
2.0 STUDY DESIGN

The EPA Region 4 South Florida Ecosystem Assessment Project, was designed to permit synoptic sampling of the canals and marshes of South Florida. The design was based on the sample survey procedures developed in the EPA Environmental Monitoring and Assessment Program (EMAP). The study design is presented in this chapter. In addition, four marsh transects were sampled along known nutrient gradients, and canal water control structures were sampled for Hg in water on a biweekly schedule for 3 years (i.e., February 1994 through February 1997) (Figure 2.1).

2.1 Design Rationale

There are two distinct paradigms for evaluating regional ecological status and trends. The two paradigms are rooted in very different perceptions of regional evaluation and lead to correspondently different methods for making regional inferences. In one, the sites are selected based on their anticipated ability to reflect regional characteristics. The site features used in site selection may be physical characteristics, spatial pattern, expected sensitivity to stress, anticipated exposure level, or any other aspect that might influence the response of the site to known or suspected environmental stresses. The quality of the resulting data depends on the judgement of the investigator. This approach can lead to biased estimates of environmental parameters. Moreover, it is difficult to assess sampling variability using this approach. An example of this approach is the design used to sample the marsh transects along known nutrient gradients.

EMAP uses an alternative approach that requires a probability sample. A key property of a probability sample is that every element in the population has some chance of being included in the sample. If this were not the case, then some parts of the population might as well not exist, because their condition could have no influence on estimates of population characteristics. This property has a side benefit, in that it forces an explicit and complete definition of the population being described. This may seem trivial; however, in practice, it is almost never easy to tightly delimit a real, physical population. Another requirement of a probability sample is that the chance of being included in the sample is known for every element in the population. This requirement is

satisfied if there is an explicit, well-understood mechanism that incorporates a random process in the selection of the sample. Thus, a probability sample has three crucial characteristics that distinguish it from other types of samples: (1) the population being sampled is explicitly described; (2) every element in the population has some opportunity of actually being sampled; and (3) the selection is carried out by a process that includes an explicit random element. These three characteristics ensure that a probability based sample provides a degree of comprehensive coverage that is not easily achieved with other methods. Furthermore, the probability based sampling approach can lead to more precise estimates.

A probability sample does not have to be a purely random sample, and in most cases should not be. In the case of environmental samples, there is almost always a great benefit in the spatial context of the sample sites. In simplest terms, spatial context is the information required to locate a sample point on the landscape, for example, latitude and longitude. However, there is a richer connotation in all of the available landscape information that can be attached to geographic coordinates: ecoregion, land use, soil type, vegetation cover, topography, and so on. Knowing the spatial context of a sample (i.e., knowing where the samples are located and knowing their spatial relationship to one another) provides the link of proximity to admit the joint evaluation of multiple responses, and to evaluate the effects of stresses with known spatial properties. The value of a probability sample is greatly enhanced if it is structured so that the sample preserves the spatial context of the population. Although any random sample accomplishes this in a sense, there is a great advantage to placing some spatial constraints on the sample so that the spatial distribution of the sample closely matches the spatial distribution of the population (i.e., the sample should in some sense be evenly distributed over the spatial extent of the population). EMAP's sampling methodology uses a grid, as described in the next section, to accomplish spatial dispersion of the sample while retaining the essential characteristics of a probability sample. The probability based sampling approach used in this study, for example, can be used to estimate the percent of canals with TP concentrations greater than some maximum or minimum threshold, or the percent of the marsh sampled where THg concentrations are greater than a maximum concentration specified.

2.1.1 Sampling Method

An obvious means to ensure sample distribution over the extent of the population is to use a systematic sample, say regularly spaced samples down the length of a canal or at the nodes of a regular grid placed over a marsh. The major disadvantage of a systematic sample is that a rigorous, design-based estimate of variance is impossible, even if there is an explicit random placement of the grid or start of the regularly spaced linear sample. The requirement for a probability sample is that every element of the population has positive probability of being in the sample; a variance estimate is possible if, in addition, every pair of elements in the population has positive probability of being in the sample simultaneously. A systematic sample does not satisfy this latter requirement.

The design used to sample the Everglades is based on a grid in order to ensure spatial coverage, but includes some additional randomization, to ensure that a variance estimate is possible. The basic concept is to randomly locate a grid over the area to be sampled and, within each polygon defined by the grid cells, to select one point at random. The basic design is called a random tessellation stratified (RTS) design (Bellhouse 1977, Dalenius et al. 1961, Olea 1984, Overton and Stehman 1993, Stevens 1997).

The basic RTS design results in every element of the population having the same chance of being included in the sample. Extensions that allow for variable inclusion probability are discussed in Overton et al. (1990) and Stevens (1997). The concept behind the variable probability extensions is to group the grid cells together to form larger polygons that also constitute a tessellation by congruent polygons. Figure 2.2 illustrates the concept, where the tessellation hexagons of a triangular grid are joined in groups of seven to form a collection of larger tessellation polygons represented by the solid lines. Variable probability is accomplished by picking one or more points from each group, and allowing the number of points picked to vary by spatial region. Stevens (1997) gives complete details of the variable probability versions of the RTS design, and Stevens and Kincaid (1997) give an easily computable and nearly unbiased variance estimator.

The design of the sampling in the Everglades was intended to obtain samples from both canals and marshes, to extend over 2 years, and to sample in both wet and dry seasons in each

year. Thus, there would be four sampling periods (2 seasons in 2 years) in both canals and marshes. A spatially interpenetrating design given by a 4-fold decomposition of EMAP's basic triangular grid (Overton et al. 1990, Stevens 1994) was chosen. In this approach, the grid cells composing the tessellation are split into four interlocking sets, and one set of cells is sampled each period.

The marsh sample was a straightforward application of these concepts. The area to be sampled was defined by a Geographic Information System (GIS) coverage and consisted of LNWR, WCA2, WCA3, ENP, and the eastern sector of BCNP. The sample was equiprobable with the exception of BCNP, which was sampled at one-third the intensity of the rest of the Everglades. A final design criterion of a maximum of 125 sampling sites per synoptic cycle, which included consideration of logistical and laboratory efficiencies, was placed on the marsh design.

The canal sample was selected differently because of the linear, essentially one-dimensional nature of the canals. The base EMAP grid was intensified by a factor of 7, resulting in a triangular grid with hexagonal cells having an area of approximately 90.1 km^2 each. A GIS was used to extract and randomly order the canal segments within each cell, where segments were defined by confluences or cell boundaries. The grid cells, along with the associated canal segments, were then randomly ordered, using the randomization procedure based on spatial partitioning described in Stevens (1994). This randomization gave it a linear order that preserved some of the spatial proximity relationship. The result of this process was a random mapping of the entire length of canals onto a single line in such a manner that every point on the line represented a known canal location (Figure 2.2). This line was then split into four pieces using the same grid decomposition as in the marsh sample and a sample selected from each piece using a systematic sample with a random start. The random mapping prior to the systematic sample ensures that every pair of points on the canals had a positive probability of being included in the sample, and the spatial-partitioning-based randomization ensures that the sample is well-distributed over the extent of the canals. A criterion of a maximum of 50 sampling sites per synoptic cycle was placed on the canal design.

2.1.2 Sample Points

This random probability-based sampling strategy was used to select sites in the Everglades ecosystem from south of Lake Okeechobee to the mangrove fringe on Florida Bay and from the ridge along the eastern coast into BCNP on the west. The distribution of 200 canal sample sites is shown in Figure 2.3 while the distribution of 500 marsh sample sites is shown in Figure 2.4. The sample points represent the current ecological condition in over 1,200 km (750 miles) of canals and in over 7,800 km² of marsh (3,000 mi² of marsh [over 2 million acres]). The canals were sampled in September 1993, May and September 1994, and May 1995. The marshes were sampled in April and September 1995 and May and September 1996. This corresponds to two dry (April and May) and two wet (September) seasons for both systems over a 2-year period. The project sampling included water, canal sediment, marsh soil, fish, macrophytes, and algae/periphyton at each canal and marsh sampling location during each sampling period. The parameters that were measured at each site can be used to answer questions on multiple issues, including eutrophication, Hg contamination, habitat alteration, and hydropattern modification.

2.1.3 Design-Based Estimation

There are two approaches to regional inference that are roughly parallel to the two methods of sample selection. (See Hansen et al. [1983] and the following discussion for a good contrast of these two approaches.) Briefly, a “model-based” approach uses conceptual, statistical, and mathematical models to draw inferences to regional populations based on in-depth information from a limited number of sites. The model may not be explicit (e.g., it may be embodied in the set of criteria used to select representative sites). The model specifies the relationship of the sites to the regional population, and the validity of the population inferences rests on the validity of the model. Model-based inference makes inferences about parameters of the model that generates the data and not about the population itself. In contrast, a design-based program is essentially an empirical approach. The design specifies what information is to be collected where, and the design stipulates the population inference. The validity of the inference rests on the ability of the design to produce regionally representative information. In general,

design-based programs rely on the methods of statistical survey sampling (see, for example, Cochran 1977) and are valid only with a probability sample.

Both approaches draw inferences by calculating estimators of population quantities (e.g., mean values, variance, and spatial pattern), and both rely on statistical theory to describe properties of the estimators. The properties most often used are descriptions of the expected behavior of the estimators: their variance, mean square error, bias, consistency, etc. The clearest contrast between model-based and design-based approaches lies in the basis for that expectation. Model-based methods rely on an assumed statistical model that describes the error distribution (i.e., the distribution of the discrepancies between reality and model results); expectations are then averages over possible error realizations. Design-based approaches rely on the explicit randomization used in selection of sample locations; expectations are averages over all possible samples.

A model-based approach uses information from sites in the population regardless of how they came to be selected. The sites may have been purposefully selected, they may be available because of historical circumstances, or they may be the result of a designed probability sample. In any case, inference to an associated population rests on the assumption that the behavior of the selected sites reflects and is typical of the behavior of the population. In a long-term monitoring program aimed at status and nonspecific change in a spatially distributed population subjected to nonuniform stresses, that assumption does not seem tenable under a judgmental sample selection protocol.

A probability sample allows the use of both design- and model-based analyses. Moreover, even if the model-based analysis does not make explicit use of the probability structure of the sample, the model-based inference is strengthened by the characteristics of a probability sample. Model-based parameter estimates can be biased under a judgement sampling design.

Hansen et al. (1983) makes several relevant points in their discussion of design-based versus model-based inference. One that is particularly relevant for an environmental monitoring program is that a probability sample permits inferences that are free of even the appearance of subjectivity. A probability sample from an explicitly defined resource population is a means of certifying that the data collected are free from any selection bias, conscious or not.

This requirement is essential for a program such as EMAP that aims at describing the condition of our national ecological resources. Moreover, analysis methods that are as free as possible from the appearance of subjectivity are also available under a design-based protocol.

2.1.4 Variable Probability Estimation

A requirement of a probability sample is that the probability of being included in the sample must be known for every element in the population. In a continuous, extensive population like the Everglades, this knowledge is contained in an inclusion probability density function $\pi(s)$. The requirement is that $\pi(s)$ be a known function, and that $\pi(s) > 0$ for every point s in the Everglades. Besides controlling the random selection process, $\pi(s)$ plays a critical role in the inference and analysis stage. The inclusion function specifies the density of sample points; therefore, it has units with dimensions like the number of sample points per unit of area (e.g., 1 sample point in 635 km^2 [245 mi^2]). Conversely, the reciprocal of the inclusion function has units of area per sample point, and thus gives the amount of area that each sample point represents. A design-based analysis of a sample with varying inclusion density function needs to account for the different weight attached to each observation (i.e., the different amount of area represented by each point). From the above discussion, the proper way to give each observation its correct weight is to multiply each observation by the reciprocal of the inclusion density function at the observation site. For example, if we have samples at locations s_1, s_2, \dots, s_n , with corresponding observations z_1, z_2, \dots, z_n , then the estimate of the mean value of z is

$$\hat{\mu}_z = \sum_{i=1}^n \frac{z(s_i)}{\pi(s_i)} / \sum_{i=1}^n \frac{1}{\pi(s_i)} = \frac{\sum_{i=1}^n z(s_i) w(s_i)}{\sum_{i=1}^n w(s_i)}$$

where $w(s_i) = 1/\pi(s_i)$. This is the essence of the Horvitz-Thompson Theorem (Horvitz and Thompson 1952, Cordy 1993), which provides general estimating formulae along with accompanying variance expressions.

In general, the variance of a quantity estimated from a probability sample depends not only on the inclusion density function, but also on the pairwise probabilities of any two points being

included. Again, the general estimating equations are given in Horvitz and Thompson (1952), Cordy (1993), or Stevens (1997). There is, however, a simplifying assumption that leads to a simple estimate of variance. The assumption is that the sample arose from independent drawings from the population. If the population has the characteristic that values measured on sites that are close together tend to be more similar than values measured on sites that are far apart, then that simplifying assumption leads to a conservative (i.e., larger than the true value) estimate of variance. To apply this assumption to estimate the variance of $\hat{\mu}_z$, we first calculate

the quantities $d(s_i) = [z(s_i) - \hat{\mu}_z]w(s_i)$. Then $Var(\hat{\mu}_z) = \frac{nS^2(d)}{[\sum_{i=1}^n w(s_i)]^2}$, where $S^2(d)$ is the

familiar estimator of the sample variance of the d_i from a simple random sample, that is,

$$S^2(d) = \frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1} = \frac{\sum_{i=1}^n d_i^2}{n-1} \quad (\text{since } \bar{d} = \frac{\sum_{i=1}^n d_i}{n} \equiv 0). \quad \text{As noted above, this method}$$

will generally underestimate the true precision of the estimate. The degree of understatement depends on the smoothness of the sampled surface, that is how smoothly $z(s)$ changes over the population domain. If the surface is very rough in the sense that there is likely to be a little more correspondence between two adjacent points than between two widely separated points, then there should be little or no understatement. If the surface is very smooth, then the variance estimate could be too large by a factor of 2 or more (Stevens and Kincaid 1997).

The above discussion deals with estimating a population mean value and its variance. The same technique can be used to estimate the proportion of a population that meets some criteria or falls within some category. For example, we may be interested in the proportion of the Everglades covered by cattails, or the proportion of the Everglades with water concentration of TP less than x . To do this, we form a new response variable that takes on the value 1 if a sample site meets the criteria or is in the category, and 0 otherwise. We call this new response the *indicator variable* for the criteria or category. For the category {land cover = cattails}, the

indicator variable is $I_{cattails}(s_i) = \begin{cases} 1, & \text{if } s_i \text{ is covered with cattails} \\ 0, & \text{otherwise} \end{cases}$. The mean value of the indicator variable is the proportion we want, and we estimate it and its variance using the same

method as for any other mean. Thus, for example, $\hat{p}_{cattails} = \frac{\sum_{i=1}^n I_{cattails}(s_i)w(s_i)}{\sum_{i=1}^n w(s_i)}$ would give the estimated proportion of the sampled population in the land cover class “cattails”.

The indicator variable technique can be used to obtain an estimate of the entire population distribution via a function known as the cumulative distribution function (cdf). The cdf for a variable z , say $F_z(x)$, gives the proportion of the population with z value less than or equal to x . For example, if z is TP concentration in (units), then $F_{TP}(#)$ is the proportion of the population with TP concentration less than or equal to $#$ (units). The cdf of z is estimated by picking a set of levels x_1, x_2, \dots, x_k that span the range of z , and then estimating the mean values of the indicator

variables $I_{(z \leq x_j)}(s_i) = \begin{cases} 1, & \text{if } z(s_i) \leq x_j \\ 0, & \text{otherwise} \end{cases}$, so that $\hat{F}_z(x_j) = \frac{\sum_{i=1}^n I_{(z \leq x_j)}(s_i)w(s_i)}{\sum_{i=1}^n w(s_i)}$.

The concept of the indicator variable seems very simple, but it is in fact a very powerful tool for doing exploratory and comparative analyses of a complex probability sample. For example, the formulae above show how to compute the cdf for the entire population (e.g., the entire Everglades). But we can also use an indicator variable to estimate the cdf for a subset of our population. For example, suppose we want the cdf of TP concentration for only that portion of the Everglades covered with cattails. We use the “cattail” indicator variable in the cdf

estimator equation to get $\hat{F}_{TP|cattails}(x_j) = \frac{\sum_{i=1}^n I_{(TP \leq x_j)}(s_i)I_{cattails}(s_i)w(s_i)}{\sum_{i=1}^n I_{cattails}(s_i)w(s_i)}$. At any particular value

x_j , $\hat{F}_{TP|cattails}(x_j)$ would be the estimated proportion of the area covered by cattails with TP concentration less than or equal to x_j . We could also calculate the cdf for the area covered by

“sawgrass” using a “sawgrass” indicator variable, and compare the two cdf’s. One way to make a quick and informative visual comparison is to calculate the two subpopulation cdfs at the same levels of the x-variable (TP in the example), and then plot corresponding values against one another, producing a plot known as a Q-Q plot (Q for “quantile”). If the two distributions are approximately equal, then they should plot on roughly a 1:1 line.

Subpopulation analyses via indicator variables also can be used to examine associations between several variables. For example, we could split the range of TP concentration into “high,” “medium,” and “low,” and then for each corresponding subpopulation, calculate the cdf of Hg in fish tissue. We could further define several geographical areas, for example, north of Alligator Alley (AA), between Alligator Alley and Tamiami Trail (TT), and south of Tamiami Trail, and then compare Hg concentration for all nine subpopulations given by all combinations of TP concentration and geographic region. The complexity of the association one can examine, or the number of variables involved is limited only by the availability of data. In the above, we suggested comparing the cdfs. For adequate precision in the estimate, a cdf estimate should be based on 30 or more points. With fewer than 30 points in each subpopulation, it would be advisable to compare proportions or means. In this case, the subpopulation analysis could look very much like an analysis of variance (ANOVA).

2.2 Indicators

An array of indicators was selected to address the water management, habitat, eutrophication, and Hg issues under investigation that could be accommodated within the scope of the sampling design and the logistical limitations (Table 2.1 and 2.2). Surface water measurements included water depth, temperature, DO, pH, specific conductivity, redox (Eh), turbidity, TP, total nitrogen (TN), total organic carbon (TOC), total ionic sulfate (TSO_4^-), alkaline phosphatase activity (APA), chlorophyll *a* (canals only), and THg and MeHg (Table 2.3). Whenever canal sites indicated a vertical differential existed between surface and bottom measurements of temperature and DO, a vertical profile was made through the water column to define the stratification. In this report the term “soil” refers to those samples obtained from the

Table 2.1 Water and soil/sediment chemical measurements to be taken at each site with the general rationale for measurement.

| Indicator | Rationale for Inclusion |
|------------------------------|--|
| Water Quality | |
| DO | Anaerobic condition promotes methylation; impacted by eutrophication, water quality standard |
| Specific conductivity | Ionic strength influences metal toxicity, indicates water source/history |
| pH | MeHg often found in low pH systems |
| Turbidity | Particulate metal transport, reduces water clarity for primary production |
| TOC | Affects metal partitioning |
| APA | Sensitive indicator of eutrophication |
| THg | Total mercury pool, by media |
| MeHg | Biologically accumulated Hg species |
| TP | Indicator of eutrophication |
| TN | Indicator of eutrophication |
| TSO ₄ | Influences Hg methylation, microbial processes, eutrophication |
| Soil/Sediment Quality | |
| Bulk density | Measure of compaction |
| % Mineral content | Estimate of non-carbonaceous material |
| THg | part of available pool |
| MeHg | Biologically accumulated Hg species |
| TP | Indicator of eutrophication |
| pH | Low pH promotes methylation |
| Eh | Influences Hg methylation, phosphorus cycling |

Table 2.2 Physical and biotic measurements taken at each site with the general rationale for the measurement.

| Indicator | Rationale for Inclusion |
|---|---|
| Physical | |
| Site location | Spatial distribution |
| Resource class (e.g., canal, marsh) | Habitat classes of interest |
| Water depth | Affected by management; critical to ecosystem restoration; may influence anaerobic conditions |
| Temperature | Influences rates of chemical reactions and biological processes |
| Weather | Explanatory variable |
| Soil thickness | Potentially available Hg; pool affected by water management; important for marsh preservation; subsidence or accretion trends |
| Biological Quality | |
| Fish tissue contaminants (Selected species) | Hg exposure in aquatic organisms |
| Periphyton, chlorophyll <i>a</i> | Indicator of eutrophication |
| Vegetation | Indicator of resource class diversity and integrity |

Table 2.3 Analytical parameters for marsh and canal samples.

| Indicator Variable | Media | | | | |
|-----------------------------------|--------------------|-------------------|-----------------------|-------------------------|-------------------|
| | Water ¹ | Soil ² | Sediment ³ | Periphyton ² | Fish ⁴ |
| Temperature | ✓ | | | | |
| Turbidity | ✓ | | | | |
| Conductivity | ✓ | | | | |
| DO | ✓ | | | | |
| pH | ✓ | ✓ | ✓ | | |
| Eh (Rodox potential) | ✓ | ✓ | | | |
| TOC | ✓ | ✓ | | | |
| TN | ✓ | | | | |
| TP | ✓ | ✓ | ✓ | | |
| TSO ₄ | ✓ | ✓ | | | |
| APA | ✓ | | | | |
| THg | ✓ | ✓ | ✓ | ✓ | ✓ |
| MeHg | ✓ | ✓ | ✓ | ✓ | |
| Ethylmercury (EtHg) | ✓ | ✓ | ✓ | ✓ | |
| % Ash Free Dry Weight | | ✓ | ✓ | | |
| Bulk density | | ✓ | | | |
| Fish length | | | | | ✓ |
| Fish weight | | | | | ✓ |
| Fish sex | | | | | ✓ |
| Chlorophyll <i>a</i> ³ | ✓ | | | | |

¹ Marsh, canal and structure sampling

² Marsh sampling only

³ Canal sampling only

⁴ Canal and marsh sampling

marsh system while the term “sediment” is used to refer to the samples obtained from the canal substrate. The marsh soil measurements included soil thickness; type; pH; Eh; bulk density; percent organic matter; TSO₄; TP; THg; MeHg; and EtHg (Table 2.3). Sediments from the canals were analyzed for percent mineral content, TP, pH, THg, MeHg, and EtHg. THg and MeHg were measured in the discharge from the structures on a biweekly frequency.

Biological tissue samples included floating and soil periphyton in which THg, MeHg, and EtHg were analyzed. Fish tissue samples were exclusively eastern mosquitofish (*Gambusia holbrooki*). Individual whole fish were analyzed for THg and the length, weight, and sex of each fish was recorded to provide population level statistics. The eastern mosquitofish was selected as a key indicator organism for Hg bioaccumulation because it is a prey species for a number of top predators in the system. This indicator species compliments the Florida Game and Fresh Water Fish Commission (FGFWFC) long-term monitoring efforts on largemouth bass. Mosquitofish has the following advantages as a systemwide biological indicator organism: (1) ubiquitous across the system, occurring in both canal and marsh habitats; (2) short life span; (3) small home range; (4) biomagnifies Hg; (5) important in the aquatic food web; (6) omnivorous; (7) easily captured; and (8) minimal size versus Hg concentration relationships.

Plant community composition and presence or absence of cattails (*Typha spp.*) and Periphyton mats were used as qualitative indicators of marsh habitat (Table 2.2). The dominant and secondary plant communities occurring at each site were identified as well as the community sampled. In addition, the presence or absence of cattails and floating periphyton mats were made from a secondary review of two 35 mm photographs taken at different angles at each marsh sampling location. The presence of a single cattail or periphyton mat was enough to indicate presence as long as these indicators were visible in both photographs taken at each sampling point. Although no formal plot size or distance criteria were established for habitat indicators, plant species or plant types had to be identifiable in the photograph to be included in the qualitative assessment of marsh habitat.

2.3 Design Summary

The EPA Region 4 South Florida Ecosystem Assessment Project design was RTS probability sample of the canals and marshes of South Florida. The project was designed around the seven policy questions listed in Section 1.3, and focused on providing preliminary answers to assessment questions pertaining to magnitude, extent, and current conditions, and on providing a baseline for assessing future trends.

Because the sample was a probability sample, model-free inferences to the sampled population are possible. The sampled population was approximately 1,200 km (750 mi) of canals, which were sampled in September of 1993, May and September of 1994, and May of 1995, and approximately 7,800 km² of marsh that were sampled in April and September of 1995, and May and September of 1996. Design-based descriptions of this population regarding extent, magnitude, and current condition can be obtained without appeal to any additional statistical, biological, or mechanical model. In particular, a model of spatial or temporal correlation is not required. Such models may be built using the data resulting from this program, but estimates of magnitude, extent, and current condition are available from a strictly design-based approach.

The canals and marshes were sampled independently. The canal sample was equiprobable, as was the marsh sample except for the BCNP. The BCNP was sampled at one third the intensity of the remainder of the marsh. Thus, the only instance when the probability weights need to be accounted for in analyzing the data is when BCNP data is mixed with other marsh data (e.g., if a median value for the entire marsh were being estimated). Furthermore, the RTS design capitalizes on any spatial pattern that exists in the response on a scale comparable to the grid spacing in such a way as to give a more precise result than would ordinary random sampling. These two facts, the equiprobable sampling and the increased precision of the RST, mean that standard statistical analyses will yield unbiased estimates of population characteristics, and conservative estimates of precision.

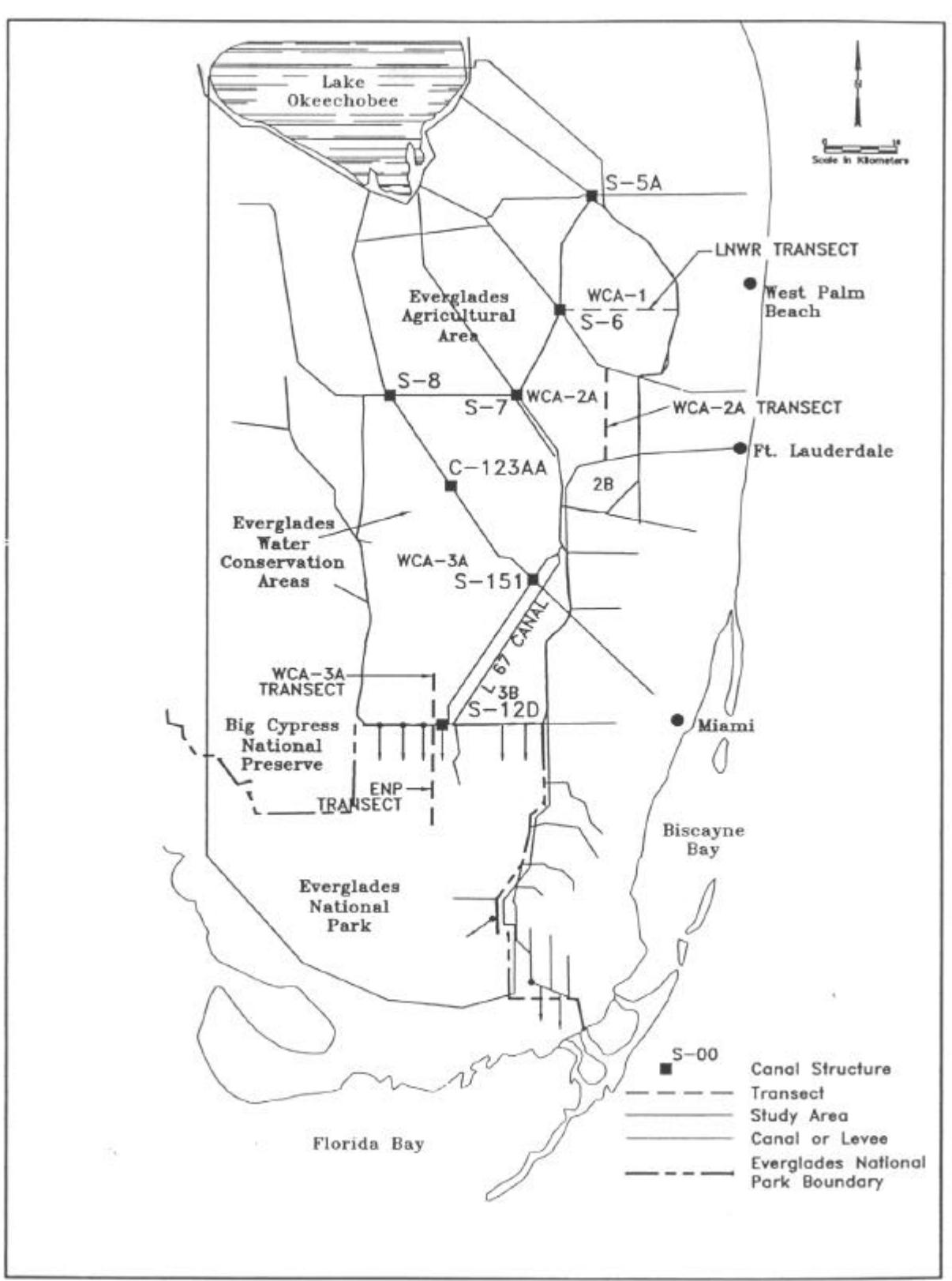


Figure 2.1 Location of four marsh transects sampled in April 1994 and canal water control structures sampled on a biweekly basis from February 1994 through February 1997.

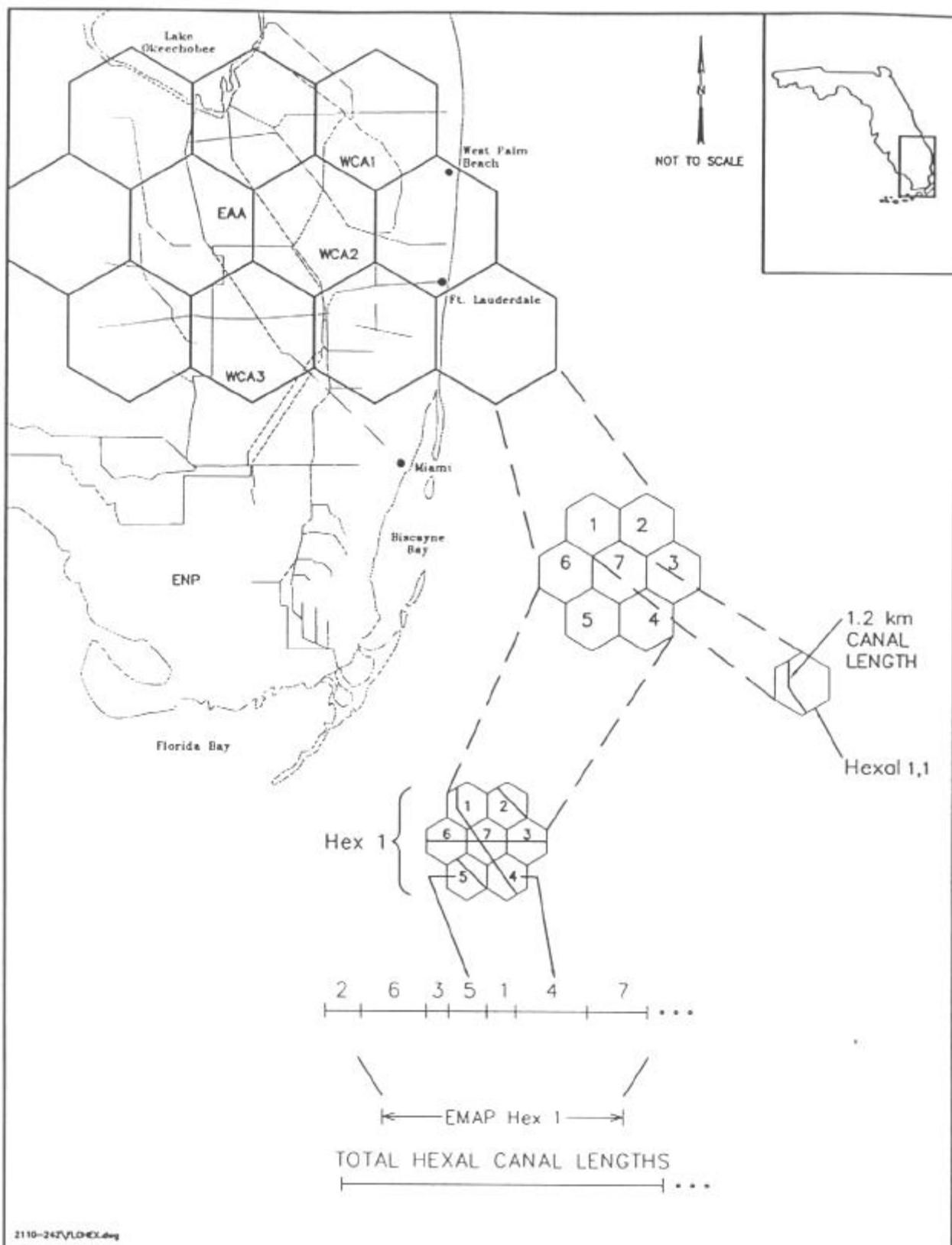


Figure 2.2 General schematic for clipping canal segments from the individual hexals and then randomly arranging them in a linear order so a systematic sample of 50 sites/cycle could be selected to sample.

CANAL SAMPLING LOCATIONS

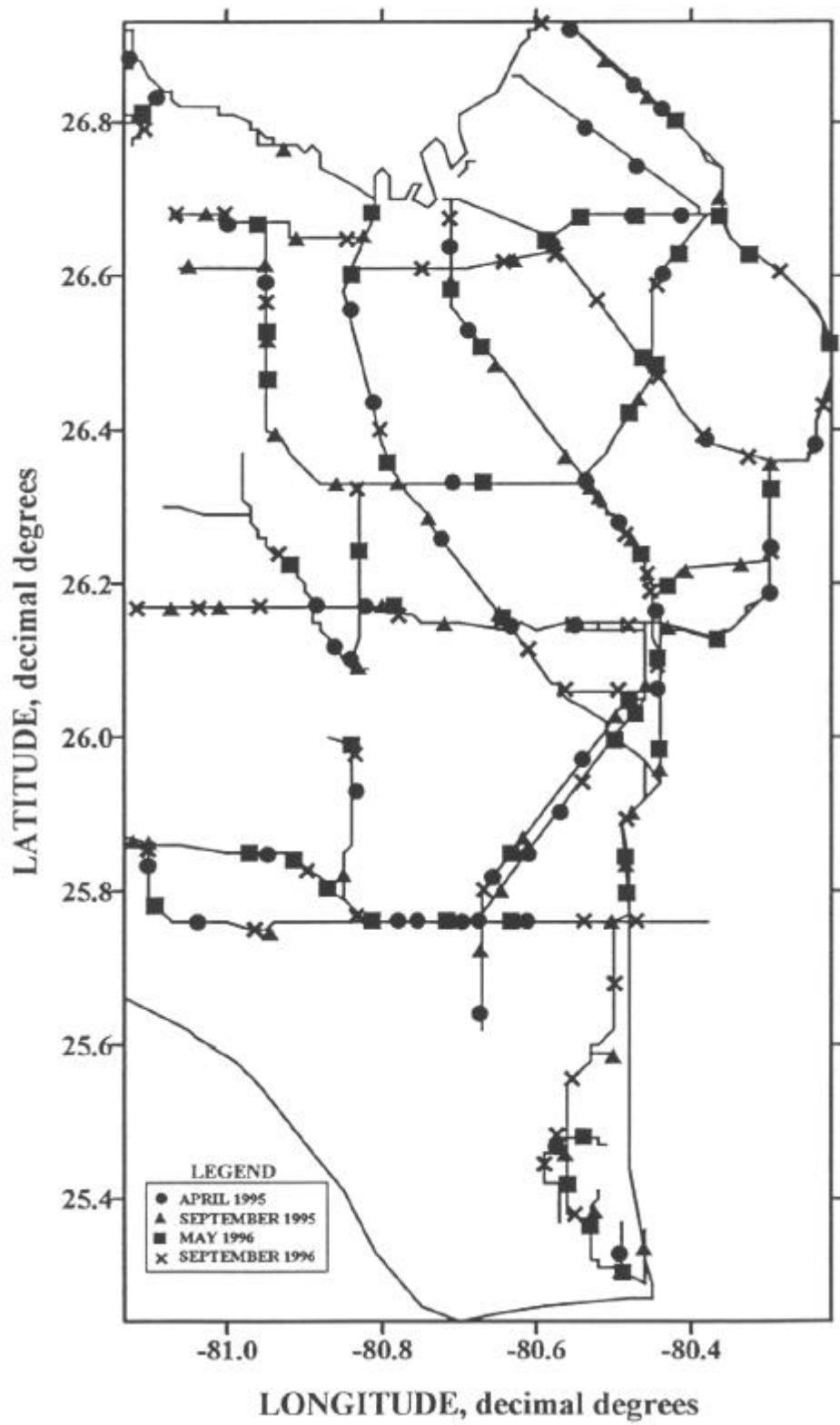


Figure 2.3 200 sampling sites are located on over 1,200 km of canals.

MARSH SAMPLING LOCATIONS

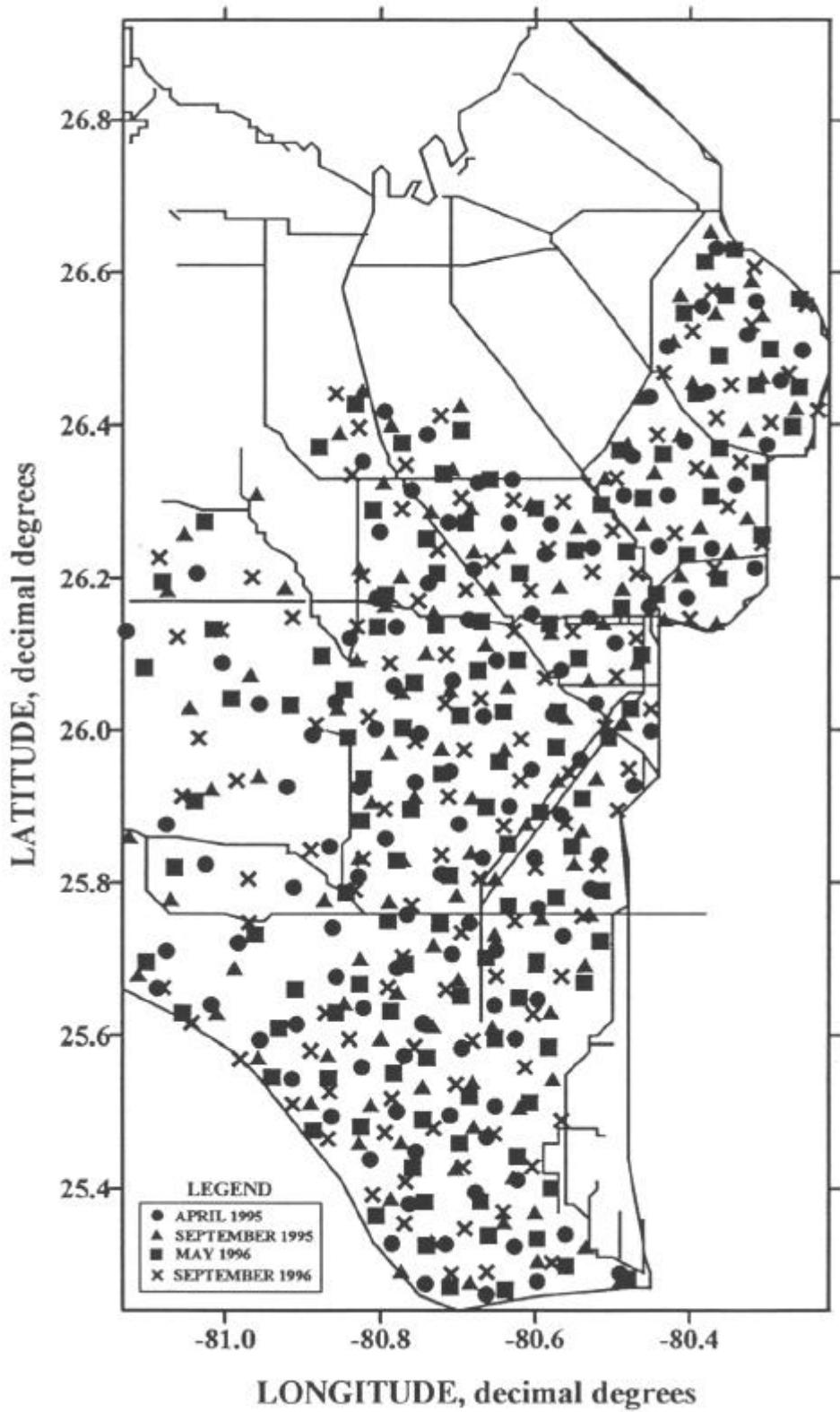


Figure 2.4 500 sampling sites are located on over 7,800 km² of marsh.