1.0 OVERVIEW OF NEW ENGLAND NUTRIENT DATABASE FOR RIVERS AND STREAMS

1.1 Purpose and Goal of Data Distribution Report

The purpose of the Data Distribution Report is to summarize and document the results of ENSR's acquisition and database entry of qualified nutrient-related data to support and facilitate the development of regional nutrient criteria for rivers/streams. This document is a deliverable under Phase III, Year 2 of the "Collection and Evaluation of Ambient Nutrient Data" Project being conducted for the New England Interstate Water Pollution Control Commission ("the Commission"). This document identifies and describes data and its sources collected and evaluated for inclusion in the New England Nutrient Database for Rivers and Streams ("the Nutrient Database"). An analogous data collection and database development was conducted for lake and ponds and resulted in the New England Nutrient Database for Lakes/Ponds/Reservoirs (NEIWPCC, 2000). One purpose of the Data Distribution Report is to evaluate the sufficiency of available data and see whether the target number of waterbodies suggested for New England sub-ecoregions are achieved or whether additional data collection may be recommended. The final project deliverable will be the Data Synthesis and Final Report ("Final Report") which will build on and expand the material presented in the Data Distribution Report. In the Final Report the finalized Database will be fully described and summarized, potential methods of developing draft nutrient criteria will be explored, and any outstanding issues that may need to be addressed further will be identified.

Specific objectives of the Data Description Report are as follows:

- Provide sufficient regulatory background as a framework for the project objectives;
- Document and describe the sources of electronic nutrient data acquired;
- Describe the basic structure and features of the draft Database;
- Describe and summarize the contents of the draft Database with regard to amount of data, number of waterbodies, parameters of interest, ecoregional coverage, etc.
- Refine the draft Database to identify waterbodies of interest (i.e., those for which sufficient nutrient data is available for comparison and analysis);
- Identify and justify selection of reference sites/conditions;
- Compare numbers of waterbodies on ecoregional basis to target numbers to identify data gaps
- Provide a strategy to address any data gaps identified; and
- Identify next steps and components for development of the Final Report.

1.2 Relationship of Nutrient Database to Regional Nutrient Criteria Development

Development of regional waterbody-specific nutrient criteria is a national priority first identified in the *National Strategy for the Development of Regional Nutrient Criteria* (U.S. EPA, 1998). The U.S. EPA has issued Ambient Water Quality Criteria Recommendations (or "reference conditions") for nutrients for rivers and lakes in the 14 national ecoregions and States must make significant progress towards adopting nutrient criteria as water quality standards by the end of 2004. U.S EPA has also issued waterbody-specific technical guidance, in the form of the *Nutrient Criteria Technical Guidance Manual Rivers and Streams* (U.S. EPA, 2000a.) For New England, U.S

EPA has established numeric nutrient criteria recommendations for rivers in Ecoregions VIII and XIV (U.S. EPA, 2000b; 2001).

This project is being conducted as part of the overall US EPA National Strategy, with the stated objective to investigate promising approaches to ecoregion-specific nutrient criteria and to assist the states in their development of implementation plans to adopt nutrient criteria. The major elements of this strategy are presented below, with those elements most relevant to this project marked in italics:

- *Use of regional and waterbody-type approach for the development of nutrient criteria***;**
- Development of waterbody-type technical guidance documents (i.e., documents for streams and rivers; lakes and reservoirs; estuaries and coastal waters; and wetlands) that will serve as "user manuals" for assessing trophic state and developing region-specific nutrient criteria to control overenrichment;
- *Establishment of an US EPA National Nutrient Team with Regional Nutrient Coordinators to develop regional databases and promote State and Tribal Involvement;*
- *Development by US EPA of nutrient water quality criteria in the form of numerical regional target ranges, which US EPA expects States and Tribes to use in implementing State management programs to reduce overenrichment in surface waters, i.e., through the development of water quality criteria, standards, NPDES permit limits, and total maximum daily loads (TMDLs***)**; and
- Monitoring and evaluation of the effectiveness of nutrient management programs as they are implemented.

To support this effort in New England, ENSR was contracted by the Commission to construct a regional database from existing Federal, State, academic and Tribal nutrient data. The development of the regional database for Rivers and Streams followed the following tasks:

- **Collection of Electronic Data** – recent vintage (i.e., 1990 or later) electronic databases of nutrient, trophic status response indicators, and ancillary water quality, flow, and watershed information will be obtained from Federal, State, Tribal sources, as well as other qualified sources (i.e., academic institutions, watershed groups);
- **Conduct QA/QC Reviews** – prior to inclusion into the regional database, information will reviewed and documented with regard to accuracy, sufficiency, representativeness, and analytical quality. Data will be separated into those to be incorporated into the database and those deferred (and broadly classified as to quality) for later consideration (see Data Gap Analysis);
- **Data Distribution Report** – based on the primary data collection efforts, a Data Distribution Report will be generated that describes the nature and extent of the qualified waterbody nutrient data, along with summary statistics and preliminary analyses. The Data Distribution Report will be examined to identify potential data gaps; and
- **Data Synthesis and Final Report** – the completed regional database will be presented with complete description of its development and a wide array of statistical comparisons to support nutrient criteria decision-making.

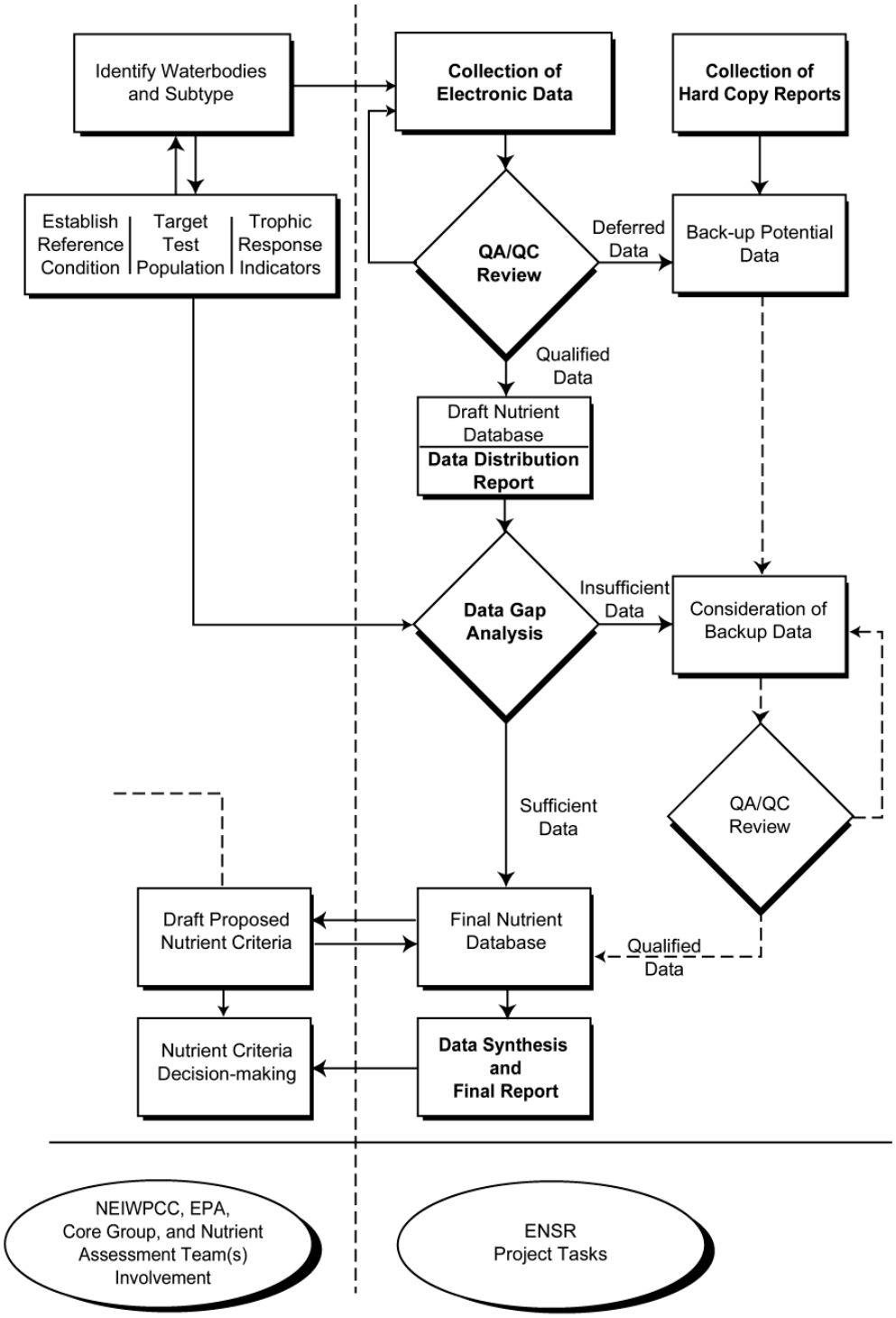
The general relationship between project tasks and the specific objectives is depicted in Figure 1-1, which provides a simplified flowchart indicating the sequence for development of the New England Nutrient Database for Rivers and Streams and its applicability to development of nutrient criteria.

The approach used for acquiring and classifying nutrient data in the database was originally described in a Technical Memorandum (April 1999) during Phase I of this project. The Technical Memorandum was distributed and presented to the project "Core Group" consisting of a selected number of state and federal agency contacts, regional technical assistance groups (RTAGs), U.S. EPA, New England Region, and the Commission. The approach described by the Technical Memorandum was reviewed, discussed, and approved in principle by the Core Group (who act as liaisons with RTAG members), U.S. EPA, and the Commission. This same approach was used in developing the R/S Nutrient Database

1.3 Organization of Report

This report is organized in the following fashion. Section 1.0 contains background material. Section 2.0 identifies the data sources for the New England Nutrient Database for Rivers and Streams. The structure and framework of the Database are described in Section 3.0. Section 4.0 contains the Development Strategy used to "refine" the initial database into a more focused and useable set of waterbodies and parameters. Section 5.0 provides initial Summary Statistics on the waterbodies and major parameters of interest. A Data Gaps Analysis to identify data needs is provided in Section 6.0 and a Summary is presented in Section 7.0.

FIGURE 1-1
Development of Nutrient Database to Support Nutrient Criteria Decision-making



M990160

2.0 SOURCES OF DATA

2.1 Data Sources

The primary goal of the project is to collect and analyze good quality data to help establish the basis and justification for regional nutrient criteria. To provide for this good quality database, nutrient data, trophic status response indicators, ancillary water quality parameters, flow, and watershed information on waterbodies in New England were acquired from a variety of qualified sources including state and Federal agencies, Tribal sources, academic institutions, watershed groups, and other sources. A list of the databases that were primary sources of data and the respective contact person are listed in Table 2-1. The data requested was for electronic databases of a fairly recent vintage (i.e., 1990 or later). A compilation of the databases used, with a brief description of the dataset, its parameters, and contact person is contained in Appendix A.

2.2 Spatial Data

The Geographical Information System (GIS) software ArcView (ESRI) was used to support some of the data mining and analysis tasks. The GIS interface was used to distinguish waterbodies from those with similar names, and to identify the correct ecoregion for each waterbody. Specific websites that were used are noted below. The spatial coverages were obtained electronically – from internet sites in most cases – and used to complete the database.

The EPA Non-Aggregated Ecoregions for United States were obtained from the EPA ftp site (ftp://cerberus.epa.orst.edu). The file covers the entire United States and shows five ecoregions within New England. Delineations of the 8-digit Hydrologic Units Codes (HUCs) were obtained from the USGS web site (www.usgs.gov). Counties and towns political boundaries were obtained from the ESRI Data CD of New England.

Additional spatial data sets were also obtained from each of the state's official GIS data web sites to help support technical analyses. These state-specific GIS sites are:

- Connecticut: http://magic.lib.uconn.edu/
- Maine: http://apollo.ogis.state.me.us/
- Massachusetts: http://www.magnet.state.ma.us/mgis/
- New Hampshire: http://nhresnet.sr.unh.edu/granit/overview.htm
- Rhode Island: http://www.edc.uri.edu/rigis/
- Vermont: http://geo-vt.uvm.edu/

Land use attributes for sampling station watersheds were derived from the USGS SPatially Referenced Regressions On Watershed Attributes (SPARROW) model. The SPARROW model is described on the web site (http://water.usgs.gov/nawqa/sparrow/).

Table 2-1 Organizational Contacts for Waterbody and Nutrient Data

Organization	Contact	Dataset Description	Example of Parameters	Years
Connecticut				
CTDEP	Chuck Lee charles.lee@po.state.ct.us	Electronic data for a limited number of lakes	ALK, TN, TP, TSS	1990-1997
CTDEP	Chuck Lee charles.lee@po.state.ct.us	Series of hardcopy reports on a number of lakes in CT	ALK, CHLA, pH, SDT, TN, TP	1989-1995
Maine				
ME-DEP	Roy Bouchard roy.j.bouchard@state.me.us (207) 287-7798	Lakes and ponds data	ALK, CHLA, DO, pH, SDT, Temperature, TP	1952-1998
ME-DEP	Paul Mitnik (207) 287-6093	River and streams data	CHLA, pH, SDT, TP, some nitrogen	1989-1998
Penobscot Indian Nation	Dan Kusnierz Pinwater@mint.net	Rivers data for the Penobscot Watershed, ME	CHLA, SDT, Temperature, TN, TP, TSS	1994-1997
Massachusetts				
MADEP	Tom Dallaire thomas.dallaire-eqe@state.ma.us	Data for brooks, rivers and streams sampled at 444 stations within 37 basins	ALK, DO, pH, Temperature, TP, TSS, some nitrogen	1994-1998
MADEP	Rick McVoy Rick.mcvoy@state.ma.us	Series of hardcopy reports on a number of lakes in MA	CHLA, pH, SDT, TN, TP	1980-1988
UMASS Acid Rain Monitoring Project	Paul Godfrey godfrey@tei.umass.edu	Predominately data for ponds and streams located in 15 different counties in MA	ALK, NO3, pH, TP	1983-1993
New Hampshire				
NHDES	Bob Estabrook, (603) 271-3357 r.estabrook@des.state.nh.us	Data for most lakes in NH.	ALK, CHLA, DO, pH, SDT, Temperature, TKN, TP, Weeds	1976-1999

Organization	Contact	Dataset Description	Example of Parameters	Years
Vermont				
VTDEC	Eric Smeltzer (802) 241-3792 eric.smeltzer@anmail.state.vt.us	Hundreds of lakes and ponds.	CHLA, phytoplankton, SDT, TN, TP	1975-1998
Rhode Island				
RIDEM	Connie Carey (401) 222-4700 ext. 7239 ccarey@dem.state.ri.us	River information	DO, pH, Temperature, TP, TSS, some nitrogen	1991-1997
URI Cooperative Extension (Watershed Watch Program)	Linda Green uriww@etal.uri.edu	Data on lakes and ponds	CHLA, SDT, TN, TP	1995-1998
RIUSGS			DO, pH, Temperature, some nitrogen	1989-1997
National Level				
STORET	Dan Parker Parker.dan@epamail.epa.gov	Stations in CT (40), MA (5), NH (11) and VT(288)	DO, SDT, Temperature, TP, some nitrogen	1990-1997
EMAP	Spence Peterson (541) 754-4457	370 lakes in Northeast Lake Survey	CHLA, SDT, TN, TP, Turbidity	1991-1994
USGS	Keith Robinson (603) 226-7809 kwrobins@usgs.gov	Data on rivers and streams		

3.0 DATABASE DESCRIPTION

The New England Nutrient Database was assembled from the data acquired from the data sources identified in Section 2.0. A description of the structure of the Database is given in Section 3.1. The main data tables are described in Section 3.2. The Quality Assurance / Quality Control measures taken in reviewing, verifying, and accepting the data are described in Section 3.3.

3.1 Database Structure

A Relational database was designed and implemented in Microsoft Access97 to accumulate and manipulate the extensive amount of available electronic data. This database was adapted from an existing one provided by national US EPA headquarters. It has been revised and adapted to meet the needs of this project. A relational database is a collection of data items organized as a set of formally-described tables that are linked into a logical structure. The New England Nutrient Database includes tables and queries. Tables are collections of data on a given topic. Their content and the relationships defined between the different tables form the core of the database applications. Queries present a certain view of the data contained in tables, or may be used to update, append or edit data records.

The data was organized into four main tables each representing one level of information, as shown in Figure 3-1. These tables contain information on the waterbody, station, sample, and water quality data,

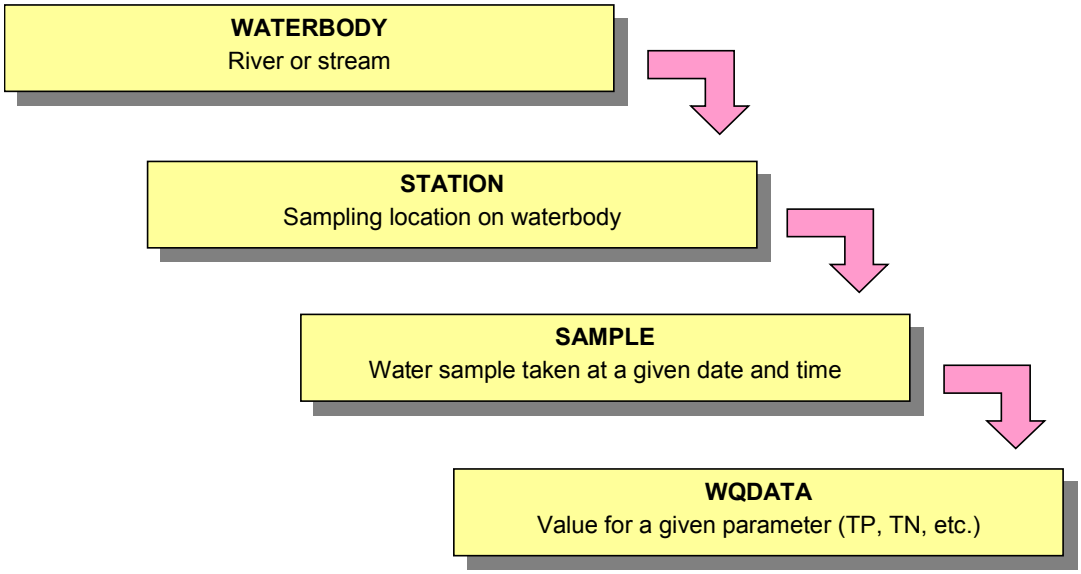


Figure 3-1: Database Main Data Tables Structure.

respectively. The tables are linked to each other through one-to-many relationships with enforced referential integrity. Referential integrity means that records in each main (or so-called “parent”) table are unique but may be associated with one or more derivative (or so-called “child”) records in other tables. As such, a given waterbody may have one or more stations, each measured at one or more

points in time, and each water sample may have been analyzed for one or more parameters. This staged structure ensures that each data item appears once only in the database, eliminating duplicate information and minimizing possible errors.

Within a given table, uniqueness of information is enforced through a single unique key field or unique combinations of fields. In the waterbody and station tables, a single field contains the identification of a unique record, the waterbody_ID and station_ID fields, respectively. In the case of the sample table, a unique record is one with a unique combination of Station ID, Sampling Date, Sampling Time, Sample Depth, and Sample Type. In table WQData, a unique record is one with a unique combination of Sample ID, Parameter, and Reported Value.

In addition to the four main data tables, a number of lookup tables have been developed to provide the possible range of values or categories for some of the fields. The relationships between the main data tables and lookup tables are indicated in Table 3-1.

Table 3-1: Relationships between Main Data Tables and Lookup Tables.

Lookup table	Main Table	Field	Source for Link Field
LTBL_AnalysisMethod	WQData	Analysis Method	Code for analysis method used.
LTBL_EPAEcoregion	Waterbody	EPA Ecoregion	Name of non-aggregated ecoregions for New England .
LTBL_Parameters	WQData	Parameter	Code for chemical/biological/physical parameter measured
LTBL_Qualifier	WQData	Reported_Qualifier	Remark on value reported. Unless specified, codes are same as used in STORET.
LTBL_Sample_Type	Sample	Sample Type	Type of sample collected (target, duplicate, etc.)
LTBL_Sampling_Conditions	Sample	Sampling Conditions	Conditions at time of sampling (dry, wet, unknown)
LTBL_Sampling_Method	Sample	Sampling Method	Sampling method used (grab, hose, composite, etc.)
LTBL_State	Waterbody + Station	State	Two-letter postal abbreviation.
LTBL_Units	WQData	Unit of Measure	Abbreviation of measurement units
LTBL_WaterbodyType	Waterbody	Waterbody Type	Code for waterbody type (P, R, S, M, O)

3.2 Main Data Tables

As noted above, the data is contained in four main data tables representing different levels of information. A listing of the fields found in each of the main data tables is provided in Appendix B. This section discusses some of the implications of the logical organization of the data.

The table Waterbody contains information that is specific to a given waterbody. A waterbody is defined as a body of water with finite, well-defined extents and relatively homogeneous physical characteristics. A waterbody can be an entire river or stream, or a specified segment or reach along a

given river or stream. The subdivision of rivers and streams into segments with relatively homogeneous characteristics is complicated, as depth, flow, and other physical characteristics are expected to change with the distance from the headwaters.

The tables Waterbody and Station contain information at two different levels of spatial extent. The table Waterbody contains overall characteristics of the waterbody while the table Station refers to a specific location on that waterbody. For example, a waterbody may have stations located in different ecoregions.

3.3 Quality Assurance and Quality Control Issues

An important part of the project was Quality Assurance / Quality Control ("QA/QC"). The following section addresses important QA/QC issues for the Database.

3.3.1 Data Import and Database Structure

The majority of the data was obtained electronically from qualified sources in the form of databases or spreadsheets. In most cases, the format of the data received needed only to be manipulated slightly to make it compatible for importing into the Nutrients Database. As such, data entry errors were assumed to be limited to those that could have taken place in the original data source.

The database enforces referential integrity of the information. For example, records can only refer to existing "parent" records (e.g., sample at existing stations). In many cases, unique identifiers were defined that prevent the duplication of information such as lake name, station ID, etc. The referential integrity check also prevents the importation of unassociated or so-called "orphan") data (i.e. data without associated sample, station, or waterbody). The use of lookup tables to provide a limited choice of valid values for some of the fields in the main tables also ensures minimal error in the content of the database. This ensures consistency of values and codes across data sources. For example, water quality parameters are limited to values listed in the Parameters lookup table.

3.3.2 Duplication of Data Between Data Sources

Because of the large number of data sources utilized, and the realization that some waterbodies potentially had measurements reported by two or more different agencies, the water quality measurements present in the draft Database was scanned for duplicates. This verification was performed by comparing the combination of waterbody, sampling date, sampling depth, parameter and value reported. In cases were more than one unique such "combination" was found for different data sources (e.g., between STORET records and a State Agency electronic file), the duplicate STORET record was flagged as non-useable, and not included in subsequent data analyses. Duplicates within a single data source were assumed to be legitimate and were identified as "DUP" in the sample type field.

3.3.3 Additional Verifications

As noted in Section 3.1, no attempt was made to verify the electronic data submitted by the agencies. However, data for selected trophic parameters within the refined Database (described in Section 4.0) were compared with a likely range of values (based on best professional judgment) to insure that the reported values were within the range of "reasonable" values. Reported values for total phosphorus (TP), total nitrogen (TN), and chl-*a* were compared to the reasonable range (Table 3-2). Reported values that were outside of the range were further investigated and verified against the original source

of the data. Negative and null concentrations were also searched for and investigated. When data was outside the expected range and there was some potential explanatory factor readily available (negative values, unit errors, etc), the data was removed from the database. On the other hand, some reported values were outside of the range, but there was no reason to question the accuracy of the data. In these cases, the values were retained in the Database.

Table 3-2: "Reasonable" Range of Values expected for Trophic Parameters

Trophic Parameter	Minimum Value	Maximum Value
Chl *a* (ug/L)	0	250
TN (u/gL)	0	5,000
TP (ug/L)	0	5,000

Many data contributors had initially estimated latitudes and longitudes of sampling stations from USGS quadrangles or road atlases. Since the Global Positioning System (GPS) has become widely available in the interim, agencies were requested to provide updated station coordinates for stations lacking coordinates or for stations with potentially inaccurate coordinates as identified with GIS and SPARROW.

4.0 DEVELOPMENT OF NUTRIENT DATABASE

This section describes the initial draft Database and its contents (Section 4.1). Due to the size of the initial draft database and the inclusion of many water quality records of lesser importance to the development of regional nutrient criteria, a subsequent "refined" New England Nutrient Database for Rivers and Streams ("Refined Database") was developed. Section 4.2 describes the strategy used to develop the Refined Database and Section 4.3 provides a summary of its contents. Section 4.4 discusses the sequence for data processing for averaging the data from an individual waterbody.

4.1 Initial Database Waterbody and Parameter Inventory

Historical water quality and ancillary data was collected from a multitude of sources that included federal and state agencies, volunteer groups and a Native American Nation. The data sources are previously discussed in Section 2.0. The initial draft New England Nutrients Database assembled from these data contained over 2,150 rivers and streams and over 172,000 water quality data records. Some of the features of the initial Database are discussed below.

4.1.1 Distribution of Data Sources

The distribution of the water quality measurements by source of data is presented in Figure 4-1. The data came from twelve sources: state and federal agencies, academic institutions, and a Native American Nation. It should be noted that this distribution represents all of the water quality records in the initial databse, and is quite different from the distribution in the Refined Database, which consists solely of data from stations where one or more of the key trophic indicators (Chl-a, SDT, TN, and TP) was sampled.

4.1.2 Period Covered

The initial draft New England Database contains data from June 1980 to August 2002, although the vast majority of data was collected from 1990 to 2001. The temporal distribution of the data for the selected trophic parameters is presented in Figure 4-2. This graphic presents the number of records available for each year for the selected trophic parameters. As indicated on Figure 4-2, the period 1990-91 provided the largest contribution of nutrient data, but significant contributions where made to the database during all periods of interest, without one period unduly over-represented.

A further breakdown of the data by season is presented in Figures 4-3 and 4-4 for the key trophic parameters TP and TN, respectively. It can be seen that TP was typically sampled in the spring (to capture spring runoff events) and in summer, with minimum sampling in winter. For TN, the summer sampling typically season predominates among data records. The seasonal distribution of data were further evaluated in the calculation of representative parameter values (see Section 4.4).

4.1.3 Water Quality Measurements

The initial database contains more than 172,000 water quality measurements. However, because of the diverse goals of the various monitoring programs that provided the information (e.g., Acid Rain Monitoring (ARM) Program), a large portion of the data reported are for parameters that are not necessarily directly related to nutrients, such as alkalinity, temperature, and pH. Whereas these

parameters may be potentially useful in allowing secondary classification of the waterbodies, they do not provide information directly applicable to the trophic status of the waterbody. Conversely, some of the nutrient data was not appropriate for application to surface waterbodies (e.g., groundwater nitrate records).

Despite these limitations, the identified critical trophic parameters of interest have a satisfactory representation within the Refined Database. This includes about 11,370 TP records, 3,880 TN records, and 1,490 Chl-*a* records. The least-represented trophic parameter is SDT, with less than 590 records.

4.1.4 Distribution of Waterbodies

The distribution by state of the of the rivers and streams contained in the initial draft Database is presented in Table 4-1. Massachusetts represented a large fraction of the sampled waterbodies in the initial Database. However, many of these waterbodies were sampled as part of the ARM program and were typically not sampled for nutrients.

Table 4-1 Number of New England Rivers and Streams in Initial Database by State.

State	Rivers and Streams
Connecticut	153
Massachusetts	1,613
Maine	46
New Hampshire	218
Rhode Island	91
Vermont	41
New England Total	**2,162**

4.2 Development of a Refined Database

At the end of the initial data collection period, the Database contained a large number of waterbodies. (>2,150) and a very large number of water quality records (>172,000). While this amount of data is impressive, it resulted in a cumbersome database that was difficult to perform standard calculations on and analyses of due to it sheer complexity. More importantly, the initial Database also contained much data not directly applicable to the issue of developing regional nutrient criteria; although, as noted above, some of the data may be useful for further correlation with and/or categorization of waterbodies. In addition there were pragmatic considerations regarding the availability of ancillary information for the selected waterbodies. For example, it was necessary to identify the spatial coordinates (i.e., latitude and longitude) for each waterbody sampling station to assign watershed and ecoregional status. Therefore, it was considered prudent to first reduce the size of the database to those waterbodies and

qualified data necessary for further analyses and investigations to support nutrient criteria development.

Accordingly, a decision was made to produce a second, smaller and more focused database. We have used the term "Refined Database" to refer to this effort since it represents a distillation of the information in the initial Database. Since the purpose of the project is to provide a database for further analyses and investigations to support regional nutrient criteria, the Refined Database contains only those rivers and streams for which information is available on the relevant trophic parameters. The purpose and strategy for development of the Refined Database was discussed and consensus reached with the Commission and US EPA Regional Nutrient Coordinator in meetings during summer 1999. This approach was presented to the Regional Nutrient Assessment Team at the September 30, 1999 meeting and is consistent with the overall goals of the program. While the approach was originally designed for application to lakes and ponds, it is considered a sound and appropriate approach for rivers and streams as well.

Briefly, the strategy acknowledges that not all waterbodies were sampled for the key trophic parameters (Chl-a, SDT, TN *and* TP). In fact, less than 2% of the waterbodies in the initial Database had information for three of these four parameters. Comparison of the number and location of these waterbodies indicated that this was an insufficient number to meet the target ranges for waterbodies discussed in the Technical Memorandum (ENSR, April 1999), as well as provide the ecoregional coverage desired. Therefore, the next step was to significantly relax the requirements for the representation of trophic parameters.

Based on the uneven availability of data, the decision was made to include those waterbodies that had data for Chl-a, SDT, TN *or* TP (see Figure 4-7). Adoption of this strategy greatly increased the number of available waterbodies, and allowed inclusion of most of the key trophic parameter data that was in the initial Database (Table 4-2). The Refined Database is composed of 569 rivers and streams. The distribution of rivers and streams across the states is shown in Table 4-3. Potential limitations to the development of the nutrient criteria from looking at this number of waterbodies will be discussed in the Data Gaps (Section 6.0) portion of this report. It will also be further considered in the Final Data Synthesis Report.

Table 4-3 Number of New England Rivers and Streams in Refined Database by State.

State	Rivers and Streams
Connecticut	149
Massachusetts	92
Maine	36
New Hampshire	182
Rhode Island	86
Vermont	29
New England Total	**569**

The Refined Database contains water quality data from rivers and streams from all six New England states. However the largest numbers of rivers and streams are located in Connecticut (149) and New Hampshire (182), with variable distribution between the other states, ranging from 29 in Vermont to 92 in Massachusetts.

The waterbodies represented by the totals in Table 4-3 were the basis of the further investigation. The spatial coordinates of each waterbody sampling station (if not electronically available) were obtained wherever possible and used to ascertain the ecoregion classification via GIS (see Section 4.3), as well as to characterize via SPARROW (through cooperation of the USGS) the land use attributes of sampling station watersheds. With the help of the respective state agencies, efforts were made to review and complete as much of the descriptive information as possible for these waterbodies and their sampling stations as to their physical characteristics, location coordinates, etc., in order to provide a complete basis for evaluation. Finally, the waterbodies in the refined Database were reviewed to determine whether they represented "reference" or "impacted" conditions (see Section 6.0).

4.3 Ecoregions and Watersheds of Interest

An important facet of the development of regional nutrient criteria is the concept of ecoregion-specific criteria. Ecoregions are generally defined as relatively homogeneous areas with respect to geomorphology, climate, ecological systems and the interrelationships among organisms and their environment (Omernik, 1995). They can be defined on a range of scales from national to very regional subdivisions.

Several potential ecoregion classification levels or schemes were identified in the course of the work. These included classifications Level 3 Aggregated and Non-Aggegated Nutrient Ecoregions proposed by Omernik), state-specific ecoregions (e.g. MA, ME), and other proposed classifications (e.g. USDA Forest Service, US EPA Region I). Following review and discussion, the EPA Level 4 Non-Aggregated ecoregions were selected as the basis for the analysis. The EPA Level 4 Non-Aggregated scheme separates New England into the five distinct regions shown in Figure 4-5. These regions are the:

- New England Highlands (NEH),
- Laurentian Plains and Hills (LPH);
- North Eastern Coastal Zone (NECZ);
- Atlantic Coastal Pine Barrens (ACPB); and
- Eastern Great Lakes and Hudson Lowlands (EGLHL) (a small portion of the around Lake Champlain, Vermont).

The ecoregions were compared in terms of their overall land use using land use cover produced by the EPA (EPA,1998). The relative percentages of the area coverage of various land use categories for each ecoregion are illustrated in Figure 4-6. As we can see, the LPH and NEH ecoregions are very similar in terms of the overall land use. The NECZ and ACPB ecoregions are characterized by their

relatively high percentage of residential land use (about 15%). However, the ACPB ecoregion differs from its NECZ neighbor by the higher proportion of wetlands and barren areas and its lesser proportion of agricultural areas. Due to the low number of ACPB watersheds available for inspection, these differences are not considered significant.

The ecoregions were used to evaluate the number of applicable waterbodies in the refined database versus the target range of waterbodies identified in the Technical Memorandum (ENSR, 1999). Due to the very limited spatial coverage of the ACPB ecoregion relative to other New England ecoregions, and its distinctive geomorphology, the river and streams in this ecoregion were deferred from further analysis in the refined Database.

4.4 Sequence for Nutrient Data Processing

The Refined Database contains a large amount of information that has to be extracted, sorted and analyzed to answer the very specific questions for the development of nutrient criteria. One of the critical decision in application of the database is to determine how trophic parameter data will be averaged to produce a representative value from the dataset of an individual waterbodies, regardless of the number of samples obtained from that waterbody. There are several ways to produce such a representative value, with potential advantages and drawbacks to each of these methods.

EPA provided the following protocol for statistical summarization of water quality parameters in the Ambient Water Quality Criteria Recommendation documents. The data are sorted by season, with the seasonal indices adjusted by aggregate ecoregion. Since New England contains rivers and streams in both aggregate ecoregions VIII and XIV and these have slightly different seasonal indices, the following definitions of seasonal indices were used for the Data Distribution Report:

- Spring – months of March to May;
- Summer – months of June to August;
- Fall – months of September to November; and
- Winter – month of December to February.

To provide a single representative parameter value for a waterbody, EPA developed a median value for all parameters within a waterbody for each of the four seasons over the period of record (US EPA, 2001). This method is used to prevent over-representation of an individual waterbody with a large amount of data vs. those with fewer data. The 25^{th} percentile for “all seasons” is calculated by taking the median of the four seasonal 25^{th} percentiles (this can be done with 3 seasons, if only those are available).

For calculation of the individual representative values for New England rivers and streams, ENSR followed the EPA protocol using the step-wise data reduction procedure outlined below.

1. All measurements for a water quality parameter made during a seasonal index period (e.g. September to November) for a waterbody are combined and the median value calculated. This produces a stream- and season-specific value;

2. All stream- and season-specific values for a water quality parameter for a particular ecoregion are pooled, the various statistical indices calculated (e.g. the 25th percentile). This produces an ecoregion- and season-specific value; and
3. The four ecoregion- and season-specific values are pooled (e.g., the four seasonal 25th percentiles from an ecoregion) and the median value taken. This produces an ecoregion and "all seasons" value.

This protocol was used to produce the statistical values described in the following chapters. This allowed direct comparison with ecoregion–specific values listed in the EPA nutrient criteria recommendation documents (U.S. EPA, 2000; 2001).

Table 4-2. Distribution of data in the initial database and refined database by state and source.

State	Data source	Min. sample date	Max. sample date	Number of data points in initial database	Percent	Number of data points in refined database	Percent
CT	CTDEP, 2002	23-Apr-97	26-Aug-02	6,520	3.8%	1,101	6.3%
CT	EMAP	15-Aug-94	15-Aug-94	13	0.0%	4	0.0%
CT	STORET	08-Jan-90	21-Mar-97	12,520	7.3%	2,801	16.1%
MA	EMAP	19-Jul-91	28-Jul-93	25	0.0%	8	0.0%
MA	MA-ARM	20-Mar-83	25-Jul-93	58,999	34.3%	0	0.0%
MA	MA-DEP	15-Jun-94	22-Apr-98	10,342	6.0%	671	3.9%
MA	STORET	08-Jan-90	25-Mar-97	1,655	1.0%	386	2.2%
ME	ME-DEP Bouchard	17-Jun-80	17-Oct-98	445	0.3%	59	0.3%
ME	ME-DEP Mitnik	02-Aug-89	21-Aug-98	5,408	3.1%	919	5.3%
ME	Penobscot Indian Nation	21-Jul-94	17-Sep-97	1,383	0.8%	546	3.1%
ME	STORET	24-Jan-90	28-Aug-96	1,097	0.6%	235	1.4%
NH	EMAP	21-Jul-91	21-Jul-91	11	0.0%	0	0.0%
NH	MODERNIZED STORET, 12/2002	28-May-90	07-Dec-00	50,656	29.4%	3,629	20.9%
NY	VT-DEC Nutrients DB	14-Mar-90	10-Nov-97	5,141	3.0%	0	0.0%
NY	VTDEC, 2002	22-Oct-96	29-Aug-01	37	0.0%	0	0.0%
RI	RI-DEM	12-Mar-91	16-Sep-97	1,762	1.0%	129	0.7%
RI	RI-USGS	03-Oct-89	07-Nov-97	760	0.4%	54	0.3%
RI	RIDEM, 2002	14-Apr-98	25-Apr-01	1,556	0.9%	447	2.6%
RI	STORET	16-Jan-90	27-Mar-97	2,359	1.4%	556	3.2%
RI	URIWW	15-Apr-95	10-Dec-98	2,961	1.7%	2,961	17.1%
VT	STORET	18-Jan-90	12-Sep-95	585	0.3%	263	1.5%
VT	VT-DEC Nutrients DB	14-Mar-90	10-Nov-97	7,756	4.5%	2,492	14.4%
VT	VTDEC, 2002	24-Jul-95	02-Jul-01	101	0.1%	90	0.5%
			Totals:	**172,092**		**17,351**	

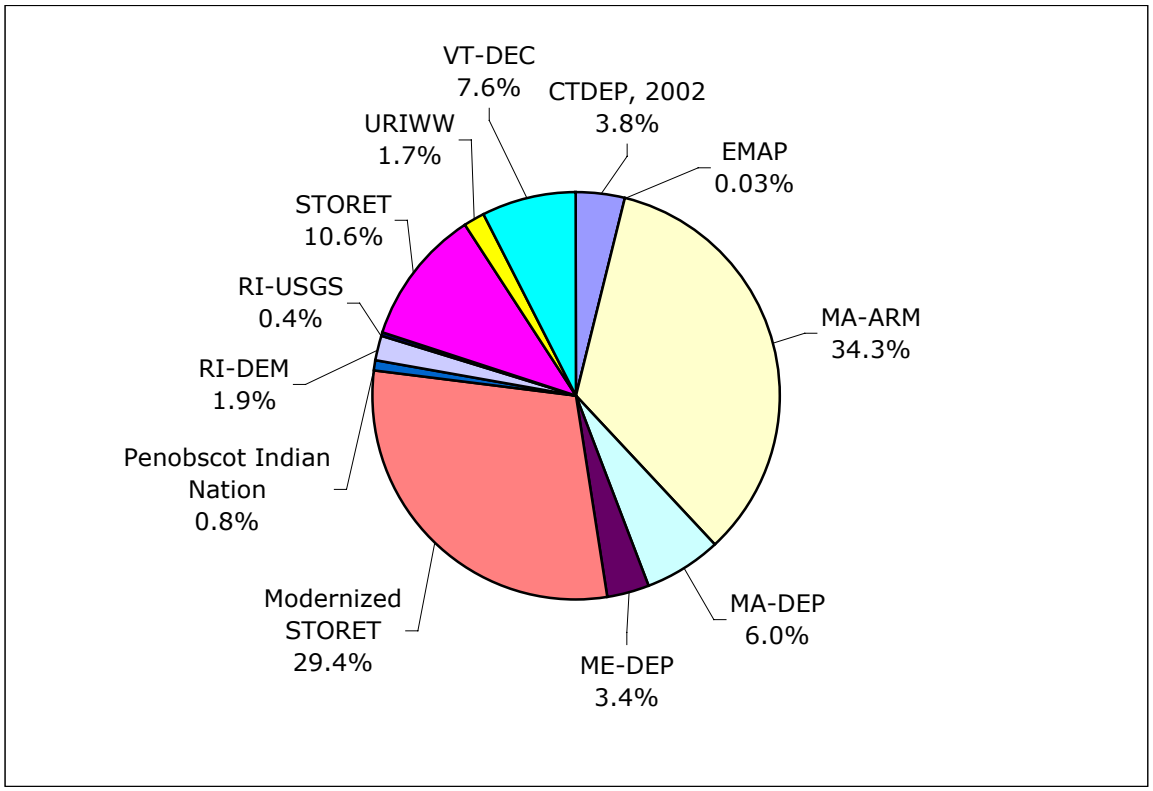
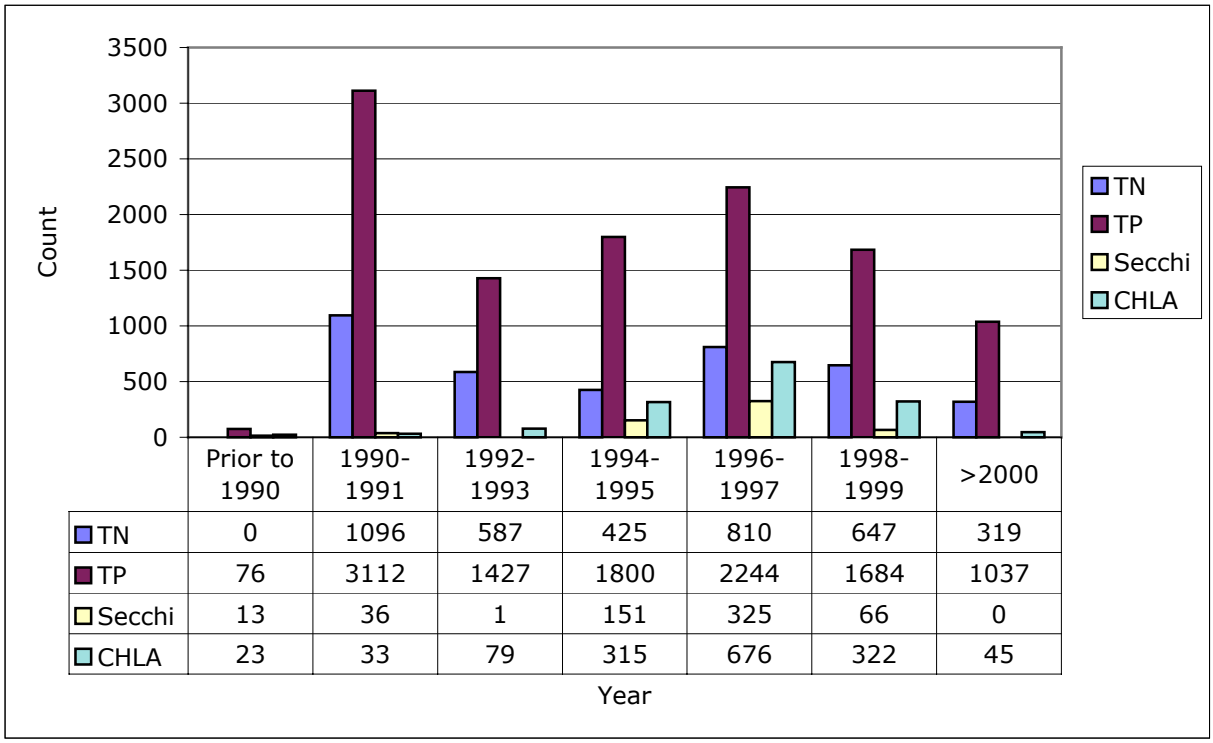


Figure 4-1. Distribution of water quality data in the initial database by source of data.



	Prior to 1990	1990-1991	1992-1993	1994-1995	1996-1997	1998-1999	>2000
TN	0	1096	587	425	810	647	319
TP	76	3112	1427	1800	2244	1684	1037
Secchi	13	36	1	151	325	66	0
CHLA	23	33	79	315	676	322	45

Year

Figure 4-2. Distribution of records in the qualified database by year for selected parameters.

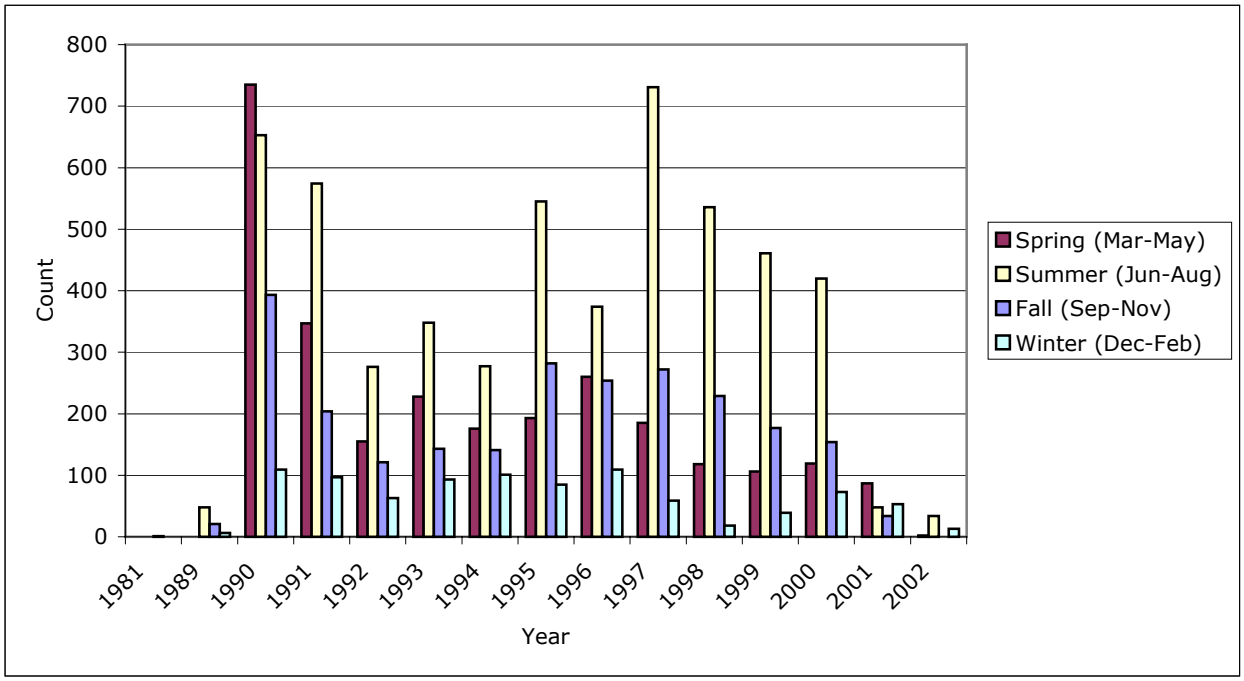


Figure 4-3: Seasonal distribution of total phsophorus records in the refined database.

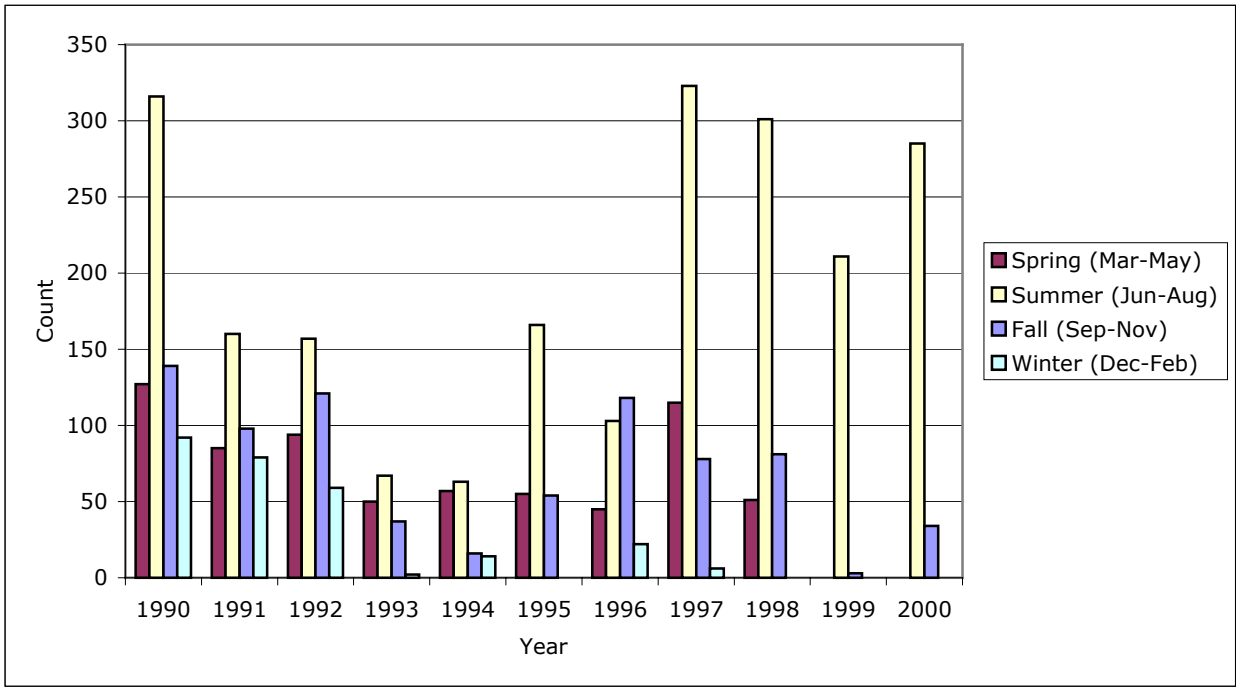


Figure 4-4. Seasonal distribution of total nitrogen records in the refined database.

Figure 4-5. New England Nutrients Ecoregions (EPA)

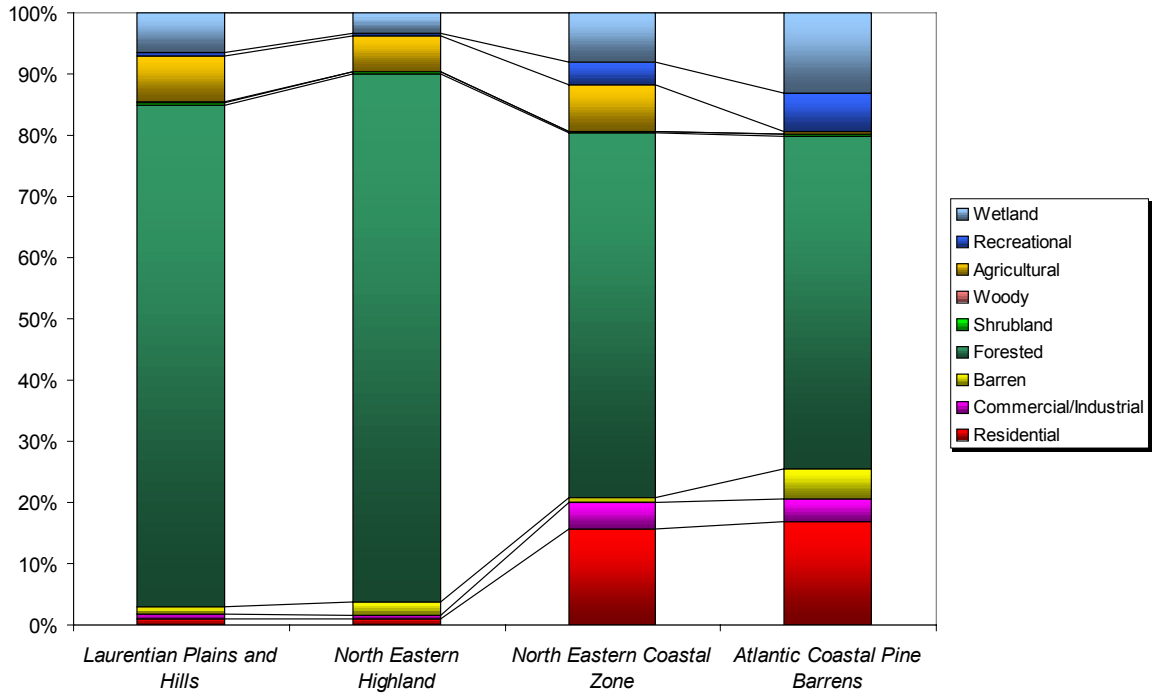


Figure 4-6 Comparison of Major New England Level 4 Ecoregions in terms of Land use.

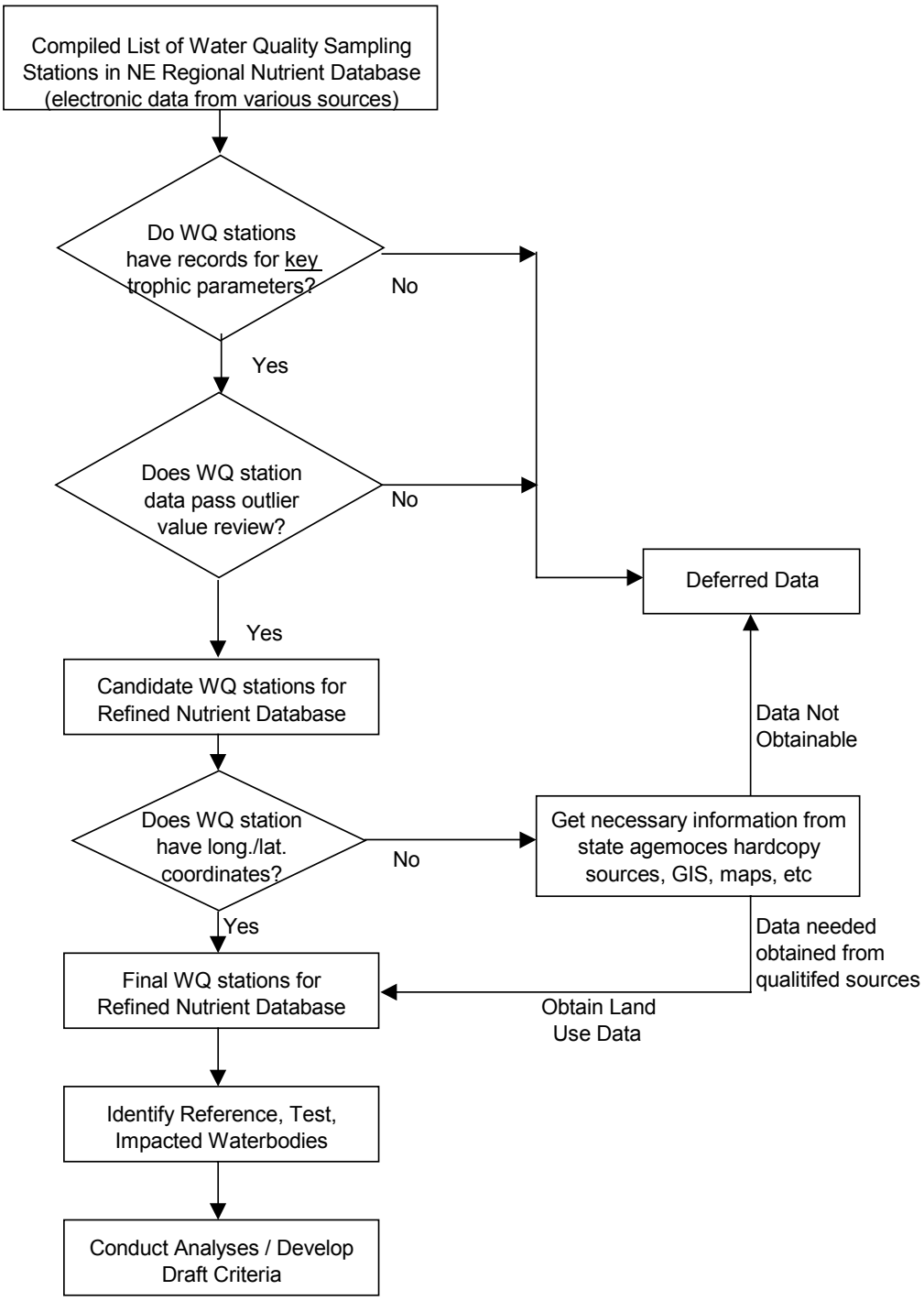


Figure 4-7. Flow Chart of WQ Sampling Location Selection for Refined NE Nutrient Database.

5.0 REFERENCES

Larsen, D.P., Stevens, D.L., Selle, A.R., and S.G. Paulsen. 1991. *Environmental Monitoring and Assessment Program, EMAP-Surface Waters: A Northeast Lakes Pilot*. Lake and Reservoir Management 7:1-12.

Liebman, M. 1999. *Development of Numerical Criteria for New England States and Tribal Governments*. EPA Region 1 Nutrient Criteria Fact Sheet, dated March 9, 1999. 6 pp.

Peterson, S.A., Larsen, D.P., Paulsen, S.G. and N.S. Urquhart. 1998. *Regional Lake Trophic Patterns in the Northeastern United States: Three Approaches*. Environmental Management 22:789-901.

Rohm, C.M., J.M. Omernik, and C.W. Kiilsgaard. 1995. *Regional patterns of total phosphorus in lakes of the northeastern United States.* Lake and Reservoir Management 11:1-14.

United States Environmental Protection Agency. 1998. *National Strategy for the Development of Regional Nutrient Criteria. Fact Sheet.* Office of Water. Washington, D.C. Document EPA 822-F-98-002.

United States Environmental Protection Agency. 2000a. *Nutrient Criteria Technical Guidance Manual. Rivers and Streams*. Office of Water. Washington, D.C. Document EPA 822-B-00-002.

United States Environmental Protection Agency. 2000b. *Ambient Water Quality Criteria Recommendations. Rivers and Streams in Ecoregion XIV.* Office of Water. Washington, D.C. Document EPA 822-B-00-022.

United States Environmental Protection Agency. 2001. *Ambient Water Quality Criteria Recommendations. Rivers and Streams in Ecoregion VII.* Office of Water. Washington, D.C. Document EPA 822-B-01-015.