Chapter 2

Methodology

EPA faced two primary challenges to achieving the short-term goals of the National Sediment Inventory (NSI) and fulfilling the mandate of the Water Resources Development Act (WRDA) of 1992, as described in the introduction to this report. The first challenge was to compile a database of consistent sediment quality measures suitable for all regions of the country. The second challenge was to identify scientifically sound methods to determine whether a particular sediment is "contaminated," according to the definition set forth in the statute.

In many known areas of contamination, visible and relatively easy-to-recognize evidence of harmful effects on resident biota is concurrent with elevated concentrations of contaminants in sediment. In most cases, however, less obvious effects on biological communities and ecosystems are much more difficult to identify and are frequently associated with varying concentrations of sediment contaminants. In other words, bulk sediment chemistry measures are not always indicative of toxic effect levels. Similar concentrations of a chemical can produce widely different biological effects in different sediments. This discrepancy occurs because toxicity is influenced by the extent to which chemical contaminants bind to other constituents in sediment. These other sediment constituents, such as organic ligands and inorganic oxides and sulfides, are said to control the *bioavailability* of accumulated contaminants. Toxicant binding, or sorption, to sediment particles suspends the toxic mode of action in biological systems. Because the binding capacity of sediment varies, the degree of toxicity exhibited also varies for the same total quantity of toxicant.

The five general categories of sediment quality measurements are sediment chemistry, sediment toxicity, community structure, tissue chemistry, and pathology (Power and Chapman, 1992). Each of these categories has strengths and limitations for a national-scale sediment quality assessment. To be efficient in collecting usable data of similar types, EPA sought data that were available in electronic format, represented broad geographic coverage, and represented specific sampling locations identified by latitude and longitude coordinates. EPA found sediment chemistry and tissue chemistry to be the most widely available sediment quality measures.

As described above, sediment chemistry measures might not accurately reflect risk to the environment. However, EPA has recently developed assessment methods that combine contaminant concentration with measures of the primary binding phase to address bioavailability for certain chemical classes, under assumed conditions of thermodynamic equilibrium (USEPA, 1993d). Other methods, which rely on statistical correlations of contaminant concentrations with incidence of adverse biological effects, also exist (Barrick et al., 1988; FDEP, 1994; Long et al., 1995). In addition, fish tissue levels can be predicted using sediment contaminant concentrations, along with independent field measures of chemical partitioning behavior and other known or assigned fish tissue and sediment characteristics. EPA can evaluate risk to consumers from predicted and field-measured tissue chemistry data using established dose-response relationships and standard consumption patterns. Evaluations based on tissue chemistry circumvent the bioavailability issue while also accounting for other mitigating factors such as metabolism. The primary difficulty in using field-measured tissue chemistry is relating chemical residue levels to a specific sediment, especially for those fish species which typically forage across great distances.

Sediment toxicity, community structure, and pathology measures are less widely available than sediment chemistry and fish tissue data in the broad-scale electronic format EPA sought for the NSI. Sediment toxicity data are typically in the form of percent survival, compared to control mortality, for indicator organisms exposed to the field-sampled sediment in laboratory bioassays (USEPA, 1994b, c). Although these measures account for bioavailability and the antagonistic and synergistic effects of pollutant mixtures, they do not address possible long-term reproductive or growth effects, nor do they identify specific contaminants responsible for observed lethal toxicity. Indicator organisms also might not represent the most sensitive species. Community structure measures, such as fish abundance and benthic diversity, and pathology measures are potentially

indicative of long-term adverse effects, yet there are a multitude of mitigating physical, hydrologic, and biological factors that might not relate in any way to chemical contamination.

The ideal assessment methodology would be based on matched data sets of all five types of sediment quality measures to take advantage of the strengths of each measurement type and to minimize their collective weaknesses. Unfortunately, such a database does not exist on a national scale, nor is it typically available on a smaller scale. Based on the statutory definition of contaminated sediment in the WRDA, EPA can identify locations where sediment chemistry measures exceed "appropriate geochemical, toxicological, or sediment quality criteria or measures." Again based on the statutory definition, EPA can also use tissue chemistry and sediment toxicity measures to identify aquatic sediments that "otherwise pose a threat to human health or the environment" because there are either screening values (e.g., EPA risk levels for fish tissue consumption) or control samples for comparison. However, EPA believes it cannot accurately evaluate community structure or pathology measures to identify contaminated sediment, based on the statutory definition, without first identifying appropriate reference conditions to which measured conditions could be compared.

For this analysis, EPA evaluated sediment chemistry, tissue chemistry, and sediment toxicity data, taken at the same sampling station, individually and in combination using a variety of assessment methods. Because of the limitations of the available sediment quality measures and assessment methods, EPA characterizes this identification of contaminated sediment locations as a *screening-level* analysis. Similar to a potential human illness screen, a screening-level analysis should pick up potential problems and note them for further study. A screening-level analysis will typically identify many potential problems that prove not to be significant upon further analysis. Thus, classification of sampling stations in this analysis is not meant to be definitive, but is intended to be inclusive of potential problems arising from presistent metal and organic chemical contaminants. For this reason, EPA elected to evaluate data collected from 1980 to 1993 and to evaluate each chemical or biological measurement taken at a given sampling station individually. A single measurement of a chemical at a sampling station, taken at any point in time over the past 15 years, may have been sufficient to classify the sampling station as having an increased probability of association with adverse effects to aquatic life or human health.

EPA recognizes that sediment is dynamic and that great temporal and spatial variability in sediment quality exists. This variability can be a function of sampling (e.g., a contaminated area might be sampled one year, but not the next) or a function of natural events (e.g., floods can move contaminated sediment from one area to another, or can bury contaminated sediment). Movement of sediment is highly temporal, and dependent upon the physical and biological processes at work in the watershed. Some deposits will redistribute while others will remain static unless disturbed by extreme events.

In this report, EPA associates sampling stations with their "probability of adverse effects on aquatic life or human health." Each sampling station falls into one of three categories (tiers): associated adverse effects are probable (Tier 1); associated adverse effects are possible, but expected infrequently (Tier 2); or no indication of associated adverse effects (Tier 3). A Tier 3 sampling station classification does not neccesarily imply a zero or minimal probability of adverse effects, only that available data (which may be substantial or limited) do not indicate an increased probability of adverse effects. Recognizing the imprecise nature of the numerical assessment parameters, Tier 1 sampling stations are distinguished from Tier 2 sampling stations based on the magnitude of a sediment chemistry measure or the degree of corroboration among the different types of sediment quality measures.

The remainder of this chapter presents a short history of how EPA developed the NSI, a brief description of the NSI data, and an explanation of the NSI data evaluation approach.

Background

EPA initiated work several years ago on the development of the NSI through pilot inventories in EPA Regions 4 and 5 and the Gulf of Mexico Program. Based on lessons learned from these three pilot inventories, the Agency developed a document entitled *Framework for the Development of the National Sediment Inventory* (USEPA, 1993a), which describes the general format for compiling sediment-related data and provides a brief summary of sediment quality evaluation techniques. The format and overall approach were then presented, modified slightly, and agreed upon at an interagency workshop held in March 1993 in Washington, DC. Following the workshop, EPA began compiling and evaluating data for the NSI. Data from several national and regional databases were included as part of the effort.

In the spring of 1994, EPA conducted a preliminary evaluation of NSI sediment chemistry data only. The purpose of the assessment was to identify sampling stations throughout the United States where measured values of sediment pollutants exceeded sediment chemistry levels of concern. The results of that assessment were then distributed to the EPA Regional offices for their review. The Regional offices were asked to review the preliminary evaluation and to:

- Verify sampling stations targeted as areas of concern.
- Identify sampling stations that might be incorrectly targeted as areas of concern.
- Identify potential areas of concern that were not targeted, but should have been.
- Inform EPA Headquarters of additional sediment quality data that should be included in the NSI to make the inventory more accurate and complete.

The EPA Regional offices completed their review of the preliminary evaluation during the winter of 1994-95. Regional comments on the results of the preliminary evaluation were incorporated into the NSI database. EPA will add new data sets identified by the Regions to the NSI and include them in the national assessment for future reports to Congress.

In April 1994, EPA Headquarters held the Second National Sediment Inventory Workshop (USEPA, 1994d). The purpose of this workshop was to bring together experts in the field of sediment quality assessment to recommend an approach for integrating and evaluating the sediment chemistry and biological data contained in the NSI. The final approach recommended by workshop participants provided the basis for the final approach adopted to evaluate NSI data for this report to Congress. Appendix I of this report provides a brief description of the workshop approach and a list of attendees.

Description of NSI Data

The NSI includes data from the following data storage systems and monitoring programs:

- Selected data sets from EPA's Storage and Retrieval System (STORET) (69 percent of sampling stations)
  - U.S. Army Corps of Engineers (USACE)
  - U.S. Geological Survey (USGS)
  - EPA
  - States
- NOAA's Coastal Sediment Inventory (COSED) (5 percent of sampling stations)
- EPA's Ocean Data Evaluation System (ODES) (6 percent of sampling stations)
- EPA Region 4's Sediment Quality Inventory (5 percent of sampling stations)
- Gulf of Mexico Program's Contaminated Sediment Inventory (1 percent of sampling stations)
- EPA Region 10/COE Seattle District's Sediment Inventory (8 percent of sampling stations)
- EPA Region 9's Dredged Material Tracking System (DMATS) (1 percent of sampling stations)
- EPA's Great Lakes Sediment Inventory (less than 1 percent of sampling stations)
- EPA's Environmental Monitoring and Assessment Program (EMAP) (2 percent of sampling stations)
- USGS (Massachusetts Bay) Data (3 percent of sampling stations)

Although EPA elected to evaluate data collected since 1980 (i.e., 1980-93), data from before 1980 are still maintained in the NSI. At a minimum, EPA required that electronically available data include monitoring program, sampling date, latitude and longitude coordinates, and measured units for inclusion in the NSI. Additional data fields providing details such as sampling method or other quality assurance/quality control information were retained in the NSI if available. Additional information about available data fields and NSI component databases is presented in Appendix A of this report.

The types of data contained in the NSI include the following:

- *Sediment chemistry*: Measurement of the chemical composition of sediment-associated contaminants.
- *Tissue residue*: Measurement of chemical contaminants in the tissues of organisms.

- *Benthic abundance*: Measurement of the number and types of organisms living in or on sediments.
- *Toxicity*: Measurement of the lethal or sublethal effects of contaminants in environmental media on various test organisms.
- *Histopathology*: Observation of abnormalities or diseases in tissue (e.g., tumors).
- *Fish abundance*: Measurement of the number and types of fish found in a water body.

The NSI represents a compilation of environmental monitoring data from a variety of sources. Most of the component databases are maintained under known and documented quality assurance and quality control procedures. However, EPA's STORET database is intended to be a broad-based repository of data. Consequently, the quality of the data in STORET, both in terms of database entry and analytical instrument error, is unknown and probably varies a great deal depending on the quality assurance management associated with specific data submittals.

Inherent in the diversity of data sources are contrasting monitoring objectives and scope. Component sources contain data derived from different spatial sampling plans, sampling methods, and analytical methods. For example, most data from EPA's EMAP program represent sampling stations that lie on a standardized grid over a given geographic area, whereas data in EPA's STORET most likely represent state monitoring data sampled from locations near known discharges or thought to have elevated contaminant levels. In contrast, many of the National Status and Trends Program data in NOAA's COSED database represent sampling stations purposely selected because they are removed from known discharges. However, many other sampling stations in the COSED database were located within highly urbanized bays and estuaries where chemical contamination was expected. These sampling stations include data from regional bioeffects assesments in which NOAA examined sediment quality in several highly urbanized areas. These surveys were region-wide assessments, not point source or end-of-pipe studies.

From an assessment point of view, STORET data might be useful for developing a list of contaminated sediment locations, but might overstate the general extent of contaminated sediment in the Nation by focusing largely on areas most likely to be problematic. On the other hand, analysis of EMAP data might result in a more balanced assessment in terms of the mix of contaminated sampling stations and uncontaminated sampling stations. Approximately two-thirds of sampling stations in the NSI are from the STORET database. Reliance on these data is consistent with the stated objective of this survey: to identify those sediments which are contaminated. However, one cannot accurately make inferences regarding the overall condition of the Nation's sediment, or characterize the "percent contamination," using the data in the NSI because uncontaminated areas are most likely substantially underrepresented.

NSI data do not evenly represent all geographic regions in the United States, nor do the data represent a consistent set of monitored chemicals. For example, several of the databases are targeted toward marine environments or other geographically focused areas. Table 2-1 presents the number of stations evaluated per state. More than 50 percent of all stations evaluated in the NSI are located in Washington, Florida, Illinois, California, Virginia, Ohio, Massachusetts, and Wisconsin. Each of these states has more than 700 monitoring stations. Other states of similar or larger size (e.g., Georgia, Pennsylvania) have far fewer sampling stations with data for evaluation. Figures 2-1, 2-2, and 2-3 depict the location of monitoring stations with sediment chemistry, tissue residue, and toxicity data, respectively. Individual stations may vary considerably in terms of the number of chemicals monitored. Some stations have data that represent a large number of organic and inorganic contaminants, whereas others have measured values for only a few chemicals. Thus, the inventory cannot be considered comprehensive even for locations with sampling data. The reliance on readily available electronic data has undoubtedly led to exclusions of a vast amount of information available from sources such as local and state governments and published reports. Other limitations, including data quality issues, are discussed in Chapter 5 of this report.

NSI Data Evaluation Approach

The methodology developed for classifying sampling stations according to the probability of adverse effects on aquatic life and human health from sediment contamination relies on measures of sediment chemistry, sediment toxicity, and contaminant residue in tissue. Although the NSI also contains benthic abundance, histopathology, and fish abundance data, these types of data were not used in the evaluation. Benthic and fish abundance cannot be directly associated with sediment contamination based on the statutory definition and currently available assessment tools, and available fish liver histopathology data were very limited.

Table 2-1. Number of Stations Evaluated in the NSI by State

Region 1	Connecticut	98	Region 6	Arkansas	107
	Maine	55		Louisiana	460
	Massachusetts	895		New Mexico	101
	New Hampshire	7		Oklahoma	286
	Rhode Island	42		Texas	662
	Vermont	5			
Region 2	New Jersey	448	Region 7	Iowa	228
	New York	618		Kansas	203
	Puerto Rico	30		Missouri	327
				Nebraska	253
Region 3	Delaware	218	Region 8	Colorado	202
	District of Columbia	4		Montana	38
	Maryland	206		North Dakota	161
	Pennsylvania	311		South Dakota	43
	Virginia	1,051		Utah	47
	West Virginia	120		Wyoming	44
Region 4	Alabama	477	Region 9	Arizona	124
	Florida	1,776		California	1,443
	Georgia	318		Hawaii	36
	Kentucky	249		Nevada	96
	Mississippi	318			
	North Carolina	612			
	South Carolina	563			
	Tennessee	646			
Region 5	Illinois	1,669	Region 10	Alaska	267
	Indiana	108		Idaho	95
	Michigan	402		Oregon	291
	Minnesota	438		Washington	2,225
	Ohio	970			
	Wisconsin	703			

The approach used to evaluate the NSI data focuses on the protection of benthic organisms from exposure to contaminated sediments and the protection of humans from the consumption of fish that bioaccumulate contaminants from sediment. In addition, potential effects on wildlife from fish consumption were also evaluated. The wildlife results were not included in the overall results of the NSI data evaluation; however, they are presented separately. Table 2-2 presents the classification scheme used in the evaluation of the NSI data. Each component, or evaluation parameter, of the classification scheme is numbered on Table 2-2. Each evaluation parameter is discussed under a section heading cross-referenced to these numbers. Figures 2-4 through 2-8 depict the evaluation parameters and sampling station classifications in flowchart format.

EPA analyzed the NSI data by evaluating each parameter in Table 2-2 on a measurement-by-measurement and sampling station-by-sampling station basis. Each sampling station was associated with a "probability of adverse effects" by combining parameters as shown in Table 2-2 and Figures 2-4 through 2-8. Because each individual measurement was considered independently (except for divalent metals, whose concentrations were summed), a single observation of elevated concentration could place a sampling station into Tier 1, (associated adverse effects are probable). In general, the methodology was constructed such that a sampling station classified as Tier 1 must be represented by a relatively large set of data or by a highly elevated sediment concentration of a chemical whose effects screening level is well characterized based on multiple assessment techniques. Fewer data were required to classify a sampling station as Tier 2. Any sampling station not meeting the requirements to be classified as Tier 1 or Tier 2 was classified as Tier 3. Sampling stations in this category include those for which substantial data were available without evidence of adverse effects, as well as sampling stations for which limited data were available to determine the potential for adverse effects.

Individual evaluation parameters, applied to various measurements independently, could lead to different site classifications. If one evaluation parameter indicated Tier 1, but other evaluation parameters indicated Tier 2 or Tier 3, a Tier 1 classification was assigned to the sampling station. For example, if a sampling station was categorized as Tier 2 based on all sediment chemistry data, but was categorized as Tier 1 based on toxicity data, the station was placed in Tier 1. This principle also applies to evaluating multiple contaminants within the same evaluation parameter. For example, if the evaluation of sediment chemistry data placed a sampling station in Tier 1 for metals and in Tier 2 for PCBs, the station was placed in Tier 1.

Recognizing the imprecise nature of some assessment parameters used in this report, Tier 1 sampling stations are distinguished from Tier 2 sampling stations based on the magnitude of a contaminant concentration in sediment, or the degree of corroboration among the different types of sediment quality measures. In response to uncertainty in both biological and chemical measures of sediment contamination, environmental managers must balance Type I errors (false positives: sediment classified as posing a threat that does not) with Type II errors (false negatives: sediment that poses a threat but was not classified as such). In

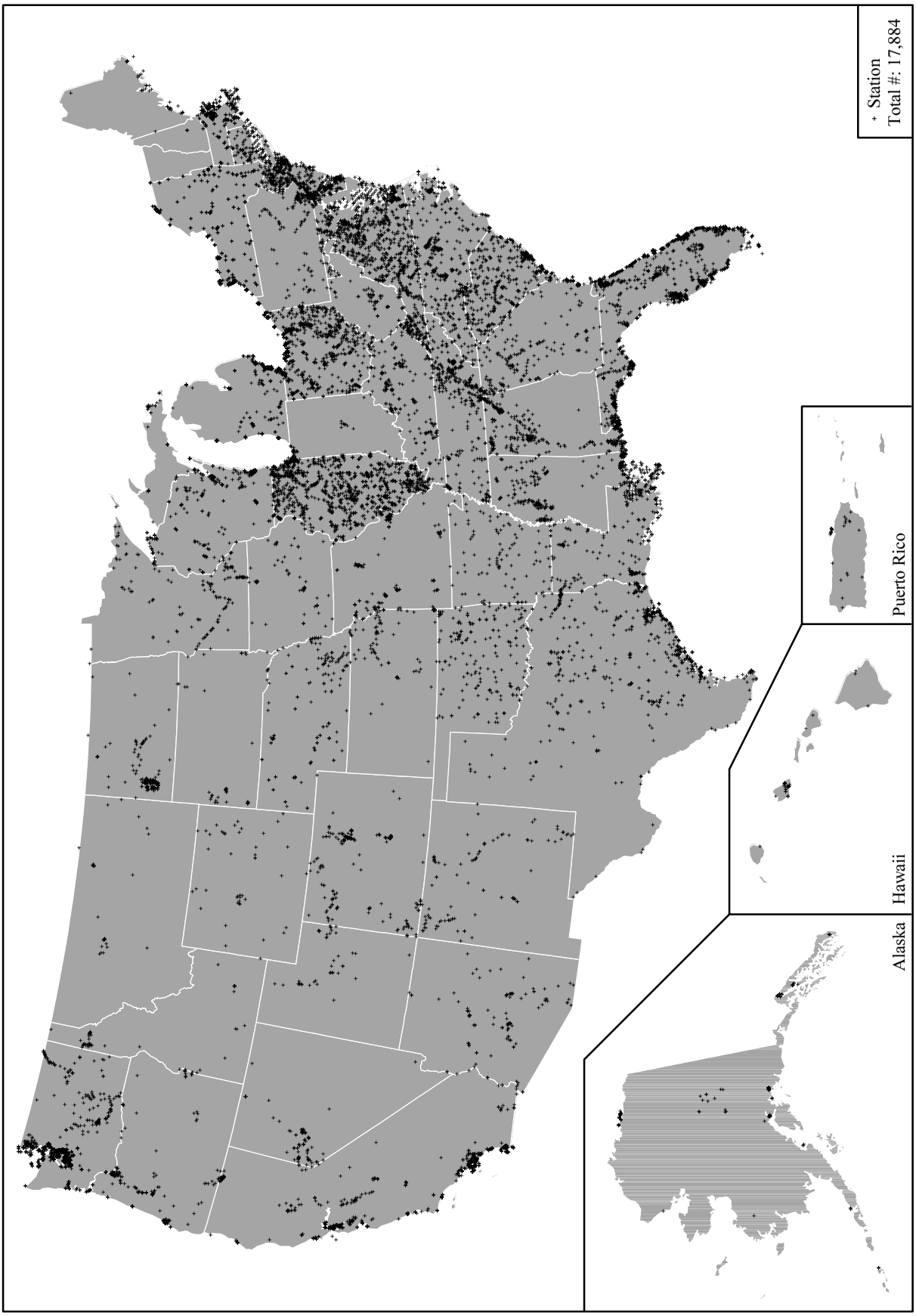

Figure 2-1. NSI Sediment Sampling Stations Evaluated.

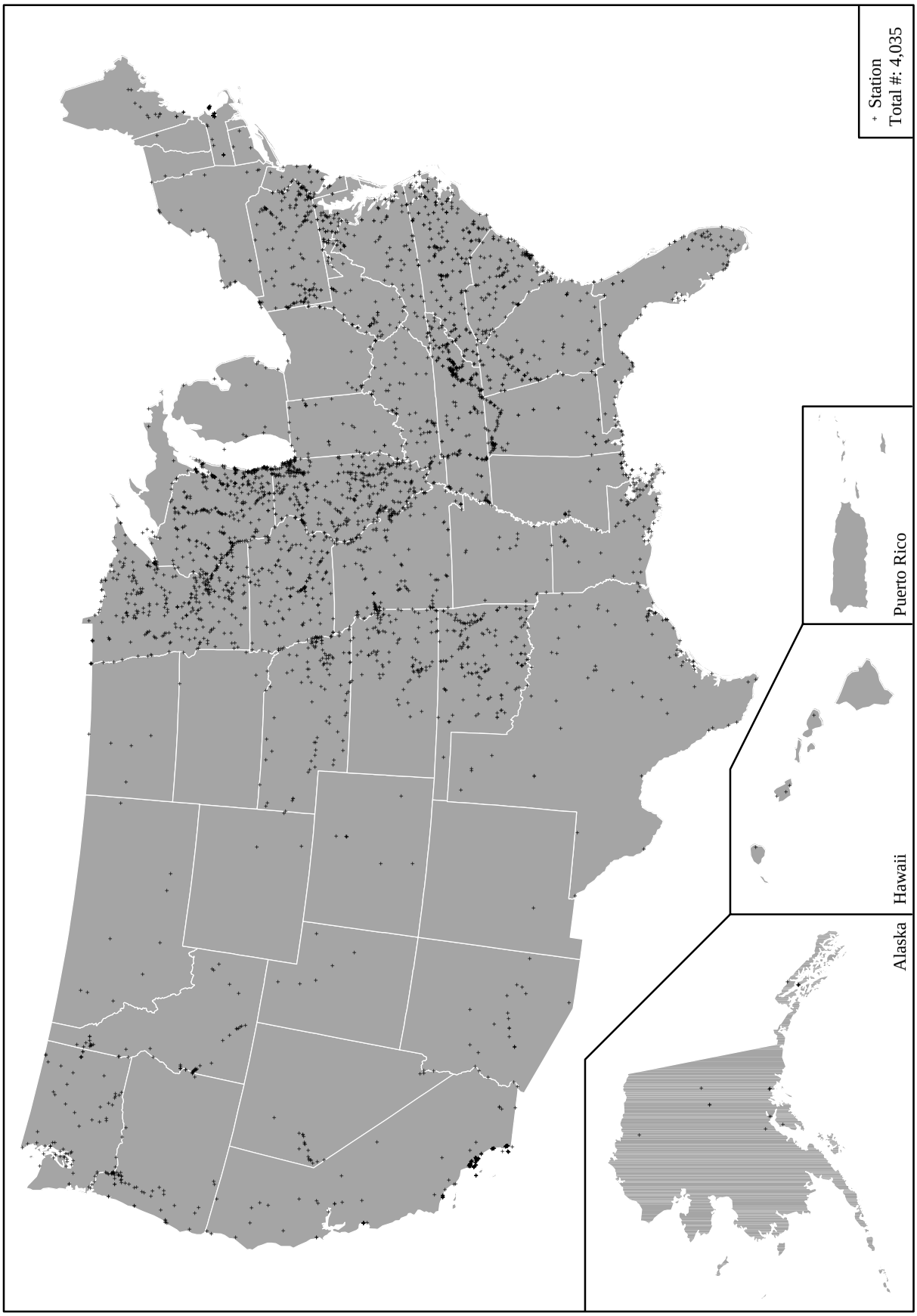

Figure 2-2. NSI Tissue Residue Sampling Stations Evaluated.

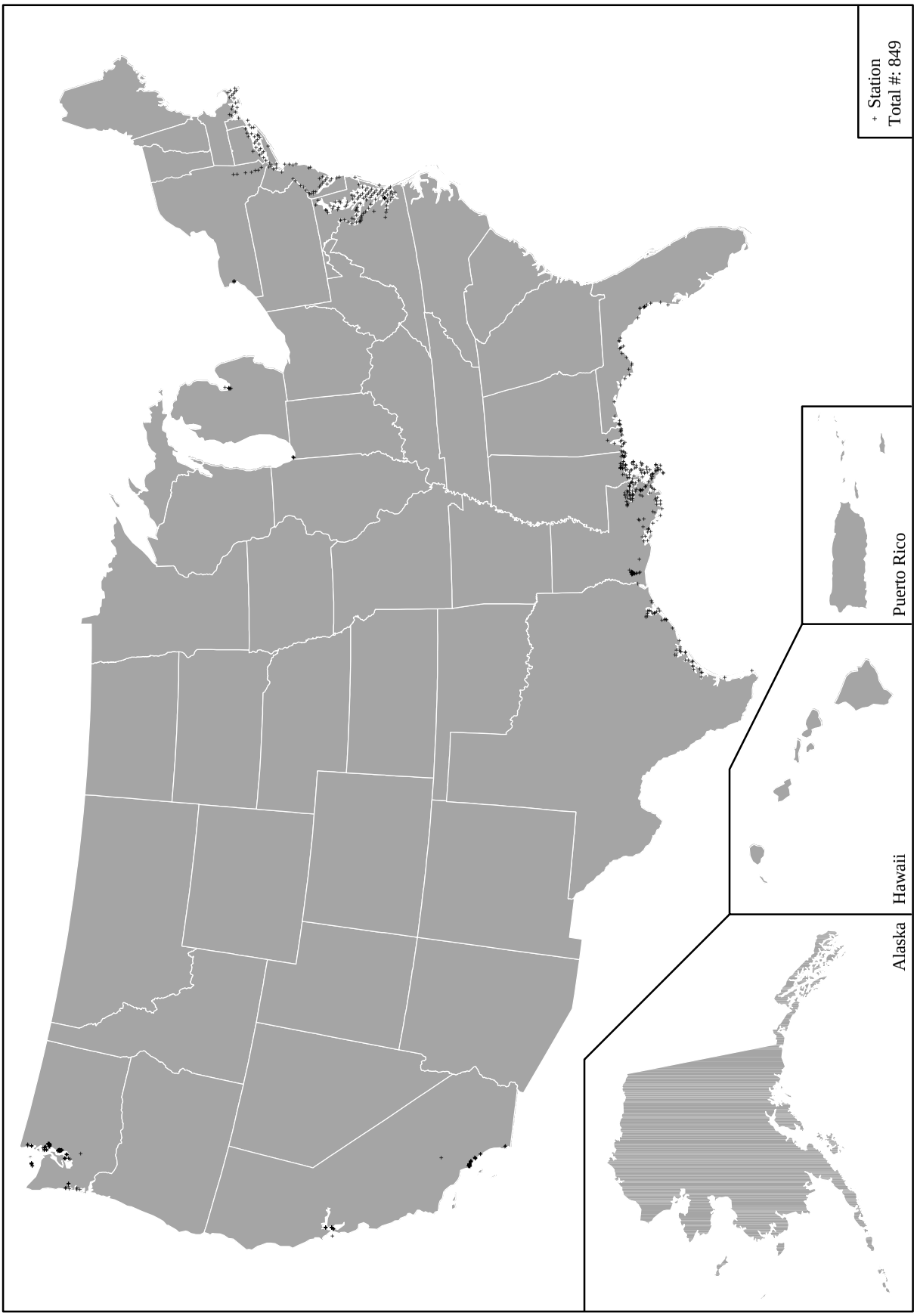

Figure 2-3. NSI Toxicity Test Stations Evaluated.

Table 2-2. NSI Data Evaluation Approach (with numbered parameters)

Category of Sampling Station Classifications	Data Used to Determine Classifications				
	Sediment Chemistry		**Tissue Residue**		**Toxicity**
Tier 1: Associated Adverse Effects to Aquatic Life or Human Health are Probable	Sediment chemistry values exceed draft sediment quality criteria for any one of the five chemicals for which criteria have been developed by EPA (must have measured TOC) **1**		Tissue levels of dioxin or PCBs in resident species exceed EPA risk levels **8**		Toxicity demonstrated by two or more nonmicrobial acute toxicity tests using two different species (one of which must be a solid-phase test) **11**
	OR [SEM]-[AVS]>5 for the sum of molar concentrations of Cd, Cu, Ni, Pb, and Zn[a] **2**	**OR**		**OR**	
	OR Sediment chemistry values exceed two or more of the relevant upper screening values (ERMs, AETs (high), PELs, SQALs, SQCs) for any one chemical (other than Cd, Cu, Ni, Pb, and Zn) (can use default TOC) **3**				
	OR Sediment chemistry TBP exceeds FDA levels or EPA risk levels **4**	**AND**	Tissue levels in resident species exceed FDA levels or EPA risk levels **9**	___	_____
Tier 2: Associated Adverse Effects to Aquatic Life or Human Health are Possible, but Expected Infrequently	[SEM]-[AVS] = 0 to 5 for the sum of molar concentrations of Cd, Cu, Ni, Pb, and Zn **5**		Tissue levels in resident species exceed FDA levels or EPA risk levels **10**		Toxicity demonstrated by a single-species nonmicrobial toxicity test **12**
	OR Sediment chemistry values exceed any one of the relevant lower screening values (ERLs, AETs (low), TELs, SQALs, SQCs) for any one chemical (can use default TOC) **6**	**OR**		**OR**	
	OR Sediment chemistry TBP exceeds FDA levels or EPA risk levels **7**				
Tier 3: No Indication of Asscociated Adverse Effects	Any sampling station not categorized as Tier 1 or Tier 2. Available data (which may be very limited or quite extensive) do not indicate a likelihood of adverse effects to aquatic life or human health.				

[a]Metals: Cd = cadmium, Cu = copper, Ni = nickel, Pb = lead, Zn = zinc.

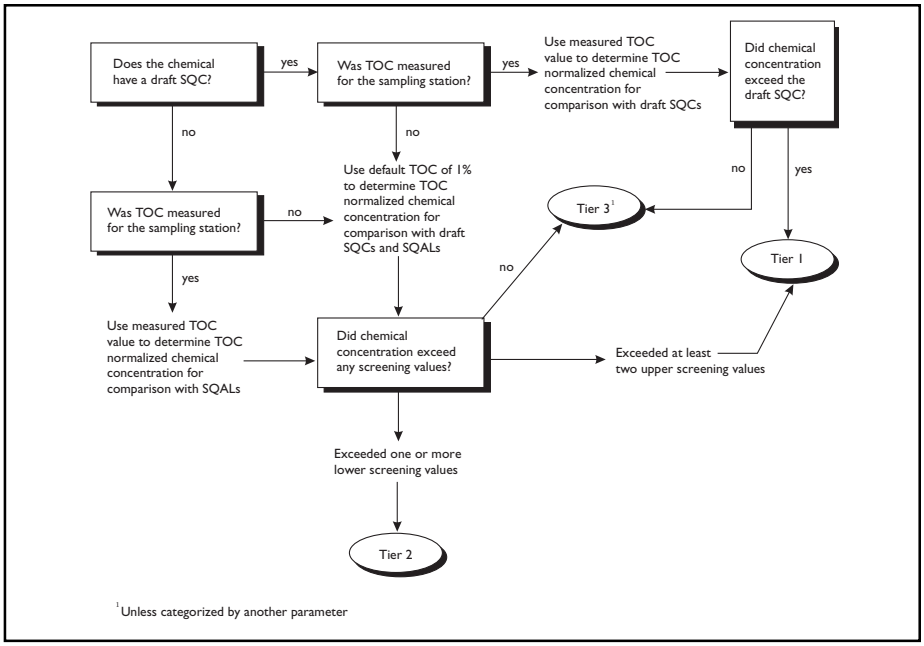

Figure 2-4. Aquatic Life Assessments: Sediment Chemistry Analysis for Organic Chemicals and Metals Not Included in the AVS Analysis.

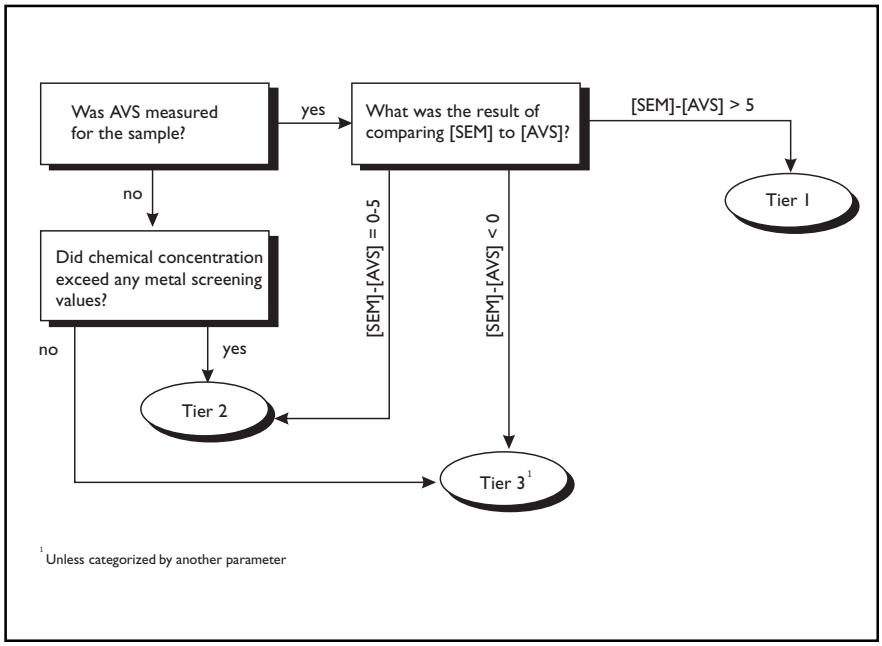

Figure 2-5. Aquatic Life Assessments: Sediment Chemistry Analysis for Divalent Metals.

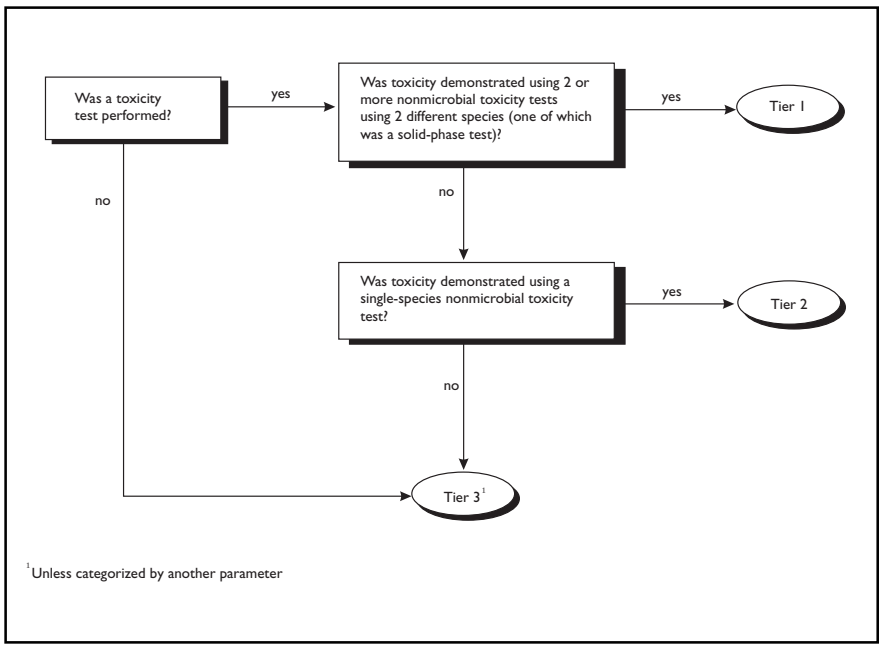

Figure 2-6. Aquatic Life Assessments: Sediment Toxicity Analysis.

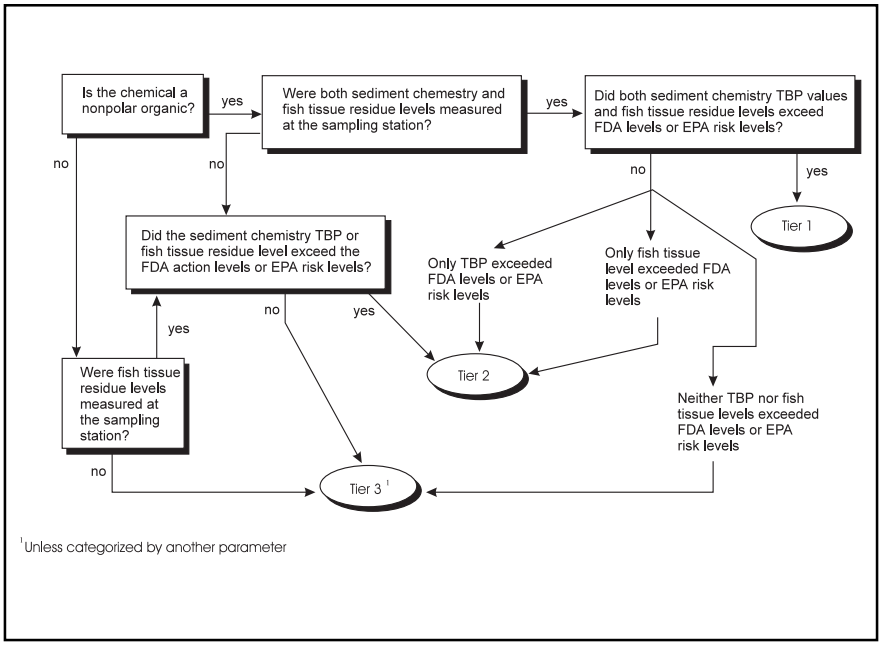

Figure 2-7 Human Health Assessments: Sediment Chemistry and Fish Tissue Residue Analysis (excluding dioxins and PCBs).

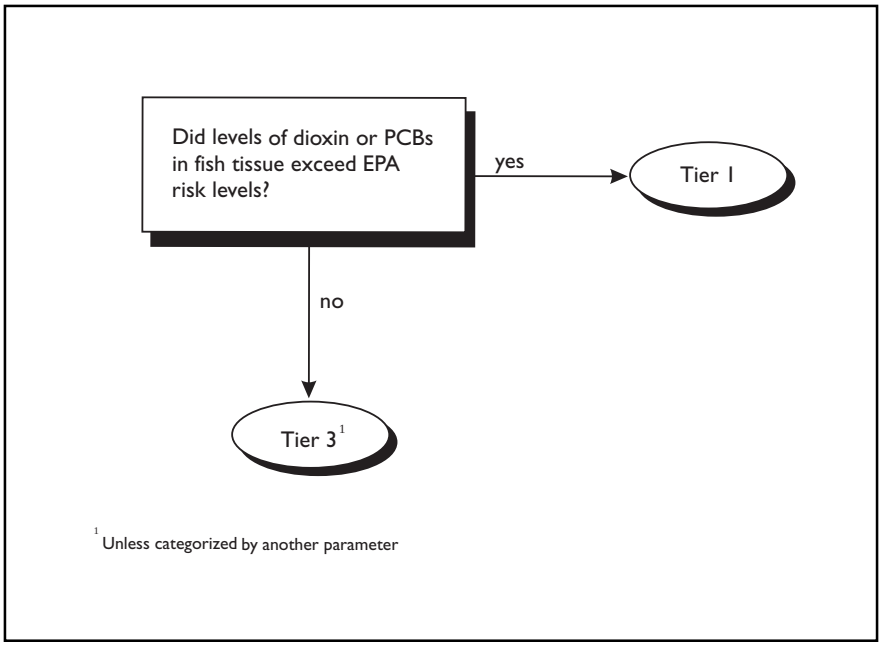

Figure 2-8. Human Health Assessments: PCBs and Dioxin in Fish Tissue Analysis.

screening analyses, the environmentally protective approach is to minimize Type II errors, which leave toxic sediment unidentified. To achieve a balance and to direct attention to areas most likely to be associated with adverse effects, Tier 1 sampling stations are intended to have a high rate of "correct" classification (e.g., sediment definitely posing or definitely not posing a threat) and a balance between Type I and Type II errors. On the other hand, to retain a sufficient degree of environmental conservatism in screening, Tier 2 sampling stations are intended to have a very low number of false negatives in exchange for a large number of false positives.

The numbered evaluation parameters used in the NSI data evaluation are briefly described below. A detailed description of the evaluation parameters is presented in Appendix B.

Sediment Chemistry Data

The sediment chemistry screening values used in this report are not regulatory criteria, site-specific cleanup standards, or remediation goals. Sediment chemistry screening values are reference values above which a sediment ecotoxicological assessment might indicate a potential threat to aquatic life. The sediment chemistry screening values used to evaluate the NSI data for potential adverse effects of sediment contamination on aquatic life include both theoretically and empirically based values. The theoretically based values rely on the physical/chemical properties of sediment and chemicals to predict the level of contamination that would not cause an adverse effect on aquatic life. The empirically based, or correlative, screening values rely on paired field and laboratory data to relate incidence of observed biological effects to the dry-weight sediment concentration of a specific chemical.

The theoretically based screening values used as parameters in the evaluation of NSI data include the sediment quality criteria, sediment quality advisory levels, and comparison of simultaneously extracted metals to acid-volatile sulfide concentrations. Empirically based, correlative screening values used in the NSI evaluation include the effects range-median/effects range-low values, probable effects levels/threshold effects levels, and apparent effects thresholds. The use of each of these screening values in the evaluation of the NSI data is described below. Another theoretically based evaluation parameter, the theoretical bioaccumulation potential (which was used for human health assessments), is also described below. The limitations associated with the use of these screening values are discussed in Chapter 5.

Sediment Chemistry Values Exceed EPA Draft Sediment Quality Criteria [1]

This evaluation parameter was used to assess the potential effects of sediment contamination on benthic species. EPA has developed draft sediment quality criteria (SQCs) for the following five nonionic organic chemicals:

- Acenaphthene (polynuclear aromatic hydrocarbon, or PAH)
- Dieldrin (pesticide)
- Endrin (pesticide)
- Fluoranthene (PAH)
- Phenanthrene (PAH)

EPA developed these draft criteria using the equilibrium partitioning (EqP) approach (described in detail in Appendix B) for linking bioavailability to toxicity. The EqP approach involves predicting the dry-weight concentration of a contaminant in sediment that is in equilibrium with a pore water concentration that is protective of aquatic life. It combines the water-only effects concentration (the chronic water quality criteria) and the organic carbon partitioning coefficient of the chemical normalized to the organic carbon content of the sediment. The draft criterion is compared to the measured dry-weight sediment concentration of the chemical normalized to sediment organic carbon content. If the organic-carbon-normalized concentration of the contaminant does not exceed the draft sediment quality criterion, adverse effects should not occur to at least 95 percent of benthic organisms. The draft SQCs are based on the highest quality data available, which have been reviewed extensively.

For the NSI data evaluation, sediment chemistry measurements with accompanying measured total organic carbon (TOC) values can place a site in Tier 1 based exclusively on a comparison with a draft SQC. The amount of TOC in sediment is one of the factors that determines the extent to which a nonionic organic chemical is bound to the sediment and, thus, the availability for uptake by organisms (bioavailability). If draft SQCs based on measured TOC were not exceeded, or if none of the five nonpolar organic chemicals that have been assigned draft SQC values were measured, the sampling station was classified as Tier 3 unless otherwise categorized by another parameter. Appendix B discusses the assumptions

and limitations associated with the use of draft SQCs. If a sample for any of the five contaminants for which draft SQCs have been developed did not have accompanying TOC data, the measured concentration was compared to the draft SQC based on a default TOC value of 1 percent. In these instances, the draft SQC was treated like other sediment quality screening values described later in this section.

The assumption that the percent TOC for samples without measured TOC is equal to 1 percent is based on a review of values published in the literature. TOC can range from 0.1 percent in sandy sediments to 1 to 4 percent in silty harbor sediments and 10 to 20 percent in navigation channel sediments (Clarke and McFarland, 1991). Long et al. (1995) reported an overall mean TOC concentration of 1.2 percent from data compiled from 350 publications for their biological effects database for marine and estuarine sediments. Ingersoll et al. (1996) reported a mean TOC concentration of 2.7 percent for inland freshwater samples. Based on this review of TOC data, EPA selected a default TOC value of 1 percent for the NSI evaluation. Consistent with the screening level application, this value should not lead to an underestimate of the bioavailability of associated contaminants in most cases.

Comparison of AVS to SEM Molar Concentrations [2, 5]

The use of the total concentration of a trace metal in sediment as a measure of its toxicity and its ability to bioaccumulate is problematic because different sediments exhibit different degrees of bioavailability for the same total quantity of metal (Di Toro et al., 1990; Luoma, 1983). These differences have recently been reconciled by relating organism toxic response (mortality) to the metal concentration in the sediment interstitial water (Adams et al., 1985; Di Toro et al., 1990). Acid-volatile sulfide (AVS) is one of the major chemical components that control the activities and availability of metals in interstitial waters of anoxic (lacking oxygen) sediments (Meyer et al., 1994).

A large reservoir of sulfide exists as iron sulfide in anoxic sediment. Sulfide will react with several divalent transition metal cations (cadmium, copper, mercury, nickel, lead, and zinc) to form highly insoluble compounds that are not bioavailable (Allen et al., 1993). It follows in theory, and with verification (Di Toro et al., 1990), that divalent transition metals will not begin to cause toxicity in anoxic sediment until the reservoir of sulfide is used up (i.e., the molar concentration of metals exceeds the molar concentration of sulfide), typically at relatively high dry-weight metal concentrations. This observation has led to a laboratory measurement technique of calculating the difference between simultaneously extracted metal (SEM) concentration and acid volatile sulfide concentration from field samples to determine potential toxicity.

To evaluate the potential effects of metals on benthic species, the molar concentration of AVS ([AVS]) was compared to the sum of SEM molar concentrations ([SEM]) for five metals: cadmium, copper, nickel, lead, and zinc. Mercury was excluded from AVS comparison because other important factors play a major role in determining the bioaccumulation potential of mercury in sediment. Specifically, under certain conditions mercury binds to an organic methyl group and is readily taken up by living organisms.

Sediment with measured [SEM] in excess of [AVS] does not necessarily exhibit toxicity. This is because other binding phases can tie up metals. However, research indicates that sediment with [AVS] in excess of [SEM] will not be toxic from metals, and the greater the [SEM]-[AVS] difference, the greater the likelihood of toxicity from metals. Analysis of toxicity data for freshwater and saltwater sediment amphipods (crustaceans) from EPA's Environmental Research Laboratory in Narragansett, Rhode Island, revealed that 80 to 90 percent of the sediments were toxic at [SEM]-[AVS] > 5 (Hansen, 1995; see also Hansen et al., 1996). Thus, EPA selected [SEM]-[AVS] = 5 as the demarcation line between Tier 1 and Tier 2. For the purpose of this evaluation, where [SEM]-[AVS] was greater than 5, the sampling station was classified as Tier 1. If [SEM]-[AVS] was between zero and 5, the sampling station was classified as Tier 2. If [SEM]-[AVS] was less than zero, or if AVS or the five AVS metals were not measured at the sampling station, the sampling station was classified as Tier 3 unless otherwise classified by another parameter. Appendix B discusses the assumptions and limitations associated with the [SEM]-[AVS] approach.

Sediment Chemistry Values Exceed Screening Values [3, 6]

Several sets of sediment contaminant screening values, developed using different methodologies, are available to assess potential adverse effects on benthic species. The screening values selected for comparison with measured sediment levels are the draft SQCs using a default TOC of 1 percent (for those samples which do not have accompanying TOC data), sediment quality advisory levels (SQALs) for freshwater aquatic life (developed using the equilibrium partitioning approach discussed previ-

ously for the development of draft SQCs), the effects range-median (ERM) and effects range-low (ERL) values developed by Long et al. (1995), the probable effects levels (PELs) and threshold effects levels (TELs) developed for the Florida Department of Environmental Protection (FDEP, 1994), and the apparent effects thresholds (AETs) developed by Barrick et al. (1988). The assumptions and approaches used to develop these screening values are discussed in detail in Appendix B.

The draft SQCs and SQALs were both developed using the same EqP approach. However, the data used to derive SQALs were not compiled from an exhaustive literature search, nor were the toxicity data requirements as extensive as specified for draft SQCs. Toxicity values used for SQAL development include final chronic values from EPA ambient freshwater quality criteria and secondary chronic values derived using EPA's Great Lakes Water Quality Initiative "Tier II" water quality criteria methodology. The data used to develop the latter values were taken primarily from quality-screened studies in published literature. The development of SQALs is discussed in further detail in Appendix B of this report. EPA has also prepared a document describing the derivation of the SQALs (USEPA, 1996). The chemicals for which SQALs have been developed are identified in Appendix D of this volume.

The ERLs/ERMs, PELs/TELs, and AETs relate the incidence of adverse biological effects to the sediment concentration of a specific chemical at a specific sampling station using paired field and laboratory data. The developers of the ERLs/ERMs define sediment concentrations below the ERL as being in the "minimal-effects range," values between the ERL and ERM in the "possible-effects range," and values above the ERM in the "probable-effects range." In the FDEP (1994) approach, the lower of the two guidelines for each chemical (the TEL) is assumed to represent the concentration below which toxic effects rarely occur. In the range of concentrations between the TEL and PEL, effects occasionally occur. Toxic effects usually or frequently occur at concentrations above the upper guideline (the PEL).

In independent analyses of the predictive abilities of the ERL/ERMs and TEL/PELs, the precentages of samples indicating high toxicity in laboratory bioassays of amphipod survival were relatively low (10-12 percent) when all chemical concentrations were in the minimal effects range, intermediate (17-19 percent) in the possible effects range, and higher (38-42 percent) in the probable effects range. Furthermore, the percentages of samples indicating high toxicity in any one of a battery of 2-4 tests performed, including more sensitive bioassays with sublethal endpoints, were 5-28 percent, 59-64 percent, and 78-80 percent among samples within the minimal, possible, and probable effects ranges (Long et al., in press).

The AET approach is not based on the probability of incidence of adverse biological effects. The AET is the highest concentration at which statistically significant differences in observed adverse biological effects from reference conditions do not occur, provided that the concentration also is associated with observance of a statisically significant difference in adverse biological effects. Essentially, this identifies the concentration above which an adverse biological effect always occurs for a particular data set. Barrick et al. (1988) list specific AET values for several different species or biological indicators. For the purposes of this assessment, EPA defined the AET-low as the lowest AET among applicable biological indicators, and the AET-high as the highest AET among applicable biological indicators. By the nature of how the AET is derived, less stringent values might evolve as more data sets become available.

For the NSI data evaluation, the upper screening values were considered to be the ERM, PEL, draft SQC (when using default TOC value of 1 percent), SQAL, and AET-high for a given chemical. The lower screening values were considered to be the ERL, TEL, draft SQC (when using default TOC of 1 percent), SQAL, and AET-low for a given chemical. Because they are not based on ranges of effects, the single freshwater aquatic life draft SQC and SQAL values for a given chemical served as both the high and low screening values.

For a sampling station to be classified as Tier 1, a chemical measurement must have exceeded at least two of the upper screening values. If a sediment chemistry measurement exceeded any one of the lower screening values, the sampling station was classified as Tier 2. If sediment concentrations at a sampling station did not exceed any screening values or there were no data for chemicals that have assigned screening values, the sampling station was categorized as Tier 3 unless otherwise categorized by another parameter.

Under this approach, a sampling station could be classified as Tier 1 from elevated concentrations of cadmium, copper, lead, nickel, or zinc based only on a comparison of [SEM] to [AVS]; that is, sampling stations could not be classified as Tier 1 based on an exceedance of two upper screening values for any of the five metals. However, sampling stations were classified as Tier 2 for these five metals based on an exceedance of one of the lower screening values if AVS data were not available.

Sediment Chemistry TBPs Exceed Screening Criteria [4, 7]

This evaluation parameter addresses the risk to human consumers of organisms exposed to sediment contaminants. The theoretical bioaccumulation potential (TBP) is an estimate of the equilibrium concentration (concentration that does not change with time) of a contaminant in tissues if the sediment in question were the only source of contamination to the organism. At present, the TBP calculation can be performed only for nonpolar organic chemicals. The TBP is estimated from the concentration of contaminant in the sediment, the organic carbon content of the sediment, the lipid content of the organism, and the relative affinity of the chemical for sediment organic carbon and animal lipid content. This relative affinity is measured in the field and is called a biota-sediment accumulation factor (BSAF, as discussed in detail in Appendix C). In practice, field measured BSAFs can vary by an order of magnitude or greater for individual compounds depending on location and time of measurement. For this evaluation, EPA selected BSAFs that represents the central tendency, suggesting an approximate 50 percent chance that an associated tissue residue level would exceed a screening risk value.

In the evaluation of NSI data, if a calculated sediment chemistry TBP value exceeded a screening value derived using standard EPA risk assessment methodology or the Food and Drug Administration (FDA) tolerance/action or guidance level, and if a corresponding tissue residue level for the same chemical for a resident species at the same sampling station also exceeded one of those screening values, the station was classified as Tier 1. Individual chemical risk levels were considered separately; that is, risks from multiple contaminants were not added. Both sediment chemistry and tissue residue samples must have been taken from the same sampling station. If tissue residue levels for the same chemical for a resident species at the same sampling station did not exceed EPA risk levels or FDA levels or there were no corresponding tissue data, the sampling station was classified as Tier 2. If neither TBP values nor fish tissue residue levels exceeded EPA risk levels or FDA levels, or if no chemicals with TBP values, EPA risk levels, or FDA levels were measured, the sampling station was classified as Tier 3 unless otherwise classified by another parameter. A detailed description of the methods used to develop TBP values and to determine the EPA risk levels used in this comparison is presented in Appendix B.

Tissue Residue Data [8, 9, 10]

Tissue residue data were used to assess potential adverse effects on humans from the consumption of fish that become contaminated through exposure to contaminated sediment. Only those species considered benthic, non-migratory (resident), and edible by human populations were included in human health assessments. A list of species included in the NSI and their characteristics is presented in Appendix F.

Sampling stations at which human health screening values for dioxin and PCBs were exceeded in fish tissues were classified as Tier 1. For these chemicals, corroborating sediment chemistry data were not required. If human health screening values for dioxin or PCBs in fish tissue were not exceeded or if neither chemical was measured, the sampling station was classified as Tier 3 unless otherwise classified by another parameter.

For other chemicals, both a tissue residue level exceeding an FDA tolerance/action or guidance level or EPA risk level and a sediment chemistry TBP value exceeding that level for the same chemical were required to classify a sampling station as Tier 1. If tissue residue levels exceeded FDA levels or EPA risk levels but corresponding TBP values were not exceeded at the same station (or there were no sediment chemistry data from that station), the sampling station was classified as Tier 2. If neither fish tissue levels nor TBP values exceeded EPA risk levels or FDA levels, or if no chemicals with TBP values, EPA risk levels, or FDA levels were measured, the sampling station was classified as Tier 3 unless otherwise classified by another parameter.

Toxicity Data [11, 12]

Toxicity data were used to classify sediment sampling stations based on their demonstrated lethality to aquatic life in laboratory bioassays. Nonmicrobial sediment toxicity tests with a mortality endpoint were evaluated. Toxicity test results that lacked control data, or had control data that indicated greater than 20 percent mortality (less than 80 percent survival), were excluded from further consideration. The EPA has standardized testing protocols for marine and freshwater toxicity tests. A review of several protocols for sediment toxicity tests suggests that mortality in controls may range from 10 to 30 percent, depending on the species, to be considered an acceptable test result (API, 1994). Current amphipod test requirements indicate that controls should have less than 10 percent mortality (API, 1994; USEPA, 1994b).

For the NSI data evaluation, EPA considered significant toxicity as a 20 percent difference in survival from control survival. For example, significant toxicity occurred if control survival was 80 percent and experimental survival was 60 percent or less.

For this evaluation parameter, corroboration of multiple tests was considered more indicative of probable associated adverse effects than the magnitude of the effect in a single test. Lethality demonstrated by two or more single-species tests using two different test species (at least one of which had to be a solid-phase test) placed a sampling station in Tier 1. A sampling station was classified as Tier 2 if toxicity was demonstrated by one single-species nonmicrobial toxicity test. If lethality was not demonstrated by a nonmicrobial toxicity test, or if toxicity test data were not available, the sampling station was classified as Tier 3 unless otherwise classified by another parameter.

Incorporation of Regional Comments on the Preliminary Evaluation of Sediment Chemistry Data

Several reviewers from different EPA Regions and states provided comments on the May 16, 1994, preliminary evaluation of sediment chemistry data. The comments included more than 150 specific comments identifying additional locations with contaminated sediment that had not been identified in the preliminary evaluation. Since the preliminary evaluation, the final NSI methodology has been developed and implemented. The updated methodology has been refined significantly to include tissue residue and toxicity data as well as revised screening values. Data corresponding to any additional comments that required further review were divided into two categories: (1) data that incorrectly identified contaminated sediment and (2) additional water bodies that contain areas of sediment contamination. The first category primarily addressed sampling stations identified in the preliminary assessment as exceeding sediment chemistry screening values for specific contaminants that reviewers stated were located in water bodies that are not contaminated from the chemical(s) in question.

EPA examined all NSI sampling stations that had been identified in the preliminary evaluation as exceeding a sediment quality screening value, but were located in water bodies that reviewers of the preliminary evaluation identified as not being contaminated by that specific contaminant or contaminants. If the sampling station in question was classified in this final evaluation as Tier 1 based only on the specific contaminant(s) identified by the reviewer as not being a problem, the sampling station was removed from the Tier 1 category and placed in the Tier 3 category. Only a few sampling stations were moved from the Tier 1 category to the Tier 3 category as a result of this procedure. Stations identified in the NSI evaluation as Tier 1 based on other chemicals not identified by the reviewer or because of toxicity data were not removed from Tier 1.

Additional water bodies that reviewers identified as potential areas of significant contamination were evaluated to determine whether sampling stations along those water bodies were classified as Tier 1 based on the final NSI data evaluation. Locations or water bodies identified by reviewers as potential areas of significant contamination are discussed separately in the results (Chapter 3).

Evaluation Using EPA Wildlife Criteria

In addition to the evaluation parameters described above and presented in Table 2-2, EPA conducted an assessment of NSI data based on a comparison of sediment chemistry TBP values and fish tissue values to EPA wildlife criteria developed for the Great Lakes. This evaluation, however, was not included with the results of evaluating the NSI data based on the other parameters. The results of evaluating NSI data based on wildlife criteria are presented in a separate section of Chapter 3. Wildlife criteria based solely on fish tissue concentrations were derived for EPA wildlife criteria for water that are presented in the *Great Lakes Water Quality Initiative Criteria Documents for the Protection of Wildlife* (USEPA, 1995a). EPA has developed wildlife criteria for four contaminants: DDT, mercury, 2,3,7,8-TCDD, and PCBs. The method to adjust these wildlife criteria for the NSI data evaluation is explained in detail in Appendix B.