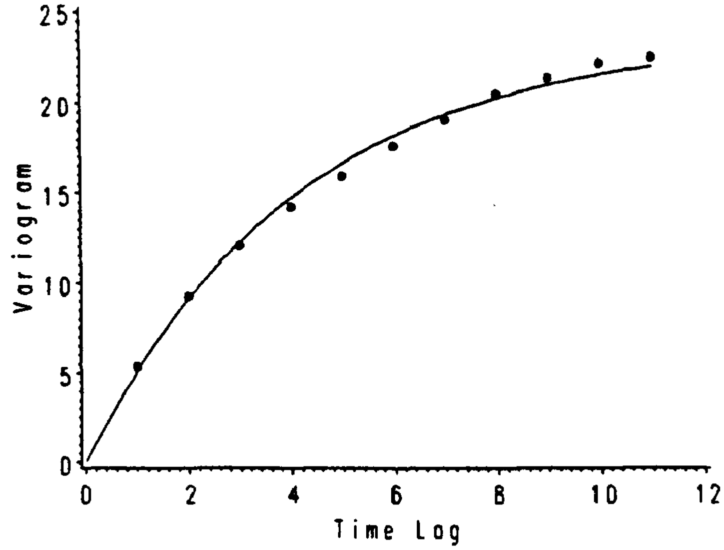


Figure 16: Fitted temporal variogram model.

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.

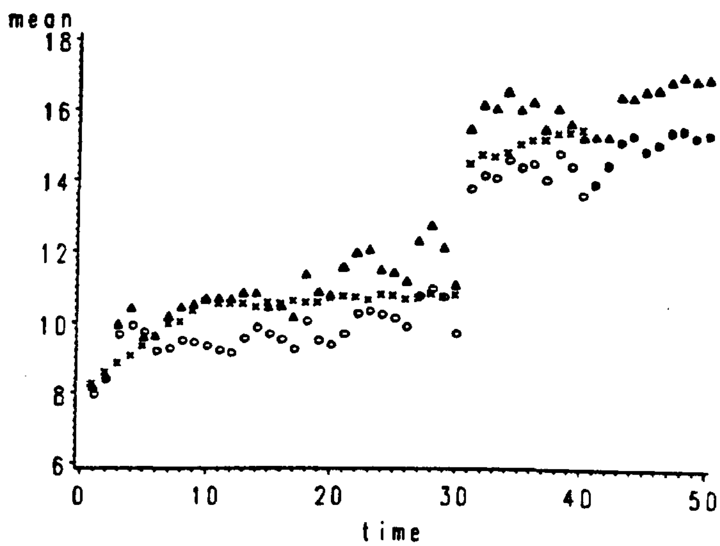


Figure 17: Comparison of predicted (x's) and unobserved (open circles) sample means for probability sites in years 1 to 40. In addition, observed annual means for judgment (triangles) and probability (open circles) sample sites are given.

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

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

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 $\hat{Z}_j(\mathbf{u}_k, t)$ is given by

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

The subscript j is included in $\hat{Z}_j(\mathbf{u}_k, t)$ 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

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

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

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

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.

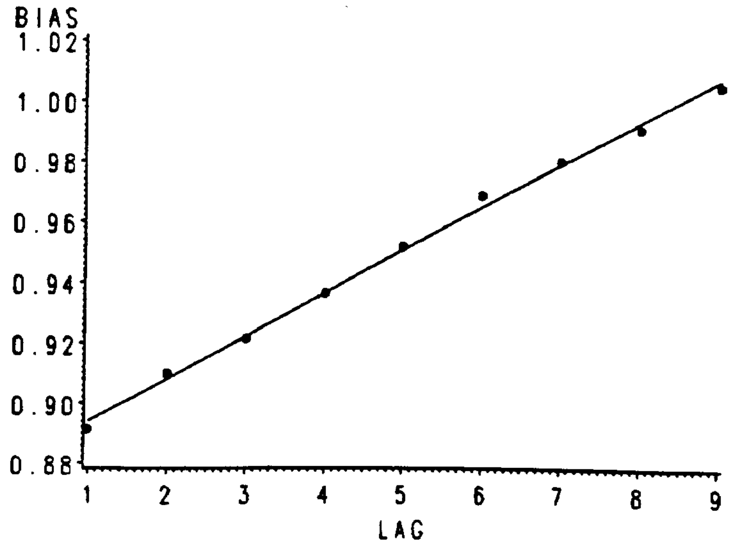


Figure 18: Adjusted mean bias plotted against time lag.

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

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

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

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

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

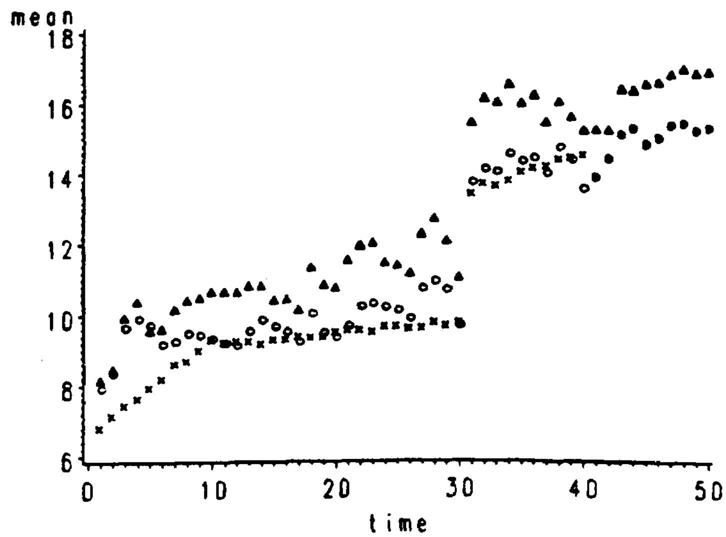


Figure 19: Comparison of bias corrected predicted (x's) and unobserved (open circles) sample means for probability sites in years 1 to 40. In addition, observed annual means for judgment (triangles) and probability (closed circles) sample sites are given.

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.

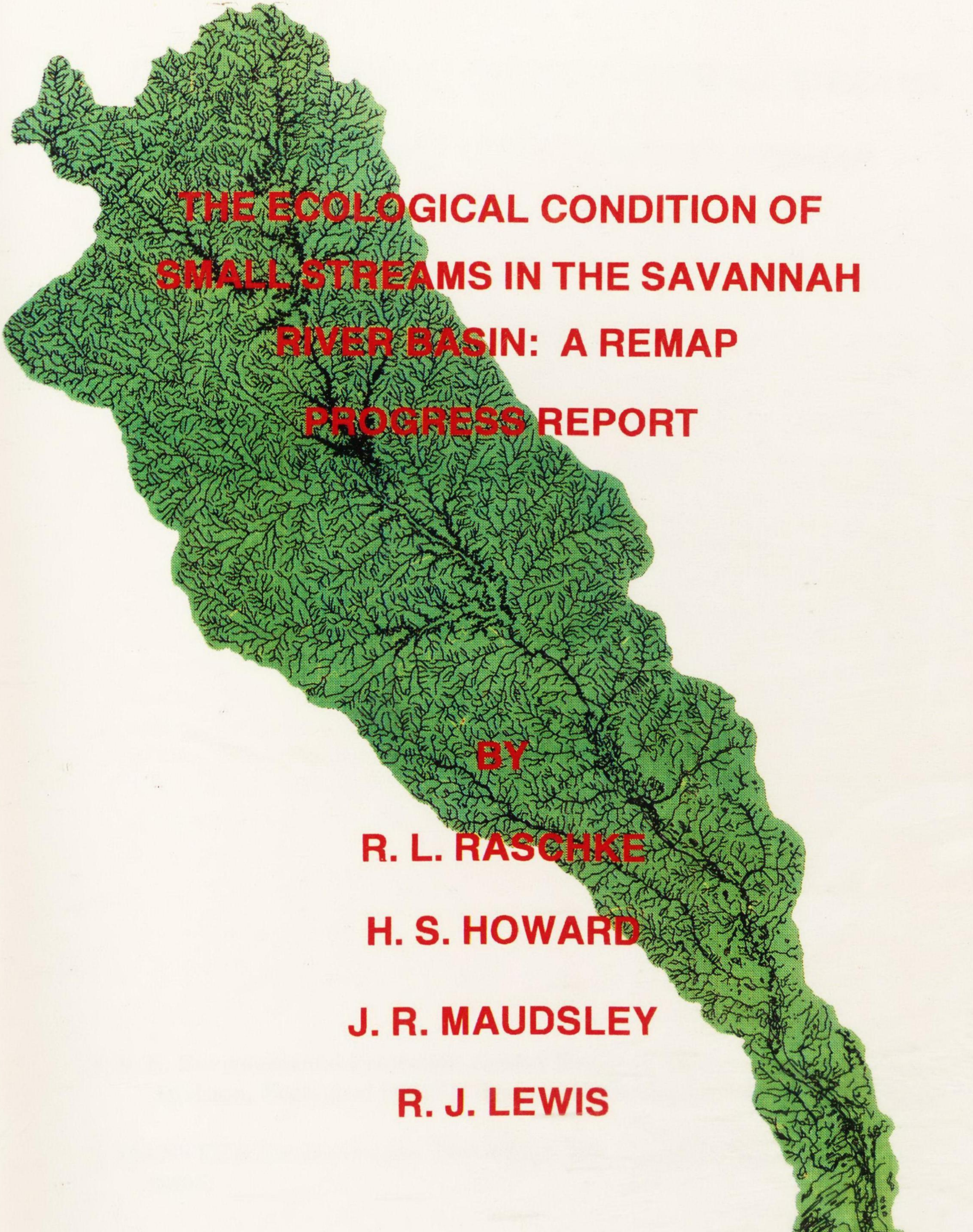
Thompson, S.K. 1992. *Sampling.* Wiley, New York.

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.

H

Appendix H

The Ecological Condition of Small Streams in the Savannah Basin: A REMAP Progress Report
By
Raschke et al.



THE ECOLOGICAL CONDITION OF SMALL STREAMS IN THE SAVANNAH RIVER BASIN: A REMAP PROGRESS REPORT

BY

R. L. RASCHKE

H. S. HOWARD

J. R. MAUDSLEY

R. J. LEWIS

THE ECOLOGICAL CONDITION OF SMALL STREAMS IN THE SAVANNAH RIVER BASIN: A REMAP PROGRESS REPORT

BY

[1]R. L. RASCHKE
[1]H. S. HOWARD
[2]J. R. MAUDSLEY
[2]R. J. LEWIS

1. U. S. Environmental Protection Agency Region 4, Environmental Services Division, Ecological Support Branch, 625 Bailey St., Athens, GA 30605

2. MANTECH Environmental Technology, Inc., Research Road, Athens, GA 30605.

SUMMARY

In response to the needs of the states of Georgia and South Carolina and their policy-relevant questions, the Ecological Support Branch of EPA Region 4 provided a monitoring strategy to help them effectively and efficiently sample Savannah River Basin waters. During the first two summer seasons (1994 & 1995) of a four-year cycle, 64 sites on wadeable streams were monitored in a systematic random manner to evaluate the status of ecological condition in the basin. By sampling fish, insects and algae, and evaluating the habitat, investigators found that water quality of most stream miles were in good condition with respect to nutrient content. However, 38% of the stream miles were affected by poor habitat, and 33% to 52% of the insect and fish communities respectively were in poor ecological condition.

Although the Branch has just begun to explore the potential use of the Geographical Information System (GIS), two areas encompassing several counties in Georgia and South Carolina seemed to have clusters of sites of poor ecological condition. Besides poor habitat, two other potential causes of poor conditions, wastewater treatment plants and animal feeding operations, may be negatively affecting the condition of insect and fish communities. Further refinement of the data analysis and eventual rechecks of sites in this area will be necessary before any permanent conclusions can be drawn. The cluster of poor sites in South Carolina at this time, are attributed only to habitat effects like, sediment erosion, deposition of sediments, and stream bank failure.

TABLE OF CONTENTS

INTRODUCTION

The southeast's population growth has brought pressure on natural resources. Responding to the challenge of adapting to more people and balancing multiple uses of natural resources, the Environmental Protection Agency (EPA) instituted the Watershed Protection Approach (WPA). WPA is a program for identifying and preventing environmental problems, setting priorities, and developing solutions through an open, inclusive process with the people (stakeholders) who live in a geographical setting. Consideration of economic prosperity and environmental well-being is the corner stone of WPA.

The Savannah River Basin was selected for environmental protection because of high population growth, known environmental problems, its susceptibility for further degradation, and the likelihood of successfully enhancing quality of life in the basin. Through the WPA program, EPA - Region 4 brought together stakeholders of varying interests who developed a comprehensive strategy known as the Savannah River Basin Watershed Project. Part of that strategy included a monitoring component, the Regional Environmental Monitoring and Assessment Program (REMAP).

REMAP represents a fundamental change in environmental-appraisal. It produces representative measurements of overall status and trends of environmental condition. Its goal is to measure cumulative effects with a known degree of certainty, provide decision makers with sound ecological data, and measure the effectiveness of environmental protection efforts.

The Environmental Services Division (ESD) of EPA Region 4 was asked by the Savannah River Watershed Project Policy Committee to implement the REMAP strategy as a demonstration project for the states of South Carolina and Georgia. These states were interested in reducing sampling and analyses, having the ability to reduce or increase sampling density, responding quickly to emerging environmental problems, and maintaining representative coverage of environmental resources through systematic yet random means of sampling.

Before the monitoring study, a set of questions was posed by the states of Georgia and South Carolina to provide direction for the monitoring design. The following policy-relevant questions were identified to guide the development of a plan of study and subsequent monitoring efforts.

- What is the status of condition of the water resources of the Savannah River Basin?
- What proportion of the Savannah River Basin surface waters are attaining designated uses?
- What are the changes of ecological condition over time?
- What factors might be associated with changes?
- Is there a tendency for distribution of condition in a specific direction (spatial gradient) over the basin landscape? What are the possible reasons for these gradients?
- What resources are at risk in the Savannah River Basin?

In response to the needs of the states and the policy-relevant questions posed, the Ecological Support Branch developed the following study objectives with the concurrence of the Policy Committee of the Savannah River Watershed Project.

- Estimate the status and change of the condition of water resources in the Savannah River Basin;
- Identify water quality spatial gradients that exist within the Savannah River Basin and

associate current and changing condition with factors that may be contributing to this condition and spatial gradients;

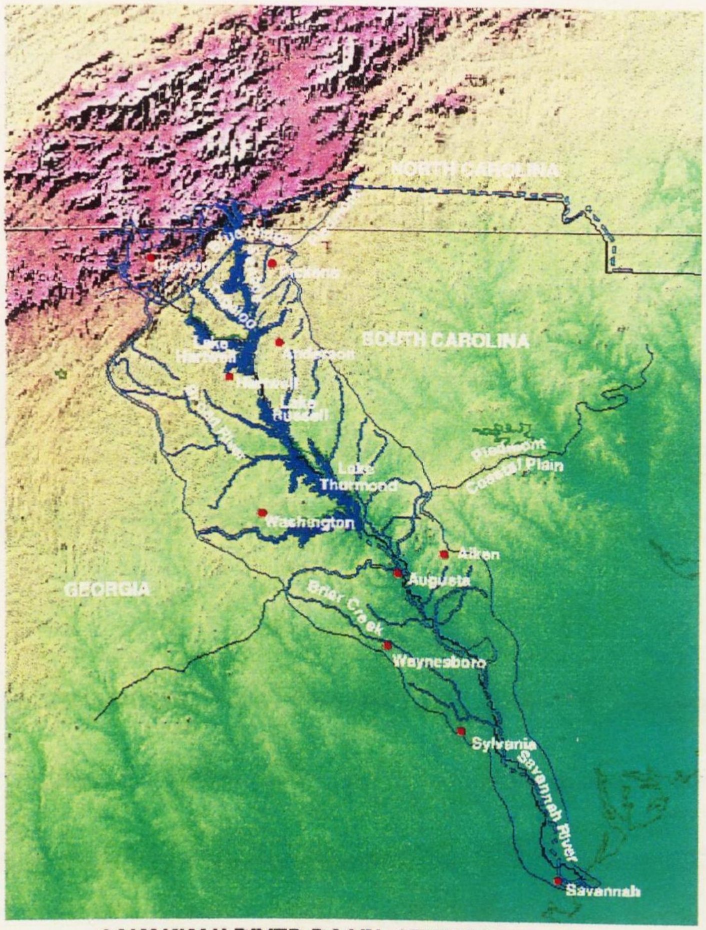
- Demonstrate the utility of the REMAP approach for watershed and river basin monitoring and its applicability for state monitoring programs;
- Incorporate the REMAP approach in the formulation and accomplishment of River Basin Management Plans; and
- Provide baseline information required to conduct comparative risk assessments in the Savannah River Basin.

Beneficiaries of the study are basin resource managers in state and federal agencies and local governments. This information will help identify stressed areas and suggest strategies for addressing high priority problems. In addition, the study will provide baseline information for comparing trends in condition throughout the Savannah River Basin and assessing the effectiveness of cumulative management efforts on protecting and managing these ecological resources.

SAVANNAH RIVER BASIN

The Seneca and Tugaloo Rivers begin on the slopes of the Blue Ridge Mountains in North Carolina. These two rivers join forming the Savannah River near Hartwell, Georgia and Anderson, South Carolina. The river flows in a southerly direction forming the boundary between South Carolina and Georgia. Eventually, the Savannah River empties into the Atlantic Ocean at the port city of Savannah (Figure 1).

Within the basin's 10,579 square miles, there are17,354 stream miles. One thousand five hundred and three of those stream miles or 5.4% are wadeable (first through third order) stream miles. The basin consists of three different land forms or physiographic provinces: the Blue Ridge, Piedmont, and Coastal Plain. The Blue Ridge is characterized by mountains covered naturally with Appalachian oak. Forests and ungrazed woodlands are the predominant land uses with some cropland and pastures. The Piedmont is characterized by gently sloping hills and smooth to irregular plains. This province is underlain naturally with nutrient poor soils supporting oak/hickory/pine and southern mixed forests. Land use is a mixture of croplands and pasture, woodlands with some urban areas. Flat plains dominated naturally by oak/hickory/pine forests, pocasin (pine, holly) forests, southern floodplain forests (oak/tupelo, bald cypress), and southern mixed forests (beech, sweetgum, magnolia, pine and oak) are characteristic of the Coastal Plain.



SAVANNAH RIVER BASIN, STUDY AREA MAP

Figure 1. Savannah River Basin with physiographic provinces.

Within the three physiographic provinces there exists distinct ecosystems based on the interrelationships between organisms and their environment. These distinct ecosystems are defined as ecoregions. While physiographic provinces may prove suitable for regional or national assessments, definition of ecoregions among broad physiographic areas is necessary to accurately assess ecological condition or health. Ecoregions are distinct areas grouped by climate, soils, land forms, and vegetative cover. The Blue Ridge physiographic province stands alone as a separate ecoregion as does the Piedmont physiographic province. However, the Coastal Plains physiographic province is composed of three distinct ecoregions: the Fall Line Hills (or Sand Hills), the Southeastern Plains and Hills, and the Coastal Plains.

MONITORING DESIGN

Objectives of the monitoring design provide information about the ecological situation and eventually trends in condition of the natural resources. One resource of interest is all perennial wadeable streams. The design strategy selects wadeable stream sampling points that provide valid estimates of general basin-wide stream condition.

One can study conditions of streams in two ways. The first is by consensus, which entails examining every point on the streams. This method is impracticable. A more practicable approach is to examine some points systematically to ensure adequate coverage of the basin, and randomly to prevent bias in selection of stream points. For example, we would not obtain a good estimate of the percent of all students in a region with hepatitis if we polled only students in small towns of less than two thousand people. This preferential or biased sample would most likely include a much lower proportion of students with hepatitis than the general population of students. Similarly in a stream study preferential sampling occurs if the sample includes only sites, for example, downstream of sewage outfalls where sewage outfalls affect only a small percentage of total stream length. This kind of sampling program may provide useful information about conditions downstream of sewage outfalls, but it will not produce estimates that accurately represent conditions of the whole basin. Preferential selection can be avoided by collecting random samples.

Randomization can be thought of as a kind of lottery drawing to determine which points are included in the sample. Randomization is important. When used, it is possible to estimate condition of streams with a known degree of confidence. In REMAP, hexagons are used to add the systematic element to the design. The hexagonal grid is positioned randomly over the basin map, and sampling points from within each hexagon are selected randomly. The grid ensures spatial separation of selected sampling points.

This design's sampling requirements reduce sampling locations to a logistically and economically feasible number. It allows fewer sites to be sampled annually, but provides for sampling of all randomly selected sites over a rotating year period. Currently, this rotation period for the Savannah REMAP project is four years. However, the period can range from two to five years. This report is an assessment of wadeable stream condition based on a sampling of 64 sites after the second year of a four-year cycle. It is expected that an additional 72 sites will be sampled for a total of 136 sites by the end of the four-year cycle.

INDICATORS

REMAP monitors ecological indicators to assess condition and trends. Indicators are defined primarily as any characteristic that estimates the condition of natural resources. The challenge is to decide which indicators to monitor. One approach for selecting indicators starts with those attributes valued by society and then decide what indicators might be associated with these values. Upon consideration of the type of streams (wadeable) to be investigated and after extensive discussions, an initial set of societal values and concerns were identified. They are: **Biological Integrity and Trophic Condition** (Table 1).

Table 1. Values, Indicators, and Measures Used to Evaluate Wadeable Stream Conditions in the Savannah River Basin.

VALUE	INDICATOR	MEASURES
Biological Integrity	**Stream Insect RBP[1] II**	**Habitat Score** **EPT[4] Index**
	Fish RBP[2] V	**Fish IBI[5]**
Trophic Condition	**AGPT[3]**	**Average Maximum Yield in Dry Weight of Algal Cells per Liter**

1. Stream insect Rapid Bioassessment Protocol based on identification of insects to family level.
2. Fish Rapid Bioassessment Protocol based on identification of different kinds of fish.
3. Algal Growth Potential Test; a measure of nutrient enrichment.
4. An index based on the identification of pollution sensitive insects known as stoneflies, mayflies, and caddisflies.
5. An Index of Biological Integrity based on twelve different characteristics of fish.

Biological integrity incorporates the idea that all is well in the community. That is, the different groups are stable and working well with little if any external management of the community, whether it is a township, coral reef, or stream.

Trophic condition is a measure of water quality condition based on different levels of available nutrients. When nutrients are in excess, overabundance of algae and larger green plants results in nuisance conditions. Millions of dollars are spent annually to control the growth of algae and other plants. Overabundant growth of plants can affect biological integrity but also human uses like fishing, boating, swimming etc.

The challenge is to identify ecological indicators that can be related directly to those societal values held by the public. There are two general types of ecological indicators; condition and stressor. A condition indicator is any characteristic of the environment that estimates the condition of natural resources and is conceptually tied to a societal value. Stressor indicators are suspected to elicit a change in the condition of the natural resource. The indicators selected to address biological integrity are stream insects and fish assemblages.

Insects represent the first consumer level in streams. They are important as processors of organic matter, like leaves and sewage, that find their way into a stream. By fragmenting or breaking down this organic matter, stream insects prepare it for decomposition by bacteria that attach or colonize the organic matter. In turn, bacteria may serve as a food source for other stream insects that seek out and graze on the organic matter. Because of their limited mobility and relatively long life span, stream insects provide a "window" of cumulative impacts on ecological or resource condition. This community is sensitive to changes; they have for many years been used as a reliable barometer of water quality conditions. Some groups of insects are very sensitive to stresses, like manmade pollution while others are tolerant. By focusing on the presence or absence of different groups of insects, an aquatic biologist is provided insight about the ecological health of a stream. Sometimes pollution effects may stem from discharges of chemicals, pesticides, or nutrients that are of a manmade origin. Often, sediments from erosion and attributable to land clearing or silvaculture practices may adversely affect the stream habitat. The materials that constitute a stream bottom are very important to both fish and stream insects. For example, very fine sediments, like silt, clay, or very fine sand, are detrimental to the reproduction of fish and eliminate preferable habitat for stream insects. Silt, especially, can interfere with a fish's or stream insect's ability to breathe. Assessment of the insect community was accomplished by using a standard field survey technique known as Rapid Bioassessment Protocol II(RBPII). With the RBP II protocol, most sites can be surveyed with relative limited time and effort in the field and laboratory.

Habitat is an important consideration when evaluating aquatic systems. To examine the quality of the habitat, habitat evaluations are conducted at each stream station. These evaluations focus on parameters such as substrate (bottom sediments) characteristics, flow regimes, impacts to the stream channel (eg., channelization, deposition), impacts to streamside vegetation, stability of the stream banks, and available cover. Ecoregional reference sites provide a basis for the best attainable conditions for all streams with similar physical dimensions for a given ecoregion. Presently, there are two reference sites per ecoregion except for the coastal plain ecoregion. The process of reference site identification is still ongoing.

Fish were chosen primarily for their societal value and role as a top consumer in streams. Fish are relatively easy to identify and with minimal training most fish can be collected, sorted, and identified at the field site and then released unharmed. They are an important part of the food web. Fish are found in the smallest of streams and some are even found in heavily polluted streams. They occupy positions throughout the food web and include groups that represent a variety of feeding types. Their diet often consists of food derived from both inside the stream and outside the stream. Fish serve as one of the major predators of stream insects. Changes in the stream insect community often result in a change in the fish community. Like stream insect communities, fish communities will respond to environmental change, whether it is chemical or physical. Some fishes are very sensitive to environmental change while others are not. By examining all fish groups that live in a stream, the general condition of a stream can be assessed. For example, if there are only one or two groups of fish in a stream who are very tolerant to pollution, and there are no groups that are sensitive to environmental change, then impairment is suspected because of environmental change that has eliminated the sensitive groups.

The Environmental Protection Agency's Rapid Bioassessment Protocol V (RBP V) is an index used to assess stream condition based on the fish community. The index consists of twelve measures scored to assess changes in the fish community compared to a reference stream, or a stream with least impact. For example, one of the measures assesses the proportion of fishes in a stream considered to be tolerant to environmental change. If the proportion of tolerant groups are high compared to the reference stream, then this would result in a lower score for that measure. Another measure looks at the number of fish groups. If the number of fish groups collected is similar to that of the reference stream then this would result in a high score. After all twelve measures have been given a score, the scores are totaled and the condition of the fish community is then characterized as either good, fair, or poor depending on how far the total score deviates from that of a reference stream.

The primary indicator selected to address trophic condition in streams is the algal growth potential test (AGPT). The AGPT is based on the premise that maximum yield of plants (e.g. algae) is limited by the amount of nutrients available to the test alga. With higher algal growth concentrations (AGPT), there is good likelihood that obnoxious plant growths can occur in a stream.

The test was selected as the indicator of choice to assess trophic condition primarily because of its specific sensitivity, reliability and the ease and economy of using it as a monitoring tool.

MONITORING ASSESSMENT

One of the objectives of this study is to estimate ecological condition in the Savannah River Basin. The task is to establish action levels (index score or concentration). These levels are then used to decide if a stream segment is in good, fair, or poor ecological condition related to a particular societal value or issue of concern. Approaches used to establish action levels included assessment of reference sites, data analysis, and field experience. Development of action levels for indicators used in this study provides the opportunity to estimate miles of wadeable streams in poor, fair, or good condition. Conforming to the adage "that a picture is worth a thousand words," estimates of the percent of wadeable stream miles in a certain condition are made easier to understand by the cumulative distribution curve (CDF). These curves show the percent of wadeable stream miles equal to or less than some specified concentration or index number plus or minus a confidence level. For the purposes of this study, we have set a confidence level of 95%. This means that we are 95% sure that the present stream miles estimated to be equal or less than a given index score or concentration is within the bounds of our confidence lines on the graph (Figure 2). There is only 1 in 20 chances (5% error) that the true or real percent of stream miles affected at a particular concentration or a score is not within the confidence bounds.

Habitat evaluation identifies degradation at a given site that may be of a physical nature (i.e., stream erosion) rather than water quality. A good portion of the wadeable stream miles (70%) had degraded habitat ranging from fair to poor (Figure 2). Based on on-site habitat evaluations, most of the degradation could be attributed to non-point source sediment erosion, deposition of sediments, and stream bank failure due to loss of bankside vegetation. This finding agrees with other reports from state resource agencies in the southeastern United States that have identified non-point source pollution, and especially sedimentation as a major concern.

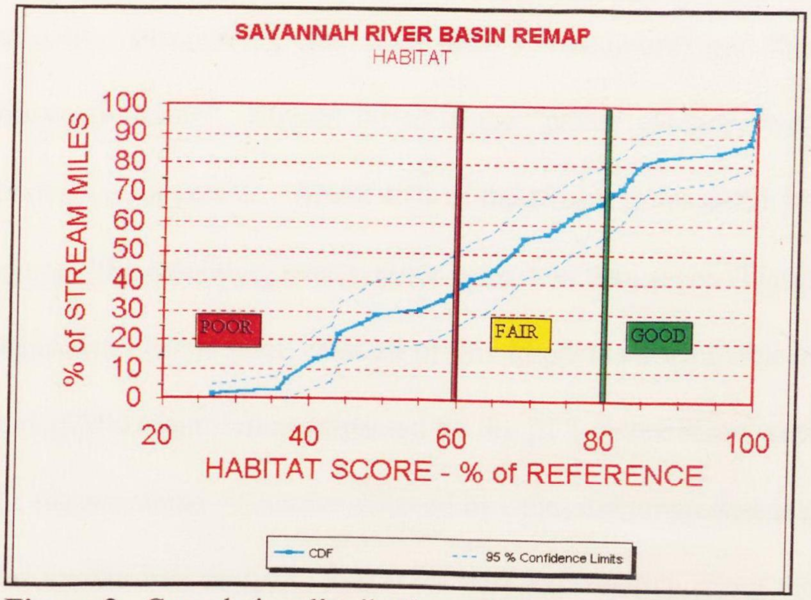


Figure 2. Cumulative distribution of habitat score.

The stream insect EPT Index, through the course of the Savannah River REMAP, has emerged as a valuable indicator for interpreting stream biological integrity. This indicator is simply a summation of the total number of pollution-sensitive stream insects. These pollution-sensitive stream insects are

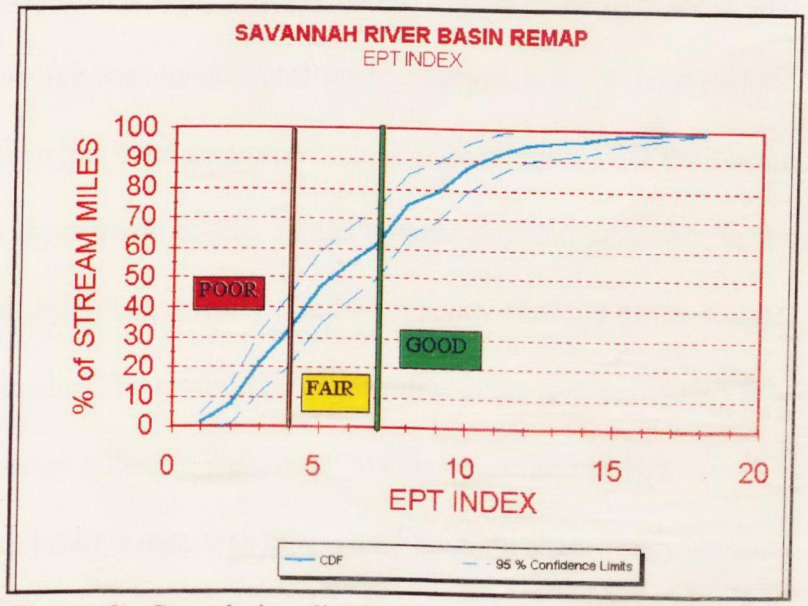


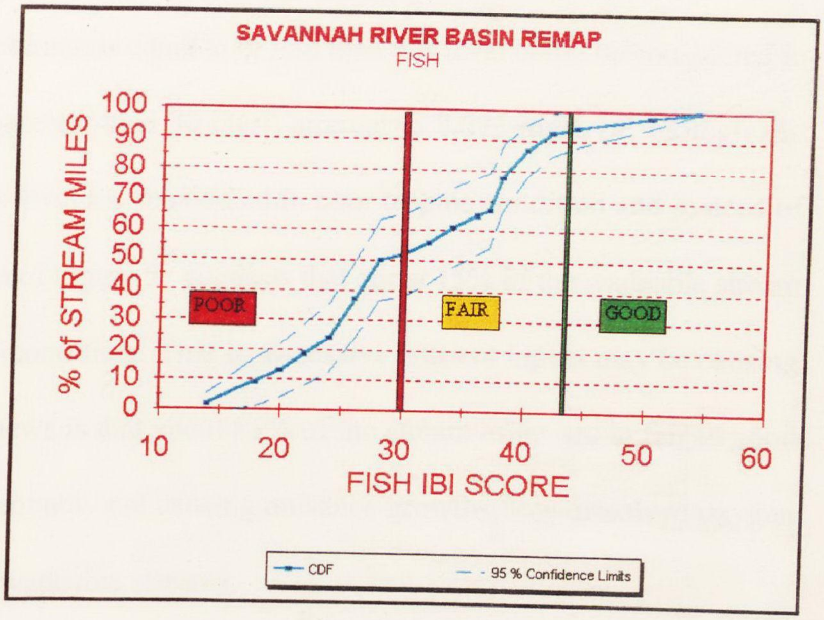
Figure 3. Cumulative distribution of EPT index.

the mayflies (Ephemeroptera), stoneflies (Plecoptera), and caddisflies (Trichoptera) and their presence or absence suggests good water quality. Stream miles in the "good" category were characterized by an EPT Index score of seven or greater. About 40% of the stream miles exhibited EPT index scores in the "good" category. The remaining stream miles were less than seven (Figure 3). That is, biological integrity ranged from fair to poor. Not all of this impact is attributable to habitat quality. Over half of the sites (53%) identified as impacted by the EPT Index score were limited by fair to poor habitat quality, the remaining 47% were affected by other unknown stresses.

Of the twelve measures used for the fish analysis, four were very informative about the condition of the fish community. The four measures were; (1) the proportion of omnivores (fish that eats both plants and animals), (2) the total number of darters (small fish that inhabit the bottom and primarily feed on stream insects), (3) the total number of suckers (a group of small to relatively large fish that inhabit the bottom and feed on stream insects), and (4) the total number of different kinds of fish collected. The proportion of omnivores reflect change in the food web of the stream. If the food web is disrupted in some way, fish adapted to eat anything will dominate over those who are more specialized feeders. Darters and suckers are sensitive to changes in the habitat and the stream insect community. If the habitat has been disrupted by increased siltation or if the stream insect community is damaged, then these kinds of fish would become reduced or absent in the stream. The total number of different fish found reflects all environmental changes to the stream. The results of analyzing these four measures alone would be very similar to the results of analyzing all twelve measures. However, all twelve measures were used to determine the fish IBI.

The RBP V fish analysis with all twelve measures have a total possible score of 60. Streams with greater scores than 43 are considered in "good" condition. Streams whose scores are between

30 and 43 are considered in "fair" condition, and streams whose scores were below 30 were considered in "poor" condition. Based on the 1994-1995 data, only 7% of the fish communities in the basin were in good condition, 93% were impacted. Poor sites represent 52% of the impacted basin streams with the remaining 41% in fair condition (Figure 4).



The Ecological Support Branch (ESB) has conducted many AGPT's related to studies in the southeast. Based on experience, literature review, and data analysis ESB set an action level of 5mg/L test algal (dry weight) that would reasonably assure protection from nuisance plant growths and fish kills in southeastern lakes. Their experience of applying this test to stream waters is limited, but they have suggested AGPT levels that are conservative yet would be protective of basin streams. AGPT dry weights of equal to or less than 20mg/L converts to 0.036mg/L of phosphorus. They believe that

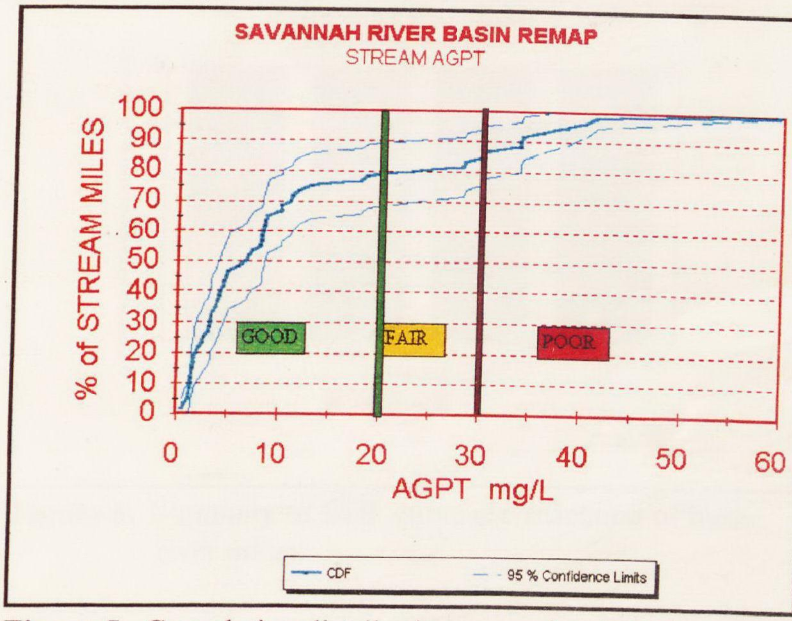


Figure 5. Cumulative distribution curve for AGPT.

streams containing phosphorus concentrations equal to or less than this level could be considered in good condition. Dry weight yields greater than 30 mg/L amount to 0.072 mg/L of biologically available phosphorus. Streams at this level are considered in poor trophic condition and in need of further investigation. An examination of Figure 5 suggests that about 15% of the wadeable stream miles in the basin are in poor trophic condition. That is, excessive nutrient inputs may be causing unsightly algal growths. The good news is that about 85% of the stream miles are in fair to good condition. Thus, nutrient levels are probably not causing nuisance growths, low dissolved oxygen problems, or fish kills in most of the wadeable streams.

Figure 6 presents a summary of stream classification (good, fair, poor) based on all ecological indicators. Based on the first two years of sampling, most of the wadeable streams in the basin were in good condition with respect to the present evaluation of enrichment as measured by the growth response of the algae.

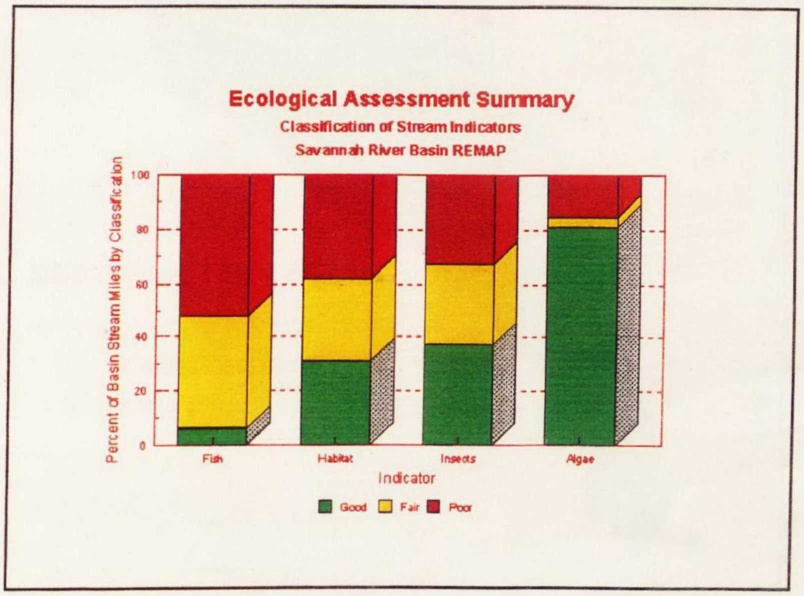


Figure 6. Summary of CDF curve classification of basin river miles.

Stream miles classified as "poor" were more prevalent among the ecological indicators that provide a measure of stream ecological condition over time (fish, insects, and habitat). For example, miles classified as poor were as follows: fish = 52%, insects = 33%, and habitat = 38%. Only 15% of the sampling sites were classified as "poor" based on the algae (AGPT). This ecological indicator is a measure of enrichment potential during the actual sampling event and is sensitive to inputs from non-point source pollution generated by storm runoff. Consequently, the percentage of miles classified as "poor" by algae would increase during wet periods.

We have just begun to explore stress indicators like landscape changes and their association with condition indicators like the EPT index by using information in the Branch's Geographical Information System (GIS). Landscape records ranging from non-point source features like coverages of industrial waste sites were examined for association with poor ecological areas. On-going data analysis continues in further refining GIS coverage and possible cause-effect relationships via association analyses. However, it was observed that many sites in "poor" ecological condition seemed to cluster in two areas of the basin, one in Georgia and one in South Carolina (Figure 7).

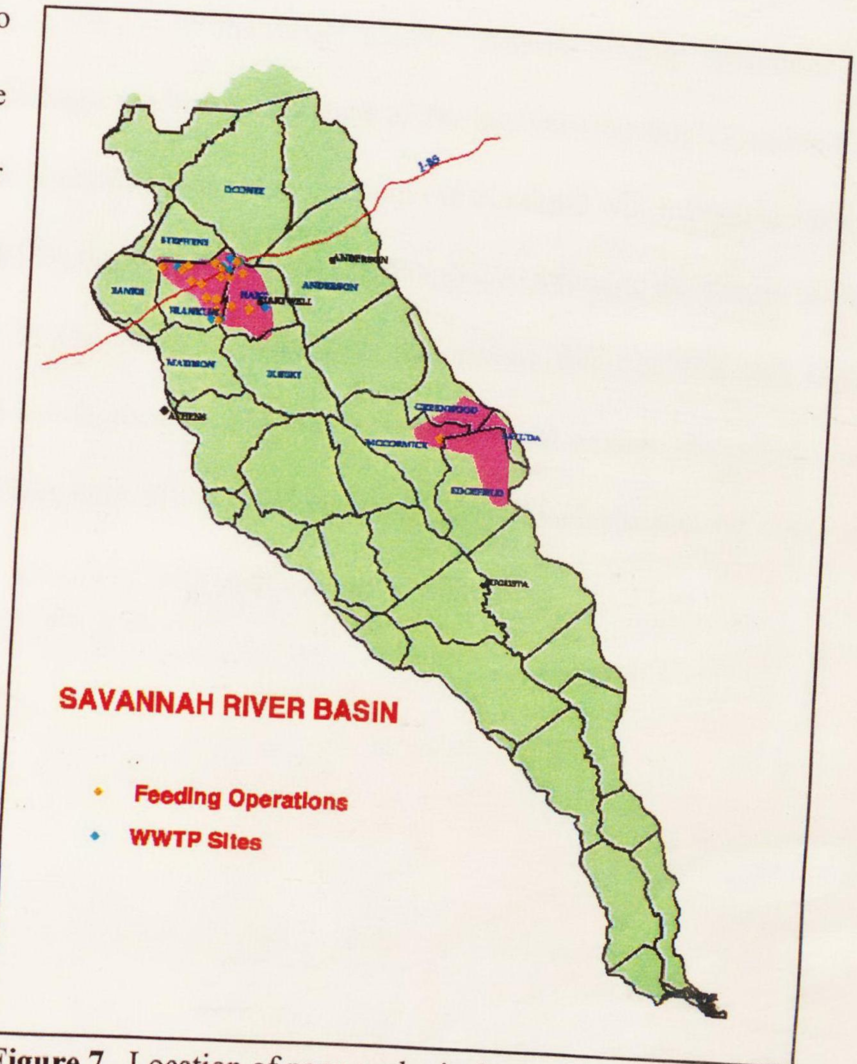


Figure 7. Location of poor ecological areas.

Degradation of habitat (Figure 6) and two landscape stressors may partially explain some causes of poor ecological condition in the Georgia area that covers six counties and ten watersheds. The stressors are wastewater treatment plants (WWTP) and animal feeding operations. Wastewater treatment plants were upstream of several sites in the Georgia area. They may be the chief cause of enrichment in Hart County as evidenced by the AGPT results. Animal feeding operations are plentiful in the northwestern Georgia piedmont. Because of the intensive feeding of animals in confined spaces, heavy organic pollution from water running off farmland will influence stream trophic condition and community biological integrity. Some thing(s) are affecting the stream insect and fish community integrity in the South Carolina area that covers five counties and eight watersheds. Besides noted habitat degradation (Figure 6) from sediment erosion, deposition, of sediments, and stream bank failure, none of the landscape intelligence available showed obvious clustering of landscape features associated with poor ecological condition.

BIBLIOGRAPHY

Anonymous. 1995. Savannah River Basin Watershed Project: Initial Assessment and Prioritization Report for the Savannah River Basin. Volume I. Prepared by the Management Committee of the Savannah River Basin Watershed Project.

FTN Associates, Ltd., R. L. Raschke, H. S. Howard. 1994. Plan of Study for Characterizing Ecological Condition in the Savannah River Basin. U. S. Environmental Protection Agency Region 4, Environmental Services Division, Ecological Support Branch, 625 Bailey St. Athens, GA 30605.

Klemm, D. J., Q. J. Stober, and J. M. Lazorchak. 1993. Fish Field and Laboratory Methods for Evaluating the Biological Integrity of Surface Waters. EPA/600/R-92/111.

Larsen, D.P., D. L. Stevens, A. R. Selle, and S. G. Paulsen. 1991. Environmental Monitoring and Assessment Program: EMAP - Surface Waters, a Northeast Lakes Pilot. Lake and Reservoir Management 7(1):1-11.

Hunsaker, C. T., and D. E. Carpenter, eds. 1990. Ecological Indicators for the Environmental Monitoring and Assessment Program. EPA 600/3-90/060.

Plafkin, J. L., M. T. Barbour, K. D. Porter, S. K. Gross, and R. M. Hughes. 1989. Rapid Bioassesment Protocols for Use in Streams and Rivers: Benthic Macroinvertebrate and Fish. EPA/444/4-89-001.

Raschke, R. L., and D. A. Schultz. 1987. The Use of the Algal Growth Potential Test for Data Assessment. Journal Water Pollution Control Federation 59(4):222-227.

Volstad, J. H., S. Weisberg, D. Heimbuch, H. Wilson, and J. Seibel. 1995. Answers to Commonly Asked Questions about REMAP Sampling Designs and Data Analyses. U. S. Environmental Protection Agency, Research Triangle Park, NC.

I

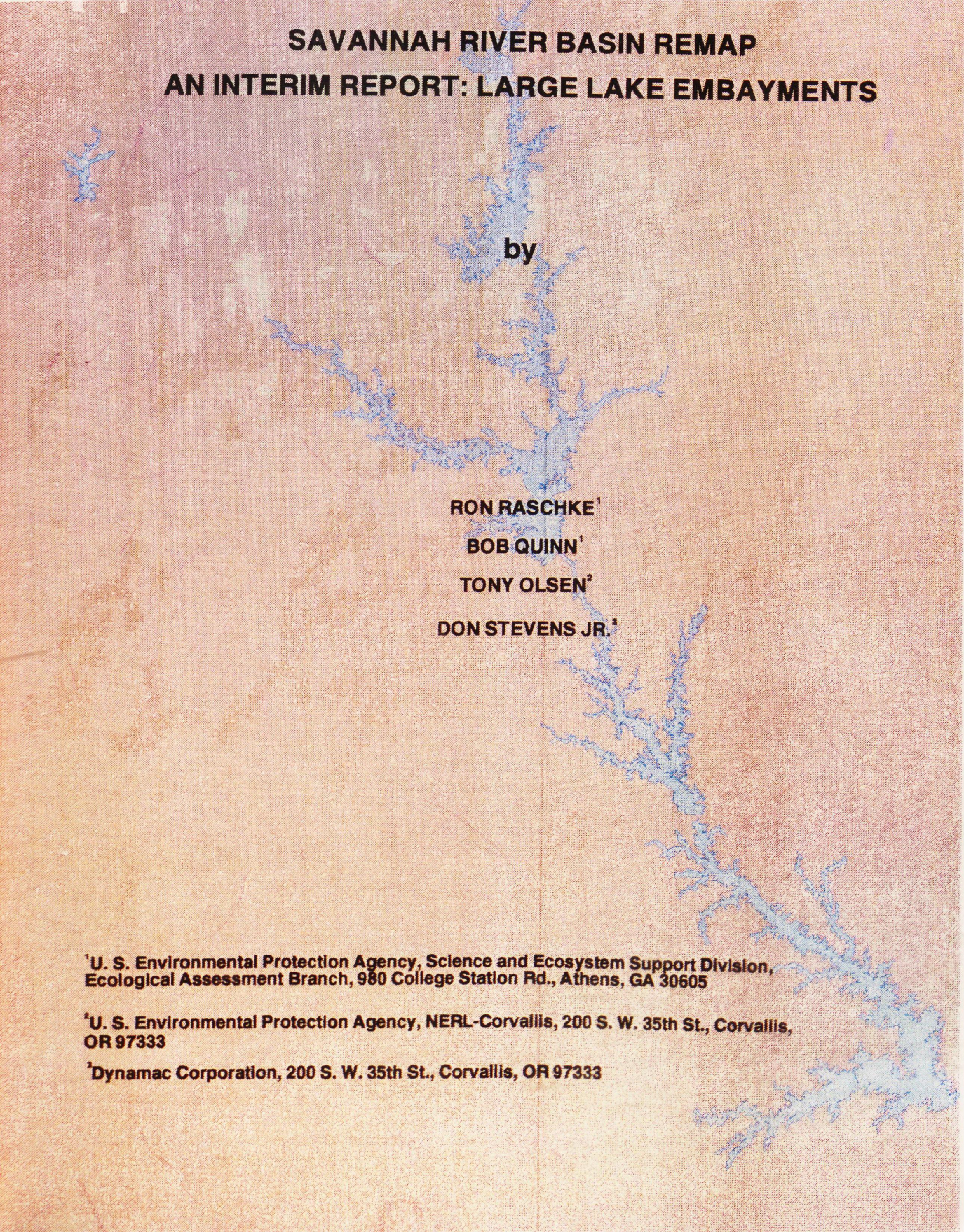
Appendix I

Savannah River Basin REMAP Interim Report: Large Lake Embayments

By

Ron Raschke et al.

SAVANNAH RIVER BASIN REMAP
AN INTERIM REPORT: LARGE LAKE EMBAYMENTS



by

RON RASCHKE[1]

BOB QUINN[1]

TONY OLSEN[2]

DON STEVENS JR.[3]

[1]U. S. Environmental Protection Agency, Science and Ecosystem Support Division, Ecological Assessment Branch, 980 College Station Rd., Athens, GA 30605

[2]U. S. Environmental Protection Agency, NERL-Corvallis, 200 S. W. 35th St., Corvallis, OR 97333

[3]Dynamac Corporation, 200 S. W. 35th St., Corvallis, OR 97333

CONVERSION FACTOR TABLE

Multiply	By	To Obtain
Meters	3.281	Feet
Hectares	2.469	Acres
Meters Square	2.47×10^{-4}	Acres

TABLE OF CONTENTS

CONTENTS **PAGE**

SUMMARY

As part of a proposed new monitoring strategy (REMAP) demonstration for the states of Georgia and South Carolina, a systematic-random sampling of embayments of major reservoirs (lakes) in the Savannah River Basin was initiated in 1995. A focus was placed on tributary embayments because they are the first portion of a reservoir to exhibit adverse nutrient impacts from man-induced changes. This report covers an assessment of trophic condition over a two-year period (1995-1996).

Trophic indicators showed that generally embayment acreage was in good condition, and it was not immediately threatened with unsightly nuisance phytoplankton blooms. Chlorophyll a was less than 12 ug/L, a concentration that meets guidelines for multiple uses, including drinking water supply. Approximately 99% of the algal growth potential standing crop was equal to or less than 5 mg/L, an action level that assures protection from fish kills and nuisance blooms. Total phosphorus ranged from 2 to 60 ug/L. Eighty-seven percent of the embayment acreage contained less than or equal to 10 ug/L of total phosphorus. With respect to water clarity, Secchi disc depth showed that only 2.6% of the acreage was less than desirable. Based on total suspended solids concentration, 0% of the waters were classified as muddy and only 3% of the acreage was classified intermediate. The low percent of acreage under less than desirable conditions were associated with near-shore stations receiving wind fetches at the time of sampling.

1.0 INTRODUCTION

1.1 PURPOSE

Responding to increased growth and demands for multiple uses of natural resources, The Environmental Protection Agency (EPA) established the Watershed Protection Approach (WPA) in 1991 (EPA, 1991; 1996). The Savannah River Basin was one of two areas selected in 1993 for the WPA in Region 4 because of its high public use, known environmental problems, susceptibility for further degradation, interest in participation by the users, and the likelihood of success. Through the WPA initiative, EPA Region 4 brought together scientists and stakeholders who developed a stategy to provide an ecological focus for resolving problems. This strategy gave birth to the Savannah River Basin Watershed Project (SRBWP) (Management Committee, 1995). The goal of the SRBWP is to develop and implement a multiagency environmental protection management project which incorportates the authorities and expertise of all interested parties in an effort to accomplish the vision of conserving, restoring, enhancing, and protecting the basins ecosystems in a way that allows the balancing of multiple uses. Further details on objectives and issues within the basin can be found in Volume I of the "SRBWP Initial Assessment and Priorization Report" by the Management Committee (1995). Part of the SRBWP strategy included a monitoring component, The Regional Environmental Monitoring and Assessment Program (REMAP)(FTN et al., 1994).

1.2 MONITORING

Environmental monitoring programs have developed in response to specific needs, such as compliance monitoring by regulating agencies responsible for the condition of surface waters, or fixed-station monitoring networks that primarily address indicators of exposure and stress. Some of the monitoring programs are driven by mandates in the Clean Water Act (CWA). The reports required by Sections 305(b) and 314 of the CWA are an example. Programs that collect data on other ecosystem types have also been established. For example, the U. S. Department of Agriculture (USDA) National Agricultural Statistical Survey collects data for agricultural resources; The Forest Service's inventory and analysis surveys of forest resources; and the U. S.

Geological Survey's National Water Quality Assessment (NAWQA) program monitors water quality in selected basins. None of the programs, however, have adopted a uniform approach for national and regional assessments across and among ecosystem types. The Environmental Monitoring and Assessment Program (EMAP) and its counterpart, REMAP is intended to fill that gap by providing the U. S. EPA Administrator, Congress, and the public with statistical data summaries and periodic interpretative reports on ecological status and trends. Because knowledge about uncertainity is important for interpreting quantitative environmental data, EMAP is designed to make rigorous uncertainity estimates as well (Larsen et al., 1991).

The REMAP was developed as a partnership between EMAP, EPA's Regional Offices, and States to promote the use of EMAP science. The objectives of REMAP follow:

1. **To evaluate and improve EMAP concepts for State and local use.**

2. **To assess the applicability of EMAP indicators and the EMAP approach at differing spatial scales, and**

3. **To demonstrate the utility of EMAP for resolving issues of importance to EPA, Regions, and States.**

The REMAP strategy lends itself to the benefits of a full partnership between states and federal agencies because both national and state monitoring needs can be met in a cost-effective manner. The EMAP approach can provide a cost effective approach for assessing ecological data and reporting estimates of status and trends in indicators of condition with known confidence. State reporting requirements under several sections of the Clean Water Act (CWA) can be accomplished using an EMAP monitoring approach. Section 305(b) of the CWA requires states to submit biennial reports that include analysis of water quality data of all navigable waterways to estimate environmental impacts. The Clean Lakes section 314 requires states to submit biennial reports that identify, classify, describe and assess status and trends in water quality of publicly owned lakes. REMAP projects are being designed to provide meaningful information to decision-makers within a 1- to 2-year period.

1.3 POLICY RELEVANT QUESTIONS

The Science and Ecosystem Support Division (SESD) of EPA Region 4 was asked by the Savannah River Watershed Project Policy Committee to implement the REMAP strategy as a demonstration project for the states of South Carolina and Georgia. These states were interested in reducing sampling frequency and analyses, having the ability to reduce or increase sampling density, responding quickly to emerging environmental problems, and maintaining representative coverage of environmental resources through a systematic yet random means of sampling. Before the monitoring study, a set of questions was posed by the states of Georgia and South Carolina to provide direction for the monitoring design. The following policy-relevant questions were identified to guide the development of a plan of study and subsequent monitoring efforts.

- **What is the status of condition of the water resources of the Savannah River Basin?**
- **What proportion of the Savannah River Basin surface waters are attaining designated uses?**
- **What are the changes of ecological condition over time?**
- **What factors might be associated with changes?**
- **Is there a tendency for distribution of condition in a specific direction (spatial gradient) over the basin landscape? What are the possible reasons for these gradients?**
- **What resources are at risk in the Savannah River Basin?**

1.4 OBJECTIVES

In response to the needs of the states and policy-relevant questions posed, The Ecological Assessment Branch (EAB) of the SESD developed the following study objectives with the concurrence of the Policy Committee of the Savannah River Watershed Project.

- **Estimate the status and change of the condition of water** resources in the Savannah River Basin;

- Identify water quality spatial gradients that exist within the Savannah River Basin and associate current and changing condition with factors that may be contributing to this condition and spatial gradients;

- Demonstrate the utility of the REMAP approach for watershed and river basin monitoring and its applicability for state monitoring programs;

- Incorporate the REMAP approach in the formulation and accomplishment of the State River Basin Management Plans; and

- Provide baseline information required to conduct comparative risk assessments in the Savannah River Basin.

1.5 DESCRIPTION OF THE SAVANNAH RIVER BASIN

The Savannah River originates in the mountains of Georgia, South Carolina, and North Carolina and flows south-southeasterly 312 miles to the Atlantic Ocean near the port city of Savannah Georgia (Figure 1). The Savannah River is formed at Hartwell Reservoir by the Seneca and Tugaloo Rivers. Above the confluence of the Seneca and Tugaloo Rivers, the headwater streams of the Seneca River are the Keowee River and Twelve Mile Creek. The Tugaloo River

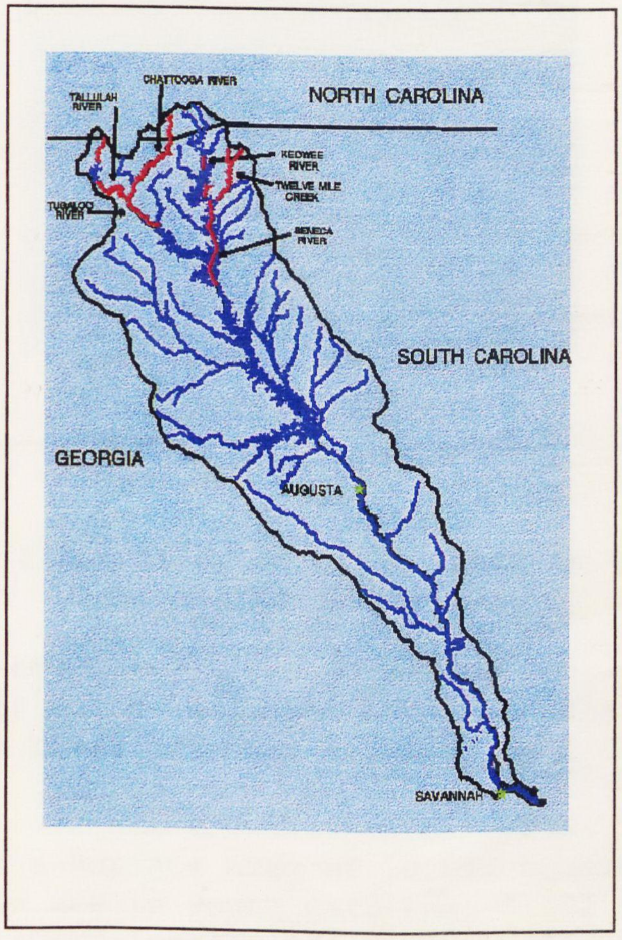


Figure 1. Savannah River Basin.

is formed by the confluence of the Tallulah and Chattooga Rivers. The Savannah River flowing in a south-southeasterly direction forms the border between the states of Georgia and South Carolina. The rivers' entire length of 312 miles is regulated by three adjoining Corps of Engineers multipurpose reservoirs, each with appreciable storage. The three lakes, Hartwell, Russell, and Thurmond, form a chain along the Georgia-South Carolina border 120 miles long. Six power developments that are part of the Georgia Power Company hydropower network exist upstream of Hartwell Lake on the Tugaloo River system; Yonah and Tugaloo lakes on the Tugaloo River, and Tallulah Falls, Rabun, Seed, and Burton lakes on the Tallulah River. Upstream of Lake Hartwell, on the Seneca River, is Duke Power Company's Keowee-Toxaway Project. The project is composed of two adjoining reservoirs, the most downstream of which is Keowee Lake, and the other two, Jocassee and Bad Creek Lakes are pump storage projects.

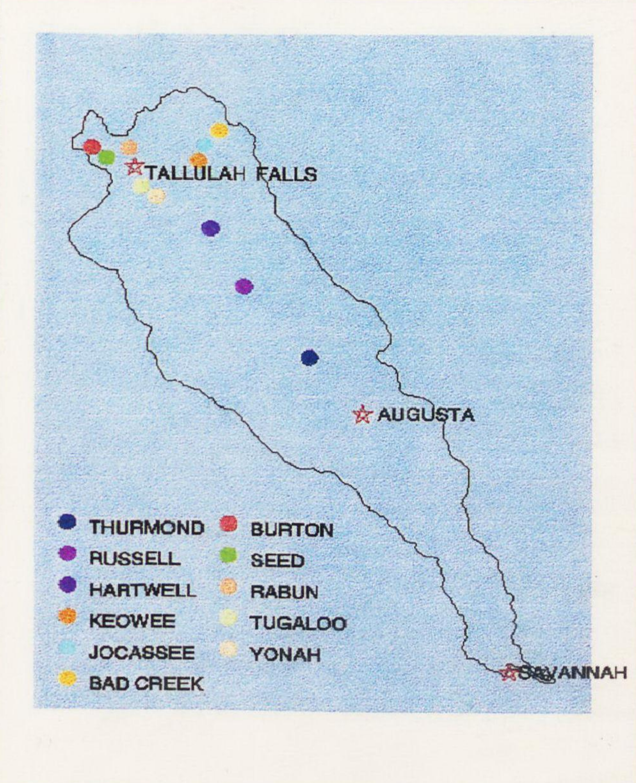


Figure 2. Location of Major Lakes in the Savannah River Basin.

The Savannah River Basin has a surface area of 10,577 square miles, of which 4,581 square miles are in South Carolina, 5,821 square miles are in Georgia, and approximately 175 square miles are in North Carolina. Like other basins of large rivers in the Southeast which flow into the Atlantic Ocean, the Savannah River Basin embraces three distinct areas: the Mountain Province, the Piedmont Province, and the Coastal Plain. The mountains and

Piedmont are part of the Appalachian area. The division between the Mountain and Piedmont is an irregular line extending from northeast to southwest, crossing the Tallulah River at Tallulah Falls. The Fall Line, or division between the Piedmont Province and the Coastal Plain, also crosses the basin in a generally northeast to southwest direction, near Augusta, Georgia. Elevations within the Mountain Province of the basin vary from 1,500 feet National Geodetic Vertical Datum (NGVD) on the Tallulah River to 5,030 feet NGVD for the highest peak, Little Bald Mountain, in North Carolina along the watershed divide. The Piedmont Province, due to its great width of over a hundred miles, is truly piedmont only in the upper parts, and gives way to a midland area before reaching the Coastal Plain. Exclusive of river valleys, its elevation generally varies from 500 feet NGVD at the Fall Line to about 1,800 feet NGVD at its upper extremity. Elevations within the Coastal Plain vary from 500 feet NGVD at the Fall Line to sea level at the Atlantic Ocean.

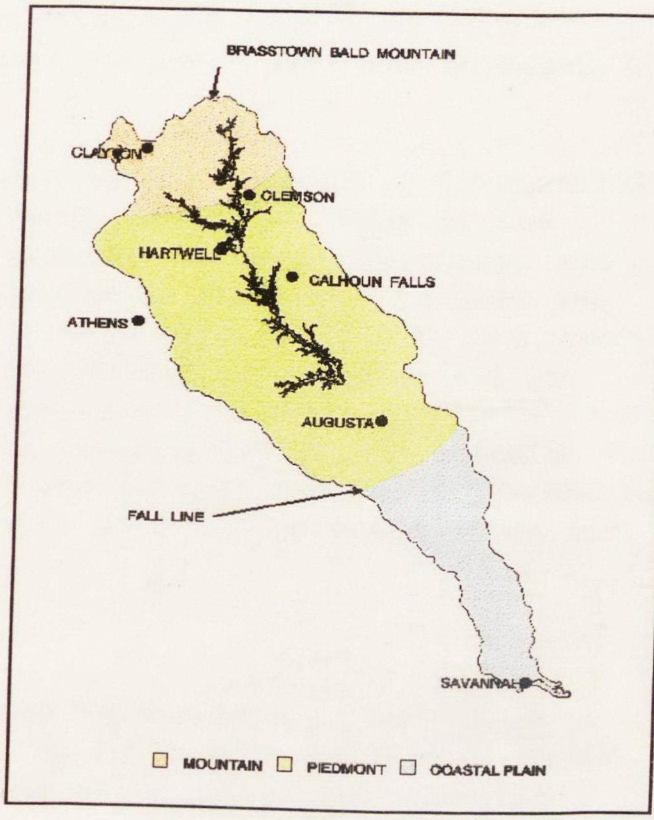


Figure 3. Physiographic Provinces of the Savannah River Basin.

Land use in the basin is agriculturally oriented. Sixty-six percent of the basin is considered timberland and 34.1% is nonforested. The number of acres farmed remains constant. Between 1987 and 1992 there was little change in the total farm acreage in the basin. However, Georgia had 330 fewer farms and lesser acreage in 1992 than in 1987 while South Carolina had an increase of 931 farms and an increase of 110,134 acres in farm land. There was a shift over the same five-year period in the types of crops grown. An increase in the number of acres cultivated have occured in corn (18%), cotton (86%), peanuts (12%), and tobacco (31%). These gains have been made with corresponding decreases in primarily wheat (-30%) and soybeans (-32%).

The Savannah River Basin contains all or part of 43 counties in Georgia, South Carolina, and North Carolina. Four of the counties are in North Carolina, thirteen in South Carolina, and twenty-six in Georgia. The population of the basin in 1990 was about 1,500,000 and is expected to grow to 1,800,000 by the year 2030. About 53% of the population resides in Georgia, 42% in South Carolina, and 5% in the headwaters located in North Carolina. Four metropolitan areas contain 62% of the basin's population. Savannah, Georgia is the largest city with 137,560 persons followed by Augusta, Georgia with a population of 44,619.

1.6 RESERVOIRS (LAKES)

Reservoirs herein called lakes are defined as bodies of water that have a surface area of at least 4 ha, with a depth of *at least 1 meter and at least 1,000 m^2 of the surface area* in open water.

Tributary embayments *of six major lakes* were studied over a two-year period (1995 & 1996). These lakes were Burton, Jocassee, and Keowee located in the Mountain Province. The other three lakes, Hartwell, Russell, and Thurmond were located in the Piedmont Province.

Lake Burton, controlled by Georgia Power Company, is located near Clayton Georgia. It is an old reservoir impounded in 1919. The lake has a shoreline length of 62 miles surrounding 2,775 acres containing 1,000,080 acre-feet of water.

Hartwell Lake is 7 miles east of Hartwell, Georgia. A dam is located at river mile 305.0. When the lake level is at elevation 660 ft. NGVD, the top of the conservation pool, the lake extends 49 miles up the Tugaloo River, Georgia, and 45 miles up the Seneca and Keowee Rivers, South Carolina covering 55,900 acres. The shoreline at elevation 660 is about 962 miles long, excluding island areas. The lake has a total storage capacity of 2,550,000 acre-feet below elevation 660. Hartwell dam was closed in 1963.

Russell dam is at River Mile 275.2 in Elbert County, Georgia and Abbeville County, South Carolina. The dam is 18 miles southwest of Calhoun Falls, South Carolina, and 40 miles northeast of Athens, Georgia. At the top of conservation pool elevation of 475 NGVD, the lake has a useable storage capacity of 126,800 acre-feet and a shoreline of 523 miles encompassing 26,000 acres. Operation of the project began in Jaunuary, 1984.

Thurmond Lake is 22 miles upstream of Augusta Georgia. At elevation 330 NGVD, at the top of the lake pool, the lake extends 40 miles up the Savannah River and about 30 miles up the Little River in Georgia. The lake has about 1,050 miles of shoreline, excluding island areas. At the top of the flood control pool (elevation 335 NGVD), the lake has an area of 78,500 acres with a total storage capacity of 2,510,000 acre-feet.

The three-project system is authorized and operated for fish and wildlife, flood control, hydropower, navigation, recreation, water quality, and water supply.

Duke Power Company built and controls Lakes Jocassee and Keowee. The uppermost lake, Jocassee, was built in 1973. It contains an area of 7,318 acres holding 1,077 acre-feet of water with a shoreline length of 75 miles. Lake Keowee built in 1971 has a shoreline length of 300 miles encompassing 18,373 acres with a storage holding capacity of 955 acre-feet.

1.7 STUDY DESIGN

1.7.1 Resources of Interest

The statistical population of interest included all tributary embayments >20 hectares associated with lakes >500 hectares. A tributary embayment is defined as a body of water associated with, but offset from, the main lake that has a permanent, blue-line stream at its headwaters. The embayment begins at the plunge point, the stream stretch where the inflow water density is greater than the density of the lake surface water, and it joins the main body of the lake at the plane created by intersecting break points of the shoreline of the embayment with the main body. Tributary embayments are associated only with lakes that have a shore line development ratio >3.0 and a surface area >500 hectares (FTN et al., 1994).

Shore line development is the ratio of the actual length of shore line of a lake to the length of the circumference of a circle the area of which is equal to that of a lake. If a lake had a shoreline in the form of a circle, the shore line development would be 1.0 (Welch, 1948).

1.7.2 Statistical Sampling Design

A probablistic sampling survey strategy was used to characterize the lake embayments of the Savannah River Basin. The sampling design was derived from the approach used in the EMAP (Messer et al., 1991; Overton et al., 1990; Stevens et al., 1992).

Probability sampling designs use randomization in the sample selection process. Probability sampling is the general term applied to sampling plans in which

- **every member of the population (i.e., the total assemblage from which individual sample units can be selected) has a known probability (>0) of being included in the sample;**

- **the sample is drawn by some method of random selection consistent with these probabilities; and**

- **the probabilities of selection are used in making inferences from the sample to the target population (Snedecor and Cochran, 1967),**

One advantage of probability-based surveys is their minimal reliance on assumptions about the underlying structure of the population (e.g., normal distribution). In fact, one of the goals of probability-based surveys is to describe the underlying structure of the population. Randomization is an important aspect of probability-based surveys. Randomization ensures that the sample represents the population. Without probability sampling, each sample often is assumed to have equal representation in the target population, even though selection criteria clearly indicate this is not the case. Without the underlying statistical design and probability samples, the representativeness of an individual sample is unknown. Drawing inferences from samples selected without randomization and without incorporating inclusion probabilities can yield misleading conclusions. To provide policy-relevant information, not only is the ecological condition of the target population important, but also the proportion of the resource that is in a particular state of condition. Very different policy and management alternatives might be evaluated if 50% rather than <10% of target embayments are hypereutrophic.

1.7.3 Frame Material

A sampling frame is an explicit representation of a population from which a sample can be selected. The sampling frame for the lake embayments is the USGS 1:100,000-scale map series in digital format (DLGs) and the modification of the DLGs represented by the U.S. EPA River Reach File (RF3), which established edge matching and directionality in the DLG files. From this, we used all lake areas identified as tributary embayments.

1.7.4 Sample Site Selection

The survey design follows the general design strategy proposed for EMAP (Overton et al.,1990; Messer et al., The EMAP sampling design (Overton et al., 1990) achieves comprehensive coverage of ecological resources through the use of a grid structure. White et al. (1992) describe the construction of the underlying triangular point grid and its associated tessellation of hexagonal areas. The EMAP base grid has a point density of one point per 635 km2. The base grid is intensified through a 7 x 7 fold enhancement, a 49-fold increase in grid density (White et al., 1992) which results in one point per 13 km2 hexagonal area. The hexagonal tessellation was randomonly located over the area covered by the embayment population. Within each hexagon, a point was randomonly selected. If the point fell within one of the embayments, then that point became a sample point. The selection process ensures that each location in the embayment population is equally likely to be sampled, and that the set of sites are spatially distributed throughout all embayments. Stevens (1997) defines this sampling process as a random tessellation stratified design. Since the study extended over two-years, two independent samples were selected - one for each year. A total of 111 sample locations were selected such that 52 were allocated in 1995 and 59 in 1996.

1.8 INDICATORS

REMAP monitors ecological indicators to assess status, trends, and changes in the condition and extent of the Region's ecological resources (Bromberg, 1990, Hunsaker and Carpenter, 1990; Hunsaker et al., 1990). Indicators are defined as any characteristic of the environment that estimates the condition of ecological resources, magnitude of stress, exposure of a biological component to stress, or the amount of change in condition.

Ecological principles state that ecosystem responses and condition are determined by the interaction of all the physical, chemical, and biological components in the system. Because it is impossible to measure all these components, REMAP's strategy emphasizes indicators of ecological structure, composition, and function that represent the condition of ecological resources relative to societal values. The challenge is to determine which ecological indicators to monitor. One approach for selecting these indicators starts with those attributes valued by society and determines which indicators might be associated with these values.

1.8.1 Societal Values

To be effective, information from the monitoring program must prompt action when required. This means the information produced must be related to perceptions of aquatic health and represent issues and values of concern and importance to the public, aquatic scientists and decision makers. The selection of these societal values drives the selection of appropriate indicators. After extensive discussions with aquatic resources managers, decision makers and the scientific community by members of the EMAP - Surface Waters Resource Group (Larsen and Christie 1993), an initial set of societal values and concerns were identified for evaluation in EMAP. These values are:

- **Biological Integrity,**
- **Trophic Condition, and**

▸ **Fishability**

Biological integrity can be defined as the ability to support and maintain a balanced, integrated, adaptive community with a biological diversity, composition, and functional organization comparable to those of natural lakes and streams of the region (Frey 1977; Karr and Dudley 1981) and includes various levels of biological, taxonomic and ecological organization (Noss 1990). Waters in which composition, structure and function have not been adversely impaired by human activities have biological integrity (Karr et al. 1986). Karr and others (1986) also defined a system as healthy "when its inherent potential is realized...and minimal external support for management is needed." This value or ethic differs considerably from values oriented toward human use or pollution that are traditionally assessed in water quality and fisheries programs, in which production of a particular species of game fish is the goal (e.g., Doudoroff and Warren, 1957), and may conflict with these definitions (Callicott 1991; Hughes and Noss 1992; Pister 1987).

Fishability is defined as the catchability and edibility of fish and shellfish by humans and wildlife (Larsen and Christie 1993). Fish represent a major human use of an aquatic ecosystem product. Protecting fish is the goal of many water quality agencies, and fish drive their water quality standards.

Trophic condition has been defined in EMAP as the abundance of production of algae and macrophytes (Larsen and Christie 1993). Trophic condition involves both aesthetic (water clarity) and fundamental ecological (production of plant biomass) components. It is a key aspect in determining both a lake's relative desirability to the public, its production of fish and its ecological character or classification by limnologists (e.g., eutrophic or oligotrophic). Because of limited resources, a decision was made to concentrate on trophic condition indicators for lakes.

1.8.2 Types and Selection of Indicators

EMAP defines two general types of ecological indicators, condition and stressor indicators. A condition indicator is any characteristic of the environment that estimates the condition of ecological resources and is conceptually tied to a value. There are two types of condition indicators: biotic and abiotic. Condition indicators relate to EMAP's first and second objectives: estimating the status, trends, and changes in ecological condition; and the extent of ecological resources.

Stressor indicators are characteristics of the environment that are suspected to elicit a change in the condition of an ecological resource, and they include both natural and human-induced stressors. Selected stressor indicators are monitored in EMAP only when a relationship between specific condition and strssor indicators are known, or a testable hypothisis can be formulated. Monitoring selected stressor and condition indicators addresses the third EMAP objective of seeking associations between selected indicators of stress and ecological condition. These associations can provide insight and lead to the formulation of hypotheses regarding factors that might be contributing to the observed condition. These associations can provide direction for other regulatory, management, or research programs in establishing relationships.

We focused on condition indicators related to trophic condition because of limited resources. The original study plan (FTN et al., 1994) proposed sampling for fishability indicators, Fish Health Index and Fish Tissue Residues; biological integrity, phytoplankton and zooplankton identification and counts; and one other trophic condition indicator, zeaxanthin, a marker pigment for blue-green algae. Work is continuing on this pigment, but the information was not sufficient for inclusion into this report.

The trophic condition indicators measured during this study were corrected chlorophyll a, total phosphorus (TP), algal growth potential (AGP), Secchi disc transparency, and total suspended solids (TSS). These indicators were selected because they provide different insights into the condition of the embayment waters.

Corrected chlorophyll a is commonly used to estimate the degree of phytoplankton bloom conditions that can affect aesthetics, fishing and swimming quality, taste and odor of fishes and drinking water, and the health of fish, waterfowl, and livestock. Chlorophyll is a measure of instantaneous standing crop, whereas TP and AGP indicate potential for blooms. Total phosphorus reveals insights about nutrient input and the potential for serious bloom conditions if we assume all of it is available. However, much of the TP is not normally available. The AGP can show how much of the TP is available for algal growth and the potential, under optimum conditions, for blooms. Secchi disc transparency is related to swimming conditions. Total suspended solids is related to transparency, but it also can be used to indicate effects upon fish production.

2.0 METHODS/QA

Standard operating procedures (SOP) of the Ecological Assessment Branch and Analytical Support Branch of EPA's Region 4 SESD were followed as the principle means of monitoring appropriate QA and QC checks on sample collections, physical measurements, chemical analyses, data gathering and processing. Data were subject to verification and validation. Verification included range checks and internal consistency checks. Validation consisted of a review of the data from a data users perspective for consistency based on known numerical relationships.

All lake sampling and measurements took place the weeks of 7/17 to 7/21, 1995 and 6/21 through 7/5, 1996. Eighty-two stations were sampled over the two-year period. This annual sampling window was selected because it is a time of maximum recreational use, and maximum water supply use.

Secchi disc transparency was measured by R.L. Raschke. Prior repetitive test measurements of Secchi depth in a variety of water bodies showed that the coefficient of variation (CV) for Secchi depth ranged from 5% to 15% among several samplers, including Raschke. In waters >1.0 meter Secchi depth transparency, the CV's were from 5 to 10%, but in waters of <1.0 meter, variability in measurements sometimes increased up to 15%.

Water samples were collected as depth integrated samples throughout the photic zone. Samples were collected for total phosphorus (TP), total suspended suspended solids (TSS), algal growth potential tests (AGPT), and chlorophyll a. Field duplicates were taken at least once in every ten samples.

For chlorophyll a, 100 to 250 mL of sample was filtered through a 24 mm diameter Whatman GF/F glass fiber filter. Samples were filtered in triplicate. The filter was folded, blotted dry, enclosed in aluminum foil, labeled and stored in a cooler containing dry ice and returned to SESD for analyses. Chlorophyll samples were extracted in 90% acetone and measured by visible spectrophotemetry and by high performance liquid chromatography (HPLC). The results given in Table 1 are from HPLC analyses as many of the chlorophyll levels were generally too low to be determined spectrophotometrically. For the field duplicates, the coefficient of variation (CV) ranged from 1.9 to 53.8%, with the average CV equal to 16.2%. According to the method (APHA, 1995), the laboratory precision ranges from 0.6 to 6.0%.

For 1995 (cycle 1), total phosphorus (TP) was analyzed using EPA method 365.1 (EPA, 1983). The results of all duplicate analyses, both laboratory and field, were below the minimum detection level of 0.02 mg/L for this method. In 1996 (cycle 2), a modification of method 365.1 was used (EPA, 1992). This modification allowed for detection at 0.006 mg/L. The CV for field duplicates at these low levels ranged from 9.4 to 71.2% with an average of 31.0%. This level of precision was not unexpected as method 365.1 gives the standard deviation at the 0.04 mg/L phosphate level as 0.019 which translates to a CV of 47.5%. The laboratory precision was much better ranging from 0 to 9.37%. The accuracy of both methods was good with spike recoveries ranging from 82 to 103% recovery. The average recovery was 98.3 %.

Total suspended solids were determined by using Method 160.2 (EPA, 1983). The CV for laboratory duplicates ranged from 0 to 19.8% with an average CV of 7.8%. Precision could not be determined for the field duplicates because all sets analyzed were below the detection limit of 4 mg/L. EPA-EMSL quality control samples analyzed by EPA Region 4 over a five year period had an average recovery of 95.8%.

Mean standing crop (MSC) and limiting nutrient were determined using the standard AGPT (APHA, 1995) as modified by

Schultz et al. (1994). The CV ranged from 1.3 to 53.1% with the average at 19.3%. This variability is similar to values listed for the method (Miller et al., 1978).

3.0 FINDINGS

The distribution of data for each variable can be characterized by its cumulative distribution frequency (cdf). These curves show the percent of embayment acreage in the basin equal to or less than some specified measurement plus or minus a confidence level. For the purpose of this study, we have set a confidence level of 95%. This means that we are 95% sure that the sampled acreage estimated to be equal to or less than a given measurement is within the bounds of our confidence lines on the graph (Fig. 4). There is a 1 in 20 chance (5% error) that the true or real percent of acreage affected at a particular measurement is not within the confidence bounds.

Chlorophyll a ranged from a low of 0.84 ug/L at Lake Hartwell to 11.56 ug/L at the most downstream lake, Lake Thurmond (Table 1).

Table 1. Range of Values for the Savannah River Lakes, 1995 & 1996.

LAKES	CHL. A UG/L	AGPT MG/L	LIMITS	TP UG/L	SD METERS	TSS MG/L
THURMOND	0.98-11.56	0.66-11.0	NP-N	2-50	1.2-4.8	2-27
RUSSELL	1.88-5.47	0.39-2.01	NP-N	10-60	0.7-2.6	2-32
HARTWELL	0.84-6.84	0.55-2.27	NP-N	3-30	1.7-10	2-6
KEOWEE	0.91-2.03	1.11-5.08	N-N	3-6	2.4-4.6	2
JOCASSEE	1.35-2.59	0.68-1.95	NP-N	3-10	3.3-6.0	2-34
BURTON	1.60	1.62	N	6	2.2	2

This range of concentrations at the times of sampling exhibit trophic conditions related to classical lake classifications of oligotrophic to mesotrophic(Olem and Flock, 1990). Chlorophyll a was less than 12 ug/L over the entire basin's large lakes (Fig. 4). Based on experience (Raschke, 1994) over the past 30 years, generally, when chlorophyll a ranges from 0 to 10 ug/L, there is no discoloration of the water and no problems. At a range of 10 to 15 ug/L, waters can become discolored and algal scums could develop. Between 20 to 30 ug/L, the water is deeply discolored, scums are more frequent, and

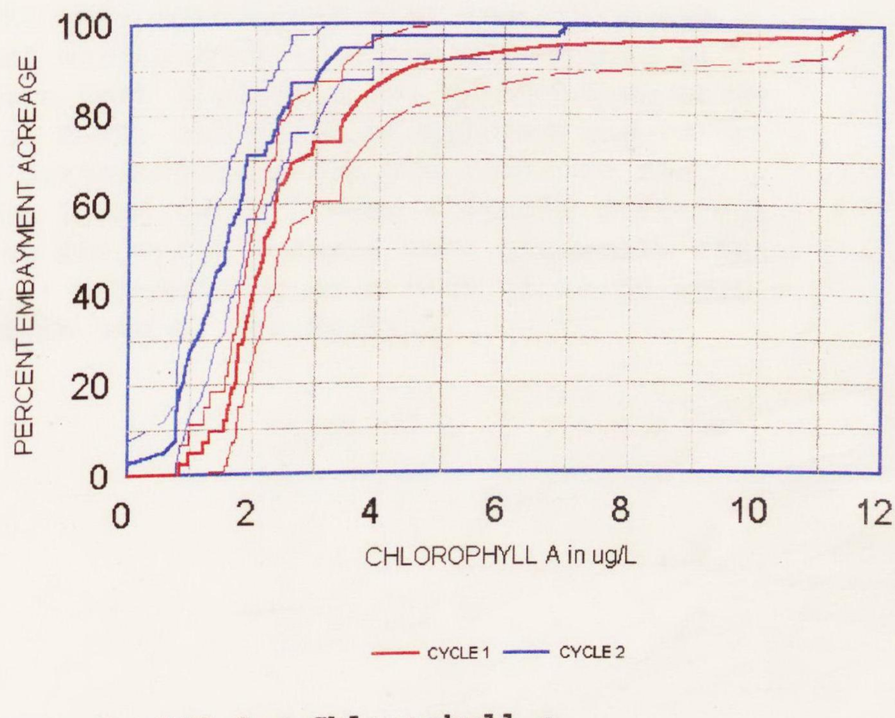


Figure 4. Cdf for Chlorophyll a.

matting of algae can occur (Raschke, 1993). EPA Region 4 (Raschke, 1993) has shown that a mean photic zone growing season average of equal to or less than 15 ug/L of chlorophyll a should satisfactorily meet multiple uses, including drinking water supply.

One of the objectives of the Savannah River REMAP is to detect trends in important environmental variables over both time and space. One means of comparison is through the testing of the null hypothesis that the population's distributions from two or more annual cycles are identically distributed. This can be accomplished through use of the Cramer von Mises test statistic (W) which is founded on design-based methods of statistical inference. For design-based statistical inference, the source of random variation is the random selection of sample sites. This is in contrast to model-based statistical inference, where the source of random variation is in the assessed statistical model (e.g., a regression model). Thus, designed-based statistical inference has the advantage that no model assumptions are required. The distribution of a population can be characterized through its cumulative distribution function (cdf). This is equivalent to testing the null hypothesis that the cdf's are identical. A test of cdf differences at alpha .05 (Table 2) using the Cramer-von Mises test statistic (W) showed that three variables, chlorophyll a, AGPT, and total phosphorus had significantly different distributions from one cycle to the other. Chlorophyll cycle 1 and cycle 2 were slightly different (W=1.26, k=2). The curve for cycle 1 rises more gradually than that of cycle 2 (Figure 4) culminating in a high of 11.56 ug/L thus suggesting the mean is higher for cycle 1.

Table 2. Cramer-von Mises Tests for Equality of Cumulative Distribution Functions for the Savannah River Basin Embayments. Equality of Cumulative Distribution Functions Between Cycles (Years) is Tested.

VARIABLE	W
CHLOROPHYLL A	1.26*
AGPT	5.84*
TOTAL PHOSPHORUS	1.86*
SECCHI DISC	0.25
TOTAL SUSPENDED SOLIDS	0.15

*Significant at alpha=.05

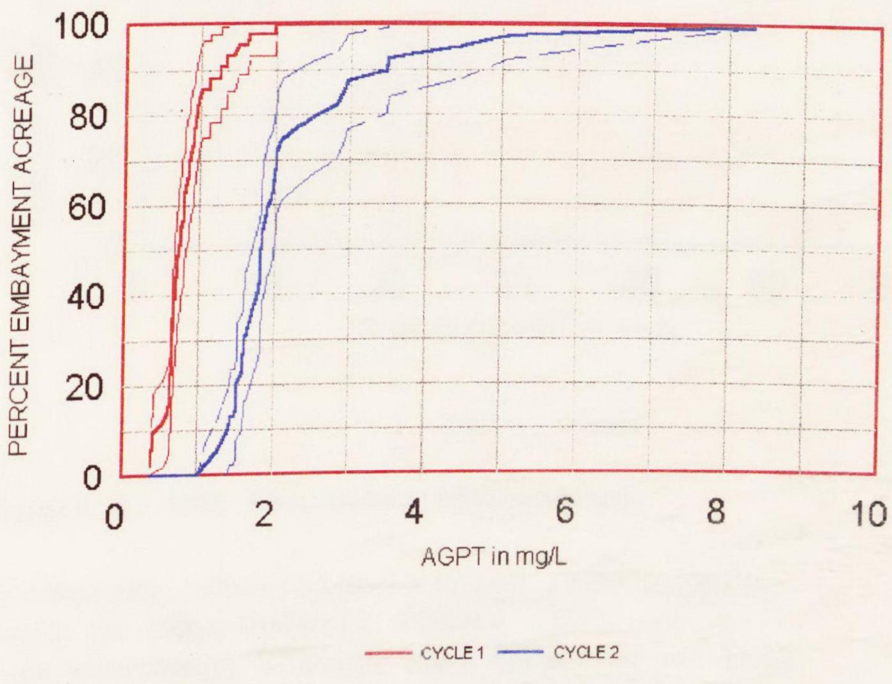


Figure 5. Cdf for AGPT.

Chlorophyll a represents phytoplankton standing crop or yield at given time periods, whereas AGPT is representative of the potential phytoplankton production, given optimum conditions of sufficient nutrients, light, time and temperature. Algal growth potential ranged from 0.39 mg dry weight (DW)/L at Lake Russell to 11.0 mg DW/L at Lake

Thurmond (Table 1). Approximately 99.7% of the AGPT dry weights were equal to or less than 5 mg/L (Fig. 5), an in-lake action level that will reasonably assure protection from nuisance algal blooms and fish kills in southeastern lakes (Raschke and Schultz, 1987). The 5 mg/L of dry weight translates to a potential chlorophyll standing crop of approximately 57 ug/L of chlorophyll a based on the following equation:

$$\mathrm{Log}_{10}\ \mathrm{chl}\ a = 1.15\ \mathrm{Log}_{10}(\mathrm{DW}) + 0.95$$ (Raschke and Schultz, 1987).

The sampled maximum chlorophyll a of 12 ug/L is much lower than the 57 ug/L of chlorophyll a derived from the 5mg DW/L AGPT action level. Figure 5 depicts the AGPT cdf's for cycles one and two. The curve for cycle two rises more gradually then that for cycle one suggesting the mean AGPT is not only higher in cycle one, but also shows greater variability within this cycle. The Cramer-von Mises test statistic confirms that the difference between the two cycles at the alpha .05 level is statistically significant (W=.5.84; k=2).

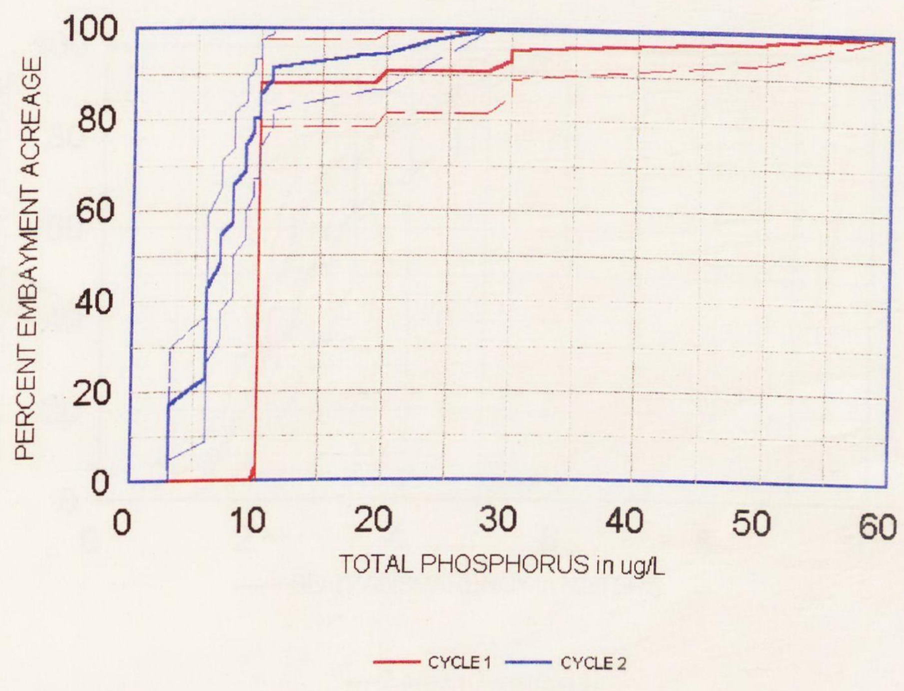


Figure 6. Cdf for Total Phosphorus.

Total phosphorus (TP), another indicator like AGPT of potential production, ranged from 2 ug/L in Lake Hartwell to 60 ug/L in Lake Russell (Table 1). Approximately 87.0% of the embayment acreage was equal to or less than 10 ug/L TP (Fig. 6). If all of the phosphorus were available for algal growth, at high values of 40 to 60 ug/L one could expect severe bloom conditions, but this was not the case as seen by the relatively low chlrophyll a values. This is not surprising; besides needing optimum conditions for maximum growth, the phytoplankton need sufficient nutrients that are bioavailable to them. Generally, not all of the TP in lakes is available for phytoplankton growth. Peters (1981) estimated that bioavailable phosphorus (BP) is 83% of TP in natural lakes and 18 to 57% in rivers. Since our lakes are reservoirs and thus an extension of a river system one would expect bioavailability to be much less than that found in natural lakes. Previous work on the 18 Mile Creek embayment of Lake Hartwell showed that the average percent of BP to TP was 38% (Raschke et al., 1985). Sometimes the BP portion of TP can be as low at 3% (Raschke and Schultz, 1987). At the alpha .05 level there was a significant difference (W=1.86; k=2) between cycles 1 and 2, but higher values were observed in cycle 1 (Fig. 6). The significant differences between cycles for chlorophyll, AGPT, and TP suggests that other than normal variables are influencing differences from one cycle to the other. We are not in a position with two years of data to focus on particular stress indicators at this time. Samples were collected from two to three weeks after rainfall events in the basin. Thus rainfall or unusually high stream flows would not seemingly cause the differences observed between cycles with respect to these three phytoplankton growth related indicators. The assumption of non-rainfall effects was confirmed by the non-significant differences between cycles for Secchi Depth (SD) (W=0.25; k=2) and total dissolved solids (TSS) (W=0.15; k=2), indicators of sedimentation effects from rainfall events. Presumably the cyclic differences were caused by internal lake

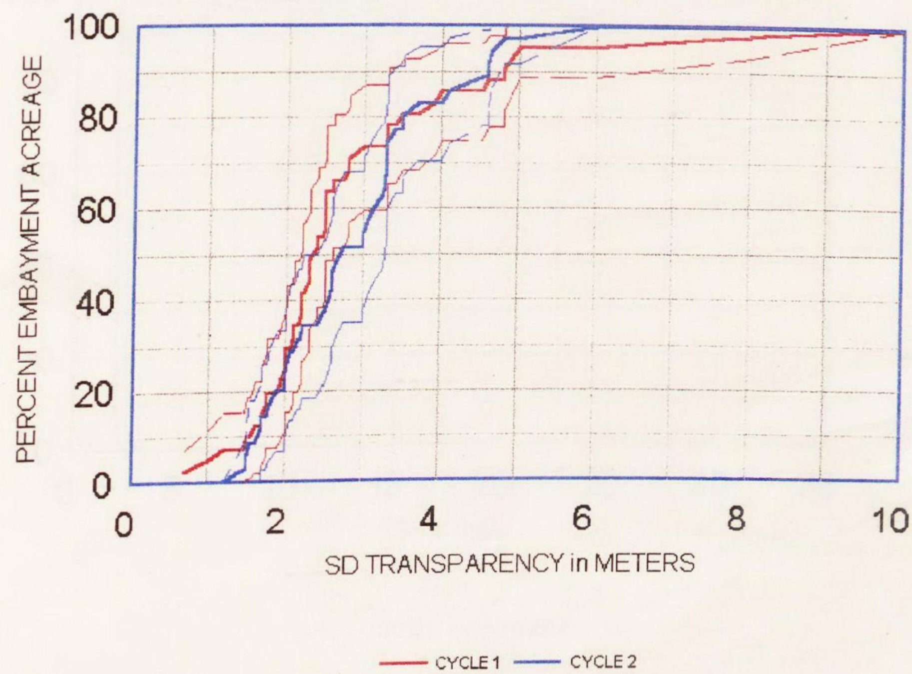


Figure 7. Cdf for Secchi Disc.

influences like internal nutrient cycling. Even these differences may be within the normal suite of variability experienced in a natural setting.

For water supply, a mean growing season average Secchi disc (SD) transparency of equal to or greater than 1.5 meters is desirable (Raschke, 1993). For non-water supply embayment situations a mean SD of greater than 1 meter is acceptable for fishing and swimming (Raschke, 1993). Secchi disc transparency ranged from 0.7 meters at Lake Russell to a high of 10 meters at Lake Hartwell (Table 1). An examination of Figure 7 shows that about 2.6% of the time less than desirable conditions existed in embayment waters for recreation purposes and only 5.3% of the time were they less than the water supply criterium of equal to or greater than 1.5 meters. Where SD was less than one meter, measurements were located near shore or at the upper end of the tributary embayments.

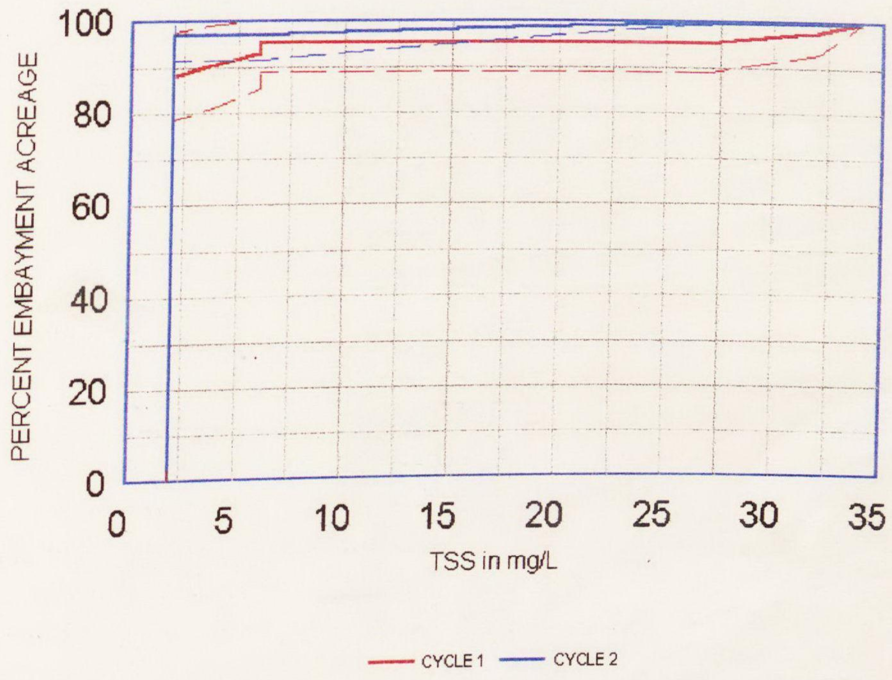


Figure 8. Cdf for Total Suspended Solids.

Buck (1956) divided impoundments into 3 categories: clear with total suspended solids (TSS) less than 25 mg/L; intermediate with TSS 25-100 mg/L; and muddy with TSS greater than 100 mg/L. The mean harvest of game fish was 162 lbs/acre for clear lakes, 94 lbs/acre in intermediate lakes, and muddy lakes only yielded 30 lbs/acre. The TSS ranged from a low of 2 mg/L at all lake embayments to a high of 34 mg/L at Lake Jocassee, the uppermost lake in the Savannah Chain of lakes (Table 1). Again these high values were attributed to near shore stations receiving wind fetch at the time of sampling. Ninety-seven percent of the embayment acreage would fall into Buck's clean category, with only 3% being intermediate with respect to water clarity (Fig. 8).

4.0 REFERENCES

APHA. 1995. Standard Methods for the Examination of Water and Wastewater, 19th Edition. American Public Health Association, Washington, D.C.

Buck, D. H. 1956. Effects of turbidity on fishing. Transactions of the American Wildlife Conference 21:249-261.

Bromberg, S. M. 1990. Identifying ecological indicators: An environmental monitoring and assessment program. Journal of the Air Pollution Control Association 40:976-978.

Callicott, J. B. 1991. Conservation ethics and fishery management. Fisheries 16:22-28.

Doudoroff, P., and C. E. Warren. 1957. Biological indices of water pollution with special reference to fish populations. Pages 144-163. In: Biological Problems in Water Pollution. U. S. Public Health Service, Cincinnati, Ohio.

EPA. 1983. Methods for the Chemical Analysis of Water and Wastes. EPA-600/4-79-020. Revised March, 1983. U.S. Environmental Protection Agency, Cincinnati, OH.

EPA. 1991. The Watershed Protection Approach, an Overview. EPA 503/9-92/002. U. S. Environmental Protection Agency, Washington, D. C.

EPA. 1992. Methods for the Determination of Chemical Substances in Marine and Estuarine Environmental Samples. U.S. Environmental Protection Agency, Cincinnati, OH.

EPA. 1996. Watershed Approach Framework. EPA 840/S-96/001. Office of Water. U. S. Environmental Protection Agency, Washington, D. C.

Frey, D. G. 1977. Biological integrity of water - A historical approach. Pages 127-140. In: R. K. Ballentine and L. J. Guarria (Eds.). The Integrity of Water. U. S. Environmental Protection

Agency, Washington, D. C.

FTN Associates, R. L. Raschke, H. Howard. 1994. Plan of Study for Characterizing Ecological Condition in the Savannah River Basin. U. S. Environmental Protection Agency, Region 4, Science and Ecosystem Support Division, Ecological Assessment Branch, Athens, Georgia.

Hughes, R. M., and R. F. Noss. 1992. Biological diversity and biological integrity: Current concerns for lakes and streams. Fisheries 17:11-19.

Hunsaker, C. T., D.E. Carpenter, and J. Messer. 1990. Ecological indicators for regional monitoring. Bulletin of the Ecological Society of America 71(2):165-172.

Hunsaker, C. T., and D. E. Carpenter. 1990. Environmental Monitoring and Assessment Program - Ecological Indicators. EPA 600/3-90/060. U. S. Environmental Protection Agency, Research Triangle Park, North Carolina.

Karr, J. R., and D. R. Dudley. 1981. Ecological perspective on water quality goals. Environmental Management 5:55-68.

Karr, J. R., K. D. Fausch, P. L. Angermeier, P. R. Yant, and I. J. Schlosser. 1986. Assessing Biological Integrity in Running Waters: A Method and its Rationale. Publication No. 15. Illinois Natural History Survey, Champaign, Illinois.

Larsen, D. P., D. J. Stevens Jr., A. R. Selle, and S. G. Paulsen. 1991. Environmental Monitoring and Assessment Program, EMAP - surface waters, a Northeast lakes pilot. Lake and Reservoir Management 7(1):1-11.

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

Management Committee. 1995. Savannah River Watershed Project: Initial Assessment and Prioritization Report for the Savannah River Basin. Volumes 1 and 2. Prepared by the Management Committee of the Savannah River Basin Watershed Project.

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

Miller, W., J. Green, and T. Shiroyama. 1978. The Selenastrum capricornutum Printz Algal Assay Bottle Test. EPA-600/9-78-018.

Noss, R. F. 1990. Indicators for monitoring biodiversity: A hierarchical approach. Conservation Biology 4:355-364.

Olem, H. and G. Flock (Eds.). 1990. Lake and Reservoir Restoration Guidance Manual. EPA 440/4-90/006. Second Edition Prepared by The North American Lake Management Society for the U. S. Environmental Protection Agency, Washington, D. C.

Overton, W. S., D. Stevens Jr., D. White 1990. Environmental Monitoring and Assessment Program - Design Report for EMAP. EPA 600/3-91/053. U. S. Office of Research and Development, Environmental Protection Agency, Washington, D. C.

Peters, H. 1981. Phosphorus availability in Lake Memphrenagog and its tributaries. Limnology and Oceanography 26:1150-?

Pister, E. P. 1987. A pilgrims progress from Group A to Group B. Pages 221-232. In: J. B. Callicott (Ed.). Companion to Sand County Almanac. University of Wisconsin Press, Madison, Wisconsin.

Raschke, R. L. 1994. Phytoplankton bloom frequencies in a population of small southeastern impoundments. Lake and Reservoir Management 8(2):205-210.

Raschke, R. L. 1993. Guidelines for Assessing and Predicting Eutrophication Status of Small Southeastern Piedmont Impoundments. U. S. Environmental Protection Agency, Region 4, Science and Ecosystem Support Division, Ecological Assessment Branch, Athens, Georgia.

Raschke, R. L., and D. A. Schultz. 1987. The use of the algal growth potential test for data assessment. Journal of the Water Pollution Control Federation 59(4):222-227.

Raschke, R. L., B. Cosgrove, and M. Koenig. 1985. Phosphorus loading effects upon phytoplankton standing crop in a tributary embayment of a southern piedmont reservoir. Pages 139-148. In: Lake and Reservoir Management: Practical Applications. North American Lake Management Society, Madison, Wisconsin.

Rathbun, S. C., R. Houghton, and C. Laurin. 1996. Cramer-von Mises Tests for Environmental Monitoring Data. A report for the U. S. Environmental Protection Agency, Science and Ecosystem Support Division, Ecosystem Assessment Branch, Athens, Georgia.

Schultz, D., R. Raschke, and R. Jones. 1994. A shortened algal growth potential test. Environmental Monitoring and Assessment 32:201-205.

Snedecor, G. W., and W. G. Cochran. 1967. Statistical Methods. Sixth Edition. The Iowa State University Press, Ames, IA.

Stevens, D. L. Jr. 1994. Implementation of a national environmental monitoring program. Journal of Environmental Management 42:1-29.

Stevens, D. L. Jr. 1997. Variable density grid-based sampling designs for continuous spatial populations. Environmetrics 8:165-197.

Stevens, D. L. Jr., A. R. Olsen, and D. White. 1992. Integrated Sampling Design: Environmental Monitoring and Assessment Program. EPA 600/xxx/xxx. U. S. Environmental Protection Agency, Environmental Research Laboratory, Corvallis, Oregon.

Urquhart, M. S., W. S. Overton, and D. S. Birkes. 1993. Comparing sampling designs for ecological status and trends: Impact of temporal patterns. In: V. Barnett and K. F. Turkman (Eds.). Statistics for the Environment. Wiley, New York.

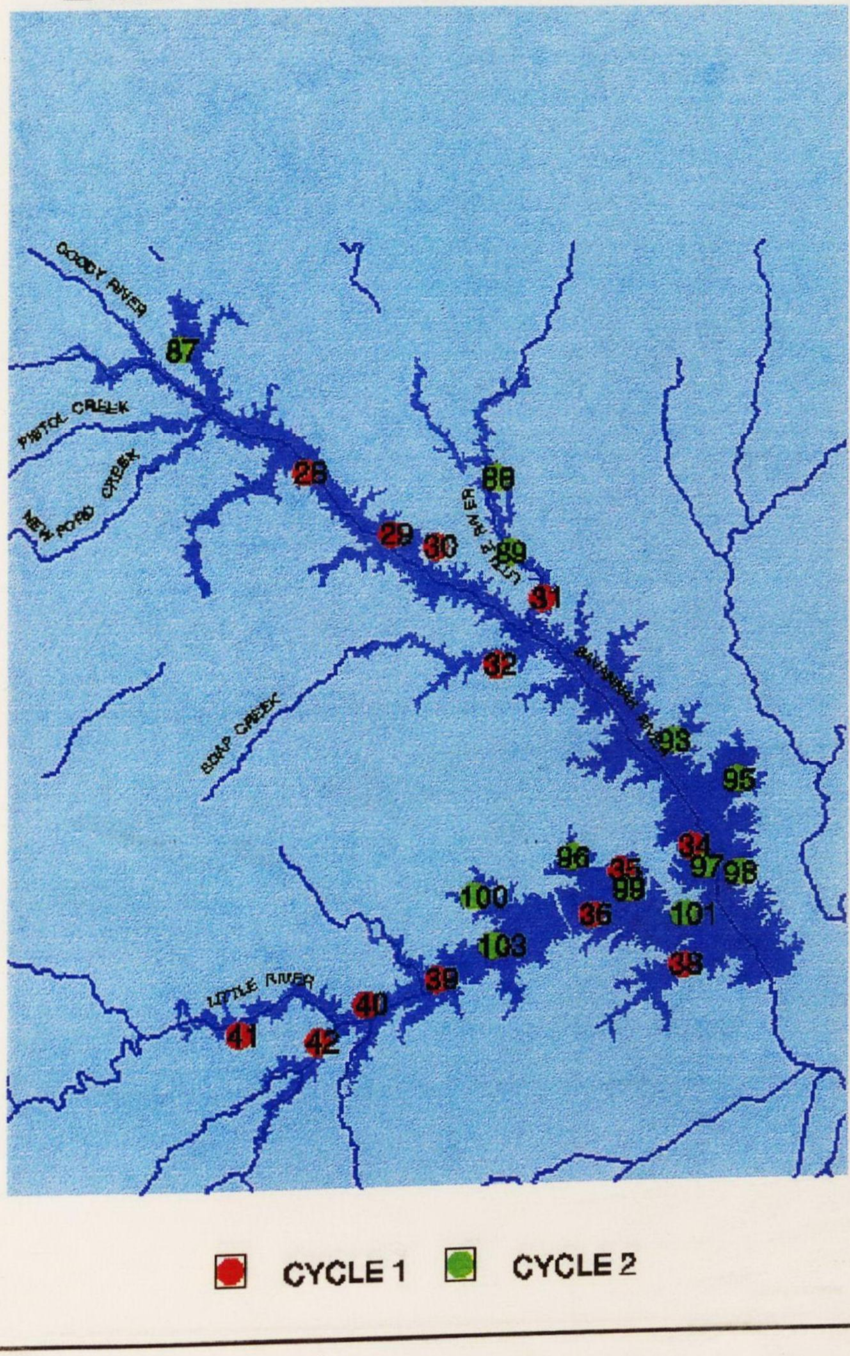
Welch, P. S. 1948. Limnological Methods. McGraw-Hill Book Co. Inc., New York.

White, D., A. J. Kimerling, and W. S. Overton. 1992. Cartographic and geometric components of a global sampling design for environmental monitoring. Cartography and Geographic Information Systems 19:5-22.

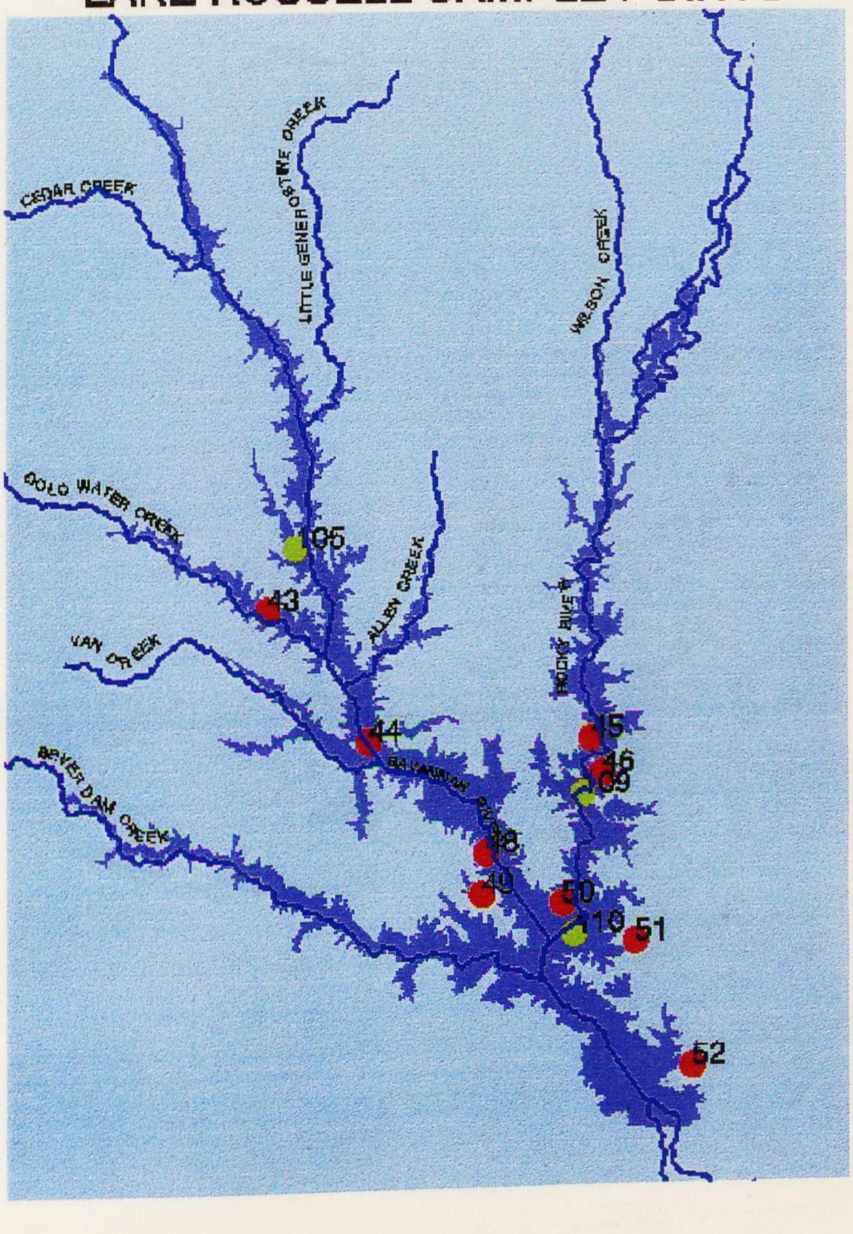
APPENDIX A

REMAP LAKE EMBAYMENT SAMPLING POINTS FOR CYCLES 1 & 2 (1995 & 1996)

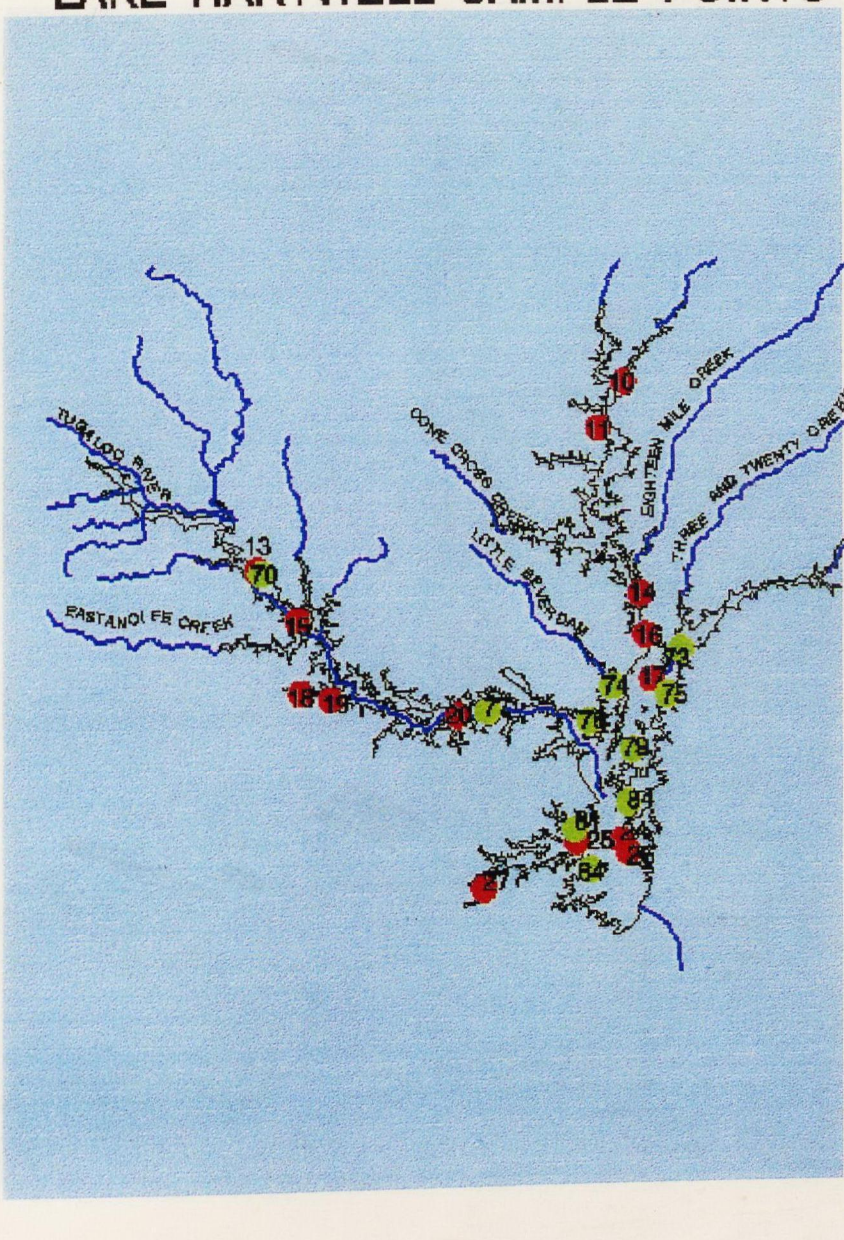
LAKE THURMOND SAMPLE POINTS

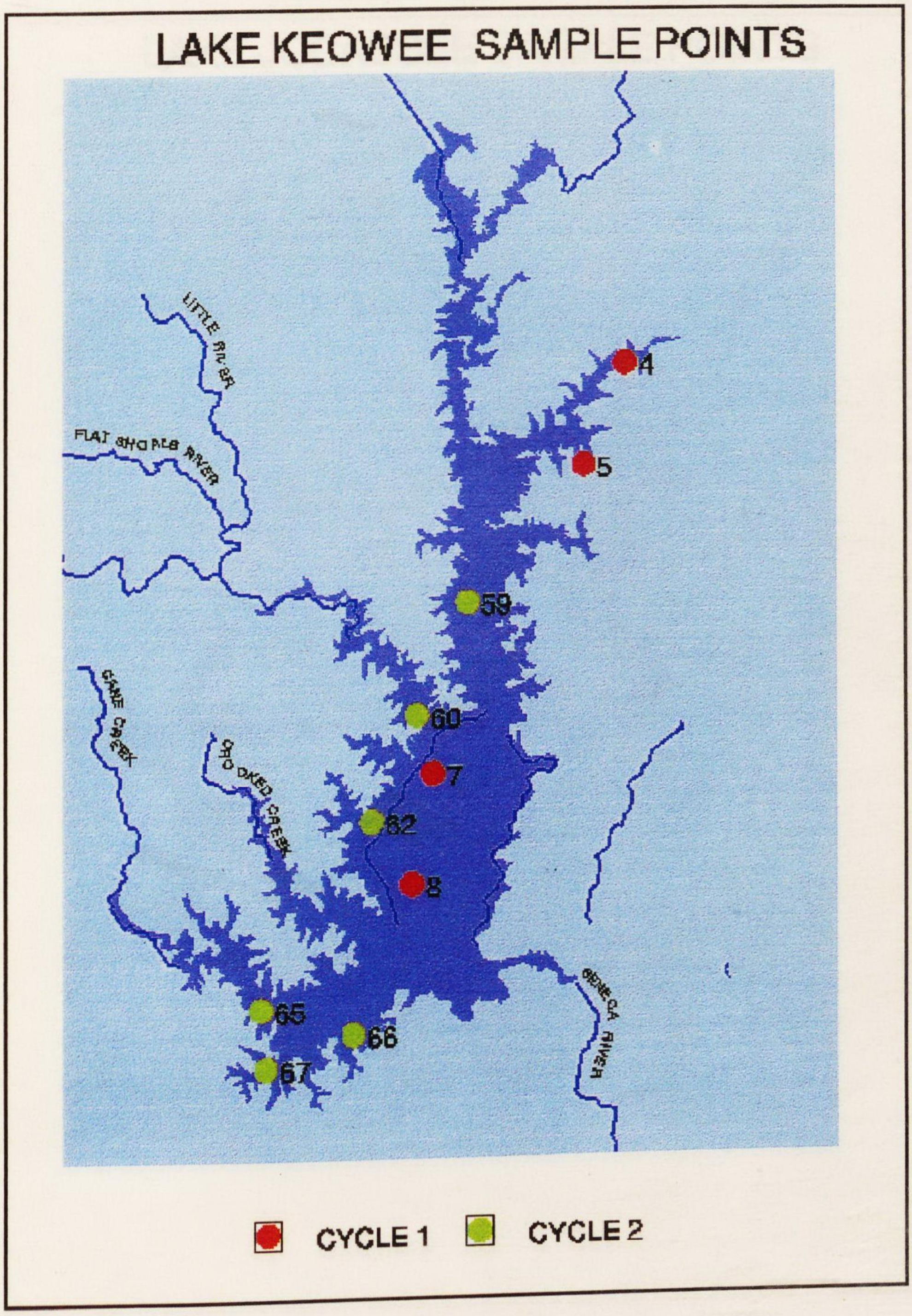


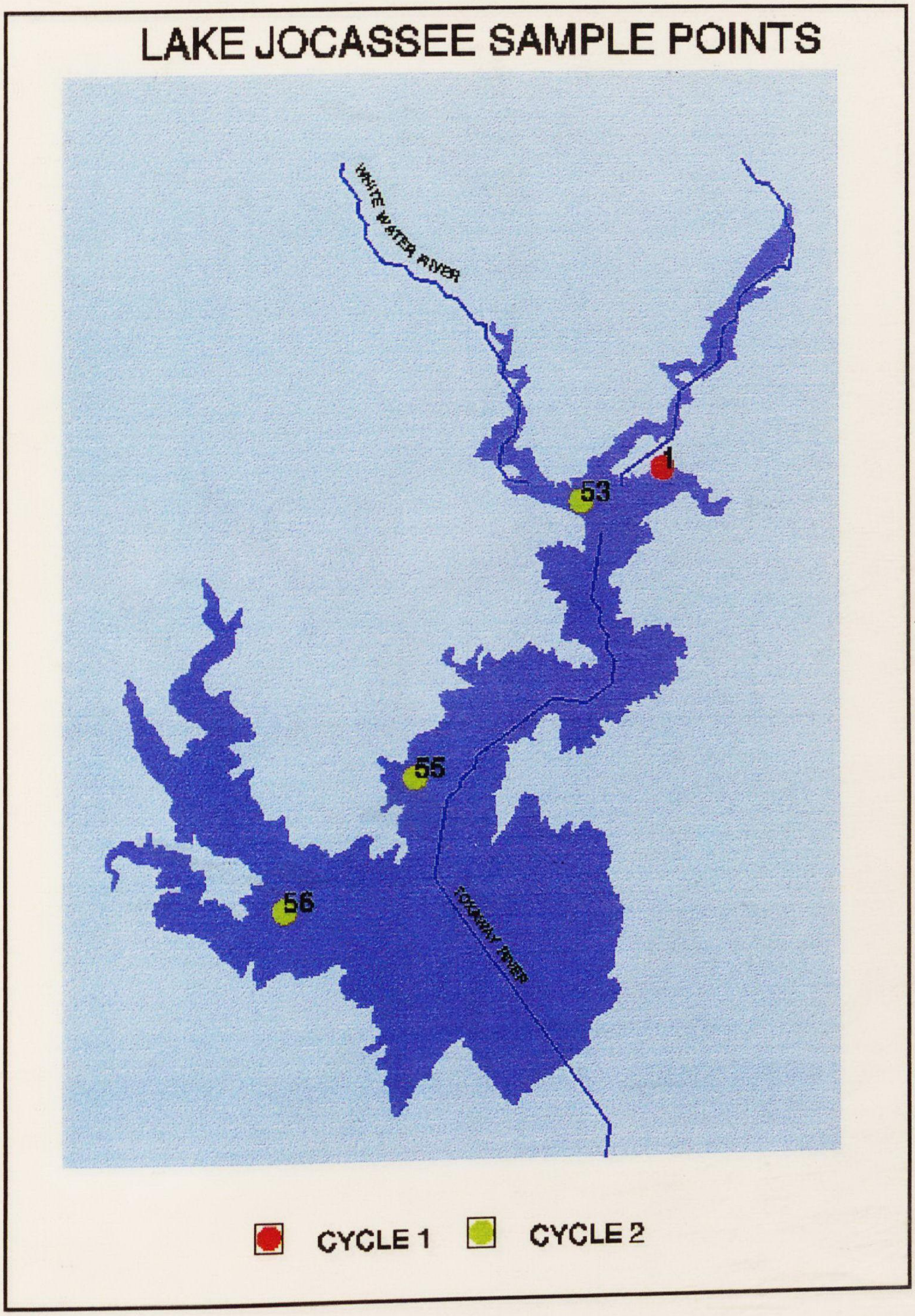
LAKE RUSSELL SAMPLE POINTS



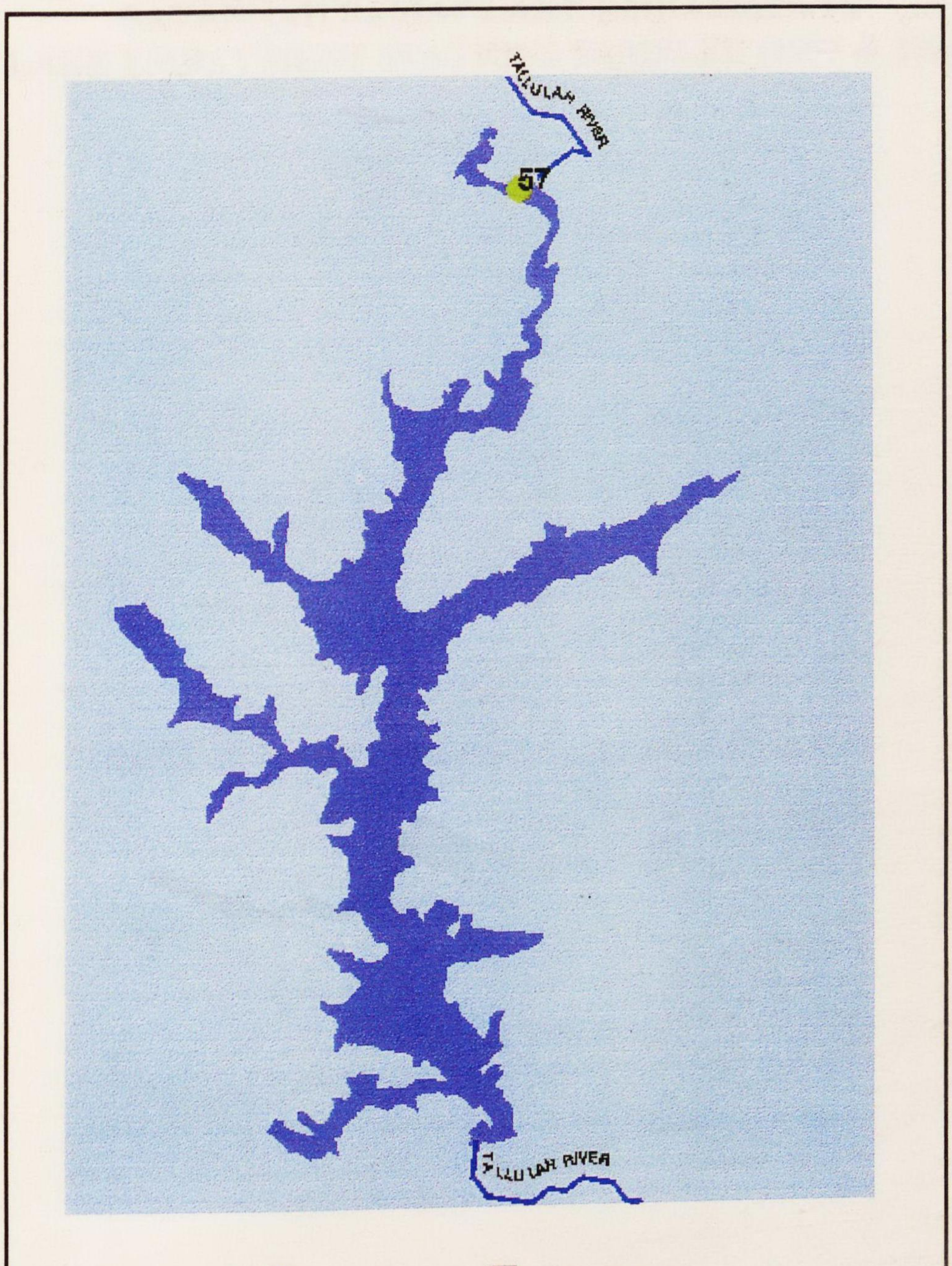
LAKE HARTWELL SAMPLE POINTS




LAKE KEOWEE SAMPLE POINTS
LITTLE RIVER
FLAT SHOALS RIVER
CANE CREEK
CROOKED CREEK
SENECA RIVER
4
5
59
60
7
62
8
65
66
67
CYCLE 1
CYCLE 2


LAKE JOCASSEE SAMPLE POINTS
WHITE WATER RIVER
TOXAWAY RIVER
1
53
55
56
CYCLE 1
CYCLE 2

LAKE BURTON SAMPLE POINTS



APPENDIX B

SAVANNAH REMAP LAKE EMBAYMENTS
LAKE EMBAYMENT STATIONS SAMPLED (1995 & 1996)

LAKE EMBAYMENT STATIONS SAMPLED, 1995 - 1996

CYCLE	STATION_ID	LONGDMS	LATDMS	LAKE
1	1	-82 54 24.530929	35 2 9.546320	Jocassee
1	4	-82 51 7.603567	34 53 45.993569	KEOWEE
1	5	-82 51 55.611809	34 52 8.208817	KEOWEE
1	7	-82 54 49.665560	34 47 11.788171	KEOWEE
1	8	-82 55 12.267418	34 45 25.877105	KEOWEE
1	9	-82 57 14.545488	34 43 25.328172	KEOWEE
1	10	-82 50 36.042482	34 41 50.982771	HARTWELL
1	11	-82 51 48.938645	34 40 1.427178	HARTWELL
1	13	-83 8 21.058132	34 34 24.386896	HARTWELL
1	14	-82 49 38.283149	34 33 34.244494	HARTWELL
1	15	-83 6 20.511658	34 32 28.952918	HARTWELL
1	16	-82 49 19.275926	34 31 56.433897	HARTWELL
1	17	-82 48 59.200669	34 30 13.216865	HARTWELL
1	18	-83 6 7.106787	34 29 38.128221	HARTWELL
1	19	-83 4 39.463439	34 29 23.852941	HARTWELL
1	20	-82 58 28.372960	34 28 47.713879	HARTWELL
1	22	-82 51 49.270112	34 27 9.136690	HARTWELL
1	24	-82 50 19.866007	34 23 54.623175	HARTWELL
1	25	-82 52 31.838827	34 23 51.677778	HARTWELL
1	26	-82 50 1.218515	34 23 29.129219	HARTWELL
1	27	-82 57 0.566973	34 22 0.973040	HARTWELL
1	28	-82 30 32.463918	33 55 29.861465	THURMOND
1	29	-82 27 11.522718	33 53 34.753482	THURMOND
1	30	-82 25 28.144621	33 53 11.871449	THURMOND
1	31	-82 21 18.751953	33 51 34.332004	THURMOND
1	32	-82 23 4.224512	33 49 30.599430	THURMOND
1	34	-82 15 26.752841	33 43 49.010464	THURMOND
1	35	-82 18 7.370846	33 43 3.904412	THURMOND
1	36	-82 19 16.008013	33 41 41.487729	THURMOND
1	38	-82 15 45.262833	33 40 6.970435	THURMOND
1	39	-82 25 10.454159	33 39 42.915041	THURMOND
1	40	-82 27 54.152732	33 38 57.791282	THURMOND
1	41	-82 32 46.441481	33 38 3.485909	THURMOND
1	42	-82 29 44.619895	33 37 50.333626	THURMOND
1	43	-82 45 17.522595	34 11 21.827095	RUSSELL
1	44	-82 42 52.970639	34 8 53.666517	RUSSELL
1	45	-82 37 41.619648	34 8 57.582352	RUSSELL
1	46	-82 37 25.496575	34 8 18.125663	RUSSELL
1	48	-82 40 7.213973	34 6 46.747787	RUSSELL
1	49	-82 40 12.296408	34 5 59.541942	RUSSELL
1	50	-82 38 19.522020	34 5 49.958644	RUSSELL
1	51	-82 36 35.699957	34 5 6.921357	RUSSELL
1	52	-82 35 13.386742	34 2 45.605296	RUSSELL
2	53	-82 55 11.608256	35 1 53.581238	Jocassee
2	55	-82 56 45.543830	34 59 45.144400	Jocassee
2	56	-82 58 0.827130	34 58 43.138440	Jocassee
2	57	-83 32 25.841760	34 52 58.890653	BURTON
2	59	-82 54 11.676408	34 49 55.812154	KEOWEE
2	60	-82 55 9.878470	34 48 7.103444	KEOWEE
2	61	-82 53 3.040507	34 47 13.030299	KEOWEE
2	62	-82 56 1.924517	34 46 25.690855	KEOWEE
2	65	-82 58 11.317878	34 43 26.490824	KEOWEE
2	66	-82 56 19.640120	34 43 2.932961	KEOWEE

LAKE EMBAYMENT STATIONS SAMPLED, 1995 - 1996

2	67	-82 58 3.955804	34 42 30.456779	KEOWEE
2	68	-83 14 49.616801	34 37 41.653244	HARTWELL
2	70	-83 8 8.974833	34 34 17.717526	HARTWELL
2	73	-82 47 36.995458	34 31 23.215728	HARTWELL
2	74	-82 50 54.886218	34 29 57.003787	HARTWELL
2	75	-82 48 12.315729	34 29 37.361079	HARTWELL
2	77	-82 57 0.235490	34 28 56.887496	HARTWELL
2	78	-82 52 4.677041	34 28 34.434387	HARTWELL
2	79	-82 49 52.508573	34 27 28.982305	HARTWELL
2	80	-82 50 3.004343	34 25 23.234650	HARTWELL
2	81	-82 52 37.963618	34 24 19.921871	HARTWELL
2	84	-82 51 50.090858	34 22 45.024973	HARTWELL
2	87	-82 35 26.567783	33 59 21.234816	THURMOND
2	88	-82 23 10.567304	33 55 20.967239	THURMOND
2	89	-82 22 36.206081	33 53 1.813908	THURMOND
2	91	-82 22 41.289075	33 49 42.901412	THURMOND
2	92	-82 25 29.984397	33 48 55.466103	THURMOND
2	93	-82 16 11.134728	33 47 8.846842	THURMOND
2	95	-82 13 41.447226	33 45 55.135352	THURMOND
2	96	-82 20 2.757306	33 43 30.367269	THURMOND
2	97	-82 14 52.721702	33 43 11.368723	THURMOND
2	98	-82 13 35.323466	33 43 0.845442	THURMOND
2	99	-82 17 52.510994	33 42 31.398682	THURMOND
2	100	-82 23 49.731993	33 42 18.312925	THURMOND
2	101	-82 15 40.402407	33 41 43.027050	THURMOND
2	103	-82 23 4.194887	33 40 45.547975	THURMOND
2	105	-82 44 38.684495	34 12 32.272620	RUSSELL
2	109	-82 37 53.290017	34 7 53.604337	RUSSELL
2	110	-82 38 2.135389	34 5 15.291914	RUSSELL
2	111	-82 36 20.648669	34 1 41.621509	RUSSELL

APPENDIX J

Peer Review Comments and the Response

Reviewer 1

Comment: You need to highlight the findings in a way that is easy to read and understand.

Response: We disagree. We reviewed our outline and the logical sequence of findings etc. Other reviewers found the report easy to follow and understand.

Comment: The descriptions through section 3 were excessive and sounded like project justification.

Response: We believe it is important to detail the approach and sampling strategy for the purpose of "fully" informing our primary client, the state monitoring and 305b coordinators, who generally are unfamiliar with the probability-based sampling design approach.

Comment: There are too many wholesale citations of SOPs which make the document difficult to stand on its own merits.

Response: We disagree. We only referred to SOPs in conjunction with field and laboratory efforts. The document is more sound when SOPs are used because the methods are based on a wealth of experience and quality control checks.

Comment: Lake Embayments - It would be helpful to estimate what you found and follow that with a discussion and interpretation section. It is hard to follow as written, but it might be a start on a scientific publication.

Response: We have no intent of publishing the lake study results in a scientific publication. We don't think the results will add anything new to science.

Comment: A lot of the figures were hard to read or missing or needed to be redrawn.

Response: Good Point. The figures were enlarged and put at the end of their respective sections.

Comment You need somewhere near the front to say what all the appendices deal with so that there is some understanding of the bulk.

Response: The titles in the table of contents and at the beginning of each appendix sufficiently describes the contents.

Comment: Important figures and tables in the appendices need to be pulled up into the main body of the report so the reader can get the message much more concisely and see what is being presented.

Response: We agree. Some tables were summarized and brought forward.

Comment: I would like to see if elevation or stream order plots of the data show the same trends as ecoregions. I am not convinced due to the disproportionate sample site distribution in the Lower Piedmont that this is the best way to parse the data. Other analytical approaches don't appear to have been explored.

Response: Our original intent was to examine the Basin as a whole (see appendices H and I). The "trend spatially" in the report is very subjective and based on few data points in some ecoregions. We noticed that there seems to be a "trend," but a new sampling design and strategy would have to be used to confirm our observations. Ecoregions provide a necessary spatial framework for monitoring ecological resources. Ecoregions represent areas of relative homogeneity. The 1991 Science Advisory Board's evaluation of the ecoregion concept said, " that the ecoregions not only provide a valuable framework for monitoring and assessment, but also provide a geographic context for defining biological criteria. Stream order and/or elevations could encompass several ecoregions.

Reviewer 2

Comment: The only substantive comment relates to recommendations for future studies. Add some more data for some of the ecoregions.

Response: We agree and there will be an opportunity in the summer of 1991 when SESD initiates the Regional REMAP study.

Comment: Add major streams to figure 1.2.

Response: We disagree. It would clutter up the figure which is intended to show the lakes that potentially could be sampled under our large lake criteria. The description in the text is sufficient.

Comment: In one place of section 5.1.1, the authors say 15 ug/L of chlorophyll A is satisfactory, but they imply that 57 ug/L of chlorophyll A when it is derived from 5 mg dryweight/L of AGP.

Response: The 15 ug/L is a growing season average based on intensive sampling of small lakes in Georgia, South Carolina, and North Carolina. The 57 ug/L is instantaneous and based on standing crop potential under optimum conditions. Since it is potential growth, a higher number derived in a laboratory setting is appropriate to initiate further investigation into a potential problem.

Reviewer 3

Comment: Overall, I think you have done an outstanding job summarizing the methods and results. The LPEI looks like a reasonable way to holistically portray the ecological information. I also like the way you answered the initial questions/objectives at the end.

Response: None.

Comment: When possible, future statistical studies should be designed to incorporate sufficient sites in each ecoregion to allow inferences to be drawn for each of he ecoregions of interest.

Response: The EMAP is designed to address ecoregion sampling. We focused at the basin scale because ecoregional sampling would have required more sampling and time. Additionally, ecoregions in the basin were not well defined at the beginning of sampling. The states of South Carolina and Georgia are in the midst of defining ecoregion boundaries and determining reference sampling sites. We and the states are in agreement with respect to the Lower Piedmont Ecoregion Boundaries.

Comment: Identification of reference areas may include subjectively selected sites if least impacted areas are under represented by the statistical sample in an ecoregion.

Response: We agree.

Comment: Further investigation of landscape/instream relationships is encouraged to build on the correlations documented here. Development of such relationships has considerable potential as a screening tool to identify potentially impaired sites.

Response: We agree. We plan to look at these relationships in an upcoming regional REMAP survey of wadeable streams.

Reviewer 4

Comment: Related to clarification and better sentence structure.

Response: Agreed with comments and expanded some sections to better explain findings.

Comment: I have concerns about the development of the LPEI and its use of the LPEI on the same data set used to develop it. Usually an index or criterion is developed on a reference set of data collected across he entire range of the target population and then applied to independent data. This data set only represents a part of the Lower Piedmont Ecoregion, and it may not capture the total range of any of the component metrics. It is truly only a Savannah Basin Lower Piedmont Index.

Response: We agree. We had not looked at the entire range (across the Lower Piedmont Ecoregion) for the individual metrics used. We only focused on the Savannah Basin. We corrected the LPEI in the text to SB-LPEI (Savannah Basin-Lower Piedmont Ecological Index). We will have an opportunity to test the index's power across many ecoregions within the Regional REMAP study beginning in the summer of 1999.

Comment: I think the appendix about locating probability sites on maps and in the field, and obtaining access permission will be very useful to us. That is exactly the stage we are at in establishing our probability network.

Response: We agree and think it is state of the art.

Comment: We have had a workshop on integration of judgement data with probability data and adequately answered state concerns. At that workshop, we were presented with some theoretical approaches for integrating data, but weren't given any procedures to use. The workshop addressed state concerns, but it didn't provide us with tools to accomplish integration. It did help illustrate the beneficial uses of probability-based designs in answering 305b and other resource-wide condition questions, and demonstrated the limitations of judgement-based designs in addressing those same questions. I think you have overstated the accomplishments of that workshop.

Response: The statement concerning the workshop was changed to reflect the reviewer's viewpoint. The follow up report in Appendix G addresses the question of merging judgement and probability data more fully.

Comment: The three-project lake system is authorized and operated...by the Corps of Engineers...for fish...etc. You mentioned who operates the other lakes, but failed to mention the COE on these major lakes.

Response: Correction noted and made by authors.

Review 6

Comment: We recognize the potential usefulness of probability sampling in our river basin sampling rotation and statewide monitoring.

Response: None

Comment: We are concerned that the results of the present report will prove difficult to fit into our 305b/303d listing process. That is, the "good," "fair," and "poor" evaluations may not provide a good fit with the 305b categories of support, partially support, and not support. For example, will fair mean partially support? The real concern is that we will probably have to take these results and fit them into 305b even though that has not been the primary purpose of the study.

Response: We agree that the primary purpose of the study was to demonstrate the feasibility of using the EMAP probability sampling approach for monitoring purposes. We believe the information gathered is amenable for inclusion into a 305b report and will work with the state on this concern.