United States
Environmental Protection
Agency

Comparison of Random and Systematic Site Selection for Assessing Attainment of Aquatic Life Uses in Segments of the Ohio River

RESEARCH AND DEVELOPMENT

EPA 600/R-06/089
September 2006
www.epa.gov

Comparison of Random and Systematic Site Selection for Assessing Attainment of Aquatic Life Uses in Segments of the Ohio River

Karen Blocksom[1], Erich Emery[2], and Jeff Thomas[2]

[1]U.S. Environmental Protection Agency
Ecological Exposure Research Division
National Exposure Research Laboratory
Cincinnati, OH 45268

[2]Ohio River Valley Water Sanitation Commission
5735 Kellogg Avenue
Cincinnati, OH 45228

U.S. Environmental Protection Agency
Office of Research and Development
Washington, DC 20460

NOTICE

The research described in this document has been funded in part by the U.S. Environmental Protection Agency. This report has been subjected to the Agency's peer and administrative review process and has been approved for publication as an EPA document.

Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

EXECUTIVE SUMMARY

Introduction

The Clean Water Act (CWA) section 305(b) requires that states report biennially on the water quality standards (WQS) attainment status of all waters. The United States Environmental Protection Agency (USEPA) has recognized that biological assessment data are very valuable and should serve as a core indicator in determining aquatic life use attainment status (USEPA 2002). However, great (or large-floodplain) rivers, such as the Ohio River, are one type of aquatic resource for which biological assessment data typically have been deficient, primarily due to difficulty in sampling (Emery et al. 2003). The Ohio River Valley Water Sanitation Commission (ORSANCO), a compact of eight states and the federal government is responsible for assessment of the Ohio River. Recognizing the abundance of stresses on this system and the lack of biological data for the river, ORSANCO developed a bioassessment program several years ago. In 1990, ORSANCO began a Long-Term Intensive Survey (LTIS) of the Ohio River to provide high sample density (one sample every 3.2-6.4 km) from selected reach segments of the river. Beginning with this effort, a fish-based index of biotic integrity (IBI) for the Ohio River was developed (Emery et al. 2003). Although the LTIS approach provided valuable data for the development of a fish IBI and biocriteria, this approach was very resource-intensive and time-consuming and was not feasible for routine sampling of the entire river. To support the development of a cost-effective program for routine monitoring, the minimal amount of effort required to produce an adequate assessment needed to be determined.

Methods

A probability-based sampling design, in which sites are randomly selected from the population of possible sites, is an effective way to reduce effort and still collect data representative of an entire resource. By simulating a probability design using the intensive survey data already collected, the number of sites required to adequately represent the condition of a pool was determined. Two approaches were used to analyze the simulation data, and both assume that the intensive survey data provide the most accurate representation of condition available. The first approach was to determine the minimum number of sites that provided a similar condition estimate to that of the full set of sites, indicating some stability in the estimate. This was achieved by calculating variables for various subsets of sites and identifying the number of sites at which distributions of variables no longer differed from those of the full set of sites. Simultaneously, the number of sites required to obtain an estimate of condition with a specific level of precision was determined. A second and more commonly used approach to determining the adequacy of sampling was based on the relative proportion of species collected. The number of 500 m reaches required to collect

80-90% of the observed taxa in the pool was calculated as a way to determine the number of sites at which a sufficient level of effort has been expended to estimate biological condition. By examining the results of both approaches for this study, a probability sampling design was developed that will still be rigorous and will provide known confidence around any estimate of condition for Ohio River navigational pools while reducing the level of effort (i.e., number of sites sampled) in the field.

Results and Recommendations

This research indicated that 15 sites (compared to 20-32 sites sampled in the LTIS) may be adequate to draw conclusions about the overall condition of a navigational pool. However, in some cases, additional sites may need to be sampled to achieve the desired level of precision around the condition assessment. Because there are a large number of pools in the Ohio River, an approach that allows ORSANCO to sample and assess more pools each year will result in a more robust assessment of the river for the 305(b) report. A suggested approach is to sample all of the navigational pools of the Ohio River over a 5-year period. In each pool, an initial sampling of 15 sites is carried out, and additional sampling is completed only if required to make a definitive assessment of the pool. This approach would limit the resources required to assess an individual pool, such that additional effort would only be required in those pools that are of more marginal condition. In addition, this approach will help ORSANCO to identify those pools in which biological condition is most impacted and to prioritize any mitigation or restoration efforts. Additional sampling of individual sites may be required to determine causes of impairment within pools but may be guided by the data acquired through the random sampling design.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

APPENDIX FIGURES

ACKNOWLEDGMENTS

We thank Mark Barath, U.S. EPA Region III in Philadelphia, PA, and Tony Olsen and Dave Bolgrien, U.S. EPA Office of Research and Development in Corvallis, OR, and Duluth, MN, respectively, for comments on an earlier draft of this report. The data used in this research could not have been collected without the tireless efforts and skill of ORSANCO field crews.

1. INTRODUCTION

The Clean Water Act (CWA) section 305(b) requires that states report biennially on the water quality standards (WQS) attainment status of all waters. Currently, an Integrated Report is prepared to fulfill requirements related to sections 303(d), 305(b), and 314 of the Clean Water Act. The United States Environmental Protection Agency (USEPA) has recognized that biological assessment data are very valuable and should serve as a core indicator in determining aquatic life use attainment status (USEPA 2002). Biological data can reflect overall ecological condition (i.e., chemical, physical, and biological), and they provide a spatially and temporally integrated measure of the aggregate effect of stressors and prevailing environmental conditions. USEPA suggests that biological criteria, or biocriteria, in WQS are most effective as numerical values, which set ranges of indicators representing acceptable conditions for attainment of the Aquatic Life Use designation of a water body (USEPA 2002). In practice, however, state WQS often include only narrative biological criteria, although numeric thresholds of biological indicators are often used to interpret the narrative criteria. In any case, the ability to report on the condition of all waters over which a state has jurisdiction is often limited because of inadequate assessment of certain types of water bodies.

Great (or large-floodplain) rivers are one type of resource for which biological assessment data typically have been deficient, primarily due to difficulty in sampling (Emery et al. 2003). However, the data that are available suggest that, despite reductions in chemical and organic pollution, the biological condition of running waters of the United States, including large and great rivers, has continued to decline since the inception of the Clean Water Act (Karr and Chu 1999). Great rivers are distinctive in that they are few in number but comprise a large and highly visible component of lotic resources in terms of volume. They are disproportionately degraded by many human actions that “pollute” them, including water withdrawals, overharvesting of fish, impoundments, and changes in the landscape that can affect natural processes (Emery et al. 2003, Karr and Chu 1999). Recognizing such stresses on the system and the lack of biological data, a bioassessment program was developed for the Ohio River. Assessment of the Ohio River is the responsibility of the Ohio River Valley Water Sanitation Commission (ORSANCO), a compact of eight states and the federal government. In 1990, ORSANCO began a Long-Term Intensive Survey (LTIS) of the Ohio River to provide high sample density (one sample every 3.2-6.4 km) from selected reach segments of the river. From these 741 fish assemblage samples, efforts began in 1991 to develop numeric biocriteria for the Ohio River (Simon and Emery 1995). A fish-based index of biotic integrity (IBI) that incorporates river kilometer (rkm) has since been developed for the Ohio River (Emery et al. 2003). Recently, biocriteria have been refined to account for habitat type and sampling date as well.

Although the intensive survey approach provided valuable data for the development of a fish IBI and biocriteria, this approach is very resource- and time-consuming and is not feasible for sampling the entire river on a regular basis. To support the development of a cost-effective program for routine monitoring, the minimal amount of effort required to produce an adequate assessment needed to be determined. A probability-based sampling design, in which sites are randomly selected from the population of possible sites, is an effective way to reduce effort and still collect data representative of an entire resource. By simulating a probability design using the intensive survey data already collected, the number of sites required to adequately represent the condition of a pool can be determined. Although the LTIS samples do not represent a complete sampling of a pool because they only consist of a 500 m reach every few km, the systematic nature of the sampling effort allows for the assumption that the data adequately characterize actual distributions in a given pool. Two approaches were used to analyze the simulation data, and both assume that the intensive survey data provide the most accurate representation of condition available.

The first approach was to determine the minimum number of sites that provided a similar condition estimate to that of the full set of sites, indicating some stability in the estimate. This was achieved by calculating variables for various subsets of sites and identifying the number of sites at which distributions of variables no longer differed from those of the full set of sites. Simultaneously, the number of sites required to obtain an estimate of condition with a specific level of precision was determined.

A second and more commonly used approach to determining the adequacy of sampling was based on the relative proportion of species collected. Previous studies have examined the level of effort necessary to produce a representative sample based on the proportion of the available species captured by that method in wadeable streams (Dauwalter and Pert 2003, Reynolds et al. 2003, Lyons 1992) and nonwadeable rivers (Meador 2005, Hughes et al. 2002, Lyons et al. 2001). These studies have been focused on determination of the appropriate continuous sampling distance at a single location, which may represent an entire stream or river. However, ORSANCO already had set the electrofishing distance at 500 m, based on work by Simon and Sanders (1999), who determined that 500 m was a sufficient distance to characterize biological integrity. For this study, the number of 500 m reaches required to collect 80-90% of the observed taxa in the pool was calculated as a way to determine the number of sites at which a sufficient level of effort has been expended to estimate biological condition. By examining the results of both approaches for this study, a probability sampling design was developed that will still be rigorous and will provide known confidence around any estimate of condition for Ohio River navigational pools while reducing the level of effort (i.e., number of sites sampled) exerted in the field.

2. METHODS

2.1. Study Area

The Ohio River begins in Pittsburgh, Pennsylvania, at the confluence of the Monongahela and Allegheny rivers (rkm 0) and flows southwesterly for approximately 1579 km through six states to the confluence with the Mississippi River (Figure 1). Currently, there are 18 high-lift and 2 low-head dams on the Ohio River, each providing a minimum of 2.75 m depth for commercial navigation. These dams define major pools on the river and significantly limit fish populations to a specific pool. For this reason, and because any watershed management actions would likely be carried out at this scale, the pool was viewed as the appropriate level of assessment in the Ohio River for this study. In addition, probability designs are not intended for assessment of individual sites but for larger areas. Thus, the pool is the most well-defined unit of assessment using a probability sampling design in the Ohio River.

From the LTIS dataset, data were selected from five pools, each sampled in one of four different years. Table 1 provides information on the pools and the samples collected in them, and Table 2 characterizes factors that may influence water quality in these pools. These five pools were selected because they represent the most intensive sampling, with a sample frequency of more than one site every 4 km (2.5 mi). Variations in condition were expected both across pools and among years, although the goal of this study was to identify patterns with sample size and not to specifically evaluate the condition of individual pools.

Table 1. The five pools used for this study, along with sampling information associated with the LTIS and selected background information.

Pool	Year of sampling	Dam location (river km)	Length (km)	Average width (m)	Number of sites sampled
R.C. Byrd	2002	449.3	67.1	352	24
Hannibal	1996	203.4	58.3	345	20
McAlpine	1997	976.5	118.3	622	32
Newburgh	1994	1249.0	89.2	755	25
Smithland	1996	1478.2	116.7	1255	30

Table 2. Selected potential influences on water quality (from ORSANCO 1994) in study pools.

Pool	Tributaries (drainage >200 mi^2 (518 km^2))	No. permitted discharges	General influences
R.C. Byrd	2	30	Intersected by heavily industrialized Kanawha River
Hannibal	2	52	Many small-moderate sized towns, industrial sites and associated barge traffic
McAlpine	2	44	Includes city of Louisville, Kentucky
Newburgh	2	53	Smaller industry, small towns, one moderate-sized city, minimal barge traffic or heavy industry
Smithland	3	10	Intersected by agriculture-influenced Wabash River, few towns, no heavy industry

2.2. Sampling Protocols

2.2.1. Fish

2.2.1.1. Electrofishing

All sampling was conducted during the low-flow, stable conditions of July through October and during water conditions meeting sampling criteria (i.e., minimum secchi depths of 38 cm; water levels within 61 cm of normal-flat-pool). Procedures for electrofishing followed that described by Emery et al. (2003). At each site, a 500 m reach was electrofished with a 5.5 m jon boat outfitted with an onboard generator. Electrofishing was conducted at night, as this is the established protocol used by ORSANCO and has been documented to provide for a more representative sample of the resident fauna in deeper rivers when compared to day electrofishing (Sanders 1992). The onboard generator supplied AC power to 150-W floodlights on the bow of the boat, and to a Smith-Root Type VI-A alternator-pulsator used to convert the AC generator output to DC and then regulate the output for electrofishing. A single stainless steel ball suspended from a bow-mounted retractable aluminum boom served as the anode, with the aluminum boat hull serving as the cathode.

Each site was electrofished proceeding downstream along the shoreline at a speed equal to, or slightly greater than the prevailing current velocity. The electrofishing time at each site generally ranged from 1800 to 5000 seconds depending on the current velocity, available cover, and the number of fish encountered. Efforts were made to capture every fish sighted by the crew.

2.2.1.2. Fish Sample Processing

Upon capture, fish were placed in an aerated, recirculating on-board live well for processing. Particular care was used in handling species of special concern. The majority of captured fish were identified to species, examined for external anomalies, weighed, measured for total length, and released in the field. Those requiring laboratory identification were preserved in buffered 10% formalin and later identified using regional ichthyological keys (e.g., Fishes of Ohio (Trautman 1981), Fishes of Missouri (Pflieger 1997), and Fishes of Tennessee (Etnier and Starnes 1993)). Fish measuring less than 20 mm in length (e.g., larval fish) were not recorded as they are difficult to identify accurately and offer data of questionable value to an assemblage assessment (Angermeier and Karr 1986).

The occurrence of external DELT (deformities, eroded fins and body parts, lesions, and tumors) anomalies was recorded following procedures outlined by Ohio EPA (1989) and refined by Sanders et al. (1999). The frequency of DELT anomalies has been shown to be a good indication of stress caused by chronic agents, intermittent stresses, and chemically contaminated sediments. As a result, it is a commonly used metric for assessment of rivers throughout the United States (Emery et al. 2003).

2.2.2. Habitat

Substrate information used in data analysis was collected in 2000 for sites sampled prior to that year, with the assumption that the basic characteristics of the habitat at a site have not changed over time. For sites sampled in 2000 and beyond, habitat was sampled within 4 months of fish sampling (i.e., during the fish sampling index period). At each site, the shoreline of each 500 m sampling zone was divided into five 100 m segments, creating 6 points of reference for the zone along the shoreline (i.e., 0 m, 100 m, 200 m, 300 m, 400 m, and 500 m). At each interval, a 6 m copper pole was used to characterize the substrate at 11 points. The first measurement was taken at the shoreline with subsequent measurements at 3 m intervals towards mid-channel (total distance = 30 m). This resulted in a total of 66 point measurements within each 500 m fishing zone. Substrate was recorded as boulder, cobble, gravel, sand, fines, hardpan, or as a combination of these substrate types and used to estimate the percentage of each sediment-type within the 500 m sample area. Habitat data for some sites was unavailable.

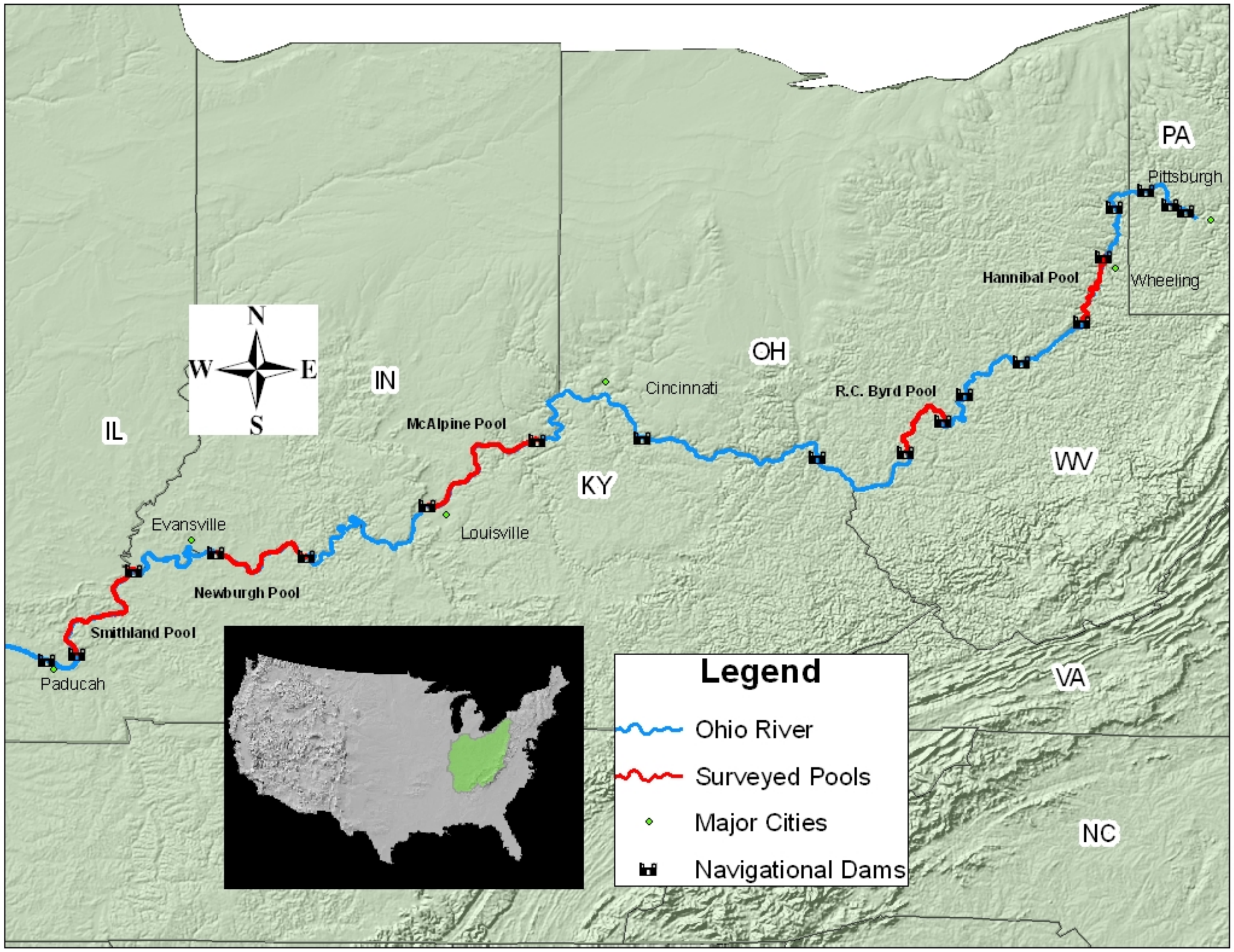

Figure 1. Locations of intensive survey pools used in this study (highlighted in red) along the Ohio River.

2.3. Data Analysis

There were three main objectives of the analysis. The first was to use a multimetric index of fish assemblage condition and the raw values of its component metrics to compare the intensive survey sampling design with subsets of those sites. The second was to compare pool-level condition estimates and precision using all of the sampling sites and subsets of sites. For this portion of the analysis, the condition estimate represented the proportion of sites in a pool for which the multimetric index score failed to meet a minimum threshold value. The final objective was to determine the number of sites in each pool at which 80% and 90% of fish species were captured, as an additional way to gauge representativeness of sampling in the pool. For all three objectives, only the sample from the first visit to each site was used in analyses.

2.3.1. IBI and Metrics

Analyses for this study involved the use of an existing multimetric index specifically developed for the Ohio River fish assemblage. The Ohio River Fish Index (ORFIn) was originally developed using LTIS data and consists of 13 metrics describing various characteristics of the fish assemblage (Emery et al. 2003) (Table 3). As part of this process, least disturbed (reference) sites were identified and defined as being at least 1 km upstream or downstream from navigational dams, at least 1.61 km downstream from any point source discharge, and at least 500 m from the mouth of any tributary (Emery et al. 2003).

Subsequent to the ORFIn's development, three habitat-based classes were developed that rely on substrate composition (ORSANCO, unpublished data): (A) cobble >= 14%; (B) cobble < 14% and sand < 70%; and (C) cobble <14% and sand >= 70%. Within each of these habitat classes, the 25th percentile ORFIn score among least disturbed sites was set as the minimum score required to be considered passing (unimpaired). This threshold for determining impairment was selected following the logic of Yoder and Rankin (1995), because the reference condition represents only least disturbed and not truly pristine conditions. The three habitat-specific thresholds for ORFIn scores were quite different from one another. For habitat A, this threshold was 39, and for habitat B, it was 33 (of a possible score of 65). For habitat C, considered to be sand flats, the condition assessment was determined to be dependent on sampling date, with higher ORFIn scores obtained later in the sampling index period. Thus, the threshold was adjusted for Julian day of the sample collection based on a 25th percentile regression using a quantile regression method (Koenker and Bassett 1978). This adjustment resulted in the following formula for the minimum adjusted score required to "Pass": (0.12*Julian Day - 2.4) (unpublished data).

For this study, each site was evaluated with regard to impairment thresholds using the ORFIn. First, ORFIn metrics and index scores were calculated for each sample. Then, each site was classified into habitat A, B, or C using the criteria above. Each site was assessed as passing or failing (impaired) based on the ORFIn score, habitat classification, and Julian day (when relevant). For sites lacking habitat data, ORFIn scores were compared to thresholds for all 3 habitat types. If the score exceeded the minimum required to pass for all habitat types, the site condition was designated as 'PASS'. If the score fell below the threshold for all 3 habitats, its condition was designated as 'FAIL'. Those sites with mixed results were assessed as 'UNKNOWN' condition.

Table 3. The metrics included in the ORFIn and their expected response to stress.

Metric	Expected response to stress
Number of species	Decrease
Number of sucker species	Decrease
Number of centrarchid species	Decrease
Number of great-river species	Decrease
Number of intolerant species	Decrease
% Tolerant individuals	Increase
% Simple lithophilic individuals	Decrease
% Non-native individuals	Increase
% Detritivore individuals	Increase
% Invertivore individuals	Decrease
% Piscivore individuals	Decrease
Number of DELT anomalies	Increase
Catch per unit effort (CPUE)	Decrease

Table 4. Frequency of samples in each habitat class by pool.

Pool	A (cobble)	B (mixed)	C (sand)	Unavailable	Total
Byrd	4	19	1	0	24
Hannibal	3	7	2	8	20
McAlpine	8	13	5	6	32
Newburgh	0	10	13	2	25
Smithland	3	12	6	9	30

2.3.1.1. Simulations

Bootstrap methods were used to simulate the random selection of sites within pools. A bootstrap approach randomly draws sites with replacement from the original dataset and assumes that the set of sites in this original dataset adequately reflects the distribution of conditions in the pool. Sampling with replacement means that during each random draw, each site has an equal probability of being selected. For example, when a subset of five sites is selected with replacement, it is possible for any particular site to be drawn all five times. This methodology is appropriate for these data because, although pools were sampled intensively for LTIS, they were not sampled in their entirety. Thus, although the systematic nature of the sampling likely resulted in a set of samples representative of the distribution of conditions in the pool, all of the possible sample locations in the pool were not specifically included in the dataset. By creating a large number of sample sets and subsets (of sites) using bootstrapping, almost any population parameter and its variance can be estimated robustly from the original dataset (Chernick 1999).

In this analysis, the general process was as follows for each pool. A number of sites equal to that in the full set of LTIS sites (e.g., 24 sites in R.C. Byrd pool) were selected with replacement from the set of LTIS sites (*original set*) to create a bootstrap set of sites (*full set*). The first five of those were combined and used to estimate certain characteristics related to condition. Then the next five were added to the first five and the characteristics re-estimated for that set of ten. This process was repeated with five additional sites added each time until the full set of bootstrap sites was included. The full set of bootstrap sites was treated during each simulation as the true measure of condition in the pool, and the characteristics for each subset of sites was compared to this full set. The entire process was repeated 500 times to obtain 95% confidence intervals for each statistic at each sample size.

2.3.1.2. Condition Measures

The cumulative distribution function (CDF) of a variable indicates the probability that any particular observation of that variable will be at or below a specified value (Sokal and Rohlf 1995). By comparing CDFs between subsets of sites and the full set of bootstrap sites, the minimum sample size at which the distributions strongly overlap and provide very similar information can be determined. For each pool, empirical cumulative distribution functions (CDFs) were calculated for each metric and the ORFIn based on various subsets of sites and for the full set of bootstrap sites. For example, for R.C. Byrd pool, there were 24 sites in the full set of data, and CDFs were generated using 5, 10, 15, 20, and 24 sites for each of the 500 simulation runs. Calculation of the empirical CDF consisted of determining the following percentiles for each run on each set of sites: 0, 1, 5, 10, 25, 50, 75, 90, 95, 99, and 100. Then, the 2.5th, 50th, and 97.5th quantiles were determined for each percentile across all 500 runs. These values provide the median CDF and its 95% confidence intervals (CI) for plotting. For each metric or the ORFIN scores and each subset size, the median CDF and its 95% CI was plotted, along with the CDF for the full set of bootstrap sites for comparison. Visual comparisons of CDFs and confidence intervals were used to evaluate the degree of similarity between the full set of sites and subsets of different sizes.

In addition to line plots of CDFs, distributions of metrics or the ORFIn were directly compared between subsets and the full set of sites using a nonparametric Kolmogorov-Smirnov (K-S) two-sample test. The test was performed using PROC NPAR1WAY in SAS with Monte Carlo estimation of exact p-values (v. 9.1, SAS Institute, Cary, NC). The K-S test assumes that the two samples being compared are independent of one another, but this assumption was not met because one sample was simply a subset of the other. This would tend to lead to an actual Type I error rate much smaller than the nominal value. To improve the ability to detect differences between distributions (power) and to achieve an actual Type I error rate closer to 0.05, a p-value of 0.20 or less was used to identify significant differences between distributions, rather than the preferred significance level of 0.05. The number of runs out of 500 in which the distributions differed significantly ($p < 0.20$) was recorded for each metric and sample size.

The habitat-specific ORFIn thresholds described previously were used to evaluate each sample as passing or failing, and the pool-level condition was determined as the proportion of samples (river km) failing. For each run, an estimate of the proportion of river km failing (P_{fail}) in the pool was calculated, and over the 500 runs, 95% confidence intervals were estimated for that proportion. This was done for each of the subsets of sites, and the full set of sites from the intensive survey was viewed as providing the 'true' assessment of the pool. For these estimates of condition, the desired precision was within 12.5 percentage points, or a total confidence interval length of 25 percentage points. A proportion of 0.25 of river km failing has been set tentatively as the threshold for assessing an entire pool as failing (for placement on the 303(d) list of impaired waters). The selection of this criterion value stemmed from the decision to set the sample-level threshold for failing at the 25th percentile of the reference distribution. Based on this threshold, it would be possible for up to 25% of sites within a pool to be considered failing, even within a pool made up entirely of reference sites (as defined by ORSANCO). Whether this variation among reference sites is due to natural factors or actual differences in disturbance level, this approach would tend to limit only the Type I error rate (i.e., incorrectly assessing a pool as failing) when assessing a pool and not the Type II error rate (incorrectly assessing a pool as passing). As a way to limit both Type I and Type II errors, a definitive assessment of a pool is attained only if the confidence interval for the proportion failing does not include the threshold. This means that proportions either above or below the threshold by a small amount may not provide a sufficient assessment of the pool as a whole.

2.3.2. Species Richness

To estimate the number of samples required to obtain approximately 80% or 90% of the total species collected within a pool, EstimateS software was used to model the species richness as a function of the number of samples (Colwell 2005). For each simulation run in EstimateS, a sample was selected from the full set of samples with replacement. The bias-adjusted bootstrap estimate of species richness (S_{boot}) was calculated based on Smith and van Belle (1984) as:

$$S_{boot} = S_{obs} + \sum_{k=1}^{S_{obs}} (1 - p_k)^m$$

where S_{obs} is the number of species observed in the pooled samples, p_k is the proportion of samples containing species k, and m is the total number of samples. From this calculation, it can be shown that the value of S_{boot} is maximized when each species occurs in only one sample, and S_{boot} equals S_{obs} when each species occurs in all samples from a pool. After calculating S_{boot}, another sample was selected with replacement and combined with the first sample, and the species richness was again estimated. This process continued until a number of samples equivalent to the original set had been selected and combined, with species richness estimated each time using the bias-adjusted bootstrap approach. The entire procedure was repeated 1000 times, and the average value and standard deviation across runs was determined. From this information, the minimum sample size (i.e., number of sites) required to collect an average of 80% or 90% of observed taxa within a pool was determined.

3. RESULTS

3.1. IBI and Metrics

Those component metrics with very small ranges tend to reflect more rare components of the fish population and contribute very little variation to the overall ORFIn score. Thus, metrics with limited ranges of values were excluded from plotting and comparing distributions (Table 5). Richness metrics with a maximum of 5 or greater and percentage metrics with a maximum of 10 or greater were included in analyses. The DELT anomalies metric was used only if the maximum value was 5 or greater. As a result of these criteria, one or more metrics was excluded from each analysis, and the metric for great-river species richness was excluded from all analyses. The ORFIN score was used in analyses for all pools. The CDFs of ORFIn scores for subsets of pools were usually very similar to those for the full set of bootstrap sites (Figures 2-6). The confidence bounds for the CDFs narrowed predictably with increasing numbers of sites relative to the maximum number of sites sampled within a pool. Within a sample size of 15 sites for Byrd, Hannibal, and Newburgh pools and 20 sites for McAlpine and Smithland pools, the confidence bounds were closely aligned between the subset and the full set of sites. Plots of CDFs of metrics can be found in Appendix A.

Table 5. Metrics used in analyses for each pool are denoted by an (X) for that metric. Metrics were excluded if they had a maximum of less than 5 for richness metrics and DELT anomalies and less than 10 for percentage metrics.

Metric	R.C. Byrd Pool	Hannibal Pool	McAlpine Pool	Newburgh Pool	Smithland Pool
Number of species	X	X	X	X	X
Number of sucker species	X	X	X	X	
Number of centrarchid species	X		X	X	X
Number of great-river species					
Number of intolerant species		X	X		X
% Tolerant individuals		X			
% Simple lithophilic individuals	X	X	X	X	X
% Non-native individuals		X		X	
% Detritivore individuals	X	X	X	X	X
% Invertivore individuals	X	X	X	X	X
% Piscivore individuals	X	X	X	X	X
Number of DELT anomalies		X	X		X
Catch per unit effort (CPUE)	X	X	X	X	X

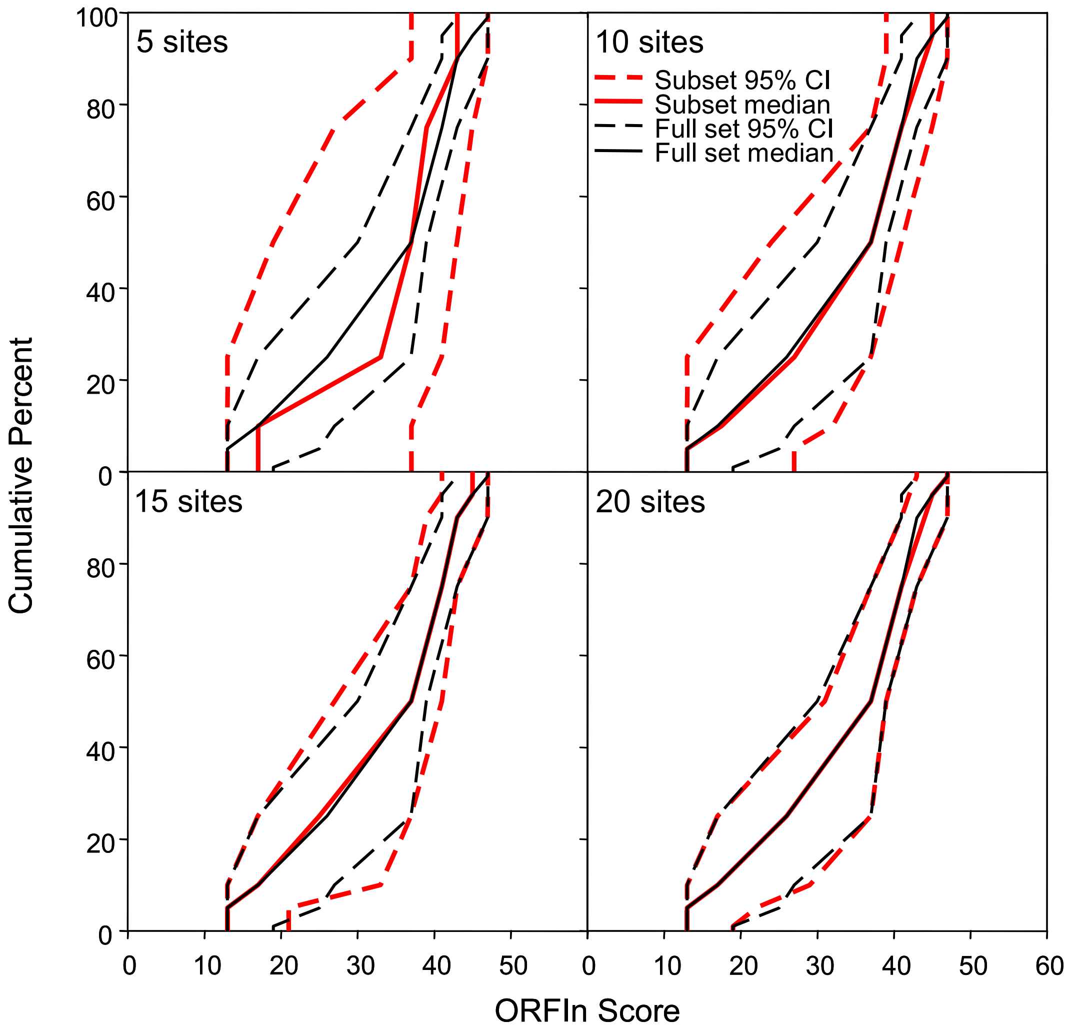

Figure 2. Cumulative distribution functions with 95% confidence bounds for ORFIn scores in R.C. Byrd pool based on bootstrapped subsets of sites (red lines) and the full set of sites (black lines, N=24).

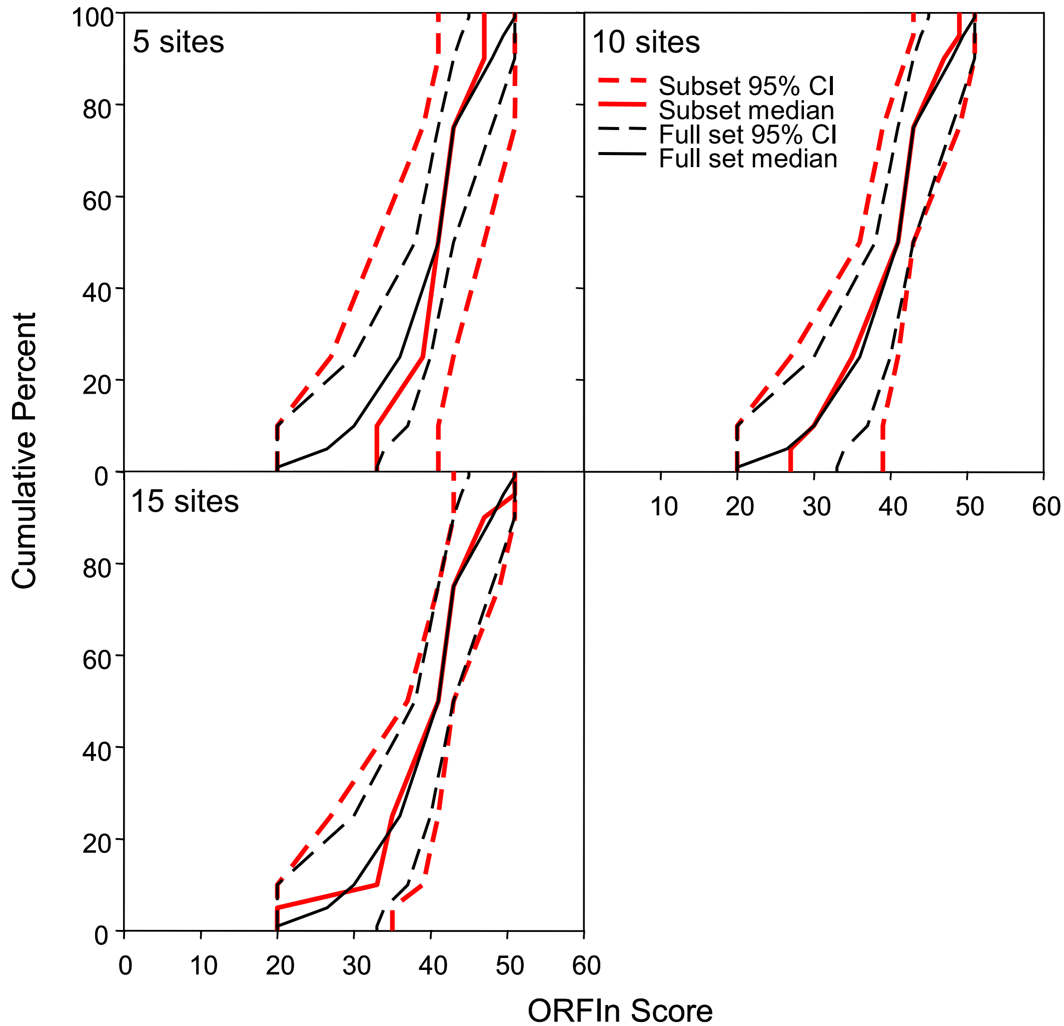

Figure 3. Cumulative distribution functions with 95% confidence bounds for ORFIn scores in Hannibal pool based on bootstrapped subsets of sites (red lines) and the full set of sites (black lines, N=20).

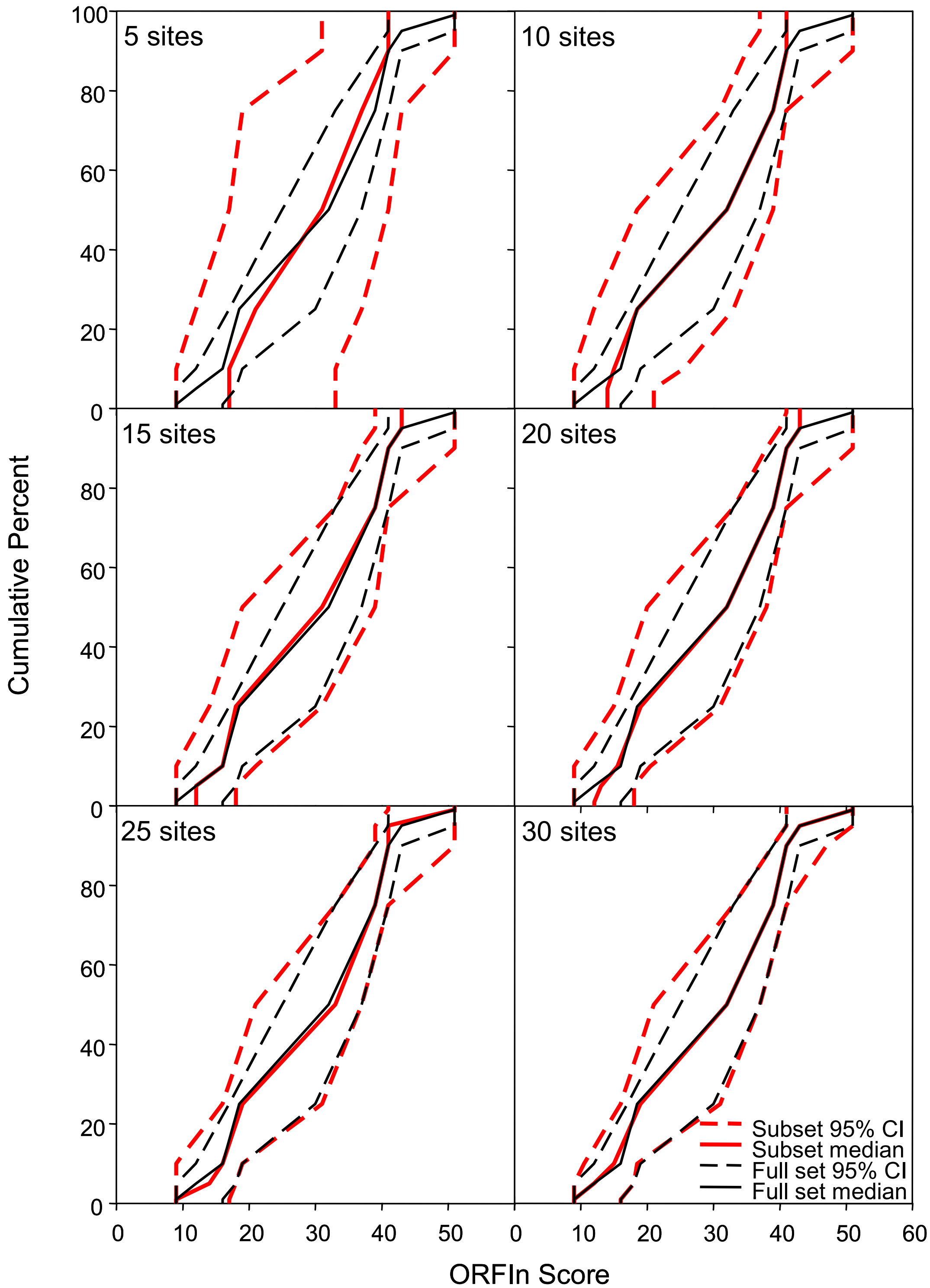

Figure 4. Cumulative distribution functions with 95% confidence bounds for ORFIn scores in McAlpine pool based on bootstrapped subsets of sites (red lines) and the full set of sites (black lines, N=32).

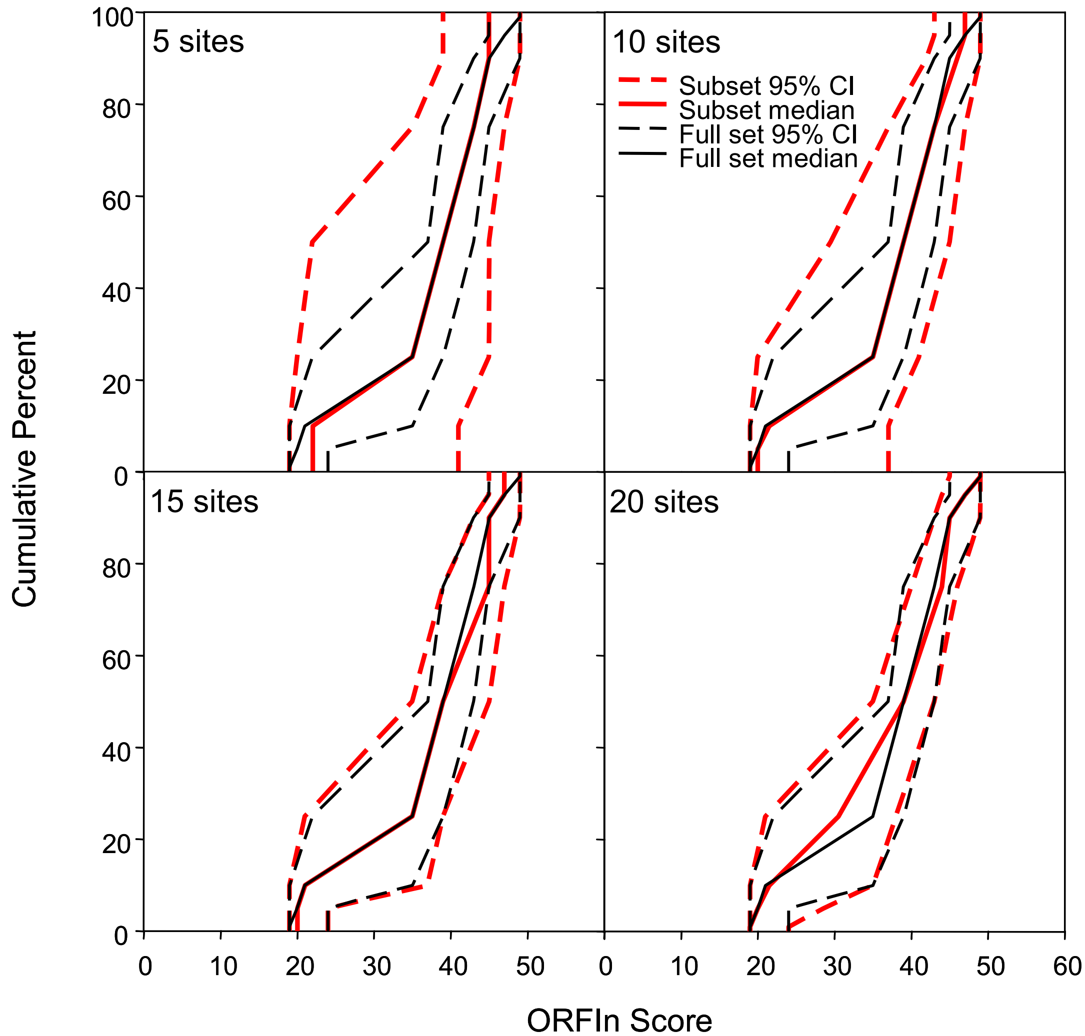

Figure 5. Cumulative distribution functions with 95% confidence bounds for ORFIn scores in Newburgh pool based on bootstrapped subsets of sites (red lines) and the full set of sites (black lines, N=25).

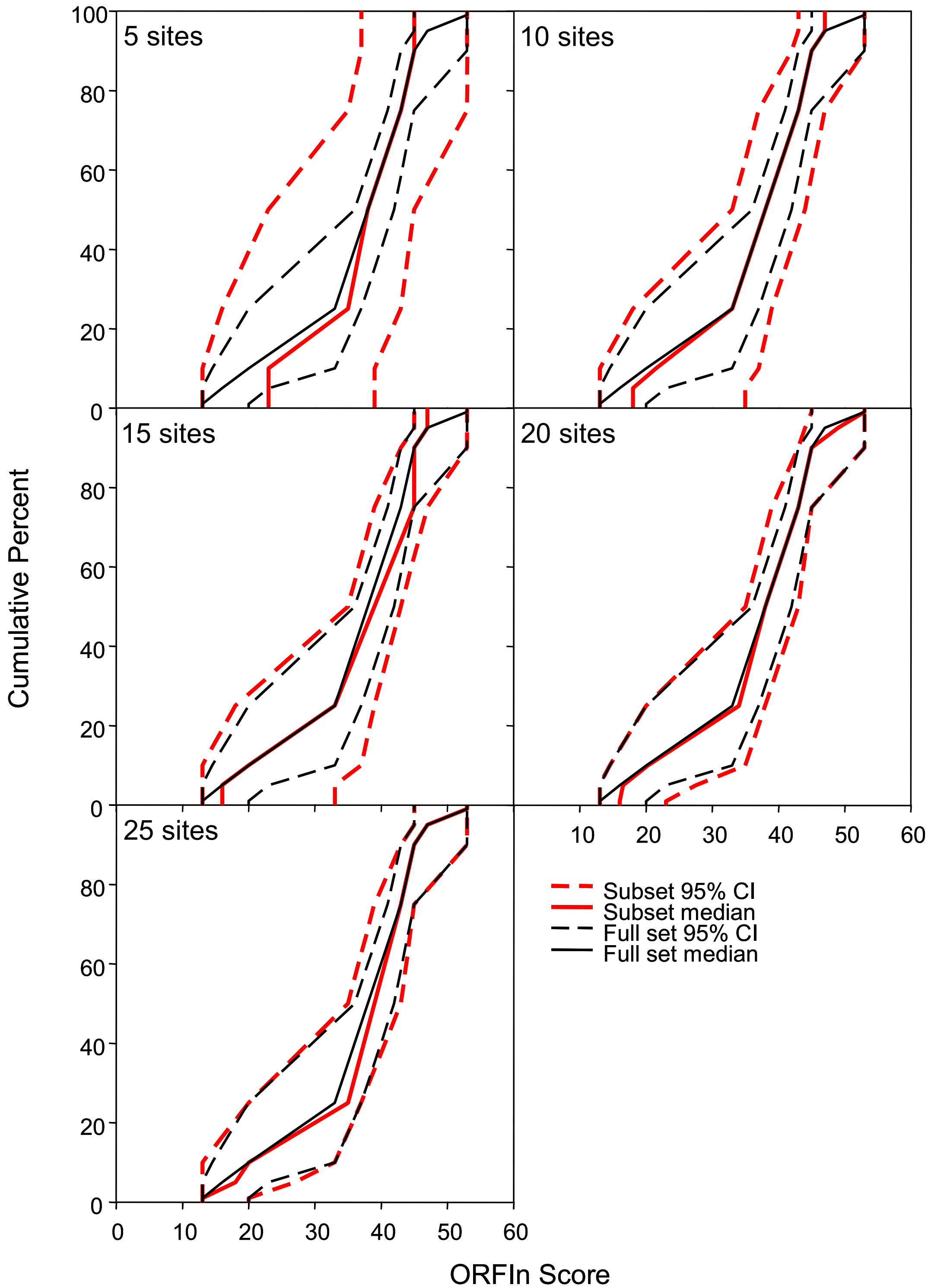

Figure 6. Cumulative distribution functions with 95% confidence bounds for ORFIn scores in Smithland pool based on bootstrapped subsets of sites (red lines) and the full set of sites (black lines, N=30).

Table 6. Number of occurrences in 500 simulations for which the K-S test of differences in distributions was significant ($p < 0.20$), shown for the ORFIn scores and metrics by pool and number of sites. If a metric was not evaluated for a particular pool (see Table 3), (NA) is listed for that cell.

Pool	No. sites	ORFIn	Species	Suckers	Centrarch.	Grt River	Intol.	% Toler.	% Simple Lith	% Non-native	% Detr.	% Invert.	% Pisc.	DELT	CPUE
R.C. Byrd	5	29	30	31	27	NA	NA	NA	24	NA	30	26	29	NA	32
	10	7	3	5	7	NA	NA	NA	2	NA	5	4	3	NA	3
	15	0	0	1	0	NA	NA	NA	0	NA	0	0	0	NA	0
	20	0	0	0	0	NA	NA	NA	0	NA	0	0	0	NA	0
Hannibal	5	8	13	18	NA	NA	13	13	8	15	13	9	11	0	14
	10	3	0	0	NA	NA	1	0	0	1	2	1	0	0	0
	15	0	0	0	NA	NA	0	0	0	0	0	0	0	0	0
McAlpine	5	30	47	39	44	NA	45	NA	37	31	32	50	46	0	48
	10	16	24	21	27	NA	16	NA	17	20	16	13	14	0	10
	15	5	0	3	6	NA	6	NA	5	9	1	2	4	0	4
	20	0	0	0	0	NA	0	NA	0	1	0	0	0	0	0
	25	0	0	0	0	NA	0	NA	0	0	0	0	0	0	0
	30	0	0	0	0	NA	0	NA	0	0	0	0	0	0	0
Newburgh	5	26	28	25	20	NA	NA	NA	25	21	29	25	27	NA	23
	10	11	4	8	9	NA	NA	NA	5	10	4	3	4	NA	5
	15	1	0	0	0	NA	NA	NA	0	0	0	0	0	NA	0
	20	0	0	0	0	NA	NA	NA	0	0	0	0	0	NA	0
Smithland	5	44	36	NA	38	NA	28	NA	33	NA	49	28	27	0	43
	10	9	3	NA	7	NA	12	NA	11	NA	12	7	13	0	12
	15	0	2	NA	1	NA	4	NA	0	NA	1	1	3	0	0
	20	0	0	NA	0	NA	0	NA	0	NA	0	0	0	0	0
	25	0	0	NA	0	NA	0	NA	0	NA	0	0	0	0	0

The K-S tests of differences between distributions of subsets of sites and the full set of sites for each pool were rarely significant, typically only for subsets of 5 or 10 sites (Table 6).

Confidence intervals around P_{fail} varied with the number of sites in the subset, but for 15 or more sites, the estimate itself was usually very close to that for the full set (Figures 7 and 8). The confidence interval length around P_{fail} tended to decrease with increasing sample size, but this pattern was not entirely consistent. Approximately 25 sites were required for Smithland and Newburgh pools and 20 sites for Hannibal pool to reach the desired 90% CI length of 25 points (0.25). This target CI length was never reached for the other two pools, even when the CI was based on bootstrapping with the full number of sites (Figure 7). Only in McAlpine pool did the confidence bounds not include the 0.25 threshold for at least some sample sizes.

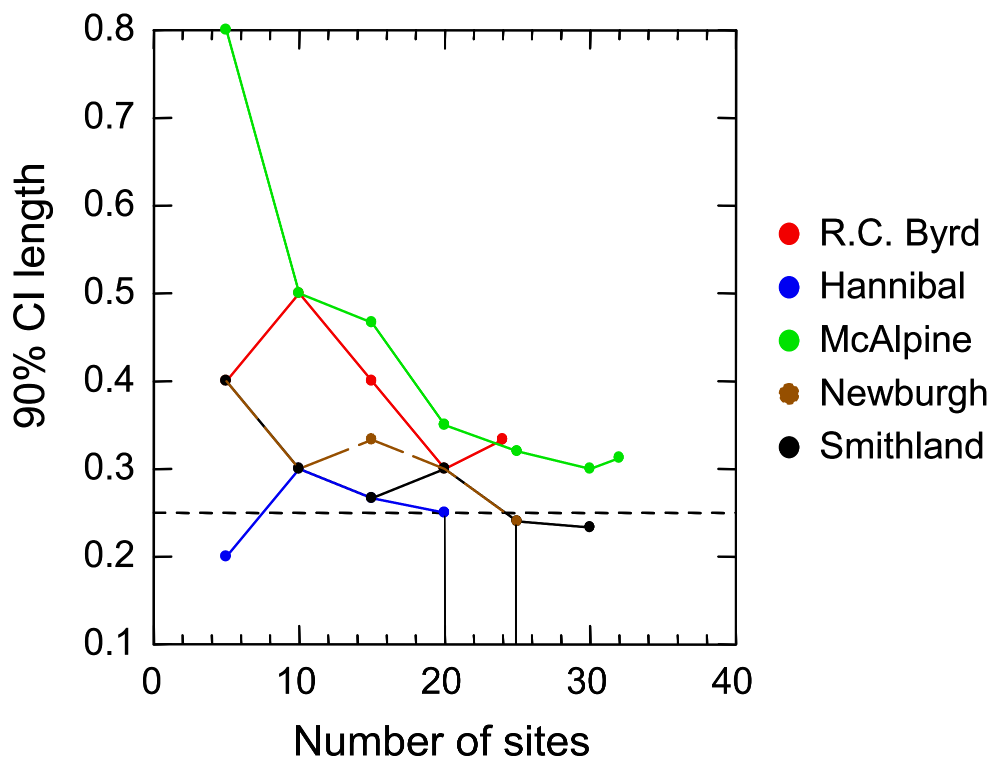

Figure 7. Confidence interval (CI) lengths for estimates of proportion of river km failing in each pool as a function of the number of sites sampled. The horizontal dotted line represents the desired maximum CI length of 0.25, and the vertical solid lines represent the approximate number of sites for which the desired CI length is achieved.

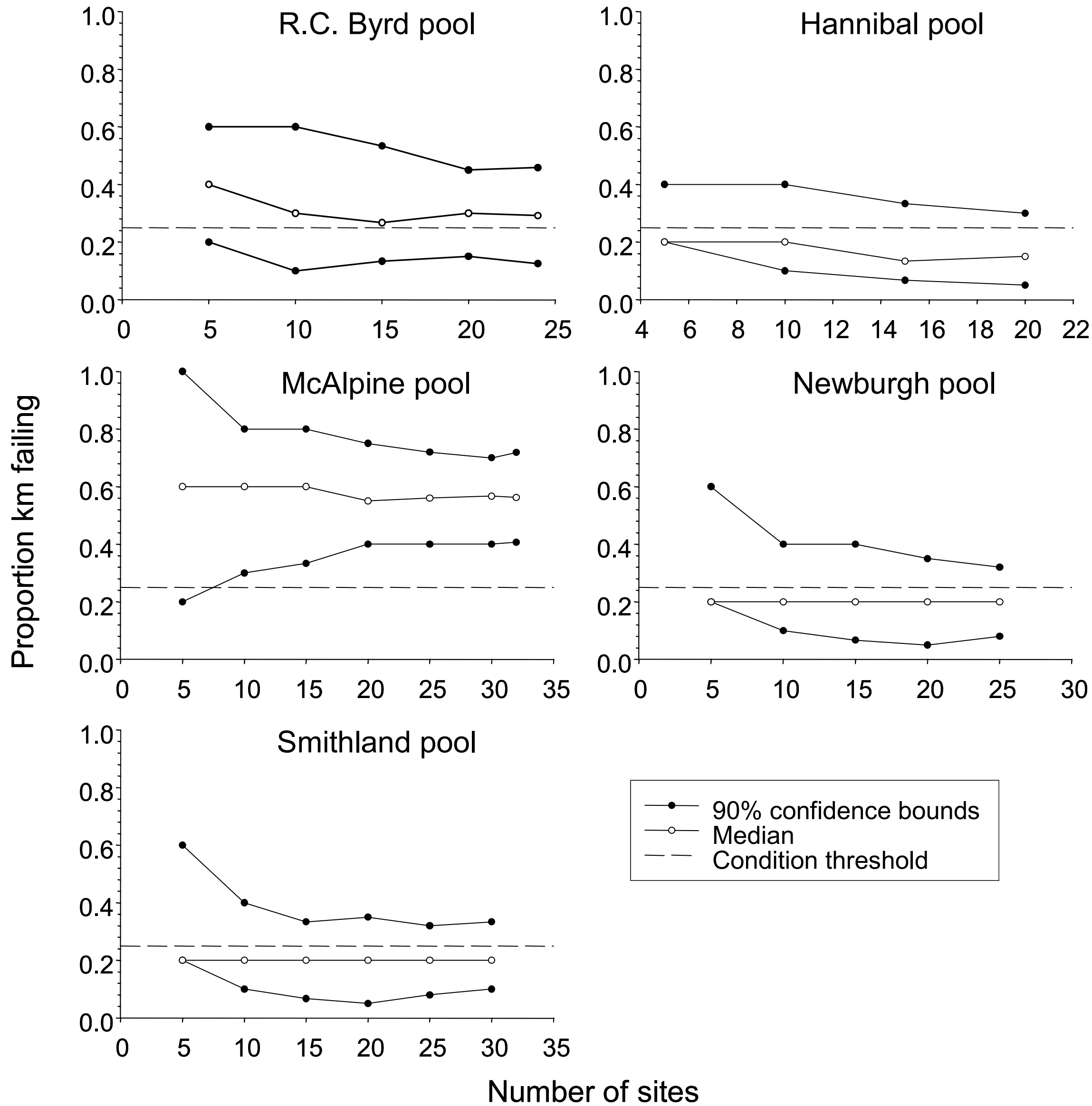

Figure 8. Estimates with 90% confidence bounds of the proportion of river km failing, based on bootstrap randomizations. The dotted line represents the threshold delineating failure of the entire pool (0.25).

3.2. Species Richness

On average, 80% of the total observed species richness in a pool were collected within 10 sites and 90% of species within approximately 15 sites (Figure 9). The two longest pools, McAlpine and Smithland, tended to require more samples than shorter pools to reach these benchmarks. Variability around species richness estimates only decreased slightly with increasing sample size.

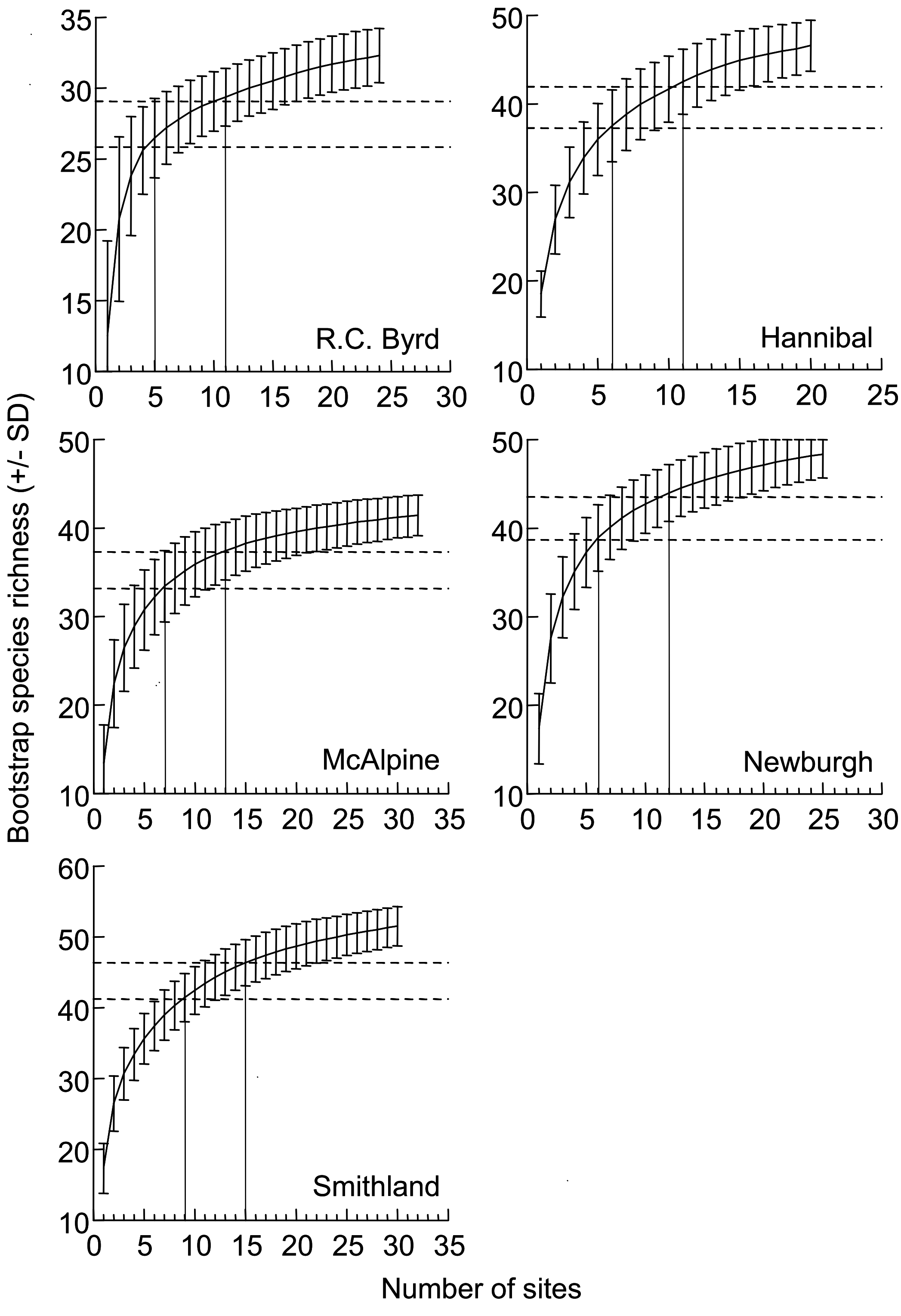

Figure 9 . Mean bootstrap species richness as a function of sample size, based on 1000 bootstrap randomizations of sites. The horizontal dotted lines represent 80% and 90% of the maximum number of species and the vertical lines represent the numbers of sites at which these are attained.

4. DISCUSSION

The use of random sampling is clearly advantageous for assessing the Ohio River. By using random sampling, many fewer samples per pool would be required for an adequate assessment, compared to the number of samples collected using the non-random intensive survey design. This reduced effort per pool would reduce the resources required to adequately assess a pool and could allow for more pools to be sampled each year. In addition, a random design brings with it certain statistical properties that result in known confidence levels around estimates of condition (Diaz-Ramos et al. 1996). The results of this study show that a random subset of samples may be able to represent condition in the pool adequately. The CDFs of both ORFIn scores and component metrics were very similar for all but the smallest subset sizes of 5 and 10 sites. The confidence intervals around the CDFs did vary with sample size, as expected, but generally were not excessively large even for sample sizes of only 10. The analysis of species richness also showed that 90% of species observed pool-wide could be captured within 10-15 samples. In addition, estimates of P_{fail} generally were very close to the estimate based on the full set for 15 sites or more.

In contrast, the confidence intervals around the estimates of P_{fail} for a given pool were generally larger than desired, sometimes even for the full set of sites. Under ideal circumstances for drawing sound conclusions about a pool's condition, the 90% confidence interval of the condition estimate would not include the threshold of 0.25. In such a situation, the size of the CI would be irrelevant, and one could assess the pool definitively as impaired or unimpaired with some known level of confidence. Based on the P_{fail} estimated from the original set of data, which can be viewed as the best representation of the true condition in the pools, only two of the five pools (i.e., Byrd and McAlpine) would be considered failing. However, the closer the actual P_{fail} is to the threshold, the more likely it is that the CI will include the threshold. For example, in R.C. Byrd pool, with a P_{fail} of 0.20, even the CI based on the full set of 24 sites included the threshold of 0.25. Thus, in cases with an estimate close to the threshold, more samples would be required to make definitive statements about the condition of the pool. On the other hand, in cases where P_{fail} was either very large or very small, a relatively small number of sites was sufficient to avoid inclusion of the 0.25 threshold in the CI (e.g., McAlpine pool, 10 sites).

Although some of the analyses may lead to different conclusions, the patterns that emerge from the data generally are consistent across pools, regardless of the level of variability within a pool. The five pools included in this study varied in potential impacts to water quality and in habitat diversity, but patterns in species richness and in variability associated with the ORFIn were similar across or seemingly unrelated to these differences. Habitat classified as a mixture of sand and cobble (habitat B) was common in all five pools, but the presence of cobble and sand habitats varied across pools (Table 4). For example, Newburgh pool had no cobble (habitat A) and mostly sandy (habitat C) sites, whereas Byrd and McAlpine pools tended to have more cobble than sand sites. Smithland had more sandy than cobble sites, and Hannibal was approximately evenly divided between cobble and sandy sites. The influences in each pool that might affect water quality (e.g., contribution by tributaries, discharges) also differed across pools (Table 2), with varying levels of industry and sizes of towns along the banks of each pool. Still, these differences do not seem to directly affect the patterns seen.

Overall, the ability of ORSANCO to report on the biological condition of the Ohio River should improve with the use of a random sampling design. Although the typical recommended minimum sample size for this type of design is usually around 50 sites (U.S. EPA Aquatic Resources Monitoring web site, http://www.epa.gov/nheerl/arm/surdesignfaqs.htm), a much smaller number of samples has the potential to provide enough information for assessment in some pools of the river. This study indicated that 15-20 sites may be adequate to draw conclusions about the overall condition of a navigational pool. However, the sufficiency of this smaller number of sites depends on the variation in condition of the fish assemblage within a given pool. The more consistent water and habitat quality are throughout a pool, the fewer samples that are needed to characterize the biological condition of that pool. Because there are a large number of pools in the Ohio River, an approach that allows ORSANCO to sample and assess more pools each year will result in a more robust assessment of the river for the Integrated Report.

5. RECOMMENDATIONS

In order to be consistent with surrounding states and maximize the data available for the Integrated Report, a five-year rotational sampling approach is highly desirable, as data up to 5 years old may be used in the report. An ideal approach would allow ORSANCO to sample all navigational pools over a 5-year period and provide an adequate assessment for each pool. Tentatively, biocriteria require less than 25% of river km in a pool be considered failing in order to consider the pool in attainment of its Aquatic Life Use designation. A requirement designed to ensure that an adequate assessment has been performed is that the confidence interval around the estimate of P_{fail} does not include the threshold of 0.25. By sampling only 15 randomly selected sites per pool in a single year, there is potential for an adequate and conclusive assessment of a pool. If, after 15 sites, a definitive assessment of the pool cannot be made, an additional 15 sites from the same design would be sampled the following year and combined with the first 15 sites. This may shift the P_{fail} but will also reduce the length of the CI so that it may no longer include the threshold. If the CI still includes the threshold, an additional 15 sites should be sampled the third year and combined with the first 30 sites. If the CI still includes the threshold at this point, the estimate of condition based on 45 samples could be used to assess the pool as above or below the 0.25 threshold, regardless of the CI. The result would be a sample size close to the recommended sample size of 50 and should lead to desirable CI lengths. This approach would limit the resources required to assess an individual pool, such that additional effort would be necessary only in those pools that are of more marginal condition. In addition, this approach would help ORSANCO to rapidly identify those pools in which biological condition is most impacted and to prioritize any mitigation or restoration efforts. To ensure that an adequate number of sites would be available to meet potential sampling needs in a given pool, an initial random draw of 60-80 sites per pool is appropriate. This approach should provide enough oversample to account for non-target, inaccessible, or otherwise unsampleable sites and still obtain up to 45 samples for assessment of condition. Finally, additional sampling of individual sites may be required to determine causes of impairment within pools but may be guided by the data acquired through the random sampling design.

6. LITERATURE CITED

Angermier, P.L. and J.R. Karr. 1986. Applying an index of biotic integrity based on stream-fish communities: considerations in sampling and interpretation. North American Journal of Fisheries Management 6: 418-427.

Chernick, M.R. 1999. Bootstrap Methods: A Practitioner's Guide. John Wiley and Sons, Inc., New York, New York. 264 pp.

Colwell, R.K. 2005. EstimateS: Statistical estimation of species richness and shared species from samples. Version 7.5. User's Guide and applications published at http://purl.oclc.org/estimates.

Dauwalter, D.C. and E.J. Pert. 2003. Electrofishing Effort and Fish Species Richness and Relative Abundance in Ozark Highland Streams of Arkansas. North American Journal of Fisheries Management. 23: 1152-166.

Diaz-Ramos, S., D.L. Stevens, Jr., and A.R. Olsen. 1996. EMAP Statistical Methods Manual. EPA/620/R-96/002. U.S. Environmental Protection Agency, Office of Research and Development, National Health and Environmental Effects Research Laboratory, Corvallis, Oregon.

Emery, E.B., T.P. Simon, F.H. McCormick, P.L. Angermeier, J.E. DeShon, C.O. Yoder, R.E. Sanders, W.D. Pearson, G.D. Hickman, R.J. Reash, and J.A. Thomas. 2003. Development of a Multimetric Index for Assessing the Biological Condition of the Ohio River. Transactions of the American Fisheries Society. 132: 791-808.

Etnier, D.A. and W.C. Starnes. 1993. The Fishes of Tennessee. The University of Tennessee Press. Knoxville. 681 pp.

Hughes, R.M., P.R. Kaufmann, A.T. Herlihy, S.S. Intelmann, S.C. Corbett, M.C. Arbogast, and R.C. Hjort. 2002. Electrofishing distance needed to estimate fish species richness in raftable Oregon rivers. North American Journal of Fisheries Management 22:1229-1240.

Karr, J.R., and E.W. Chu. 1999. *Restoring Life in Running Waters: Better Biological Monitoring.* Island Press, Covelo, California. 206 pp.

Koenker, R. and G. Bassett. 1978. Regression Quantiles. Econometrica. 46(1): 33-50.

Lyons, J. 1992. The length of stream to sample with a towed electrofishing unit when fish species richness is estimated. North American Journal of Fisheries Management 12:198-203.

Lyons, J., R.R. Piette, and K.W. Niermeyer. 2001. Development, validation, and application of a fish-based index of biotic integrity for Wisconsin's large rivers. Transactions of the American Fisheries Society 130:1077-1094.

Meador, M.R. 2005. Single-pass versus two-pass boat electrofishing for characterizing river fish assemblages: species richness estimates and sampling distance. Transactions of the American Fisheries Society 134:59-67.

Ohio EPA. 1989. Addendum to: Biological Criteria for the Protection of Aquatic Life: Volume II: Users Manual for Biological Field Assessment of Ohio Surface Waters. Ohio EPA. Columbus, OH.

ORSANCO. 1994. Ohio River Water Quality Fact Book. Ohio River Valley Water Sanitation Commission, Cincinnati, Ohio.

Pflieger, W.L. 1997. The Fishes of Missouri. Missouri Department of Conservation. Jefferson City, MO.

Reynolds L., A.T. Herlihy, P.R. Kaufmann, S.V. Gregory, and R.M. Hughes. 2003. Electrofishing Effort Requirements for Assessing Species Richness and Biotic Integrity in Western Oregon Streams. North American Journal of Fisheries Management. 23: 450-461.

Sanders, R.E. 1992. Day Versus Night Electrofishing Catches from Near-Shore Waters of the Ohio and Muskingum Rivers. Ohio Journal of Science. 92(3): 51-59.

Sanders, R.E., R.J. Miltner, C.O. Yoder, and E.T. Rankin. 1999. The Use of External Deformities, Erosions, Lesions, and Tumors (DELT Anomalies) in

Fish Assemblages for Characterizing Aquatic Resources: A Case Study of Seven Ohio Streams. Pages 225-248 *in* T.P. Simon, editor. Assessing the Sustainability and Biological Integrity of Water Resources using Fish Communities. CRC Press, Boca Raton, Florida.

Simon, T.P. and E.B. Emery. 1995. Modification and Assessment of an Index of Biotic Integrity to Quantify Water Resource Quality in Great Rivers. Regulated Rivers: Research and Management. 11: 283-298.

Simon, T.P. and R.E. Sanders. 1999. Applying an Index of Biotic Integrity Based on Great River Fish Communities: Considerations in Sampling and Interpretation. Pages 475-506 *in* T.P. Simon, editor. Assessing the Sustainability and Biological Integrity of Water Resources using Fish Communities. CRC Press, Boca Raton, Florida.

Smith, E.P., and G. van Belle. 1984. Nonparametric estimation of species richness. Biometrics 40:119-129.

Sokal, R.R., and F.J. Rohlf. 1995. Biometry. W.H. Freeman and Company, New York. 887pp.

Trautman, M.B. 1981. The Fishes of Ohio. Ohio State University Press. 782 pp.

U.S. Environmental Protection Agency (USEPA). 2002. Consolidated Assessment and Listing Methodology: Toward a Compendium of Best Practices, 1st edition. U.S. Environmental Protection Agency, Office of Water, Washington, D.C. http://www.epa.gov/owow/monitoring/calm.html.

Yoder, C.O. and E.T. Rankin. 1995. Biological criteria program development and implementation in Ohio. Pages 109-144 *in* W.S. Davis and T.P. Simon, editors. Biological Assessment and Criteria. Lewis Publishers, Boca Raton, Florida.

APPENDIX A

Plots of CDFs and 95% CI of ORFIn metrics for subsets of sites (red lines), with the CDF and 95% CI for the full set of sites (black lines).

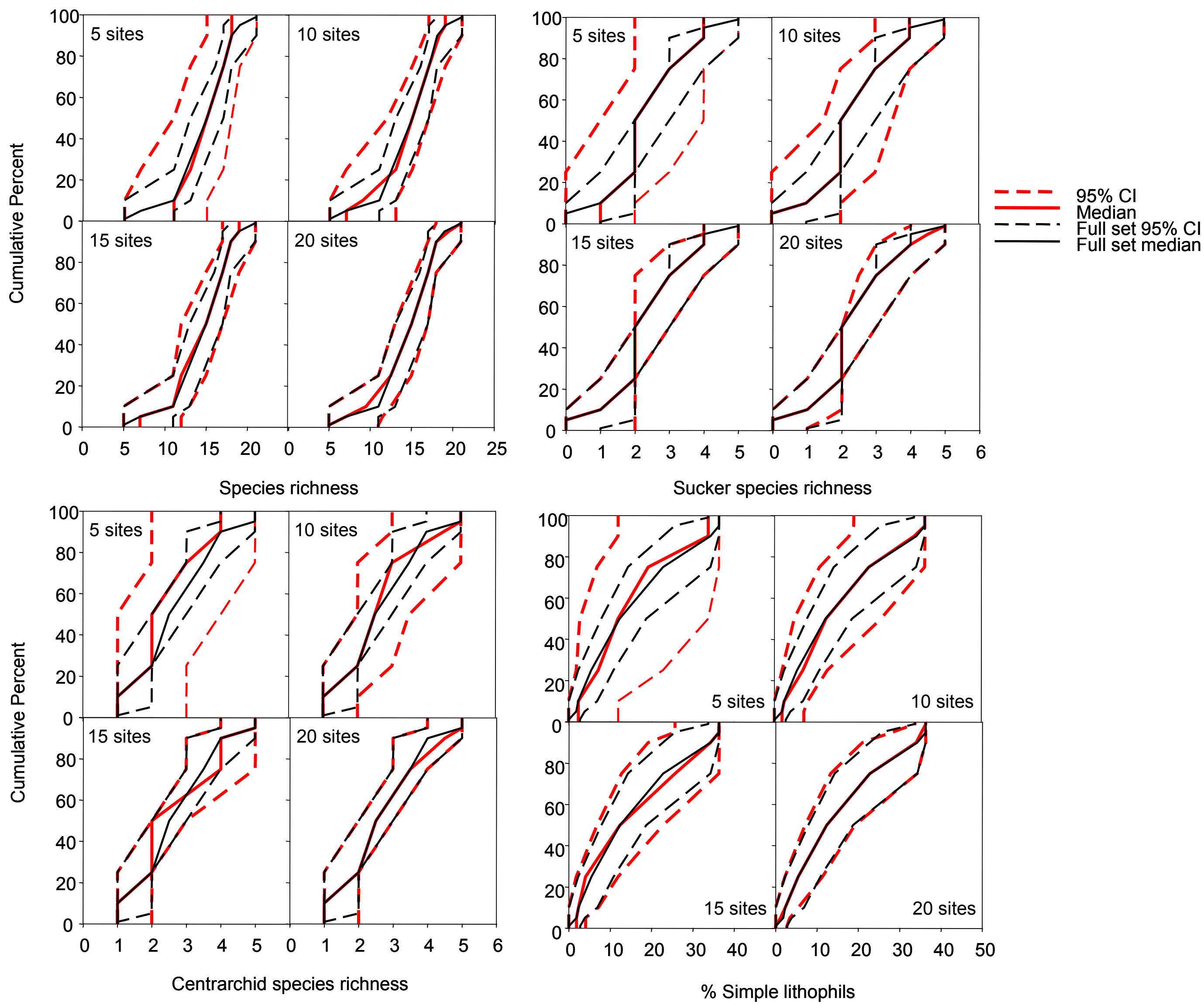

Figure A-1. Cumulative distribution functions for (a) species richness, (b) sucker species richness, (c) centrarchid species richness, and (d) % simple lithophils in R.C. Byrd pool for random subsets and the full set of sites.

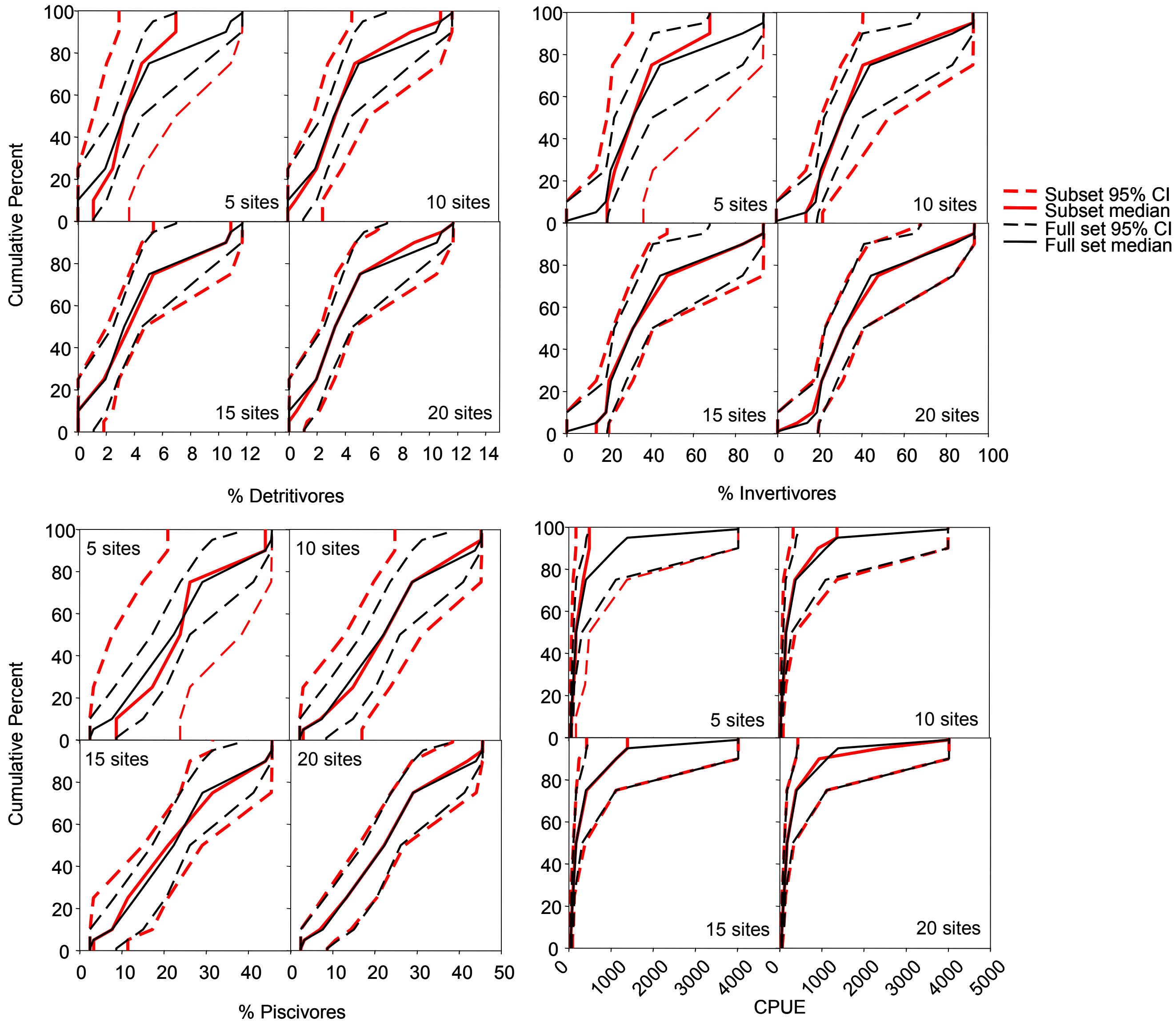

Figure A-2. Cumulative distribution functions for (a) % detritivores, (b) % invertivores, (c) % piscivores, and (d) CPUE in R.C. Byrd pool for random subsets and the full set of sites.

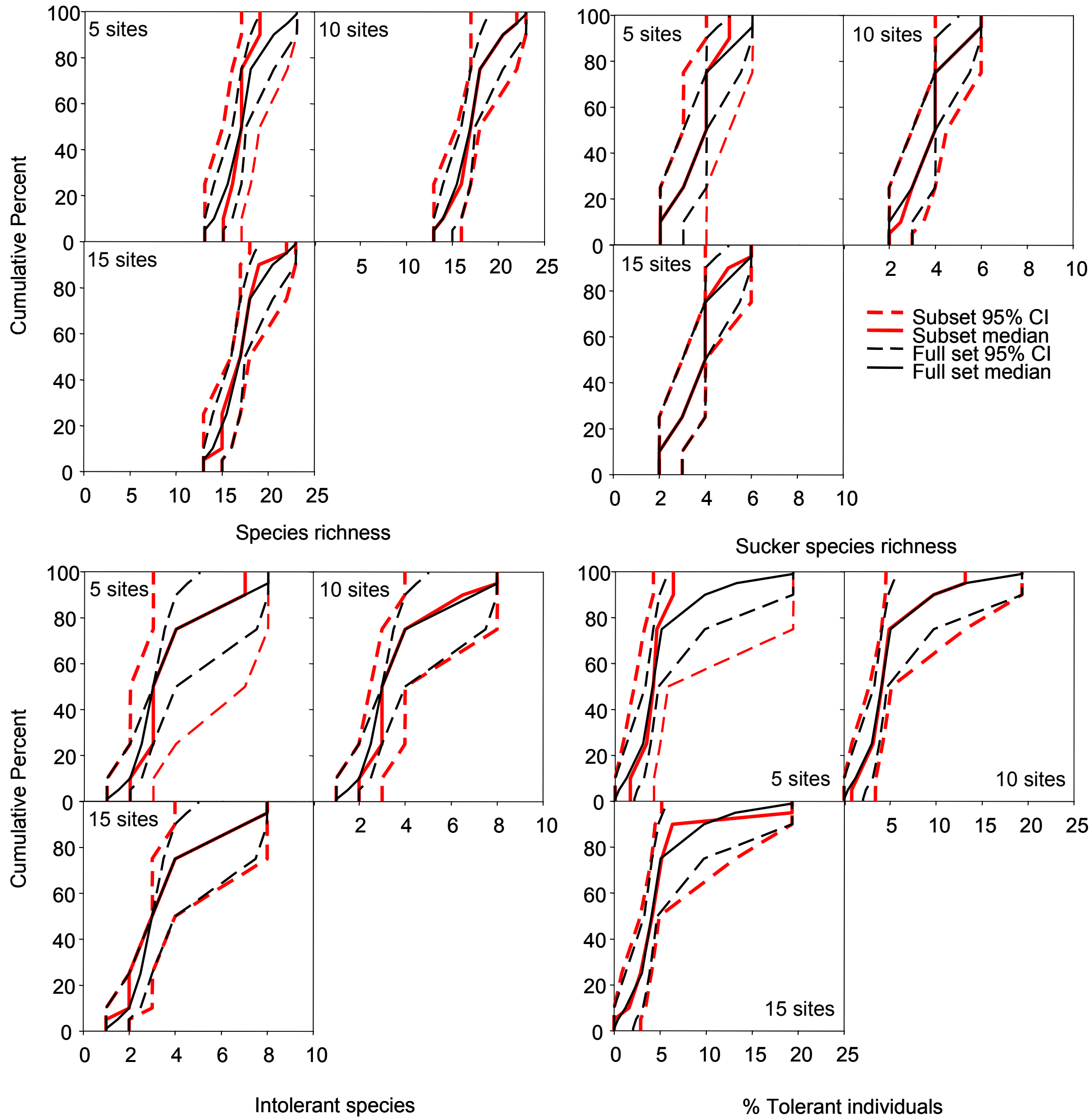

Figure A-3. Cumulative distribution functions for (a) species richness, (b) sucker species richness, (c) intolerant species richness, and (d) % tolerant individuals in Hannibal pool for random subsets and the full set of sites.

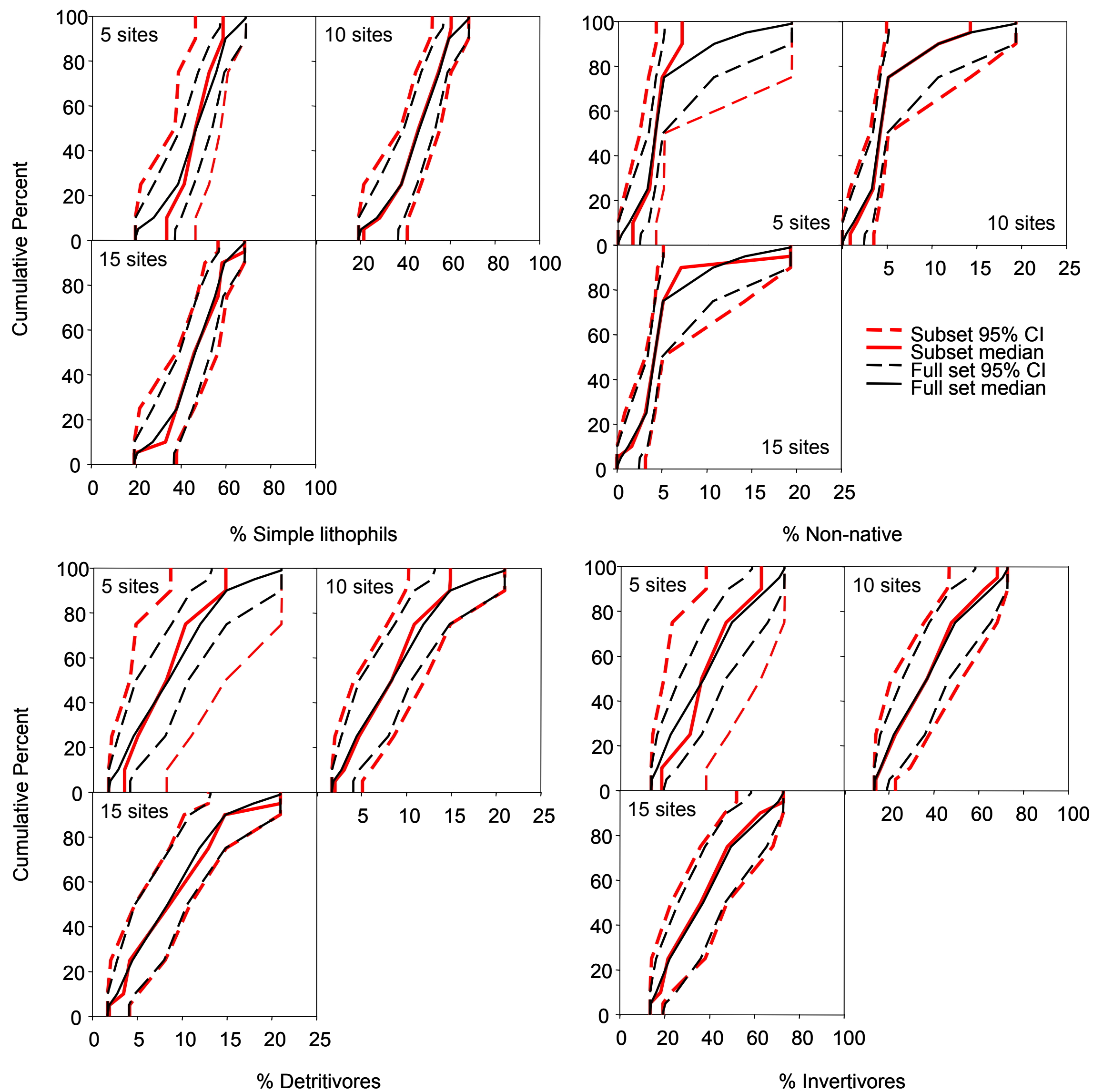

Figure A- 4. Cumulative distribution functions for (a) % simple lithophils, (b) % non-native individuals, (c) % detritivores, and (d) % invertivores in Hannibal pool for random subsets and the full set of sites.

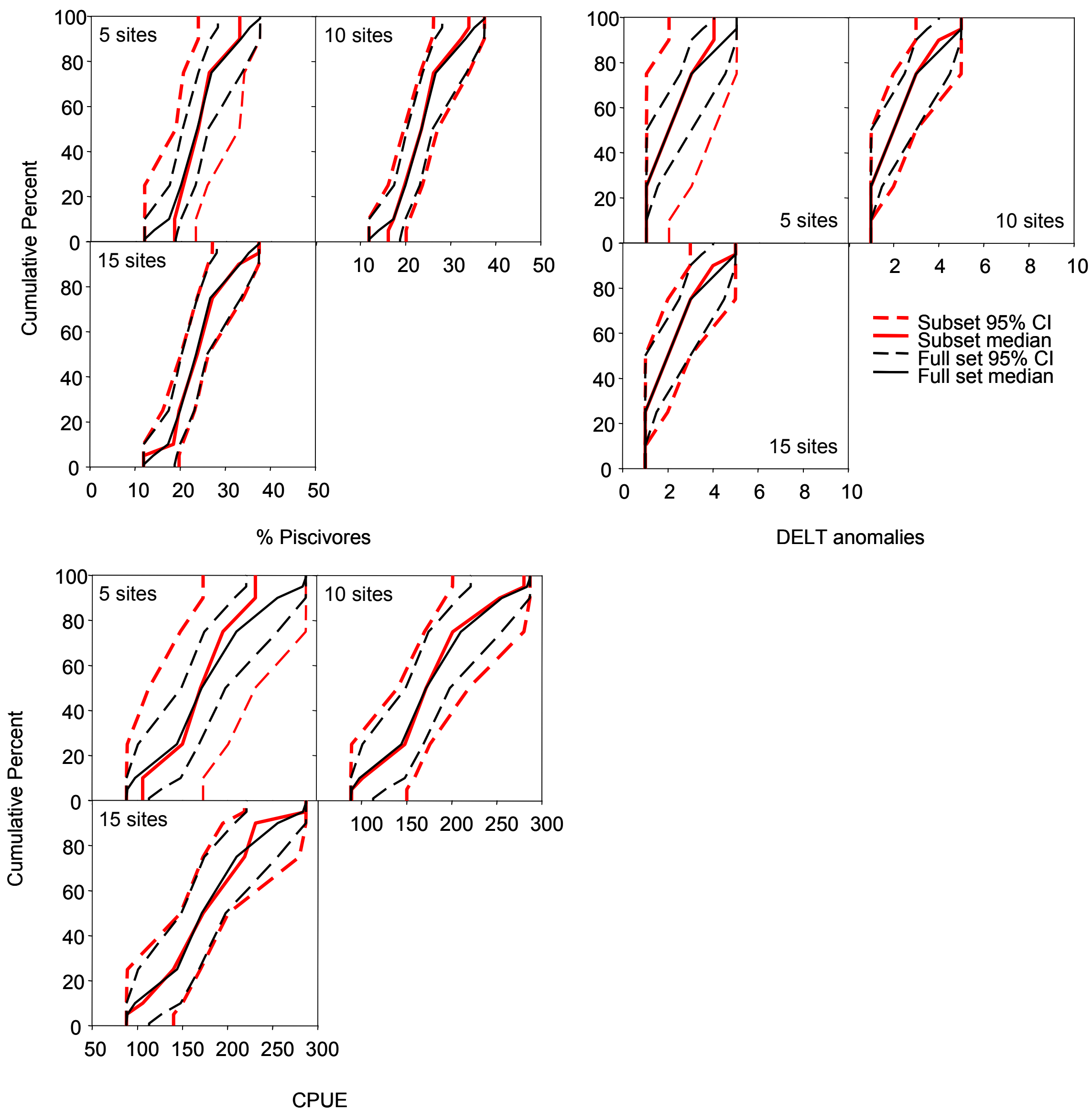

Figure A-5. Cumulative distribution functions for (a) % piscivores, (b) DELT anomalies, and (c) CPUE in Hannibal pool for random subsets and the full set of sites.

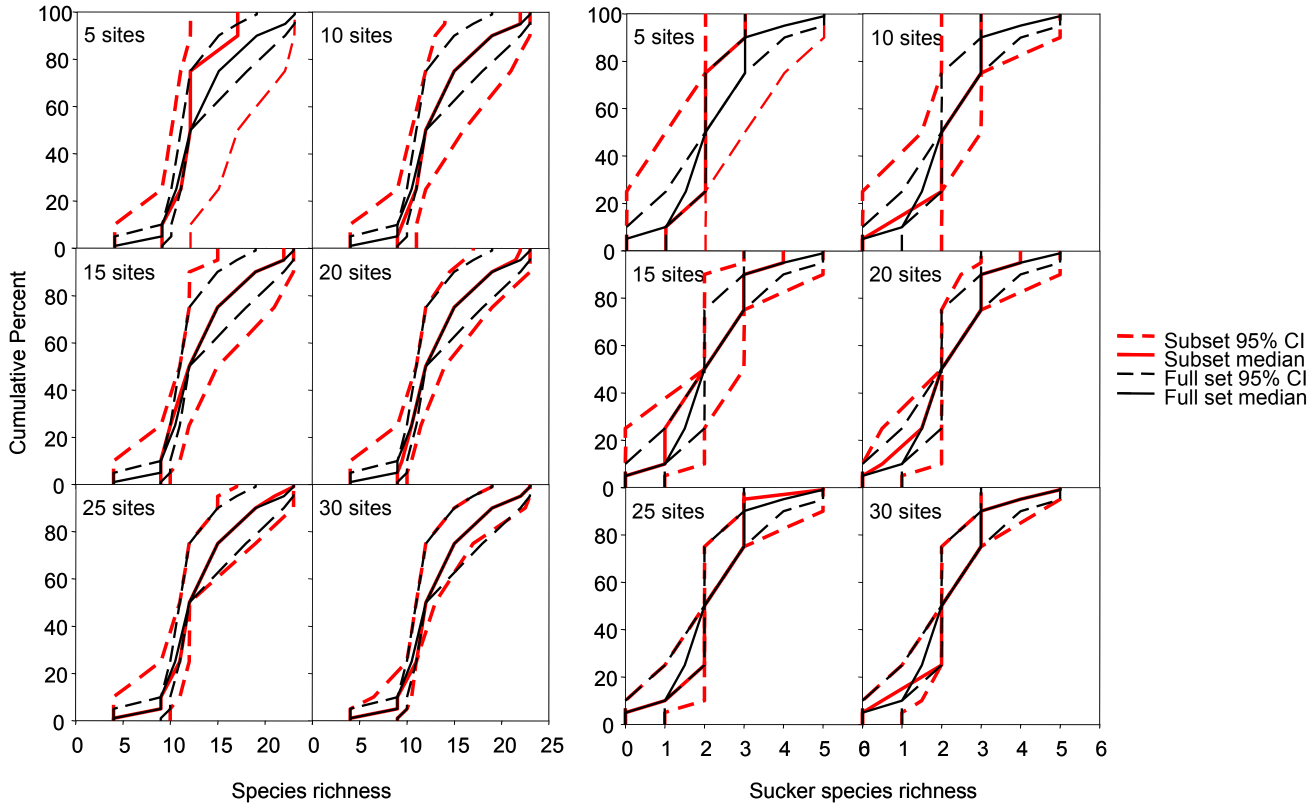

Figure A-6. Cumulative distribution functions for (a) species richness, and (b) sucker species richness in McAlpine pool for random subsets and the full set of sites.

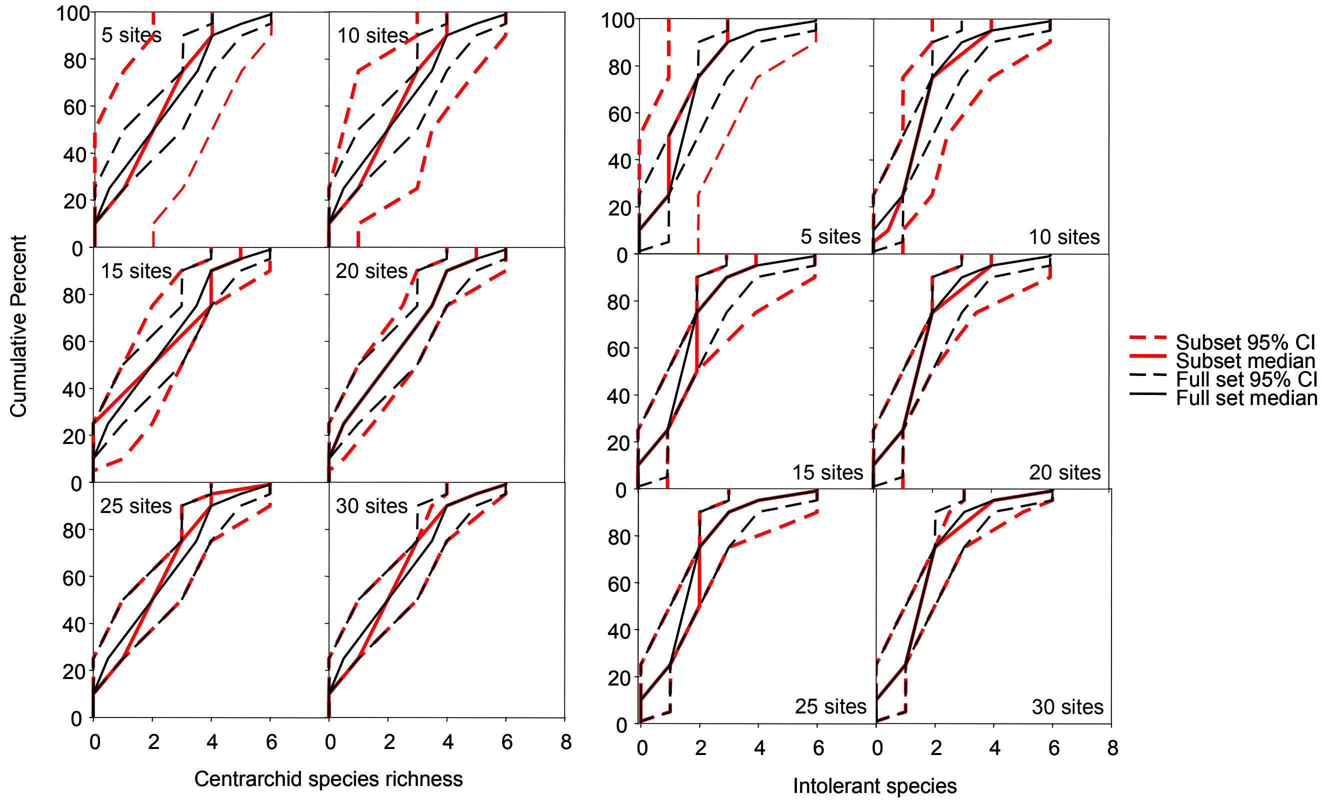

Figure A-7. Cumulative distribution functions for (a) centrarchid species richness, and (b) intolerant species richness in McAlpine pool for random subsets and the full set of sites.

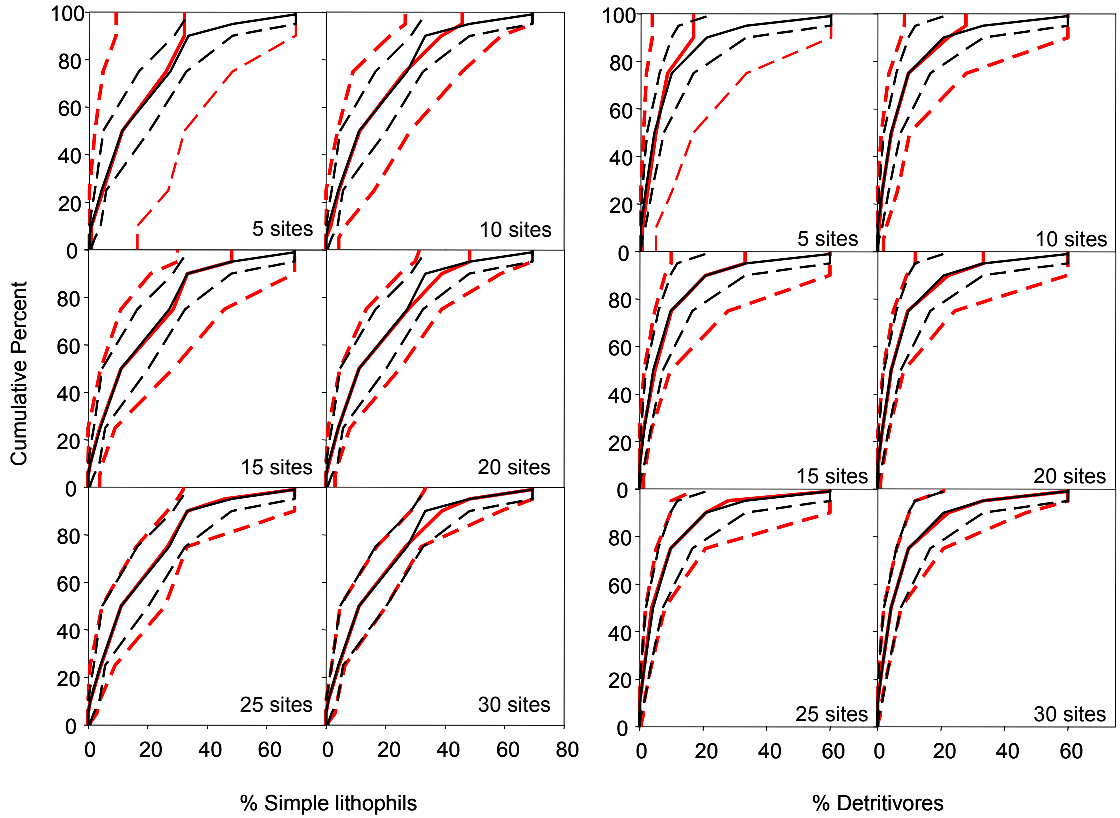

Figure A-8. Cumulative distribution functions for (a) % simple lithophils, and (b) % detritivores in McAlpine pool for random subsets and the full set of sites.

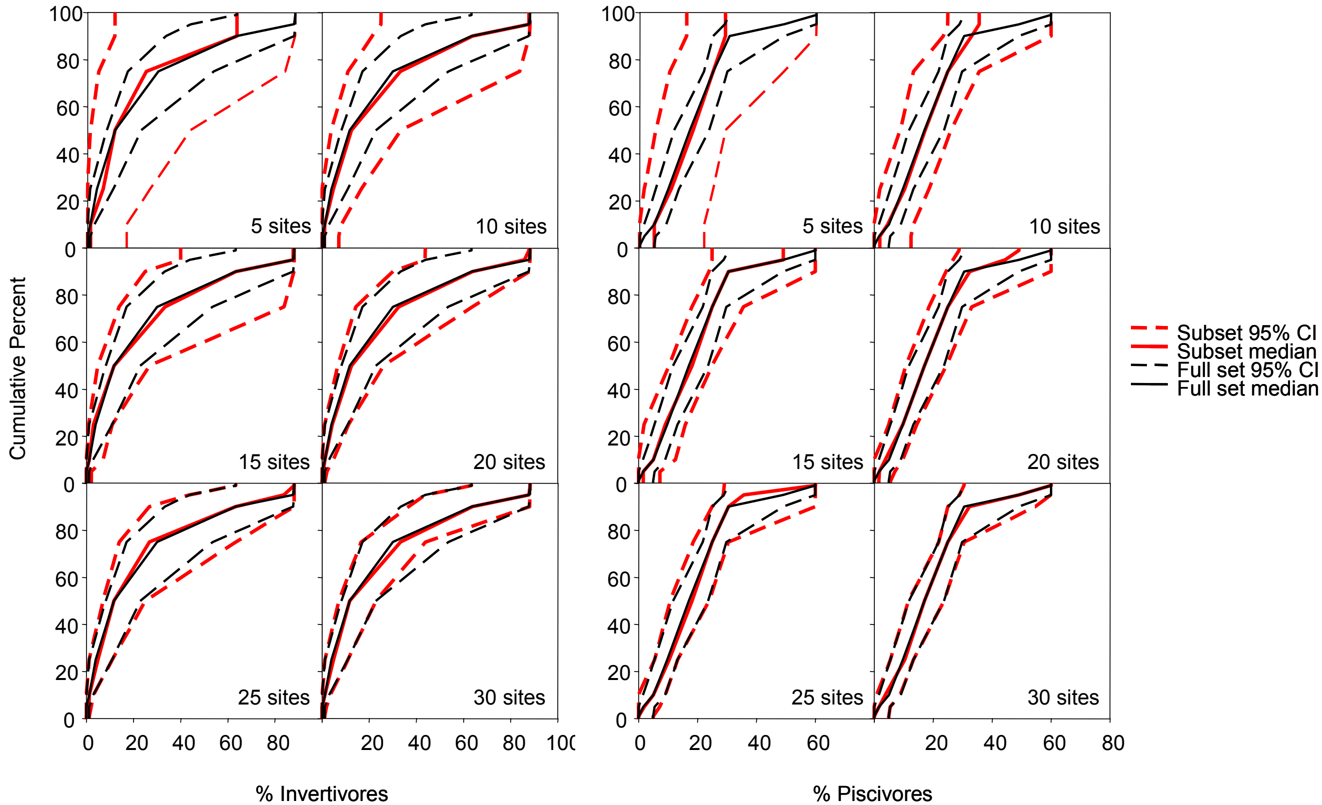

Figure A-9. Cumulative distribution functions for (a) % invertivores, and (b) % piscivores in McAlpine pool for random subsets and the full set of sites.

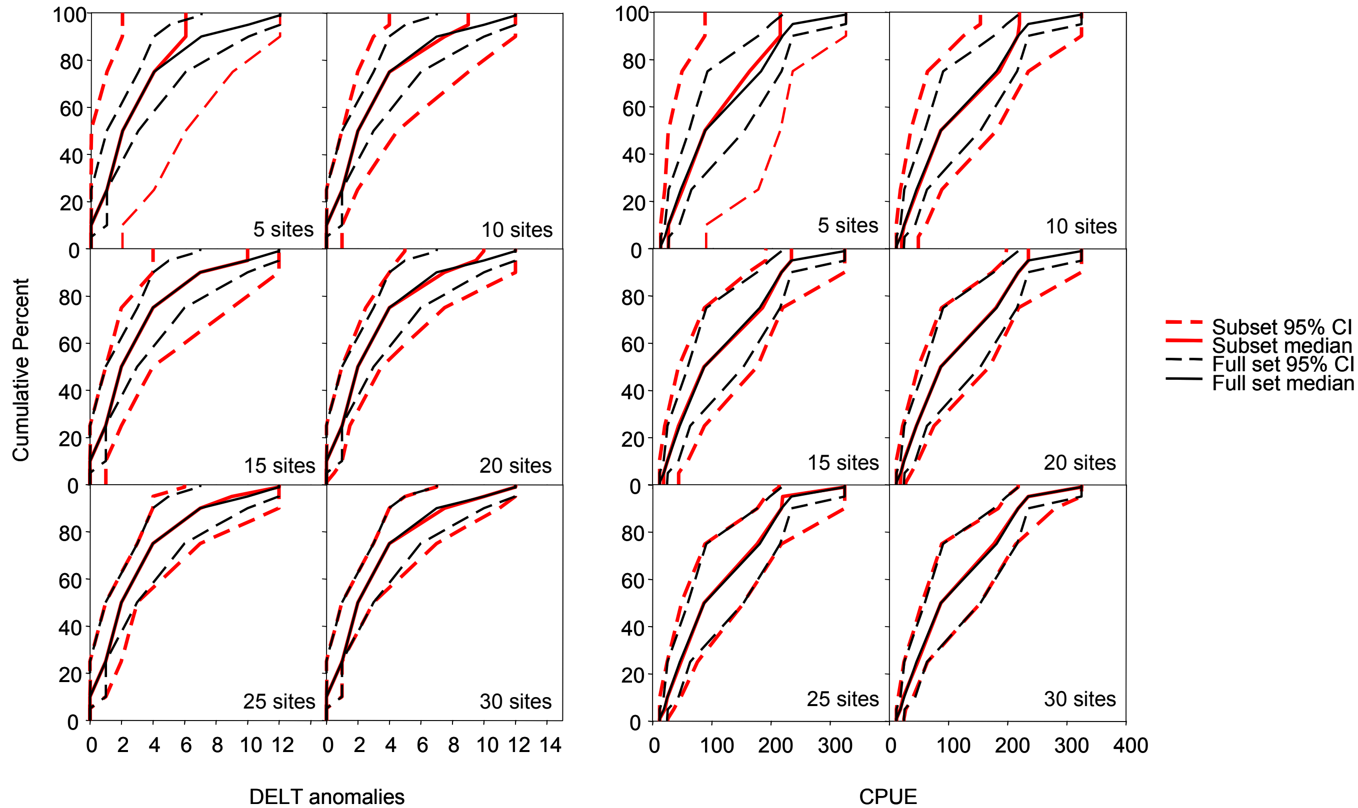

Figure A-10. Cumulative distribution functions for (a) DELT anomalies, and (b) CPUE in McAlpine pool for random subsets and the full set of sites.

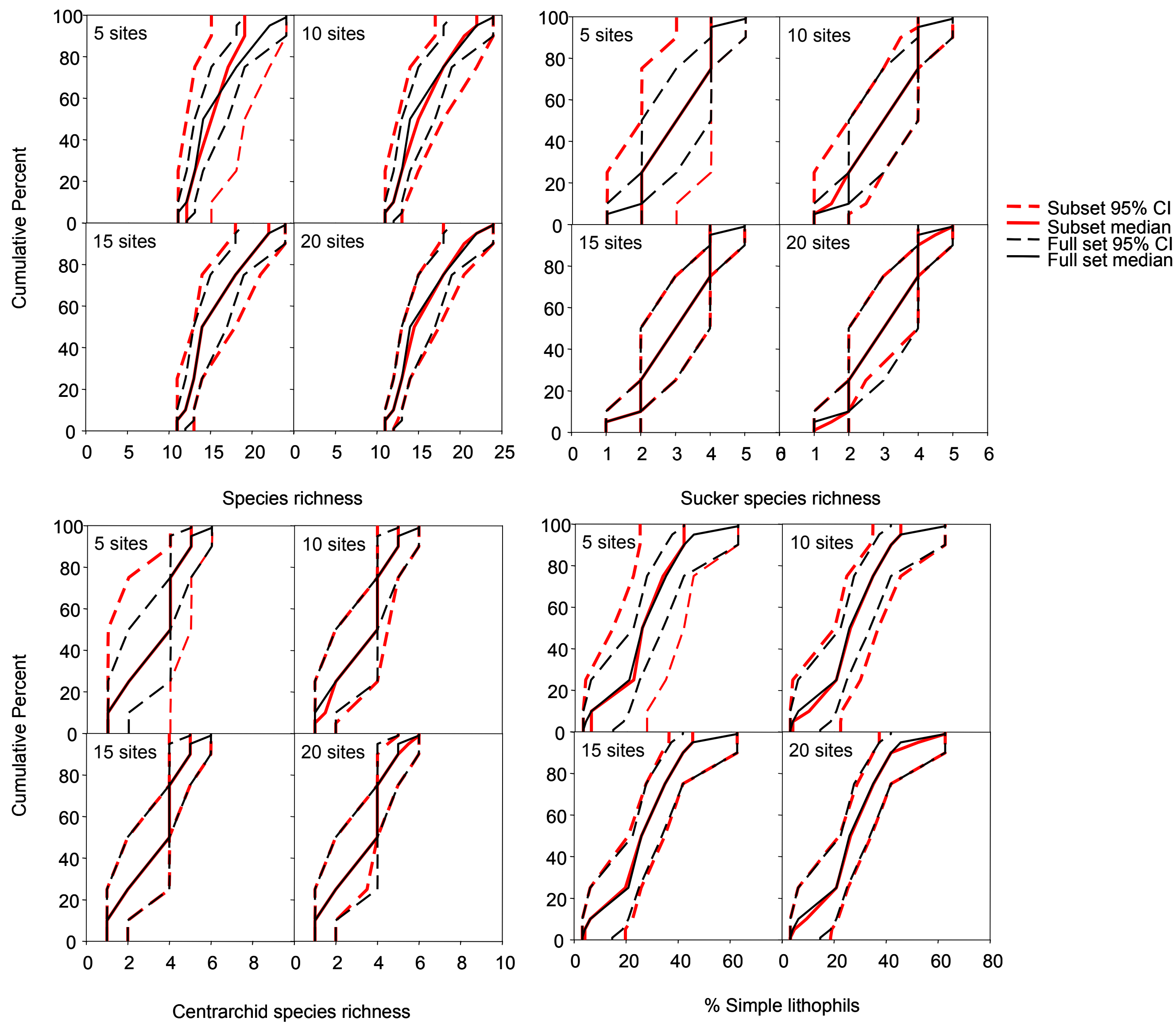

Figure A-11. Cumulative distribution functions for (a) species richness, (b) sucker species richness, (c) centrarchid species richness, and (d) % simple lithophils in Newburgh pool for random subsets and the full set of sites.

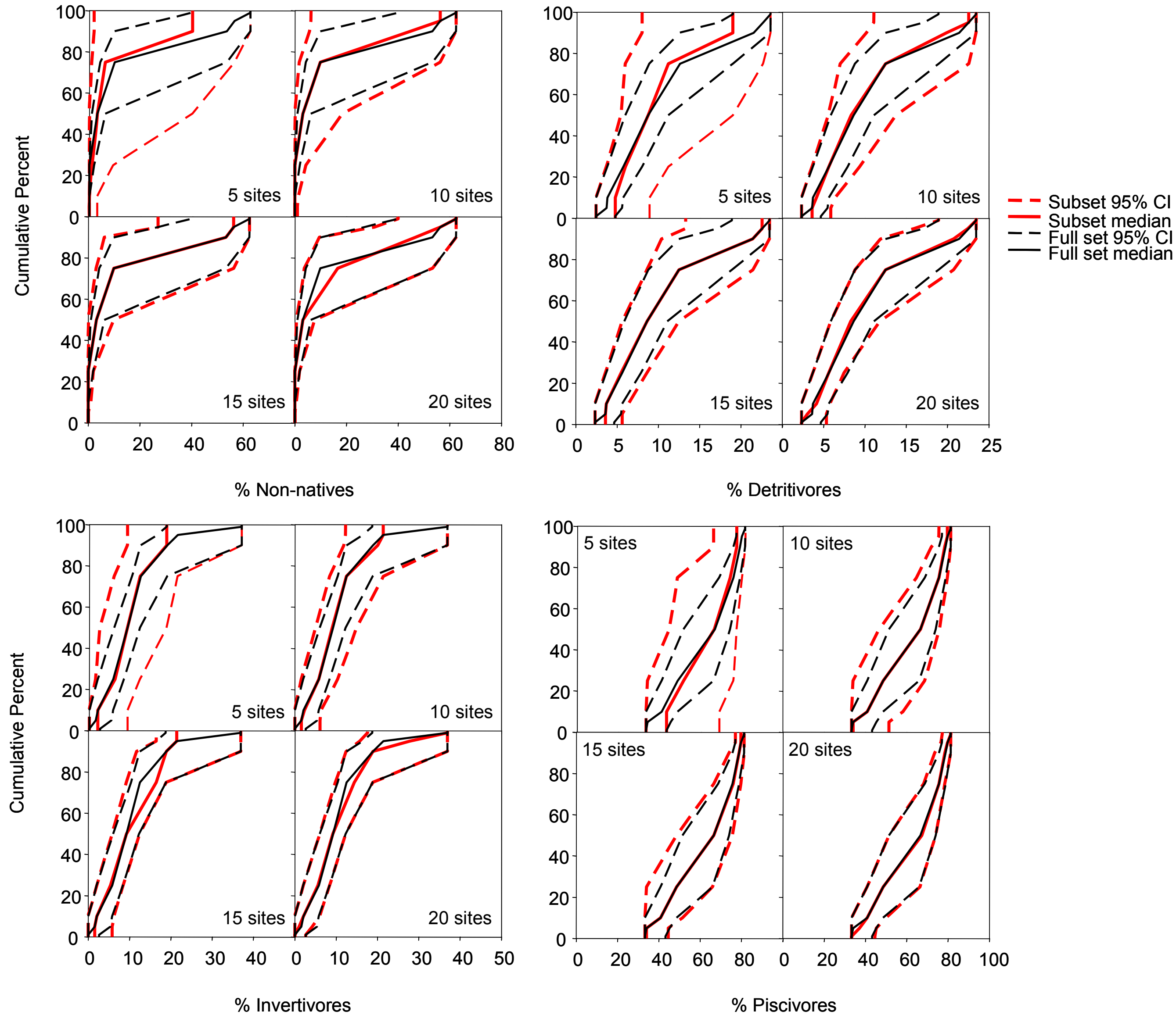

Figure A-12. Cumulative distribution functions for (a) % non-natives, (b) % detritivores, (c) % invertivores, and (d) % piscivores in Newburgh pool for random subsets and the full set of sites.

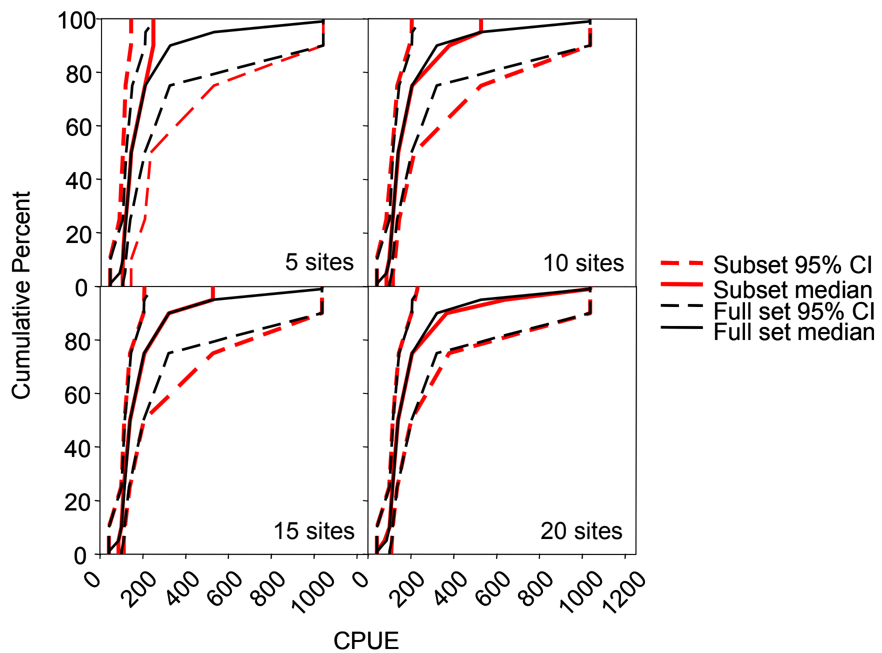

Figure A-13. Cumulative distribution functions for CPUE in Newburgh pool for random subsets and the full set of sites.

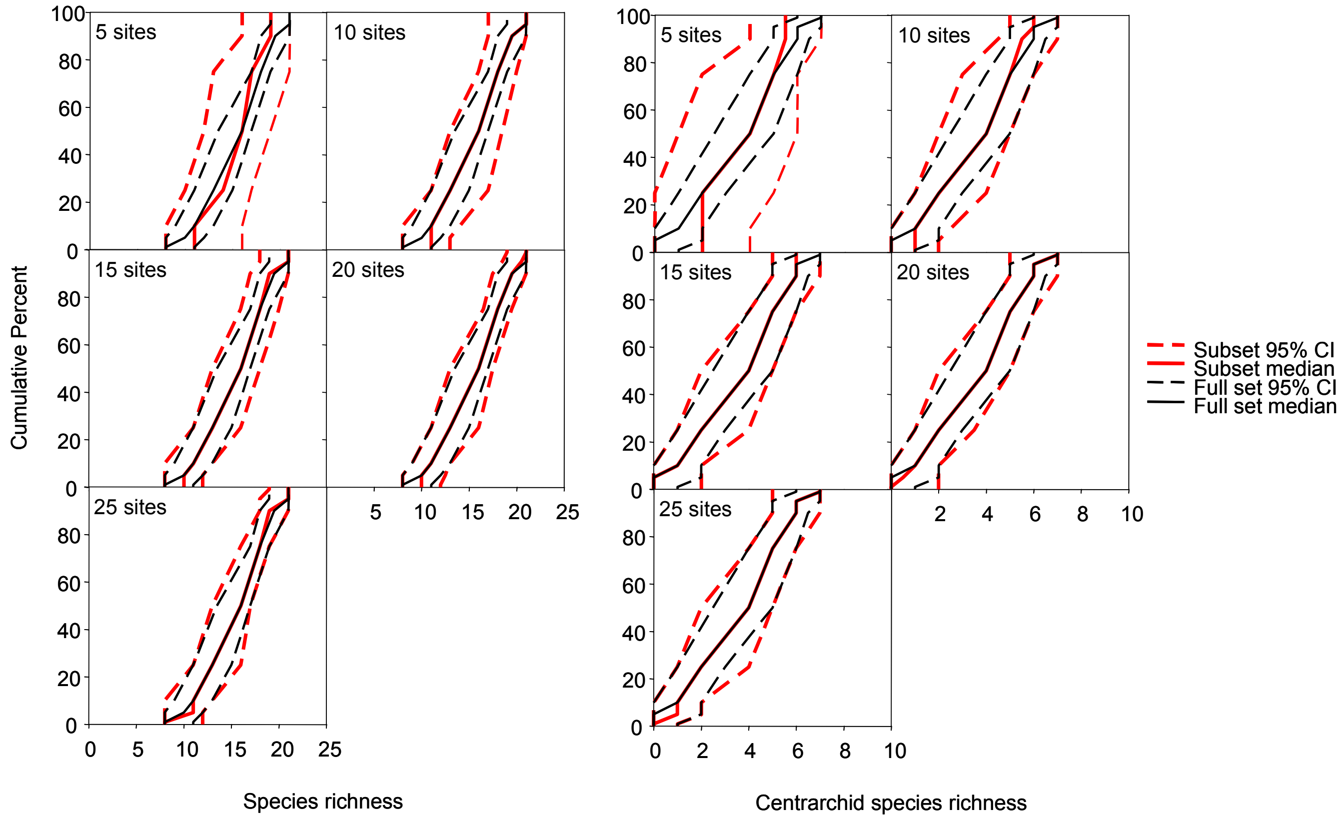

Figure A-14. Cumulative distribution functions for (a) species richness, and (b) centrarchid species richness in Smithland pool for random subsets and the full set of sites.

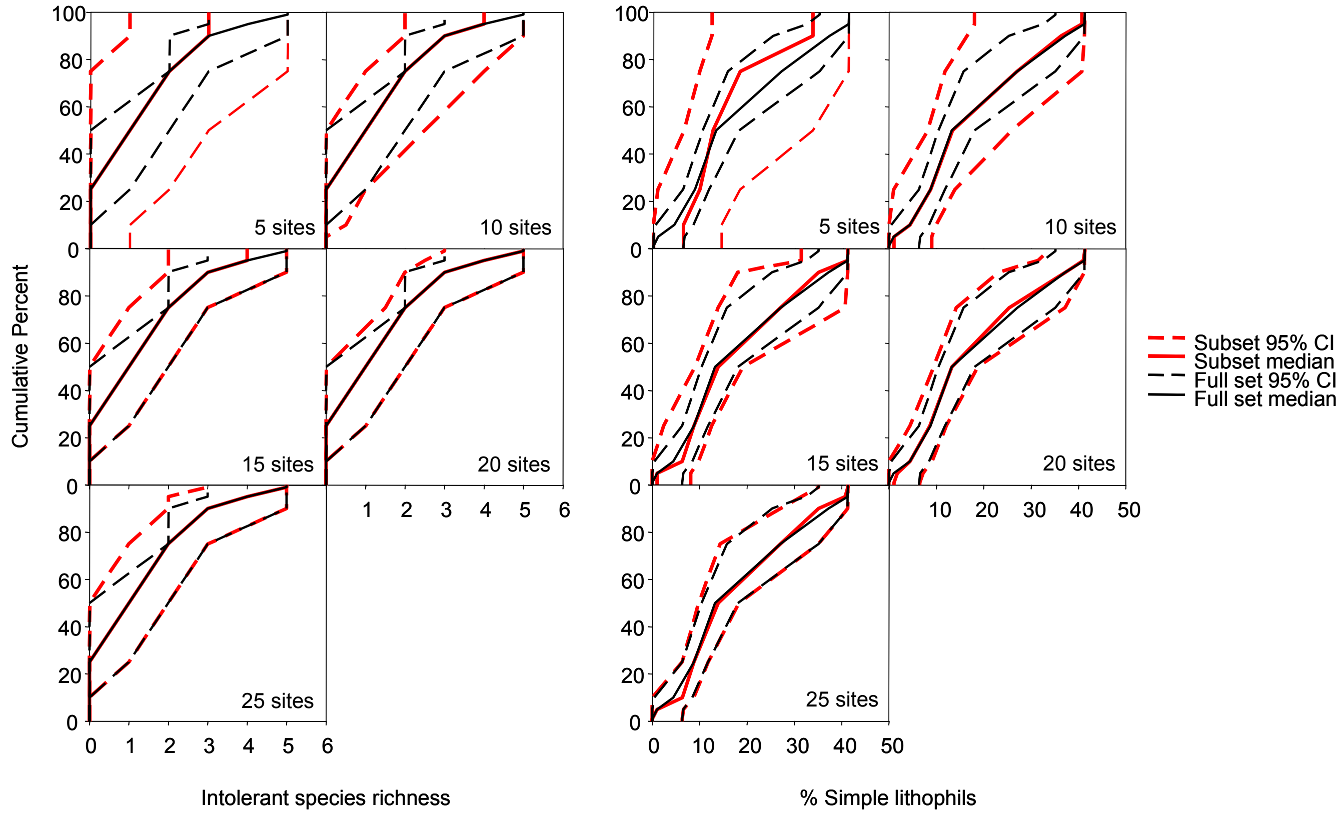

Figure A-15. Cumulative distribution functions for (a) intolerant species richness, and (b) % simple lithophils in Smithland pool for random subsets and the full set of sites.

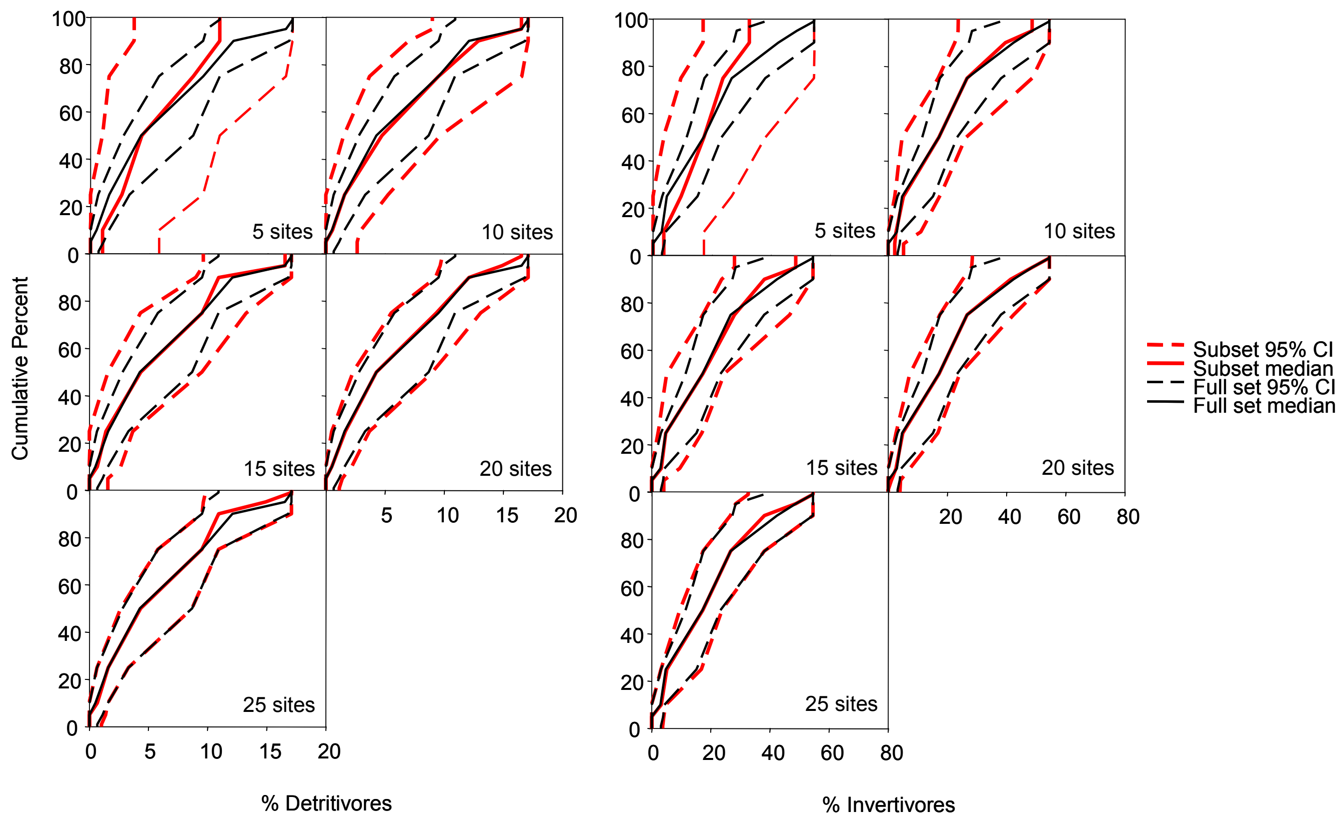

Figure A-16. Cumulative distribution functions for (a) % detritivores, and (b) % invertivores in Smithland pool for random subsets and the full set of sites.

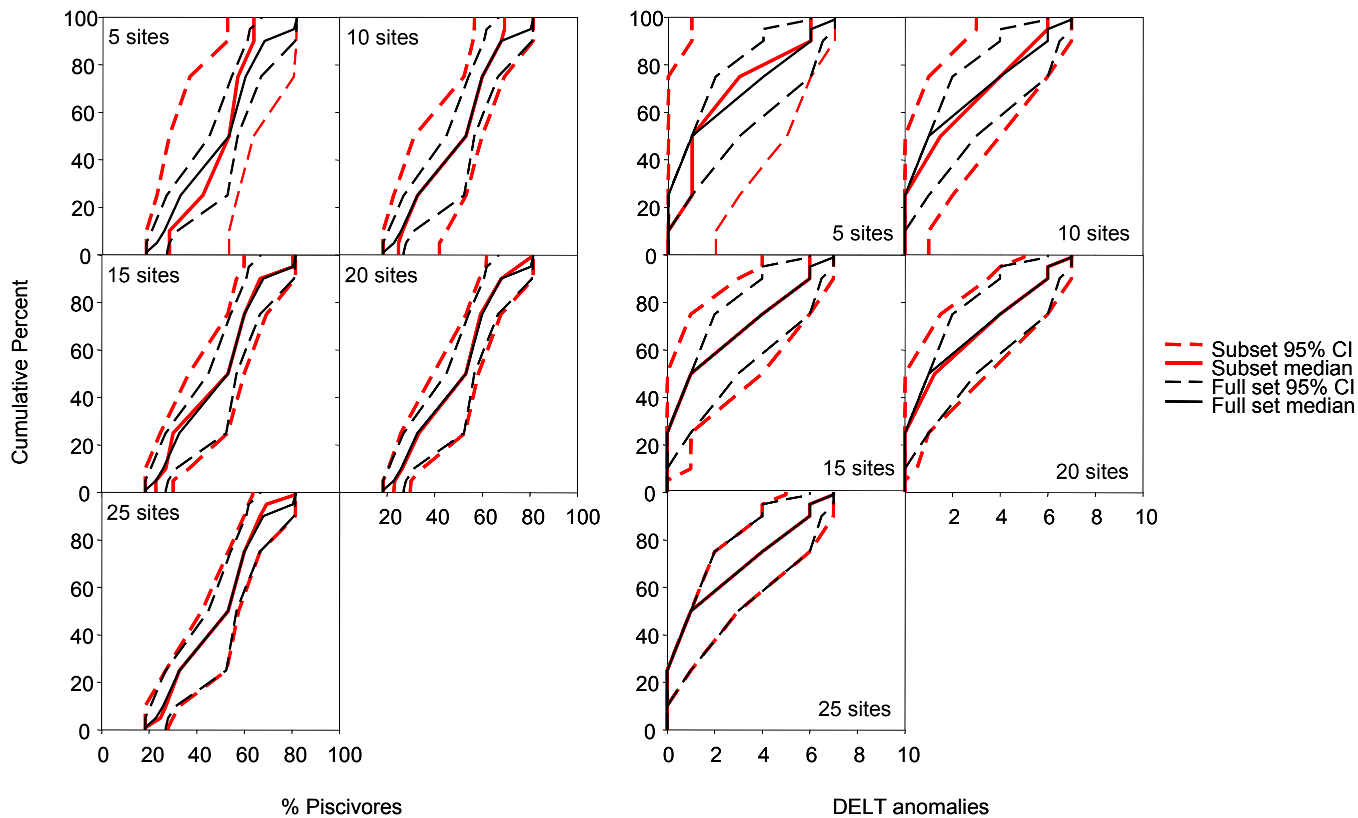

Figure A-17. Cumulative distribution functions for (a) % piscivores, and (b) DELT anomalies in Smithland pool for random subsets and the full set of sites.

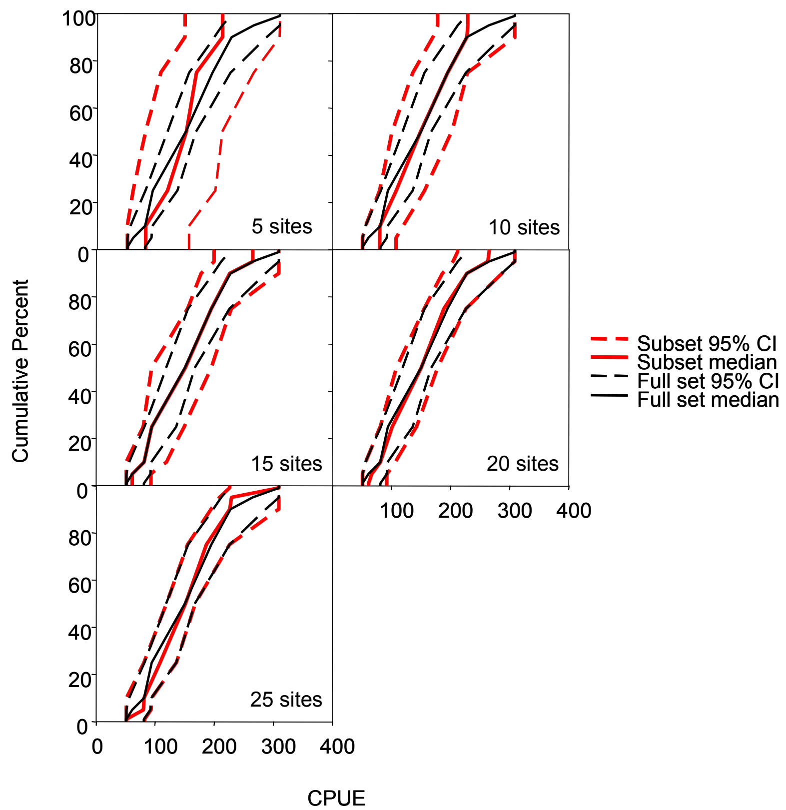

Figure A-18. Cumulative distribution functions for CPUE in Smithland pool for random subsets and the full set of sites.

United States
Environmental Protection
Agency

Office of Research
and Development (8101R)
Washington, DC 20460

EPA 600/R-06/089
September 2006
www.epa.gov