NEIWPCC

Lowell, MA

Collection and Evaluation of Ambient Nutrient Data For Lakes, Ponds, and Reservoirs in New England

Data Synthesis Report

INTERIM FINAL

____

Prepared By

____

Reviewed By

ENSR Corporation

April 2000

Document Number 8726-780-600

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Overview of New England Nutrient Database

1.1 Purpose and Goal of Data Synthesis Report

The purpose of the Data Synthesis and Final Report (the "Data Synthesis Report") is to describe and summarize ENSR's development of a nutrient-related database for New England waterbodies, and to describe how this data may be applied to develop preliminary draft regional nutrient criteria for lakes, ponds, and reservoirs. This document is the final project task deliverable of the "Collection and Evaluation of Ambient Nutrient Data" Project ("Phase 1") conducted by ENSR for the New England Interstate Water Pollution Control Commission ("the Commission"). The Data Synthesis Report incorporates information from the earlier Data Distribution Report and Data Gaps Analysis, but builds on and expands the material presented in those deliverables. The Data Synthesis Report describes and summarizes the finalized New England Nutrient Database ("the Nutrient Database"), explores potential methods of developing draft nutrient criteria, and details several outstanding issues that may need to be addressed further in the development of regional nutrient criteria.

Specific objectives of the Data Synthesis Report are as follows:

- Provide sufficient regulatory background as a framework for the project objectives;
- Document and describe the sources of data used in the Nutrient Database (also identify data which was deferred from inclusion and other potential data sources);
- Describe the basic structure and features of the Nutrient Database;
- Describe and summarize the contents of the Nutrient Database with regard to amount of data, number of waterbodies, parameters of interest, ecoregional coverage, etc.
- Describe selection of reference, test, and impacted waterbodies;
- Provide examples of preliminary draft nutrient criteria based on suggested methods from the EPA Technical Guidance Document and
- Identify potential issues regarding development of nutrient criteria that should be considered and resolved prior to issuing of draft criteria.

1.2 Relationship of New England Nutrient Database to Regional Nutrient Criteria Development

As part of the national Nutrient Strategy described in the "Clean Water Action Plan" (US EPA, 1998a), U.S EPA will establish numeric criteria for nutrients by the year 2000 and assist the states in adopting "ecoregion-specific" standards based on these criteria. This project is being conducted as part of the US EPA National Strategy for the Development of Regional Nutrient Criteria, with the stated objective of development of waterbody-type technical guidance manuals and region-specific nutrient criteria (Liebman, 1999). The water quality criteria documents will include default, or proposed, regional criteria for nutrient endpoints, specifically nitrogen and phosphorus, for three types of waterbodies -- lakes, rivers and estuaries. The nutrient criteria will reflect numerical target ranges for nutrient response parameters such as chlorophyll *a* (chl *a*) and turbidity or Secchi disk transparency depth (SDT). The major elements of this strategy are presented below:

- *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, stakeholder and Tribal nutrient data. The development of the regional database followed the overview described in the original Request for Proposals (RFP) that identifies the nature of the data to be collected and entails 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);
- **Collection of Hard Copy Reports** – in addition to electronic data, selected agency or literature reports will be acquired and incorporated into the regional database;
- **Conduct QA/QC Reviews** – prior to inclusion into the regional database, information will be 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;
- **Data Gap Analysis** – the Data Distribution Report will be examined to identify potential data gaps, with potential re-examination of collected (but deferred) data for potential inclusion; 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 and its applicability to development of nutrient criteria.

The approach used for acquiring and classifying nutrient data in the Database was described in a project Technical Memorandum distributed in April 1999 (ENSR, 1999a). The Technical Memorandum was distributed and presented to the project "Core Group" consisting of a selected number of state and federal agency contacts, the two Regional Nutrient Assessment Teams (RNAT), 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 Nutrient Assessment Team members), U.S. EPA, and the Commission.

During late summer 1999, two review draft technical guidance manuals were issued by U.S. EPA. These manuals were *Nutrient Criteria Technical Guidance Manual: Lakes and Reservoirs* (EPA 822-D-99-001, U.S. EPA, 1999a) issued August, 1999 *and Nutrient Criteria Technical Guidance Manual: Rivers and Streams* (EPA 822-D-99-003, U.S. EPA, 1999b) issued September, 1999. These draft manuals, particularly the Lakes and Reservoirs manual, provided suggested methodologies for developing nutrient criteria that served as the basis for preliminary draft nutrient criteria in the Data Synthesis Report.

A preliminary description and evaluation of the draft Database was presented to the Core Group in September 1999 to familiarize the group with the nascent database, do a preliminary data gap evaluation, and to identify issues and promote dialog regarding use of the Database to develop regional nutrient criteria. As part of the meeting follow-up, state agencies were charged with classifying (with justification) reference and impacted status waterbodies from a list of the Nutrient Database. These classifications were incorporated into the Data Distribution Report.

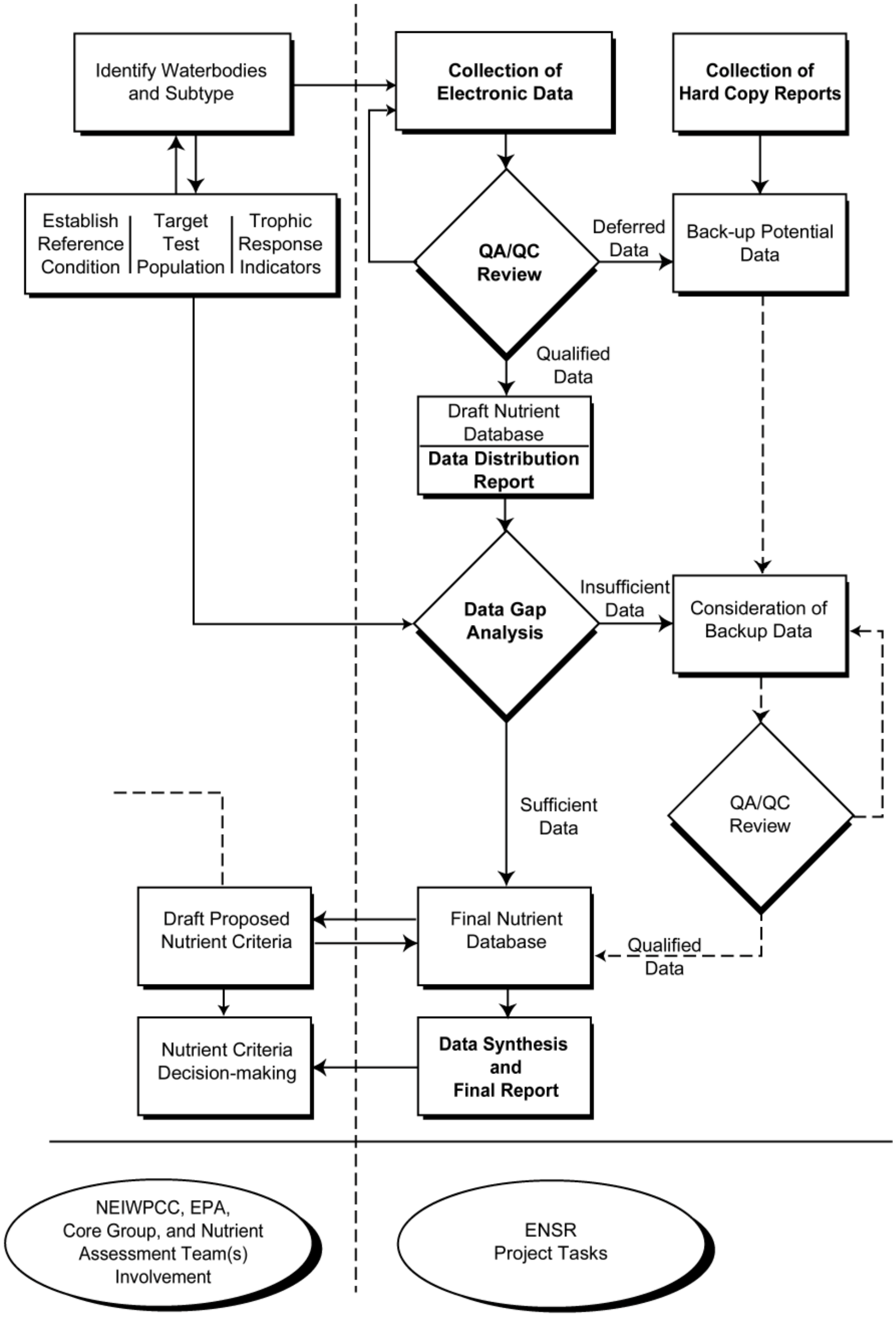
The Data Distribution Report was issued in December 1999 (ENSR, 1999b) and provided documentation of how electronic and hardcopy data were acquired, validated, and incorporated into the Nutrient Database. The Data Gap Analysis reviewed the sufficiency of the available data with regard to target ranges of waterbodies described in the Technical Memorandum (ENSR, 1999a). Based on the Data Gap Analysis, additional hardcopy report data were transcribed to augment and expand the Database for ecoregions of interest. The resultant finalized database was designated as the New England Nutrient Database – Lakes/Ponds/Reservoirs for purposes of this report.

Another result of the Data Gaps Analysis was a decision to defer consideration of the rivers/streams data to support nutrient criteria development to a later phase. Further investigation of this waterbody type will be addressed in the upcoming data collection program for Estuaries and Coastal Embayments under the Collection and Evaluation of Ambient Nutrient Data – Phase 2 ("Phase 2 work) (see discussion in Section 4.5.4).

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. The structure and framework of the Access97 ® database used to house the Nutrient Database are described in Section 3.0. Section 4.0 contains the Development Strategy used to "refine" the initial data collected into a more focused and useable set of waterbodies and parameters. Section 5.0 provides Summary Statistics on the waterbodies and major parameters of interest contained in the Nutrient Database. Section 6.0 contains application of criteria-making using methods suggested in the draft Technical Guidance Manuals with calculation of draft preliminary nutrient criteria. Section 7.0 identifies outstanding issues associated with criteria making that were identified by Core Group members during the project. These issues will need to be addressed further to achieve consensus as to the development of nutrient criteria for New England. Section 8.0 provides a summary of the Report.

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



M990160

Figure 1-1 Development of Nutrient Database to support nutrient criteria decision-making.

2 Sources of Data

2.1 Data Sources

The primary goal of the project was 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 major databases that were primary sources of electronic data and the respective contact person are listed in Table 2-1. The data requested were from electronic databases of a fairly recent vintage (i.e., 1990 or later). As noted in Table 2-1, the data received included pre-1990 data, but the majority of data (~90%) included in the draft and final Nutrient Database were primarily from 1980 to the present. In addition, a number of hardcopy data sources were noted for potential use in supplementing the electronic data on a per needs basis (as identified in the data gaps section). A compilation of the major databases used, with a brief description of the dataset, its parameters, period of sampling, associated documents, 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 extraction and analysis tasks. The GIS interface was used to distinguish waterbodies from those with similar names, assign waterbodies to the correct watersheds or hydrologic units, 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 Level 3 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. Delineation of the 8-digit Hydrologic Units Codes (HUCs) was obtained from the USGS web site (www.usgs.gov). Delineation of rivers and streams, with corresponding Reach File RF3 data, was obtained from the BASIN web site (www.epa.gov/OST/basins/). Counties and towns political boundaries were obtained from the ESRI Data CD of New England. Spatial coordinates were obtained from the USGS Geographical Names Identification System (GNIS).

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/*

2.3 Hardcopy Data Sources

In addition to the electronic sources, several potential hardcopy datasets (either in the form of unpublished data sheets, waterbody status reports, or waterbody compilations were also identified. The main types of hard copy data sources included:

- Lake and Stream Inventories or Gazetteer documents;
- State Trophic Classification Documents;
- Clean Lake Program Diagnostic/Feasibility Study Reports;
- River Monitoring Documents;
- River Waste Load Allocation Reports;
- United States Geological Survey (USGS) Monitoring Data;
- Academic Institutions; and
- Watershed Groups.

These hardcopy data sources were initially deferred from inclusion in the draft Databases. However, following the Data Gap Analysis, these hard copy data sources were used to provide information and nutrient data for additional waterbodies in particular ecoregions of interests. Based on the ecoregional distribution of waterbodies initially achieved, additional hardcopy data focused on data from the New England Coastal and Atlantic Coastal Pine Barrens ecoregions. A listing of these hardcopy sources incorporated into the Nutrient Database is provided in Appendix A.

Table 2-1 Organizational Contacts for waterbody and nutrient data.

Reserved for Table 2-1

3 Database Description

The New England Nutrient Database was assembled from the data acquired from 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 U.S. EPA headquarters for another regional project. 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 database application includes tables, queries, forms, and reports. 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. Forms constitute the "graphical user interface" to the data. They are used to enter new data, view existing data, or perform operations in a user-friendly manner.

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, 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 water depth, 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 combination 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.

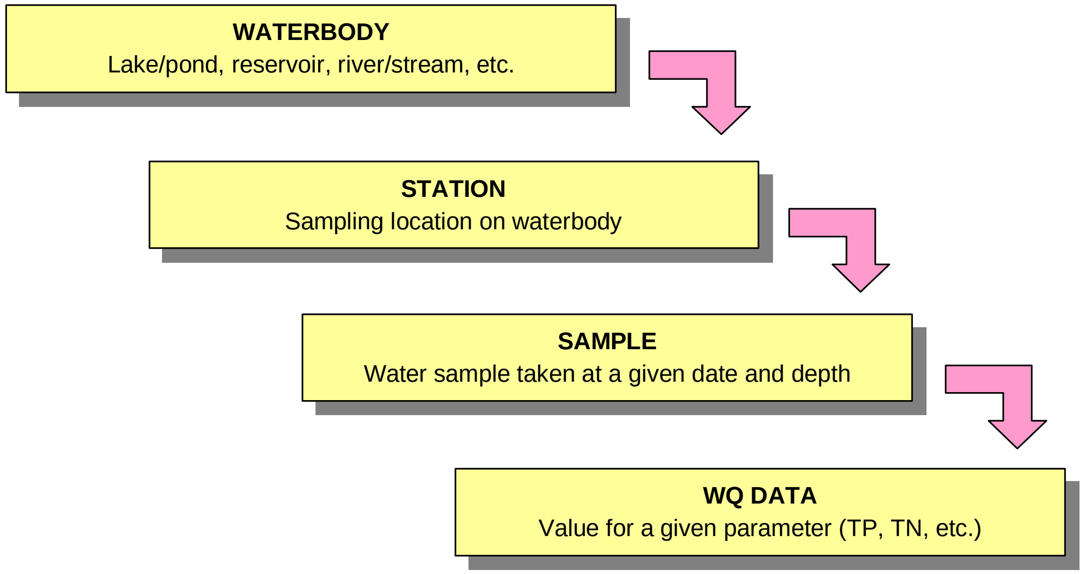


Figure 3-1 Database Main Data Tables Structure.

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. In Table 3-1, the waterbody types include Lake/Pond (P), Reservoir (R), River/Stream (S), Marine (M), and Other (O).

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 and Station Tables	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 a lake/pond, a reservoir, or a specified segment or reach along a given river or stream. [Note: reservoirs were identified through waterbody names or specific identification as such by state-specific data sources]. While the identification of lakes/ponds and reservoirs is relatively easy and non-ambiguous, the subdivision of rivers and streams into segments with relatively homogeneous characteristics is more complicated, as depth, flow, and other physical characteristics are expected to change with the distance from the headwaters. The USGS RF3 reach file provided a classification scheme that can be used for the Database.

Because of the potential differences between the waterbody types, the table Waterbody contains some fields such as the mean depth and surface area that are relevant only to lakes, ponds and reservoirs, but which are less applicable to rivers and streams. It also has a number of fields that are specific to rivers and streams (e.g., tributary_code).

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. The station may therefore have a local depth that is different from the average or maximum depth reported for the waterbody.

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 Quality for Source Data

The majority of electronic data was obtained from designated federal and state agencies who, for diverse reasons, maintain active files of water quality and waterbody information. No attempt was made by ENSR to directly verify the accuracy of this electronic information contained in these files, since the individual states are assumed responsible for the quality control of these data files. Statements regarding the QA/QC aspects of the individual state programs are described in Appendix A. Data, which was not electronically available (i.e., hardcopy reports), were transcribed to the Nutrient Database. Following transcription, these data were subject to QA/QC including verification of data from approximately 10% of the waterbodies as well as the additional verification procedures described in Section 3.3.4.

3.3.2 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 slightly manipulated to make it compatible for importing into the Nutrient Database. As such, data entry errors were assumed to be limited to those that could have taken place in the original data source. This assertion was reinforced by a later QA check on the water quality data for Vermont waterbodies contained in the database. Comparison of the data contained in the Nutrient Database to those contained in Vermont state files indicated that no loss of information had occurred.

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.

3.3.3 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 were 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 deleted. Duplicates within a single data source were assumed to be legitimate and were identified as “DUP” in the sample type field.

3.3.4 Additional Verifications

As noted in Section 3.3.1, no systematic attempt was made to verify the electronic data submitted by the agencies. However, data for selected trophic parameters within the Database were compared with an expected range of values based on best professional judgment (Table 3-2). For example, SDT measurements were compared to the maximum depth of the water body and potentially spurious values (i.e., SDT > maximum depth) were identified and verified against the original data source. Reported values for total phosphorus (TP), total nitrogen (TN), and chl *a* were similarly compared to the expected range. 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.

Another partial quality review was conducted through the inspection by the RNAT members of the state-specific listing of waterbodies and representative trophic parameters as part of the assignation of reference and impacted waterbodies. State agency staff noted a few instances when representative data was apparently at variance with their expectations of water quality for a waterbody. In all cases, these discrepancies were successfully resolved, or the data removed from the database.

Table 3-2 Expected Trophic Parameter Range.

Trophic Parameter	Minimum Value	Maximum Value
Chlorophyll a (ug/L)	0	250
Secchi Disk Trans. (m)	0	Maximum Water Depth
Total Nitrogen (ug/L)	0	5,000
Total Phosphorus (ug/L)	0	1,000

4 Development Of Nutrient Database

This section describes the development of the Nutrient Database, specifically, how it was created from review and refinement of the primary data collection effort. This refinement was required by the sheer size of the resulting database and the inclusion of many water quality records of lesser importance to the development of regional nutrient criteria (Section 4.1). The resultant draft New England Nutrient Database was developed based on a selection of lakes, ponds, and reservoirs ("L/P/R") with critical trophic parameter data. Section 4.2 describes the strategy used to develop the draft Nutrient 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 to provide a representative value for that waterbody. Section 4.5 summarizes the Data Gap Analysis conducted to identify data needs remaining existing in the draft Nutrient Database and the hardcopy reports used to supplement and produce the final Nutrient Database. An overview of this process is provided in Figure 4-1.

4.1 Primary data collection - Waterbody and Parameter Inventory

Historical water quality, waterbody morphometric, and ancillary data were collected from a multitude of sources, including federal and state agencies, academic sources, watershed stakeholders, volunteer groups and Tribal Indian Nations. The data sources have been previously identified and discussed in Section 2.0. The primary data collection assembled from these data contained over 7,000 waterbodies, 10,700 stations, 350,000 samples, and 780,000 water quality data records. Some of the features and characteristics of the primary data collection 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 4-2. As can be seen, the electronic data primarily came from fourteen data contributions from state and federal agencies, academic institutions, and a Tribal Nation (i.e., Penobscot Indian Nation). This wide spectrum of data sources is indicative of the diverse nature and intent of the data contained in the electronic files acquired. It should be noted that this distribution of data encompasses all of the water quality records, including rivers and streams, and the distribution is quite different when considering subsequent draft Nutrient Database, which focused on L/P/R waterbodies with sufficient data on a majority of the four trophic indicators (TP, TN, chl *a*, SDT) to merit inclusion.

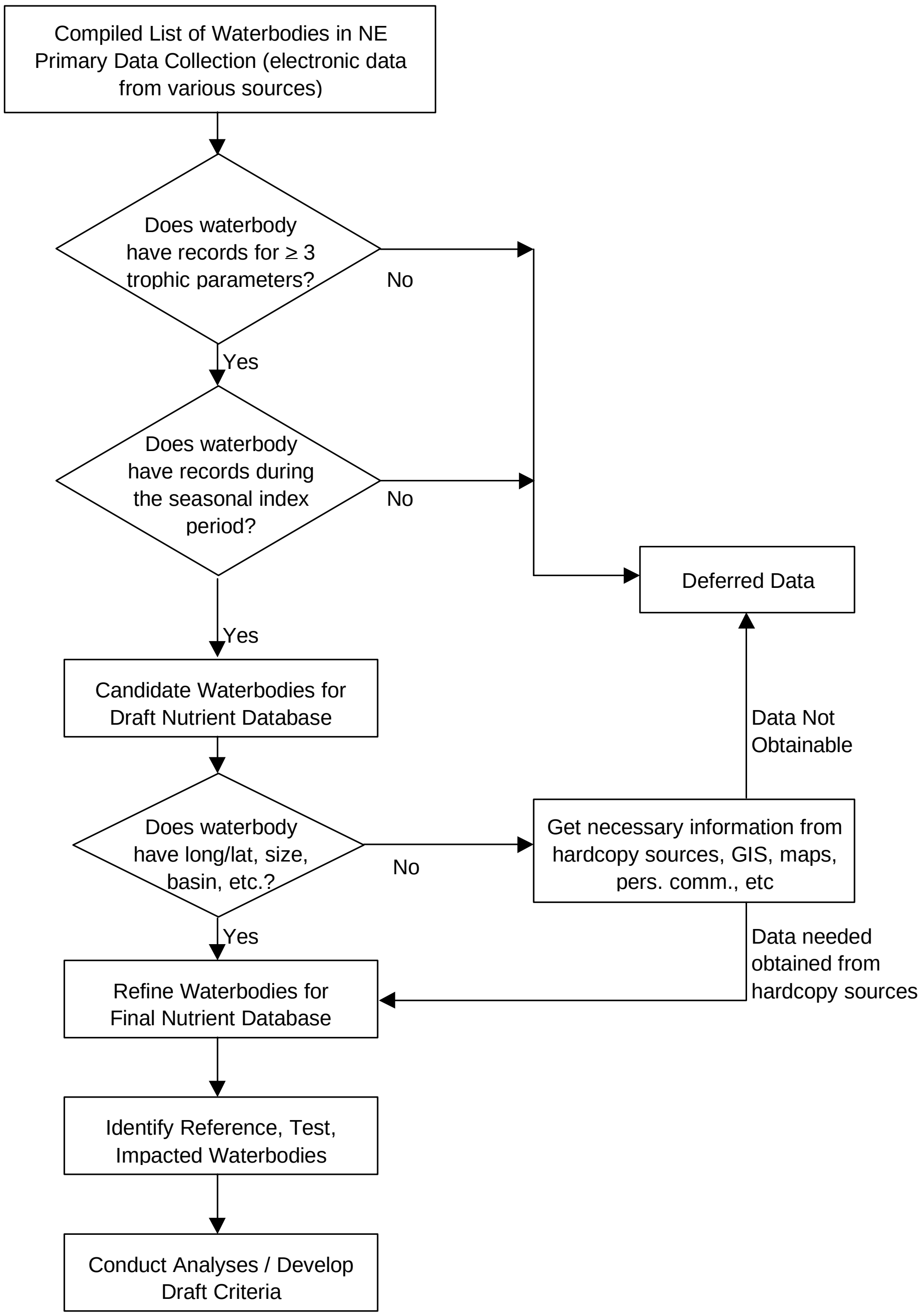


Figure 4-1 Flow chart of waterbody selection for New England Nutrient Database.

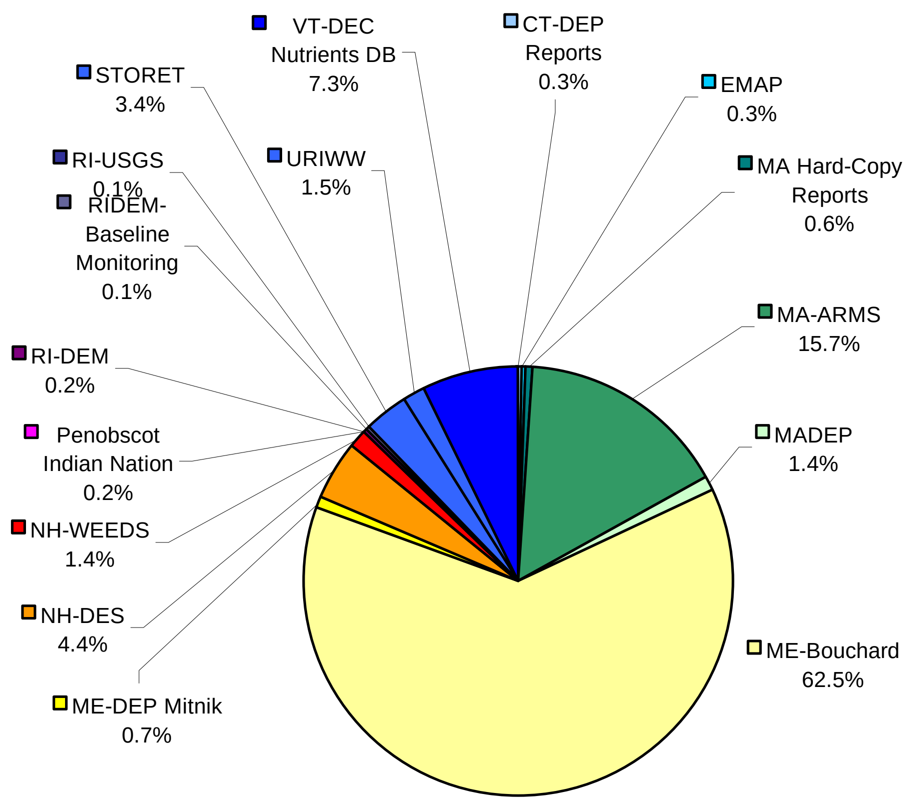
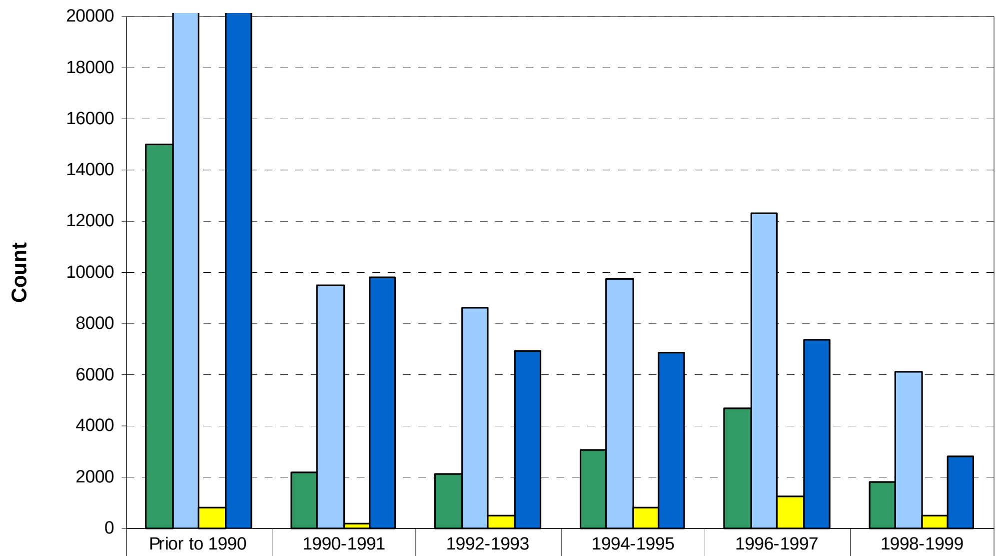


Figure 4-2 Distribution of Water Quality Measurements in Primary Data Collection by Source of Data.

4.1.2 Period Covered

The primary data collection contains data and information that spans from 1952 to 1998, although the vast majority of electronic data collected were generated during the 1990s. The range of years of data record is quite variable among trophic parameters For example, SDT records go as far back as 1952, available chl *a*, TP, and total kjeldahl nitrogen (TKN) have records going back to the 1970s, while the earliest TN record dates to 1989. The data also contain many non-nutrient data (e.g., temperature, pH, alkalinity) records as well, many of which were collected for acid rain monitoring purposes during the late 1980s and early 1990s. The distribution of the data for the selected trophic parameters is presented in Figure 4-3. The graphic presents the number of records available for each year for selected trophic parameters.



	Prior to 1990	1990-1991	1992-1993	1994-1995	1996-1997	1998-1999
CHLA	15022	2210	2096	3081	4708	1813
SECCHI	41332	9480	8655	9750	12305	6102
TN	801	163	509	831	1261	528
TP	24274	9811	6946	6859	7371	2794

Figure 4-3 Distribution of records by year for selected trophic parameters.

4.1.3 Water Quality Measurements

The primary data collection contained more than 500,000 water quality measurements drawn from lakes, ponds, reservoirs, rivers, and streams. However, because of the various goals and natures of the monitoring programs that provided the information (e.g., acid rain monitoring programs), 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 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 acquired nutrient data were not appropriate for application to surface waterbodies (e.g., groundwater nitrate records).

Despite these limitations, the identified critical trophic parameters of interest are well represented within the primary data collection. This includes about 57,200 measurements of TP, 87,300 SDT measurements, and 28,500 chl *a* values. The least represented trophic parameter is TN with fewer than 3,500 records available. The distribution of water quality measurements in the database by parameter is presented in Figure 4-4.

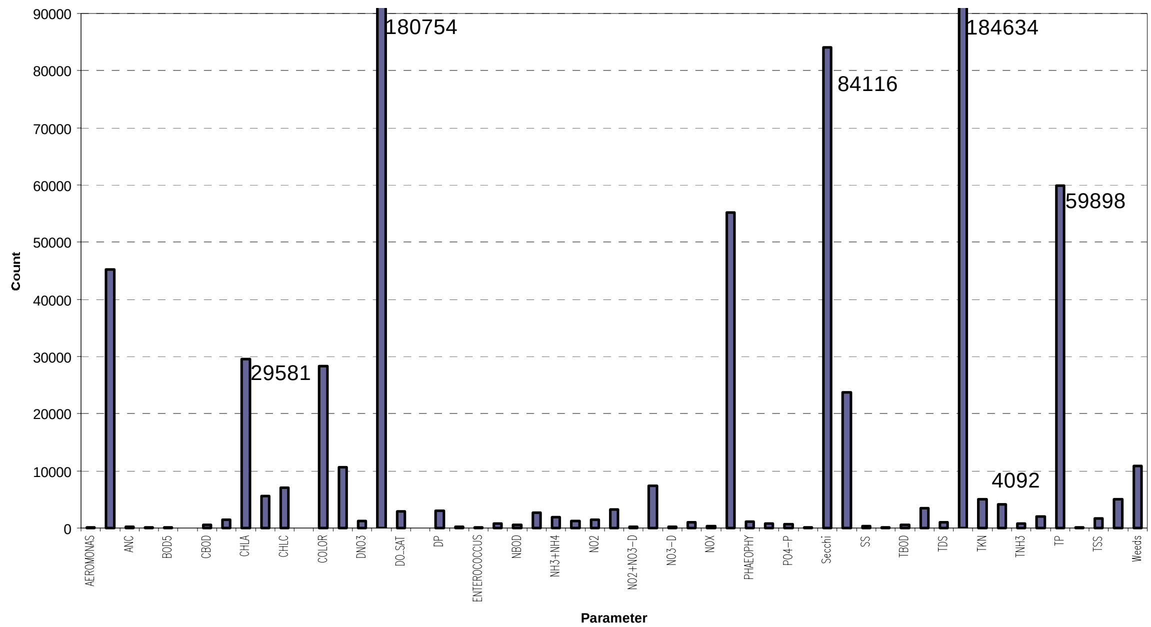


Figure 4-4 Water Quality Measurements by Parameter in Primary Data Collection.

4.1.4 Distribution of Waterbody Types

As part of the primary data collection, data were obtained on a wide variety of waterbody types. Based on the water quality and ancillary information in the data collected from the various data sources, the waterbodies in the database were classified into three types:

- *(P) Lakes and ponds;*
- *(R) Reservoirs; and*
- *(S) Rivers and streams;*

The distribution of the P, S. and R classified waterbodies contained in the primary data collection by type and state is presented in Table 4-1. Note that the number of P, S, and R waterbodies for a state differs from the amount of data contributed by that state (Figure 4-2). As can be seen, Massachusetts represented a large fraction of the waterbodies. However, a significant number of these waterbodies were sampled as part of the ARM program and were less frequently sampled for nutrients. It can also be seen that Massachusetts also had the greatest number of River/Stream samples in the primary data collection.

Table 4-1 Number of waterbodies in Primary data collection for types P, S, and R.

State	Lakes and Ponds (P)	Rivers and Streams (S)	Reservoirs (R)
Connecticut	66	43	9
Massachusetts	2363	1613	239
Maine	817	26	0
New Hampshire	659	144	25
Rhode Island	70	60	14
Vermont	831	41	10
New England Total	4806	1927	297
Total		7030	

4.2 Development of a Draft Nutrient Database

At the end of the primary data collection period, information had been obtained on a large number of waterbodies (>7,000) and a very large number of water quality records (>780,000). While this amount of data is impressive, it resulted in a cumbersome database that, due to its sheer size, was difficult to perform standard calculations and analyses on. More importantly, the primary data collection 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 necessary ancillary information for each of the selected waterbodies. For example, it was necessary to identify the spatial coordinates (i.e., latitude and longitude) for each of the selected waterbodies to assign watershed and ecoregional status. Finally, it would be necessary for state agencies to review the selected waterbodies to assign “reference” or “impacted” status. Therefore, it was prudent to reduce the size of the database to a lesser but sufficient number of waterbodies and qualified data still capable of fully supporting nutrient criteria development.

Accordingly, a decision was made and a strategy developed to produce a smaller and more focused database. We have used the term “draft Database” to refer to this effort since it represents a distillation of the information in the primary data collection. Since the original aim of the project is to provide a database for regional nutrient criteria, the draft Database contains only those waterbodies for which sufficient information was available on the relevant trophic parameters. This purpose and strategy for development of the draft 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 appears to be consistent with the overall goals of the program.

Briefly, the strategy acknowledges that not all waterbodies were sampled for the four key trophic parameters (TP, TN, chl *a*, and SDT). In fact, less than 5% of the waterbodies (or about 300) in the primary data collection had information for these four parameters. These roughly 300 waterbodies included those for which nitrogen data were available but not necessarily as the preferred TN fraction. Comparison of the number and location of these 300 waterbodies indicated that there were an insufficient number to meet the target ranges for waterbodies discussed in the Technical Memorandum (ENSR, 1999a), as well as provide the ecoregional coverage desired.

Further, there were insufficient data in the "S" (rivers and streams) waterbodies to merit their inclusion in the draft Database, since virtually none of these lotic waterbodies had the three key trophic parameters (only 14 of the 1927 waterbodies). This is an important limitation to the development of the nutrient criteria for these environments and will be discussed in the Outstanding Issues (Section 7.0) portion of this report. **Therefore, the statistical summaries and analyses in the remainder of this report pertain only to the L/P/R waterbodies.**

The next step was to relax the requirement for L/P/R waterbodies to have data for all four trophic parameters. Therefore, a decision was made to expand the database to those waterbodies with data for at least three of the four parameters of interest, specifically TP, SDT and chl *a.* This decision was based on the availability of data (e.g., the relatively small amount of TN records), the generally-held hypothesis that a majority of freshwater systems are phosphorus-limited most of the year, and the better-established correlation of TP with predictable changes in overall water quality (i.e., eutrophication effects) (U.S. EPA, 1998b).

Adoption of this strategy greatly increased the number of candidate waterbodies but still retained a great majority of the data acquired in the primary data collection for these three key parameters. For example, approximately 97% of the data for TP from the primary data collection was captured by the draft Database, for SDT > 97%, and for chl *a* > 99% of the data. This indicated that much of the trophic parameter information in the primary data collection is associated with a select subset of waterbodies.

The draft Database was composed of 1155 L/P/R waterbodies from all six New England states. The distribution of the L/P/R waterbodies throughout the New England states is shown in Table 4-2. The largest percentage of lakes/ponds is located in Maine, with fairly equal distribution among the other states.

Table 4-2 Number of waterbodies in draft Database by state for types P and R only.

State	Lakes and Ponds (P)	Reservoirs (R)
Connecticut	63	8
Massachusetts	59	7
Maine	661	0
New Hampshire	179	7
Rhode Island	62	13
Vermont	92	4
New England Total	1116	39
Total		**1155**

The waterbodies represented by the totals in Table 4-2 were the basis of further investigation. The spatial coordinates of each waterbody were obtained and entered to ascertain the ecoregion classification (see Section 4.3). 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 as to their physical characteristics, location coordinates, etc., in order to provide a more complete basis for evaluation. Finally, the waterbodies in the draft Database were reviewed to determine whether the represented "reference" or "impacted" conditions (see Section 6.2).

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, 1987). 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-Aggregated 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 3 Non-Aggregated ecoregions were selected as the basis for the analysis. These ecoregions were modified from the hierarchical framework of Omernik (1987). The EPA Level 3 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).

It should be noted that these ecoregions are not exclusive to New England and there is overlap into nearby states (e.g., New Jersey, New York, Ohio). Therefore it is important to note that the ecoregional preliminary draft nutrient criteria developed in this report pertain only to those portions of the ecoregions occurring in New England.

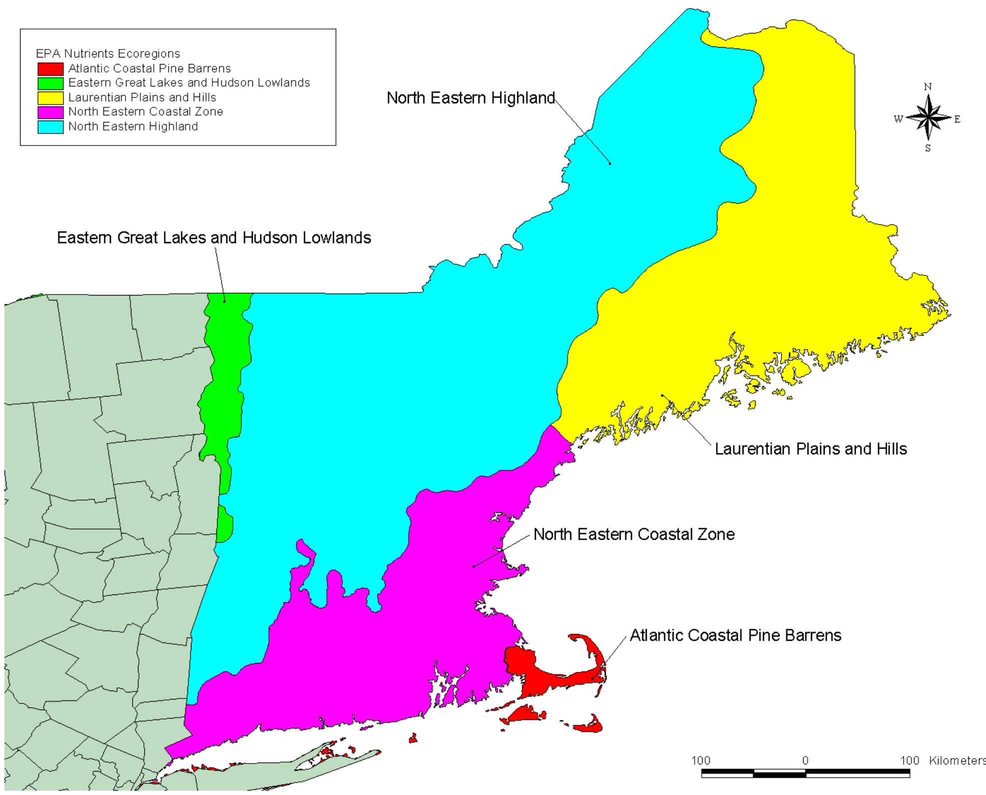


Figure 4-5 New England Nutrient Ecoregions (EPA)

The ecoregions were compared in terms of their overall land use using land use coverages obtained from the EPA (EPA, 1999c), as part of the Multi-Resolution Land Characteristics (MRLC) dataset. The MRCL dataset is produced by an interagency consortium that regroups the EPA, USGS, and NOAA. The land use coverage is distributed as a mosaic of ArcView/ArcInfo grid files with a resolution of 30 meters. Each pixel (i.e. small square) of the grid has an associated land use category. There are 18 land use categories identified, which include open water, residential area, forested, pastures, wetlands, etc. Figure 4-6 shows an example of the land use file for a selected area in Connecticut.

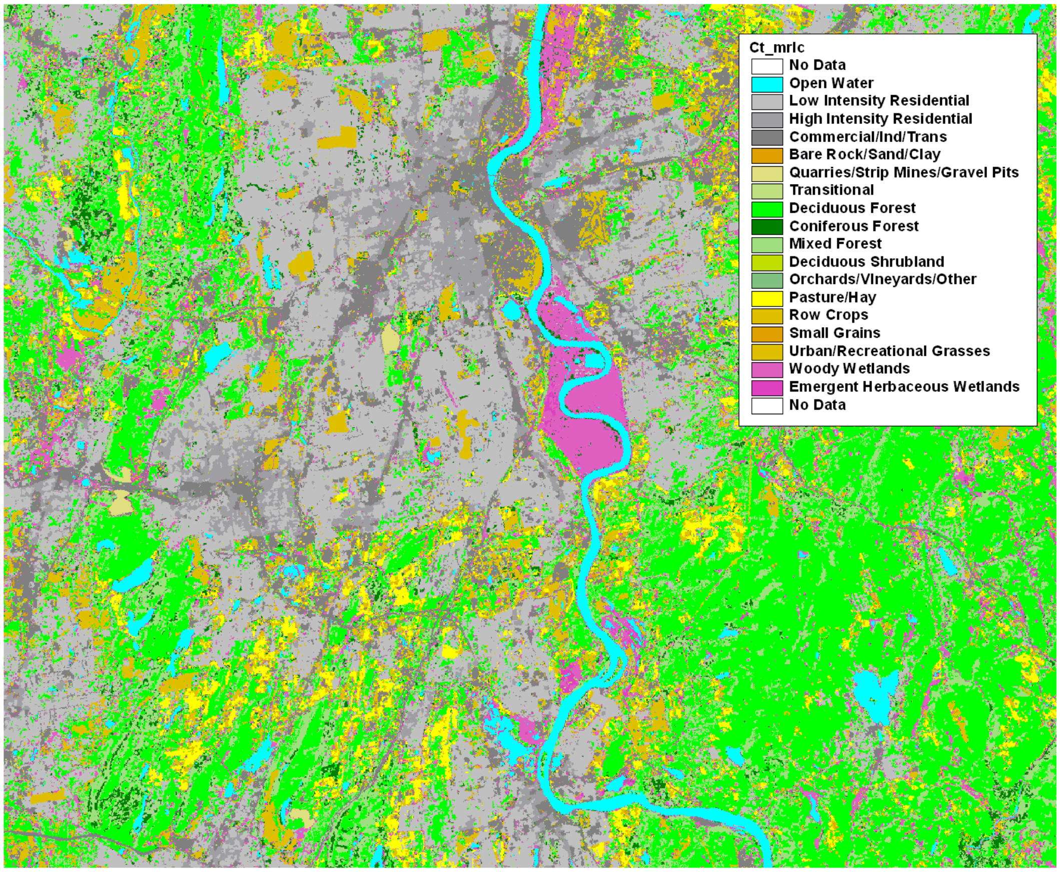


Figure 4-6 View of the land use coverage for a selected area in Connecticut.

The relative percentages of the area coverage of various land use categories for each ecoregion are illustrated in Figure 4-7. 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 higher 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 L/P/R waterbodies considered for the ACPB ecoregion, these differences are not considered significant.

The ecoregions were used to evaluate the number of applicable waterbodies in the draft Database versus the target range of waterbodies identified in the Technical Memorandum (ENSR, 1999a). Due to the very limited spatial coverage of the EGLHL ecoregion relative to other New England ecoregions, and its distinctive geomorphology (Smeltzer, pers. Comm.), the 18 lakes and ponds in this ecoregion were deferred from further analysis in the draft

Database. It has been suggested that information from the EGLHL waterbodies be considered for inclusion with similar waterbodies that may be considered by EPA Region 2.

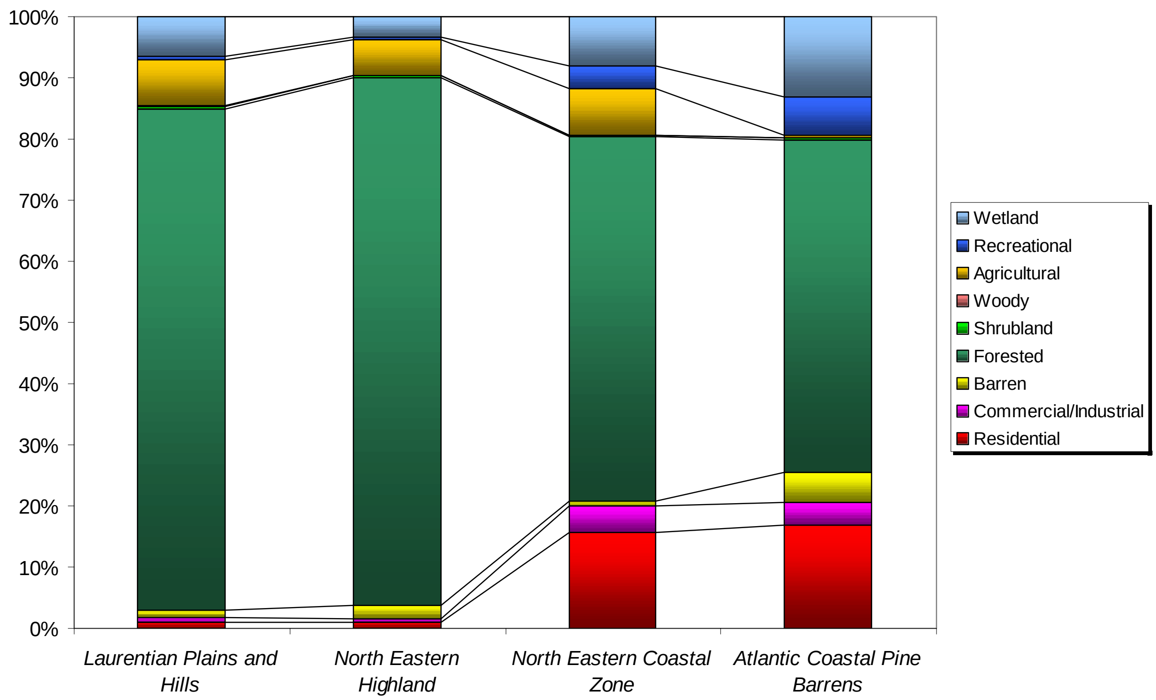


Figure 4-7 Comparison of major New England Level 3 Ecoregions in terms of land use.

4.4 Sequence for Nutrient Data Processing

The draft Database contains a very 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 decisions in application of the database is to determine how trophic parameter data will be averaged to produce a representative value (or indicator of central tendency) from the dataset of an individual waterbodies, regardless of the number of samples obtained from that waterbody. This was required since the draft Database contains variable amounts of data for individual waterbodies. There are several ways to produce such a representative value, with potential advantages and drawbacks to each of these methods. ENSR developed the following approach for processing of nutrient data from the database and presented it in the September 30, 1999 meeting. The data for each L/P/R waterbody were selected and reduced to representative value. The representative values were pooled to get a population value for each trophic parameter of interest. The geometric mean was selected as an appropriate indicator of central tendency.

The data selection and reduction sequence is as follows.

1. Average measurements made on a given station, date/time, and depth. This is done to take into account the possibility for duplicate sampling/analysis;

2. Select measurements made at each station in the upper 5 meters of the water column during the summer index period (July-September);

3. Calculate geometric mean of measurements made at a given lake/pond or reservoir during the summer index period. This minimizes the influence of outliers on the overall indicator for a given waterbody. The geometric mean was selected as an indicator of central tendency based on the apparent log-normal distribution of the parameters (see Section 5.2);

4. A population is obtained from the single geometric means of each lake/pond or reservoir. This ensures that the waterbodies contribute equally to the overall population since some waterbodies have been sampled more frequently than others.

As was discussed in the Technical Memorandum (1999a), the selection of the period of the summer index period is consistent with the timing of the more apparent manifestations of eutrophication (low transparency, nuisance algal blooms) and with greater recreational use of lakes and ponds. Discussion regarding the advisability of using summer vs. spring phosphorus concentrations to characterize lakes/ponds is contained in Section 7.1.

It should be noted that the approach followed in this project is very similar but not identical to a recently proposed draft sequence proposed by the EPA (Gibson, 1999).

4.5 Data Gaps Analysis

The overall objective of the Ambient Nutrient Data Project is to provide a sufficient database to support decision-making for ecoregional nutrient criteria development in New England. A Data Gaps analysis was conducted to identify potential data gaps with regard to numbers and spatial distribution of the waterbodies in the draft Database, with re-examination of collected (but deferred) data for potential inclusion. However, it should be recalled that the draft Database is not intended to be a comprehensive compilation of waterbodies in New England but, rather, a collection of data that provide good representation of the expected range of trophic state indicators for similar waterbodies in an ecoregion. The following section describes this process.

4.5.1 Establishment of Target Ranges for Waterbodies

A strategy for identifying the target number of waterbodies was developed as part of the Technical Memorandum (ENSR, 1999a). Based on estimated numbers of waterbodies in New England (Larsen et al., 1991, Peterson et al., 1998), a set of target goals were established to get approximately (5$\pm$%) of this estimated population for a sub-sample of the population. Five percent of the estimated lake population was assumed to be somewhere in the range of 50 to 250 lakes per ecoregion, with a target goal of 150 lakes to meet the decision-making needs of the database. For these databases, ENSR proposed that a target range of 10% of the lakes be reference lakes, with a minimum of 10 reference lakes per ecoregion. As noted in the Technical Memorandum, ENSR evaluated the achievability of these targets in the Data Distribution Report (ENSR, 1999b).

4.5.2 Comparison of Waterbodies in Draft Database with Targets

The first step in the comparison of waterbodies in the draft Database with targets was to establish the number of lake/ponds found within the four selected ecoregions. The coordinates of the lakes, ponds and reservoirs were obtained from the state contacts or from USGS GNIS (see Section 2.2). The EPA ecoregion corresponding to each waterbody was then identified automatically from its coordinates in ArcView. Eight-digit hydrological units codes (HUC) were identified in the same manner.

Figure 4-8 shows the distribution of the lakes, ponds and reservoirs by ecoregion from the draft Database assembled from available electronic data files (i.e., prior to addition of any hard-copy reports). As can be noted on that figure, all New England ecoregions are fairly well represented with the exception of the ACPB ecoregion. The number of lakes located in each of the four major New England ecoregions is given in Table 4-3. Except for the ACPB ecoregion, the number of lakes in each ecoregion clearly exceeded the target number of 150 waterbodies established in ENSR's proposed strategy (ENSR, 1999a).

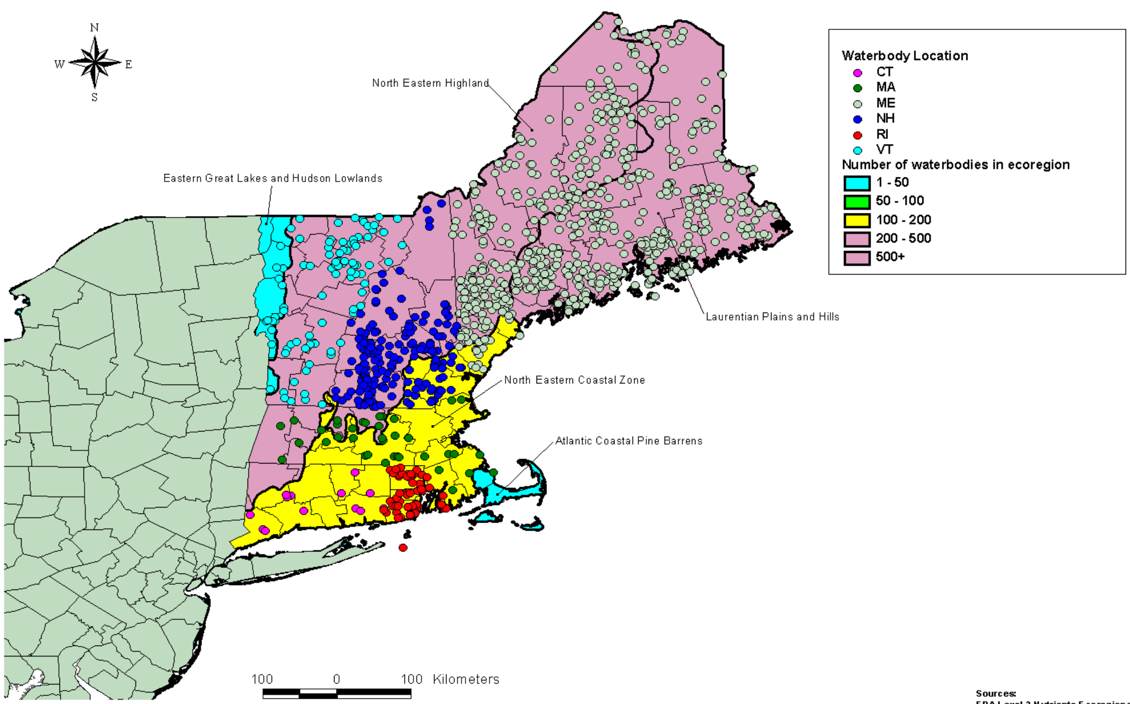


Figure 4-8 Distribution of lakes, ponds and reservoirs in Draft Database prior to adding hardcopy reports data.

Examination of the distribution of the waterbodies in Figure 4-8 indicates a wide distribution of lakes in the northern ecoregions. However, it was noted that the preponderance of lakes in the southern ecoregions were located either near the northern boundaries of the NECZ ecoregion or within Rhode Island. There was less representation of waterbodies located in Massachusetts and Connecticut, with few waterbodies located in more urbanized regions or in the western regions (i.e., Berkshire/Litchfield Hills) of these states.

The significance of these trends was that it appeared that the NEH waterbody population was less likely to fully represent the range of nutrient conditions found in the shallow lakes of urban/residential watersheds typical of many areas of southern New England. Due to their depth and watershed characteristics, urban lakes are more likely to exhibit elevated levels of nutrients and/or overabundance of aquatic macrophyte coverage. Similarly, lakes in the western regions may exhibit different responses to nutrients due to their more calcareous watersheds (Mattson et al., 1992; Carnavan and Siber, 1994). This potential spatial bias was also indicated by an apparent under-representation of eutrophic waterbodies (classified by TP concentrations) found in the draft Database relative to comparison with other New England trophic status inventories (Peterson et al., 1998). In addition, there was only 1 waterbody representing the ACPB ecoregion. These factors suggested that further waterbody data from these ecoregions should be considered for inclusion in the draft Database.

Table 4-3 Comparison of Draft Nutrient Database with L/P/R targets prior to adding hardcopy reports data.

Ecoregion	Target L/P/R	Database L/P/R
Atlantic Coastal Pine Barrens	150	1
Laurentian Plains and Hills	150	368
North Eastern Highland	150	513
North Eastern Coastal Zone	150	179
New England Total**	600	1086

**Numbers for New England include 18 waterbodies in EGLHL ecoregion and 7 waterbodies with no assigned ecoregion.

4.5.3 Inclusion of Additional Data from Hardcopy Reports

Potential sources of data for the ecoregions in question were identified from existing hardcopy sources. Qualified data for additional Connecticut lakes/ponds were readily available from two documents (USGS/CT DEP, 1995; CT DEP, 1998). Data from Massachusetts waterbodies were also available, but in a more diffuse form – existing as numerous final reports from Diagnostic / Feasibility Studies conducted under the MA Clean Lakes Program during the 1980's and early 1990's (see listing in Appendix A). These two primary sources were used to supplement the draft Database on waterbodies located in the NECZ, LPH and APCB ecoregions. Selected lake, watershed, and water quality data were transcribed from these reports for inclusion into the draft Database. Following the inclusion of this hardcopy data, the Database was compared again to the waterbody targets and examined for spatial coverage.

The hardcopy data supplemented Database was re-examined for the numerical compliance and spatial distribution among ecoregions. The coverage of the lakes/ponds in New England following the inclusion of hardcopy data is shown in Figure 4-8. Comparison between Figures 4-8 and 4-9 indicates the additional Massachusetts and Connecticut waterbodies added to the Database. While the number of waterbodies in these two states are still below those found in the Northern New England States, there is a more uniform distribution of waterbodies within the states and increased coverage of more urbanized areas (i.e. in the vicinity of Boston, Hartford) and western regions. Lakes added in these states were located mostly within the NECZ with a lesser number located in the LPH. In terms of numerical compliance with targets, the revised Database well exceeds the minimum number of waterbodies in the proposed targets (Table 4-4).

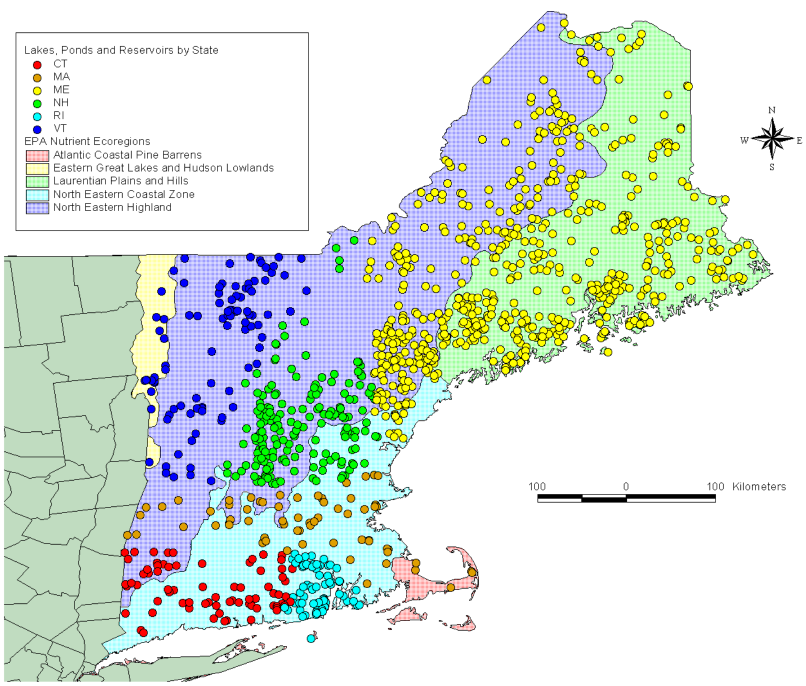


Figure 4-9 Distribution of lakes, ponds and reservoirs in Draft Database after adding hardcopy reports data.

In addition, the revised Database provides sufficient representation of reference waterbodies. The reference waterbodies indicated in Table 4-4 were identified by the Regional Nutrient Assessment Team (RNAT) members following review of water quality data within the draft Database, application of additional knowledge regarding watershed land use and development, macrophyte abundance, discharge locations, etc., and best professional knowledge. The individual states' approaches to establishing reference waterbody status are discussed in Section 6.2. The number of reference waterbodies met the target goals for 3 of the 4 ecoregions. It should be noted that in the course of the RNAT review of waterbodies for potential inclusion as reference, the overall number of waterbodies was refined (e.g., ponds that in reality were wetlands were identified and removed from the Database). Thus, both the number of waterbodies and reference waterbodies was considered acceptable for the LPH, NEH, and NECZ ecoregions. No further inclusion of additional hardcopy data was required for these areas.

Table 4-4 Comparison of Draft Database with L/P/R targets after adding hardcopy reports data.

Ecoregion	Target L/P/R	Database L/P/R	Target Ref L/P/R	Database Ref L/P/R
Laurentian Plains and Hills	150	368	15	162
North Eastern Highland	150	533	15	195
North Eastern Coastal Zone	150	223	15	38
New England Total	600	1124	60	395

In contrast, comparison of the Database to the targets shows significant under-representation of the ACPB ecoregion. Only six waterbodies were included following inclusion of the hardcopy reports and no reference waterbodies were identified. There are several potential reasons why there is insufficient data for this ecoregion. One apparent reason is that it is the smallest of the Level 3 Non-Aggregated ecoregions in New England. It is more-or-less restricted to Cape Cod, the Islands, and a small portion of southeastern Massachusetts. The geomorphology and soils of these areas also tend to reduce the size of the lakes and watersheds in this region, as many of the lakes are a direct result of glacial activity and a predominance of groundwater-fed "kettlehole" lakes exists. In addition, this area is more susceptible to low pH conditions. These may be causal factors in explaining why nutrient eutrophication is less prevalent in the region, perhaps resulting in a reduced number of investigations.

Another contribution is the lack of easily accessible data on the lakes of the Cape. The Cape Cod Commission (CCC) is in possession of considerable data, but the data is not electronically available or organized into easily accessible reports (Ed Eichner, CCC, pers. comm.). Other potential sources include studies done on or near the Massachusetts Military

Reservation (Spence Smith, pers. comm.) and data from waterbodies in or near the National Seashore park (John Portnoy, pers. comm.). This suggests that there is some additional data, but it is clear that even with this data, the ACPB data target of 150 waterbodies is unlikely to be reached. Further investigation and inclusion of additional data for this ecoregion may be warranted at some future date, but is outside the scope and level of effort in the present program. It was concluded that this data gap would not be addressed within the confines of the present program. Accordingly, the ACPB waterbodies were deferred from the Database.

4.5.4 Summary of Data Gaps Analysis Results

A Data Gaps Analysis was conducted on the draft Nutrient Database to identify its sufficiency regarding representation of ecoregions, numbers of waterbodies, and numbers of reference waterbodies. An initial evaluation of the draft Database indicated numerical deficiency for lakes/ponds in the ACPB ecoregion and for rivers and streams in general. The numerical targets for waterbodies were met for the LPH, NEH, and NECZ ecoregions, but the spatial distribution of waterbodies in the latter suggested an under-representation of lakes/ponds from Connecticut and Massachusetts's urban/residential areas and western regions. Hardcopy data were obtained for these two states and applied to the draft Database. A re-evaluation of the hardcopy-supplemented Database indicated that target numbers for waterbodies and more uniform spatial representation were achieved for the LPH, NECZ, and NEH ecoregions. In addition, reference waterbodies (as identified by Regional Nutrient Assessment Team members) for these ecoregions exceeded target numbers. Severe data deficiencies still exist for lakes/ponds in the ACPB ecoregion and for Rivers/Streams in general. Both categories were deferred from further analysis. Following the additions and refinements discussed above, the resultant Database was designated as the draft final Nutrient Database for the purposes of statistical summary (Section 5.0) and preliminary draft criteria development (Section 6.0).

The draft final Nutrient Database was distributed to RNAT members with the draft Data Synthesis Report. As part of their review, RNAT members reviewed data for L/P/R waterbodies in their states and reported any discrepancies to ENSR. Following resolution of these discrepancies, the Nutrient Database was finalized and released as the New England Nutrient Database – Lakes/Ponds/Reservoir; Version 1.0 on CD-ROM format in April 2000.

.

5 Summary Statistics

The finalized Nutrient Database for L/P/R waterbodies represents a large compilation of recent water quality data from New England lakes, ponds and reservoirs, collected from a multitude of sources that includes federal and state agencies, academic institutions, volunteer monitoring groups and Tribal Indian Nations. The nature and characteristics of the L/P/R waterbodies in the Nutrient Database are discussed in Section 5.1. Water quality characteristics for New England and the three major ecoregions of interest (LPH, NEH, NECZ) are presented in Section 5.1.2. Trophic state classification of the waterbodies is considered in Section 5.2.2.

5.1 Characteristics of Waterbodies in Nutrients Database

5.1.1 Lake Size and Depth

The distribution of morphological parameters (i.e., size and depth) for the waterbodies in the Nutrient Database are presented in Table 5-1. The size of lakes ranged from <1 ha to more than 30,000 ha, with a mean depth ranging from 20 cm to 42 m. It can be seen that the two northern ecoregions (LPH, NEH) contain populations of larger and deeper lakes (median surface area, 64-133 ha; median depth, 4.3-4.4 m as compared to the southern ecoregion (NECZ) lakes (median surface area 29 ha; median depth, 2.8 m). In addition, the largest lakes were located in the northern ecoregions. These values suggest that the northern ecoregions are more likely to contain more thermally stratified waterbodies than the southern ecoregions, but stratification status was not determined for the waterbodies in question. Potential differences in water chemistry between lakes in the northern and southern ecoregions were further evaluated below (Section 5.2).

Table 5-1 Size and depth of lakes, ponds and reservoirs in the Nutrient Database.

	Lake Area (ha)				Lake Mean Depth (m)			
Ecoregion/ Statistical Parameter	Total 3 Ecoregions	LPH	NECZ	NEH	Total 3 Ecoregions	LPH	NECZ	NEH
Arithmetic Mean	328.2	433.8	72.1	357.8	5.2	5.2	3.5	5.7
Median	66.0	133.0	28.9	64.0	4.1	4.3	2.8	4.4
Std. Dev.	1349.2	919.6	160.2	1776.4	4.1	3.5	2.6	4.7
Min	0.4	1.0	1.0	0.4	0.2	0.6	0.7	0.2
Max	30876.0	7405.0	1673.4	30876.0	42.7	24.2	16.3	42.7
No. Reported	1072	350	204	518	996	350	146	500

5.1.2 Land Use

An important determinant of water quality for a lake/pond is the land use occurring in the watershed. As shown previously, the land use differs among the four ecoregions (see Figure 4-6). Determination of land use in the watersheds of the waterbodies in the

Nutrient Database was not feasible due to the non-availability of electronically-delineated watershed maps. However, as an indicator of potential land use/land cover in the vicinity of the lakes/ponds in the Nutrient Database, a buffer of 5 km radius was defined in GIS around the center of each lake/pond. The relative importance of major land use/land cover groups was determined for that buffer area using ArcView coverages of land use in New England (EPA, 1999c). Land use/land cover distributions were compared for waterbodies in the various ecoregions to determine if differences could be noted among lakes found in each of the ecoregions in terms of typical land use/land cover in the vicinity of the lakes. The results of the analysis are presented in Figure 5-1. The figure presents the average land use/land cover distributions within the buffer zone around each of the lakes within a given ecoregions.

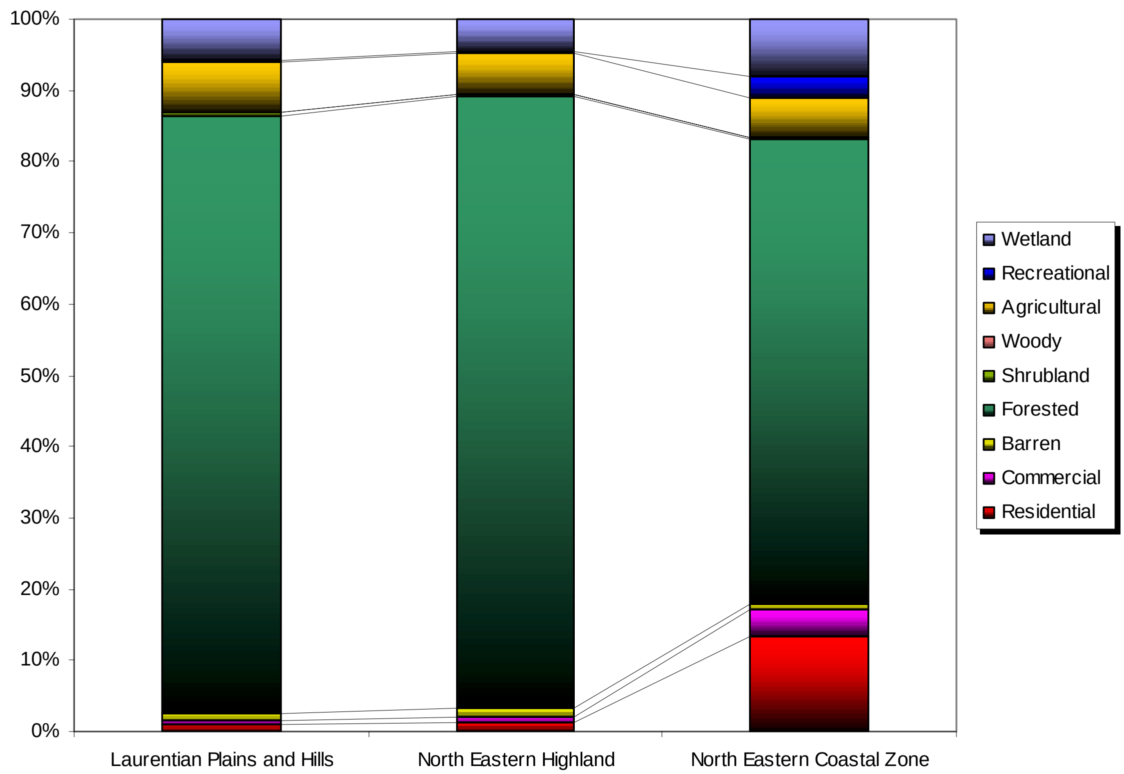


Figure 5-1 Land use in 5-km buffer zone around lakes, ponds and reservoirs of three New England Ecoregions.

Figure 5-1 shows land use/land cover distributions are fairly consistent for the LPH and NEH ecoregions but that lake buffer zones within the NECZ ecoregion contains a higher fraction of residential and commercial areas than LPH and NEH. This provides some indication of the potential determinant factors for water quality in these ecoregions.

5.2 Water Quality

The water quality data of the waterbodies contained in the Nutrient Database were characterized. The data processing sequence presented in Section 4.4 was applied to the waterbodies for derivation of representative values of the four key trophic parameters. The dataset population obtained is composed of one representative value per lake, pond or reservoir, using the geometric mean of all measurements taken in the upper 5 meters during the summer index period as the indicator of central tendency. A summation of the geometric means for th L/P/R watershed in the Nutrient Database are contained in Appendix C, Table 3.

The waterbody geometric means for the four key trophic parameters in the Nutrient Database are shown in Figures 5-2 to 5-5. These figures show the frequency distribution of geometric means of TP, TN, chl *a*, and SDT, respectively, in lakes/ponds of the Nutrient Database.

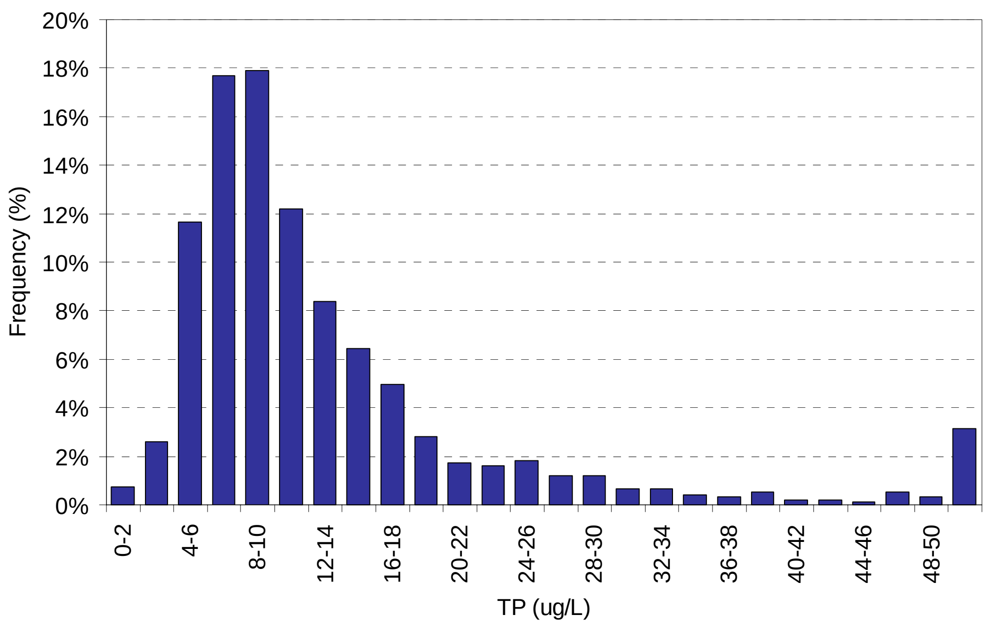


Figure 5-2 Frequency distribution of Total Phosphorus measurements in lakes, ponds and reservoirs in New England.

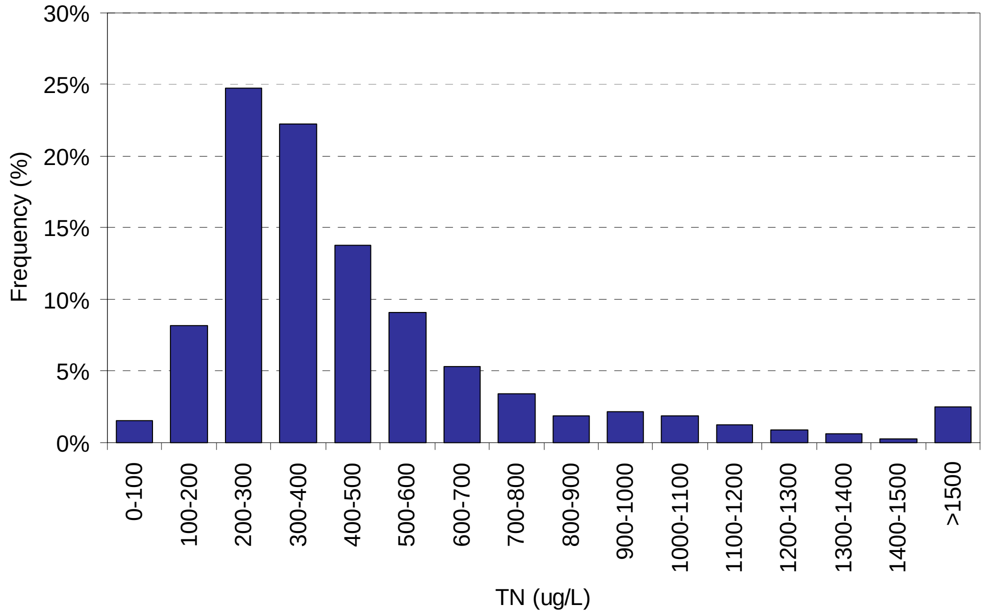


Figure 5-3 Frequency distribution of Total Nitrogen measurements in lakes, ponds and reservoirs in New England.

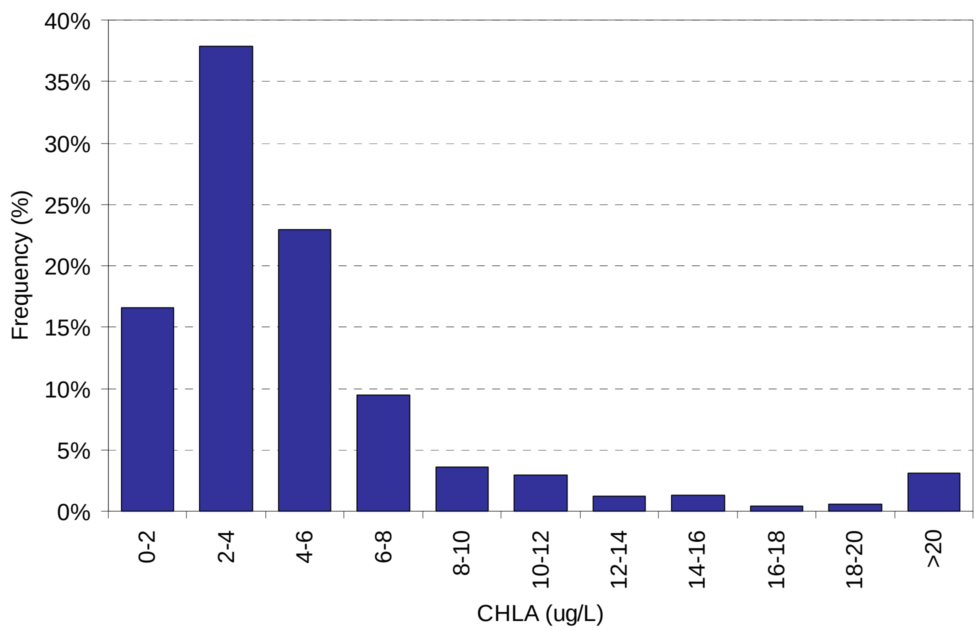


Figure 5-4 Frequency distribution of Chlorophyll a measurements in lakes, ponds and reservoirs in New England.

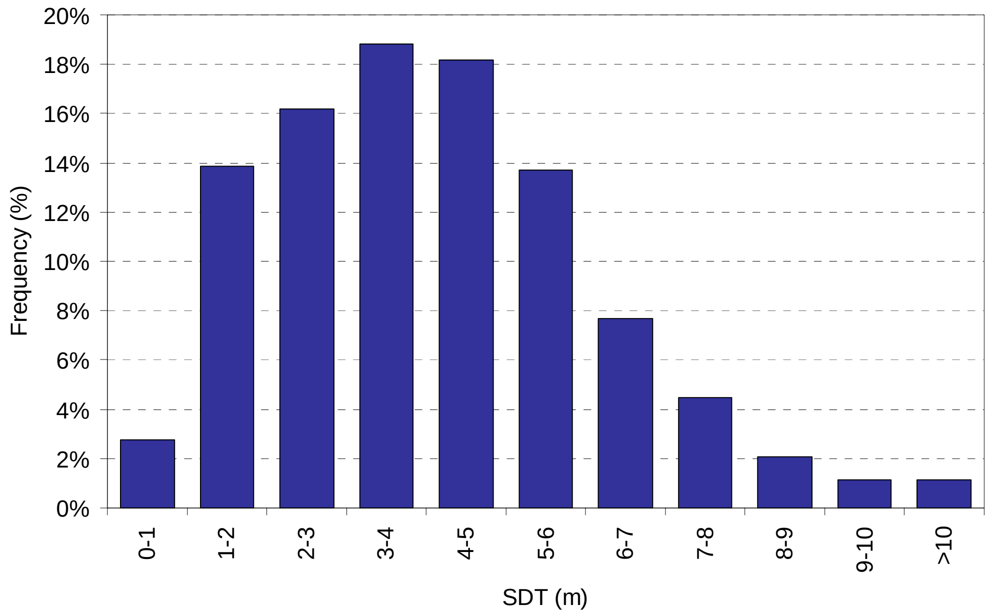


Figure 5-5 Frequency distribution of Secchi Disk Transparency measurements in lakes, ponds and reservoirs in New England.

These figures indicate that all four populations of lake/pond parameters were log-normally distributed; strongly so for the nutrients and chlorophyll, and to a lesser degree for the transparency data. The shape of the data distributions do not indicate a bimodal distribution indicative of a clear underlying division among lakes in different regions or lake size classes. Based on these distributions, it appears that differences among ecoregions or lake classes are not likely to be of an order of magnitude, as seen in some regional lake classification schemes (e.g., Minnesota lakes described by Heiskary et al., 1987; 1988). This suggested that If an underlying differentiation existed, it was likely to be a more graduated response function. Parameters were separated into different ecoregions and classes to investigate this possibility.

5.2.1 Distributions of Trophic Parameters by Ecoregion

The distribution of the key trophic parameters (TP, TN, chl *a*, SDT) was considered by ecoregion (Figures 5-6 to 5-9). These figures indicate that some parameters were strongly influenced by ecoregion while other parameters showed little separation among areas. Figure 5-6 displays the distribution of TP values across the three ecoregions with the lines indicating a computer-generated polynomial best fit solution to the data. Figure 5-6 indicates that while there is great overlap in values for the three ecoregions, the phosphorus values increase going from NEH to LPH to NECZ. In particular, the NECZ distribution is flatter and skewed to the higher values (note higher fraction of values > 50 ug/l). This pattern suggests that underlying patterns do exist for phosphorus concentrations in the waterbodies among ecoregions.

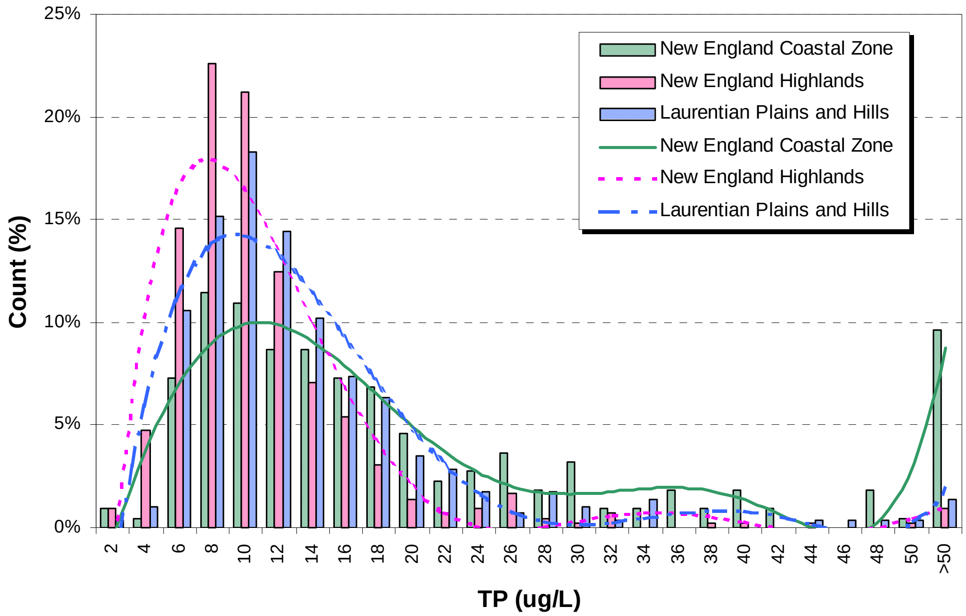


Figure 5-6 Frequency distribution of Total Phosphorus measurements in lakes, ponds and reservoirs for the three main ecoregions of New England.

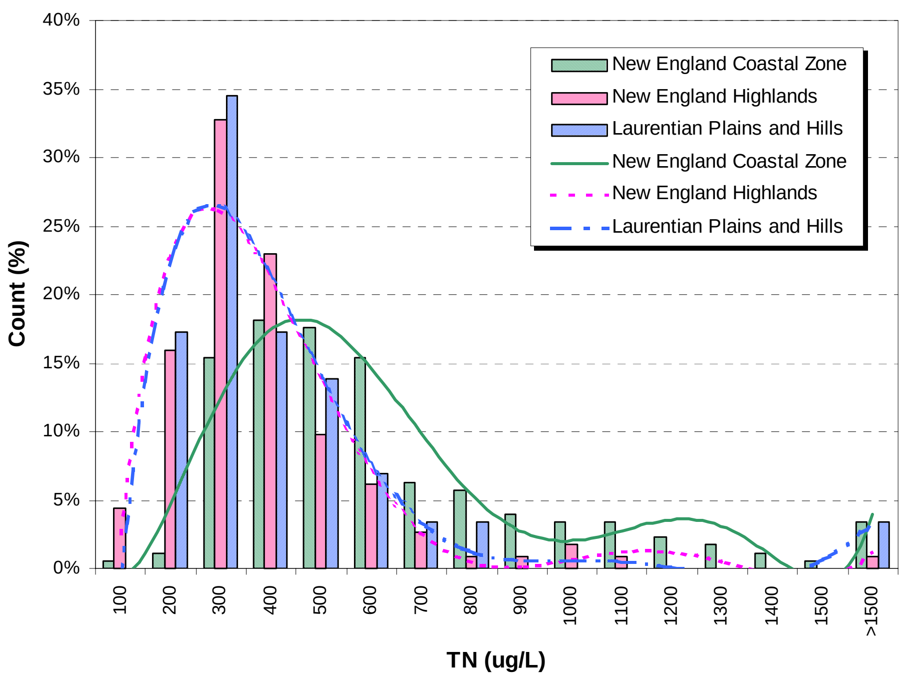


Figure 5-7 Frequency distribution of Total Nitrogen measurements in lakes, ponds and reservoirs for the three main ecoregions of New England.

A slightly different pattern is displayed for TN (Figure 5-7). In this case, the TN values between the NEH and LPH are nearly coincident, with the only observable difference between those two and the NECZ. Interpretation of this curve should be tempered by the recognition of the smaller dataset available for TN in the Nutrient Database, however.

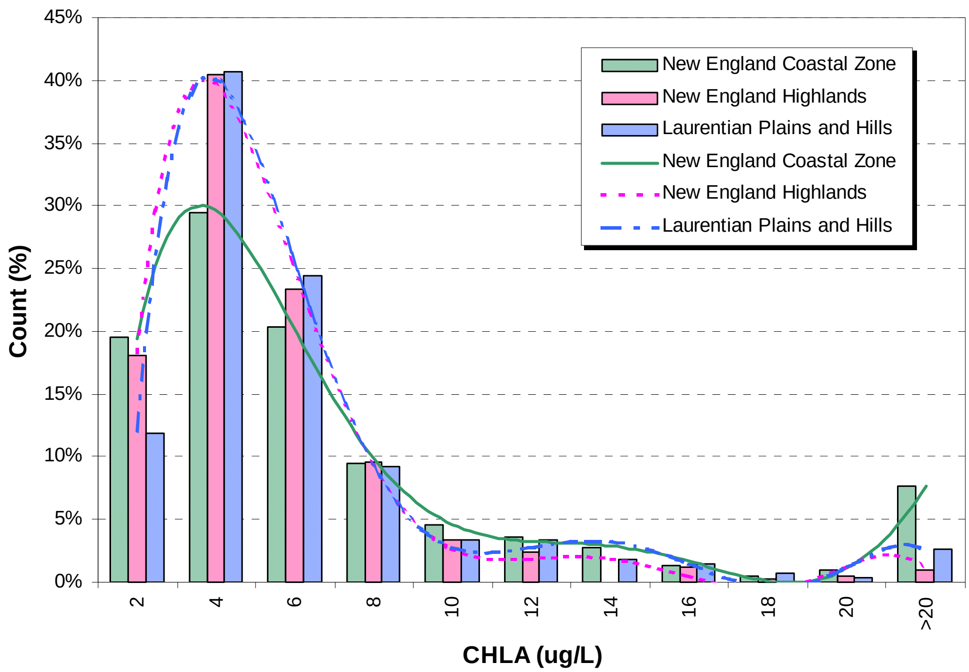


Figure 5-8 Frequency distribution of Chlorophyll-a measurements in lakes, ponds and reservoirs for the three main ecoregions of New England.

The pattern of chl *a* in New England lakes/ponds provides a contrast to the nutrient data (Figure 5-8), with little difference among the distributions of the three ecoregions. NECZ contains a higher proportion of the lowest and highest values, displaying a more complex pattern than for the nutrients.

Figure 5-9 shows the greatest contrast between regions for the trophic parameters occurs for SDT measurements. The SDT values of the NECZ zone peaking in the 2-3 m range, while the LPH and NEH medians are in the 4-5 m range. These trends are consistent with the shallower depth typically found in the NECZ lakes/ponds (Table 5-1), at least a portion of which are man-made impoundments, and because the maximum value that can be observed depends largely on a lake's depth. Figures 5-6 to 5-9 indicate some apparent differences between the ecoregions, with regard to at least some of the trophic parameters. This analysis was extended to consideration of ecoregion based statistical indicators (see Tables 5-2 through 5-5) in New England lakes/ponds. It should be noted that the numbers of waterbodies for which a particular parameter was analyzed differed between trophic parameters (i.e., most waterbodies had 3 of 4 trophic parameters, but not all).

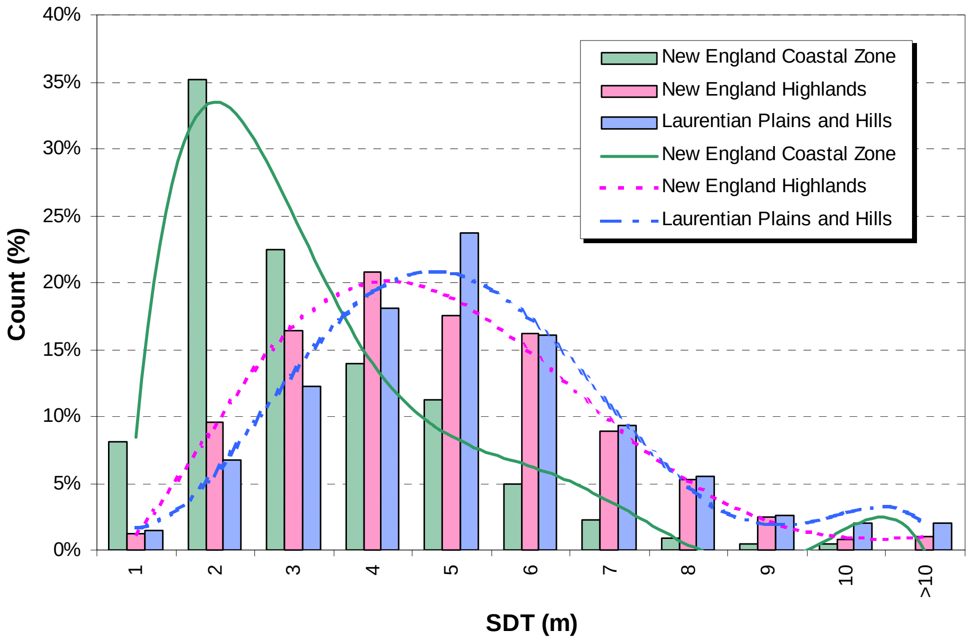


Figure 5-9 Frequency distribution of Secchi Disk Transparency measurements in lakes, ponds and reservoirs for the three main ecoregions of New England.

Table 5-2 Statistical Summary of Total Phosphorus by Ecoregion.

Ecoreregion / Statistical Parameter	Total 3 Ecoregions	LPH	NECZ	NEH
Arithmetic Mean	15.18	13.36	26.38	10.62
Median	10.00	10.81	14.35	8.78
Std. Dev.	22.37	9.93	41.27	8.01
Min	0.95	2.00	0.95	1.00
Max	376.25	87.12	376.25	77.94
No. Lakes, Ponds and Reservoirs	928	284	219	425

Table 5-3 Statistical Indicators of Total Nitrogen by Ecoregion.

Ecoreregion / Statistical Parameter	Total 3 Ecoregions	LPH	NECZ	NEH
Arithmetic Mean	484.70	378.76	588.27	349.66
Median	370.00	300.00	447.83	300.00
Std. Dev.	401.59	301.31	468.85	227.00
Min	20.00	153.00	122.92	20.00
Max	3796.77	1770.00	3796.77	1877.39
No. Lakes, Ponds and Reservoirs	319	29	177	113

Table 5-4 Statistical Indicators of Chlorophyll a by Ecoregion.

Ecoreregion / Statistical Parameter	Total 3 Ecoregions	LPH	NECZ	NEH
Arithmetic Mean	5.47	5.36	7.52	4.46
Median	3.70	3.95	4.08	3.47
Std. Dev.	8.36	5.24	14.56	4.41
Min	0.10	0.80	0.20	0.10
Max	172.25	50.57	172.25	51.65
No. Lakes, Ponds and Reservoirs	911	270	221	420

Table 5-5 Statistical Indicators of Secchi Disk Transparency by Ecoregion.

Ecoreregion / Statistical Parameter	Total 3 Ecoregions	LPH	NECZ	NEH
Arithmetic Mean	4.13	4.73	2.71	4.36
Median	3.94	4.48	2.34	4.09
Std. Dev.	2.07	2.08	1.58	1.97
Min	0.39	0.71	0.39	0.63
Max	13.70	13.70	9.14	13.24
No. Lakes, Ponds and Reservoirs	1052	340	221	491

Again it should be noted that the values displayed in Tables 5-2 to 5-5 represent the statistical profile of the set of representative values (i.e., geometric means of data in the 0-5 m profile of lakes/ponds during the July-September index period).

To evaluate the significance of the differences shown in Tables 5-2 to 5-5, the median values of the trophic parameters were statistically tested to detect significant differences between medians of different ecoregions. Use of a two-sample t-test requires that the two sampled populations be normally distributed and have equal variances, characteristics not met by the trophic parameters in the Nutrient Database. Therefore, in order to determine if a difference exists between two populations, the nonparametric analogue (Mann-Whitney) to the two-sample t-test was employed. As for many other nonparametric procedures, the actual measurements are not employed, instead the ranks of the measurements are used. The Mann-Whitney test statistic is the U metric. If the calculated U is greater than the test statistic U, the null hypothesis (no difference in the two populations) is rejected. A comparison of all possible pairing of the ecoregions was conducted.

Table 5-6 presents the results of the pair-wise Mann-Whitney test with a statistical significance of alpha equal to 0.05, and show statistical differences exist between at least some of the ecoregions. Significant differences in total phosphorus concentrations exist between all three ecoregions ($p < 0.05$). A significant difference was recorded between LPH and NECZ ($p < 0.05$) and NEH and NECZ ($p < 0.05$) in total nitrogen concentrations. There was no difference between LPH and NEH regions. Chlorophyll a concentrations differed significantly between NEH and NECZ ($p = 0.0054$) and LPH and NEH ($p = 0.0059$). No difference was noted between LPH and NECZ. Secchi disk transparency differed significantly between all three ecoregions ($p < 0.05$).

Table 5-6 Results of Mann-Whitney Test for Comparisons Among Ecoregions

	LPH vs NECZ	NECZ vs NEH	LPH vs. NEH
Total Phosphorus	0.0001*	0.0001*	0.0001*
Total Nitrogen	0.0001*	0.0001*	0.8218
Chlorophyll a	0.6325	0.0054	0.0059
Secchi Disk	0.0001*	0.0001*	0.0105

* = test is significant at p >0.0001.
Shading indicates significance at alpha = 0.05

The median values were also compared to other available data of general coverage in New England, specifically the EMAP database developed by Peterson et al., (1998) as an approximation of lake/pond conditions in New England (Table 5-7). The EMAP project also used ecoregions to differentiate areas, but did not use the EPA Level 3 Non-aggregated Ecoregions. Two of the EMAP ecoregions used - New England Uplands (NEU), and Coastal Lowland Plains (CLP), correspond approximately to a combination of the LPH and NEH, and NECZ, respectively. Comparison between these pairings shows similarity between the characteristics of the NEU and LPH/NEH ecoregions, while the CLP and NECZ show more eutrophic conditions. The higher levels found in the CLP may be due to the fact that this ecoregions extends southward along the Atlantic Seaboard and includes lakes/ponds in urbanized areas of New York and New Jersey.

Table 5-7 Comparison of NE Nutrient Database with EMAP Data

Data Source	Ecoregion	Median TP (ug/l)	Median TN (ug/l)	Median Chl a (ug/l)	Median SDT (m)
EMAP Database	NEU	10.7	341	4.6	2.7
(Peterson et al. 1998)	CLP	26.0	502	7.7	1.5
NE Nutrient Database	LPH	10.8	300	4.0	4.5
	NEH	8.8	300	3.5	4.1
	NECZ	14.4	448	4.1	2.3
	NE Total	10.0	370	3.7	3.9

Notes:
NEU = Northeastern Uplands; NEH = North Eastern Highland
CLP = Coastal Lowland and Plateau; NECZ = North Eastern Coastal Zone
LPH = Laurentian Plains and Hills

5.2.2 Trophic Classification of Waterbodies

One of the more powerful paradigms in limnology is the concept and classification of lakes as to their so-called trophic state. A trophic state classification is typically based on a generally recognized set or range of chemical concentrations and physical and biological responses. Lakes are generally classified as oligotrophic, mesotrophic, or eutrophic; the three states representing a gradient between least affected to most impacted waterbodies. Classification is based on the proximity of a lake's chemistry and biology to the list of characteristic for a specific trophic type.

Classification may be based on both quantitative (e.g., chemical concentrations, turbidity) and/or qualitative factors (e.g., presence of pollution-tolerant species, aesthetic appearance). While this system is widely accepted, there is no consensus regarding the absolute nutrient or trophic parameter value that defines a waterbody trophic state, although some guidelines have been suggested (see Section 2.0 in the *Nutrient Criteria Technical Guidance Manual: Lakes and Reservoirs* (EPA, 1999)). Indeed, it should be remembered that classification of lakes into the categories produces an arbitrary difference among lakes that may show very little differences in nutrient concentration. Despite its limitations, the trophic state concept is easily understood and widely used by limnologists, lake associations, state agencies, etc., to classify lakes and manage lakes.

Application of a trophic state classification to the New England Nutrient Database is useful for a number of reasons. It provides a simple, comprehensible index of the water quality of lakes within an ecoregion. It allows comparison with previously classified L/P/R waterbodies. Finally, it can be used as an indirect means of linking impairment of designated uses with critical nutrient levels or threshold values (i.e., the transition from one trophic state to another is likely associated with effects on designated uses).

As part of the evaluation of the Nutrient Database, waterbodies were classified according to the Carlson Trophic State Index (TSI), a widely used indicator of trophic state (Carlson 1977),. Carlson's TSI is a plant biomass-based index that relates the relationship between trophic parameters to levels of lake productivity. The TSI method provides three equations relating log-transformed concentrations of TP, chl *a*, and SDT to algal biomass, resulting in three separate TSI scores (e.g, TSI(TP), TSI(chl *a*), TSI(SDT)). The three equations are scaled such that the same TSI value should be obtained for a lake regardless of what parameter is used. Comparison of the results of the TSI system to more traditional trophic state classification identified TSI scores that are associated with the transition from one trophic state to another (Carlson, 1977).

For purposes of this report, we used a system assuming thresholds or criteria for the transition from a oligotrophic to mesotrophic state (estimated as a TSI value of 35) and for transition from a mesotrophic state to a eutrophic state (estimated as a TSI value of 50). These criteria are generally consistent with those contained in Table 7.2 of *Technical Guidance Manual* (U.S. EPA, 1999a), although that table suggests more intermediate transition states as well. Table 5-8 indicates the numeric values for TP, chl *a* and SDT that correspond with TSI criterion values of 35 and 50 used to identify potential trophic status. For example, insertion of a value of 20 ug/l of TP into the TSI equation results in a TSI(TP) score of 47.3 and a mesotrophic status.

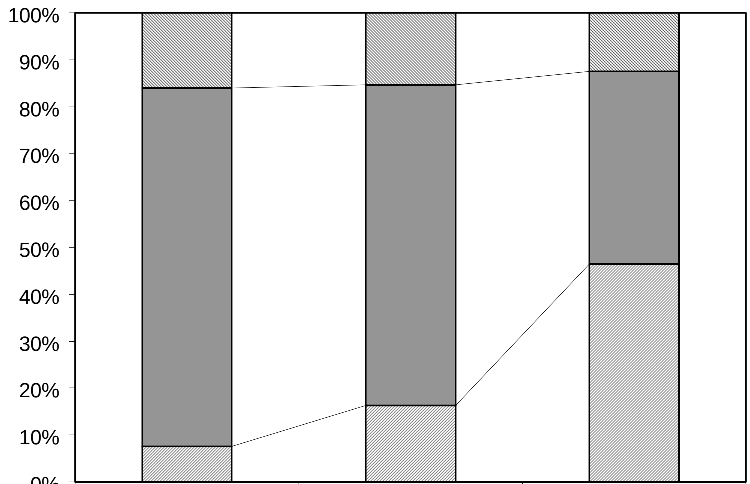
The selected TSI thresholds are based on general lake attributes and are not specific to the New England ecoregions. Alternative numeric criteria could be used equally as well, and development or refinement based on ecoregion specific information regarding trophic response and/or protection of designated uses is a specific goal in the development of final ecoregional criteria. However, Table 5-8 represents a first approximation of the trophic status for purposes of comparison between ecoregions and historic data.

Table 5-8 Trophic Status Classification based on water quality variables

Variables	Oligotrophic (TSI < 30)	Mesotrophic (30 ≤ TSI < 50)	Eutrophic (TSI > 50)
TP (ug/l)	<10	10-24	>24
Chl a (ug/l)	<1.5	1.5-7.2	>7.2
SDT (m)	>6	2-6	<2

This system was applied to waterbodies and data contained in the Database. Trophic status was estimated, based on the geometric mean of each TP, chl *a*, and SDT measurements taken during the summer index period in the upper 5 meters of the water column.

Figure 5-10 shows the distribution of trophic status in New England Lakes, Ponds and Reservoirs of the Nutrient Database for the three trophic parameters, TP, chl *a*, and SDT



	CHLA	SDT	TP
Eutrophic	155	174	119
Mesotrophic	736	778	391
Oligotrophic	73	185	443

Figure 5-10 Distribution of lakes, ponds and reservoirs by Trophic Status Class.

A breakdown of trophic status according to ecoregion is contained in Table 5-9. In this table the classifications according to each trophic parameter are indicated. No average value was generated as the TSI values are designed to be considered independently of each other. [Note: comparison can be made across ecoregions based on a single trophic parameter, however]. For comparative purposes, the percentage of reference and impacted lakes/pond identified for each ecoregion are also included on Table 5-9. Reference classifications were designated through a combination of data review and BPJ as described in Section 6.2. Impacted status was determined by listing of the waterbody on the current 303(d) list based on nutrient-related factors. The interrelationship of the trophic status for a waterbody as indicated by the three different trophic parameter TSI score was not probed in this analysis and merits further investigation. Similarly, comparison between the trophic state distributions and the designated reference and impacted L/P/R waterbodies should be further considered.

Table 5-9: Predicted Trophic State based on New England Nutrient Data (trophic state determined by TSI criterion values of 35, 50).

	Ecoregions	NE			LPH			NEH			NECZ		
Parameter	Trophic Status	O	M	E	O	M	E	O	M	E	O	M	E
TP		47%	41%	13%	40%	51%	8%	62%	33%	6%	27%	44%	29%
Chl a		8%	76%	16%	3%	80%	17%	9%	81%	10%	12%	64%	24%
SDT		16%	68%	15%	21%	72%	7%	18%	72%	10%	4%	55%	41%

Notes:
NE = New England Total
LPH = Laurentian Plains and Hills
NEH = North Eastern Highlands
NECZ = North Eastern Coastal Zone

O = Oligotrophic
M = Mesotrophic
E = Eutrophic

The trophic classification generated by the Nutrient Database was compared with the EMAP data (Peterson et al., 1998) and ecoregions considered previously in Table 5-7. Table 5-10 compares the distribution of trophic states between ecoregions. Only the TSI (TP) values were used for comparison since this was the trophic classification method used by Peterson et al. (1998). As in Table 5-7, there is good comparability between the breakdown of trophic states in the NEU and LPH/NEH and CLP and NECZ pairings. As noted earlier, the higher percentage of eutrophic waterbodies in the CLP may be due to inclusion of waterbodies located in nutrient-rich urban watersheds located in coastal New York and New Jersey.

Table 5-10: Comparison of NE Nutrient Database with EMAP Data

Data Source	Ecoregion	Oligotrophic Waterbodies	Mesotrophic Waterbodies	Eutrophic Waterbodies
EMAP Database	NEU	48.5%	48.4%	3.1%
(Peterson et al. 1998)	CLP	24.8%	30.1%	45.1%
NE Nutrient Database	LPH	40.4%	51.2%	8.4%
	NEH	61.8%	32.6%	5.6%
	NECZ	26.9%	44.3%	28.8%

Notes:
Trophic status is based on TSI (TP) values.
NEU = Northeastern Uplands
CLP = Coastal Lowland and Plateau
LPH = Laurentian Plains and Hills
NEH = North Eastern Highland
NECZ = North Eastern Coastal Zone

6 Preliminary Draft Nutrient Criteria Development

This section pertains to the derivation of preliminary draft nutrient criteria for New England lakes/ponds. Section 6.1 introduces the major approaches used to derive preliminary draft nutrient criteria. Section 6.2 contains information on the identification of reference and impacted waterbodies used to support criteria development. Section 6.3 develops preliminary draft nutrient criteria through consideration of statistical indicators of reference and general waterbody distribution. Section 6.4 discusses criteria development through consideration of designated uses and literature threshold values. This approach requires further strengthening of the linkage between nutrient levels and effects on designated uses.

6.1 General Approaches to Nutrient Criteria Development

While the need for development of nutrient criteria is clearly needed, the most appropriate method to achieve this goal has not been well established. Several regional or lake-specific approaches have been successfully implemented, but there is no clear consensus among states or federal agencies regarding the best means to accomplish this goal, due to the difficulty in defining precisely what concentrations will be protective of waterbodies' water quality as well as their designated uses. Given this level of uncertainty, a conservative way to proceed is through derivation of nutrient criteria via several methods using a "weight-of-evidence" approach to establish targets.

The *Nutrient Criteria Technical Guidance Manual: Lakes and Reservoirs* (EPA, 1999) suggests several approaches to derivation of nutrient criteria and contains a useful compendium of case studies. Two of the methods described in the Lake and Reservoir *Technical Guidance Manual* were investigated and discussed at meetings with the Regional NAT. The two methods investigated were the use of target percentile (i.e., "Statistical Method") and consideration of designated uses (i.e., "Designated Use Method"). Both methods are discussed briefly below.

The Statistical Method uses two approaches for determining candidate reference condition values for TP, TN, chl *a* and SDT, and relates these reference conditions to desired nutrient ranges. In both cases, the goal is to select the threshold value from available data for a given category of L/P/R waterbodies. The EPA defines reference a condition as that representative of the least impacted conditions or what is considered to be the most attainable conditions for lakes within a state, or ecoregion. Reference conditions were established by a variety of different methods (see Section 6.2).

The first approach of the Statistical Method consists in selecting a percentile from the distribution of measured variables (in this case geometric means of trophic parameters of interest) from known reference lakes, (i.e., the highest quality or least impacted lakes). Since these lakes are already considered to be in an ideal state or at least as close as can be reasonably achieved, the approach suggests using a higher percentile of nutrient conditions as the reference condition. The Lake and Reservoir *Technical Guidance Manual* suggests the 75th percentile, although this is an arbitrary value and could be replaced with higher or lower percentiles, as considered appropriate.

The second approach suggested by the Statistical Method consists in selecting a percentile from the distribution of measured variables for a general population that includes all lakes/pond within a region or class. In this case, the percentile level chosen should be higher since the population contains a mix of degraded (impacted) and cleaner (reference) lakes. The Lake and Reservoir *Technical Guidance Manual* recommends the use of the 25th percentile. However, if almost all reference lakes within the population are felt to be impacted to some extent, the EPA guidance document suggests that the 5th percentile should be used instead. Figure 6-1 graphically summarizes the two approaches of the Statistical Method.

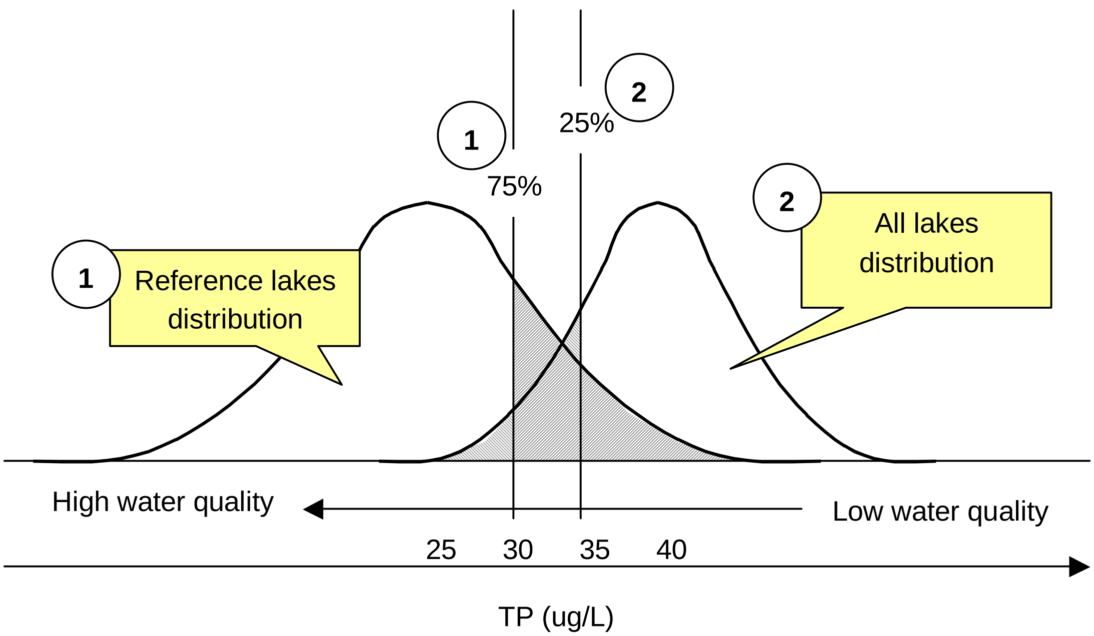


Figure 6-1   Two approaches for finding reference condition value for total phosphorus.

The second method considered in the Lake and Reservoir *Technical Guidance Manual* is not statistical in nature, but rather seeks to establish a linkage between nutrient concentrations with protection of critical waterbody functions and services, so-called designated uses, protected by water quality standards. In the Designated Use Method, literature values are used to establish nutrient levels that are expected to support water quality-related designated uses. These threshold values are those considered in the development of nutrient criteria. For purposes of this report, literature values were used based on associated trophic states and inferred effects on designated uses.   To develop ecoregional nutrient criteria, a more direct linkage between nutrient levels and effects on designated uses in New England lakes needs to be established. Developing this linkage is considered a high priority for future criteria development.

6.2 Identification of Reference, Test, and Impacted Waterbodies

The initial step in the Statistical Method is the establishment of reference waterbodies in order to establish the "reference" population used for the first approach. A related step is the identification of impacted waterbodies. Although not suggested by the Lake and Reservoir *Technical Guidance Manual*, identification of impacted waterbodies was devised as a way of checking the relative level of eutrophication in the general lakes/ponds population.

To establish reference waterbodies, RNAT members and/or designated State experts were asked to evaluate a state-specific list of waterbodies retained in the final Nutrient Database. They were asked to identify the reference lakes as those most likely to represent "ideal" or most desirable conditions based on their state-specific methods and/or best professional judgment. Water quality information provided to the experts included minimum value, maximum value, and arithmetic and geometric means of TP, TN, chl *a*, and SDT of the waterbody (see Appendix C, Table 3). Similarly, State experts were also asked to identify "impacted" waterbodies as those that show signs of nutrient-related problems. Information provided to the experts included whether the waterbody was currently listed on the 303(d) list for nutrient-related factors as indicated on Table 6-1. Accordingly, all lakes, ponds and reservoirs listed in the Nutrient Database were classified by the state experts as "reference", "impacted" or "test". The "test" category simply pertains to those waterbodies that are considered neither reference nor impacted. Table 6-2 contains a list of the number of L/P/R waterbodies identified as reference, test, or impacted in each of the three ecoregions of interest.

Table 6-1 Nutrient Related Factors Included in current EPA 303(d) Listings of Impaired Waters of New England States

Criteria	CT	ME	MA	NH	RI	VT
Ammonia (unionized)	X		X			
Algae		X			X	
Dissolved Oxygen (low)	X	X		X		X
Hypoxia	X				X	
Impaired Biologic Community		X		X	X	
Nitrates	X					
Nitrogen	X					
Noxious Aquatic Plants	X		X			
Nutrients	X	X	X		X	X
Organic Enrichment/ Low DO			X			X
Organic Enrichment Sediments						X
Phosphorus		X		X		X
Suspended Solids	X		X		X	
Turbidity	X		X		X	

The basis for assessment of reference and impacted conditions differed slightly from state to state. The assessment of the lake's conditions were generally based on consideration of a combination of factors such as the typical nutrient concentrations measured, amount of shoreline development and land use, discharges, observed macrophyte coverage, etc, as well as application of BPJ. The state-specific approaches are described below.

6.2.1 Connecticut

Connecticut generally identified reference lakes/ponds on the basis of water quality information, the results of Clean Lake Study Reports, the amount of watershed development, and BPJ. Identification of impacted waterbodies used the information above, as well as the abundance of macrophytes, presence of stormwater and/or wastewater discharges, and high levels of shoreline development. Based on the evaluation, several reference waterbodies were identified that appear to have natural mesotrophic conditions including abundant macrophytes.

6.2.2 Maine

Maine has previously developed a protocol called the "Watershed Development Ranking" (Roy Bouchard, pers. comm.) for identifying levels of anthropogenic influence on a waterbodies which are not impoundments for hydroelectric generation or run-of-river lakes with rapid flushing rates (e.g., greater than 30 flushing/year). This qualitative assessment, largely based on watershed assessment on observable topographic map features, ranks waterbodies from 1 ("very low development") to 5 ("highly developed"). This ranking is based on a set of watershed features including: watershed population density, road access to shoreline, shoreline development, accessibility of watershed via public roads, amount of agricultural activity, presence of point sources or known significant disturbances, and record of internal phosphorus recycling or known algal bloom problems (if available). The watershed development rankings were then converted. Lakes/ponds with "Very Low Development" or "Low Development" were identified as reference; lakes/ponds with "Medium development" were considered test lakes, and lakes/ponds with "Significant Disturbance" or "Highly Developed" rankings were identified as impacted.

6.2.3 Massachusetts

Massachusetts generally identified reference lakes/ponds on the basis of the water quality information, the results of Diagnostic/Feasibility (D/F) Study Reports, the amount of watershed development, and predominant water use (e.g., public water supply). Identification of impacted waterbodies used the information above, as well as the abundance of macrophytes, presence of stormwater, and high levels of shoreline development.

6.2.4 New Hampshire

New Hampshire based its reference and impacted assessments using current water quality data (chl *a* and phosphorus). Since the assessments were based on chl *a* and phosphorus, the designated uses that would be impacted by high values are drinking water and swimming. Chl *a* (phytoplankton biomass) was the only biological criterion used. Macrophytes were not considered in the development of reference conditions, as macrophyte growth is considered dependent on substrate type, water depth, etc and not on water column nutrient concentrations. [Note: macrophyte abundance is regularly considered as part of New Hampshire lake trophic classification protocol (Robert Estabrook, pers. comm.)] Impacted lakes were identified by presence of the 303(d) list.

6.2.5 Rhode Island

Rhode Island based on its reference and impacted assessments using water quality, algal blooms, macrophyte abundance, level of development in the watershed, and seasonal anoxia as evaluation factors. Some of the reference lakes were shallow or in urban settings, and identified as "good water quality for shallow pond"; good water quality for urban pond").

6.2.6 Vermont

In Vermont, reference lakes were identified on the basis of: level of watershed development (i.e., amount of developed (non-forested) land in the watershed was $\leq$ 10%; origin of the lake outlet (i.e., natural outlet with or without some artificial control; no entirely artificial impoundments); level of shoreline development (density of shoreline camps $\leq$ 1 camp per 10 acres of lake surface area); and level of disturbance (i.e., no known significant human effects on the lake from causes such as recent heavy logging in the watershed, direct highway erosion, or large water level fluctuations). The lakes assessed as "impacted" were lakes listed in Vermont's 303d list as being impaired (i.e., not meeting Vermont Water Quality Standards) because of phosphorus over-enrichment and algae problems.

Table 6-2 L/P/R waterbodies identified as Reference, Test, and Impacted by ecoregion.

	LPH	NECZ	NEH
Reference	162	38	195
Test	123	142	295
Impacted	83	43	43
Total	368	223	533

6.3 Application of Statistical Method to Develop Preliminary Draft Criteria

The reference, test, and impacted L/P/R waterbody classification scheme developed in the process described in Section 6.2 was used to develop preliminary draft nutrient criteria. One of the assumptions of this approach to nutrient criteria development is that differences exist between reference lakes and the rest of the population. In order to verify this hypothesis, the reference population formed by the lakes, ponds and reservoirs identified as "reference" lakes by the state experts was compared to the general population of lakes in New England for each of the trophic parameters. Figures 6-2 to 6-5 illustrates that comparison for TP, TN, chl *a* and SDT, respectively, where cumulative frequency distributions are given for the reference, general ("all"), lake populations. On these figures the 75th percentile for the reference lake population and the 25th percentile for the general lake population are indicated, as are the corresponding parameter values. The impacted lake population was also included for comparative purposes. Note that these percentiles are not used for SDT, where higher values denote higher transparency and better water quality. In that case, the 75th percentile for the reference and 25th percentile for the general lake population was used.

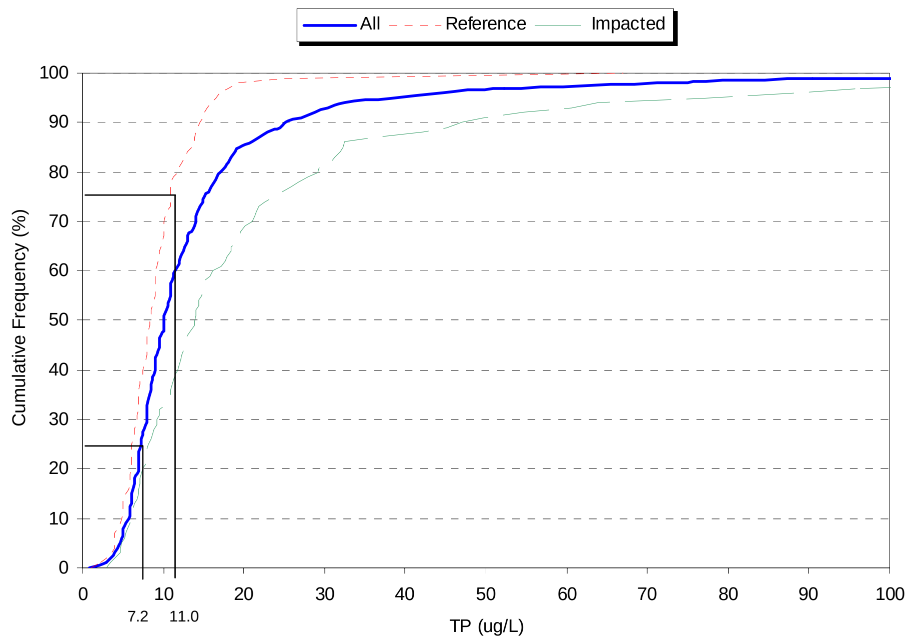


Figure 6-2 Comparison of cumulative frequency distribution of Total Phosphorus measurements for reference, impacted and all lakes populations

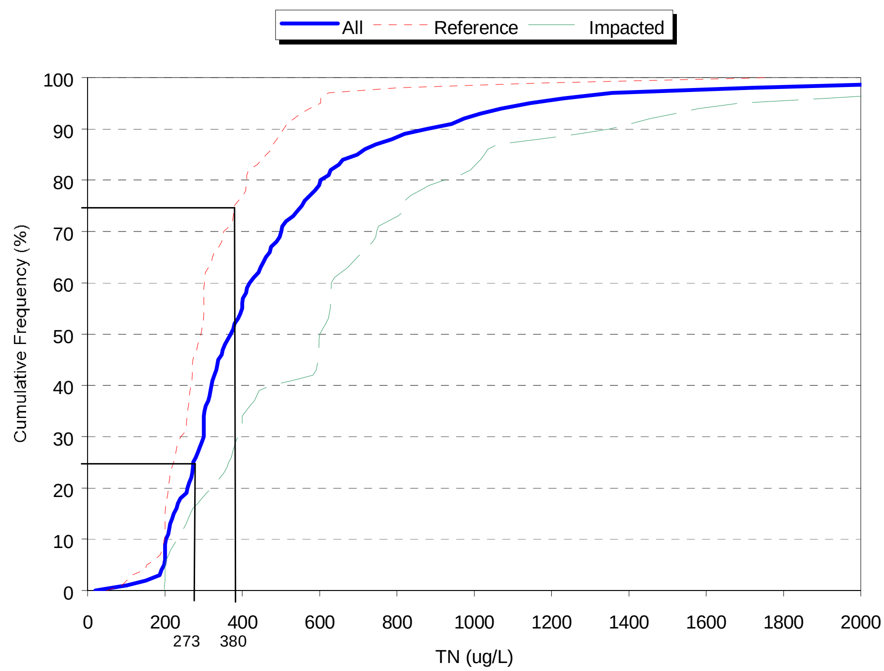


Figure 6-3 Comparison of cumulative frequency distribution of Total Nitrogen measurements for reference, impacted and all lakes populations

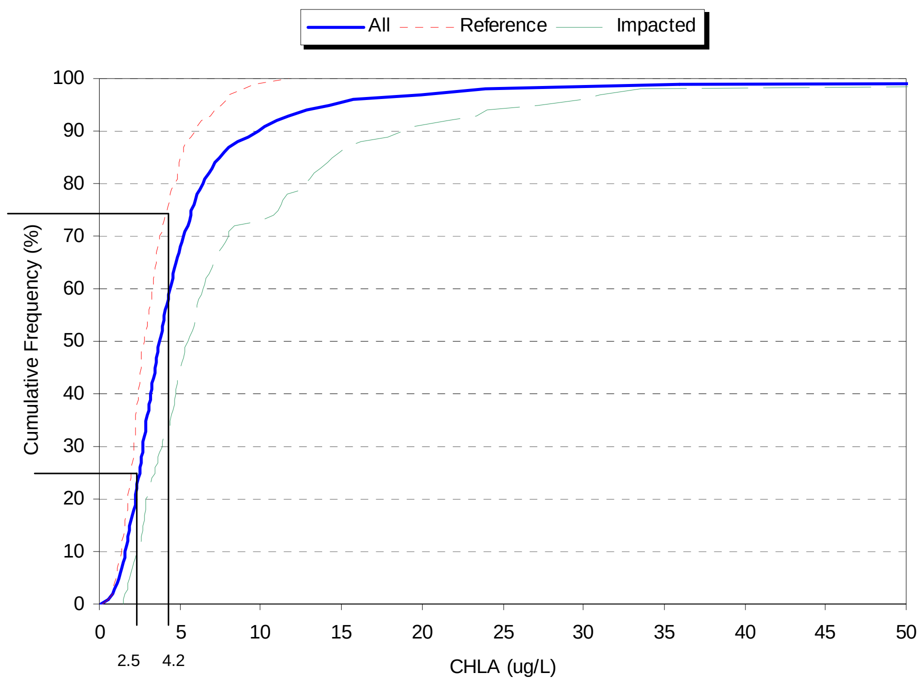


Figure 6-4 Comparison of cumulative frequency distribution of Chl a measurements for reference, impacted and all lakes populations.

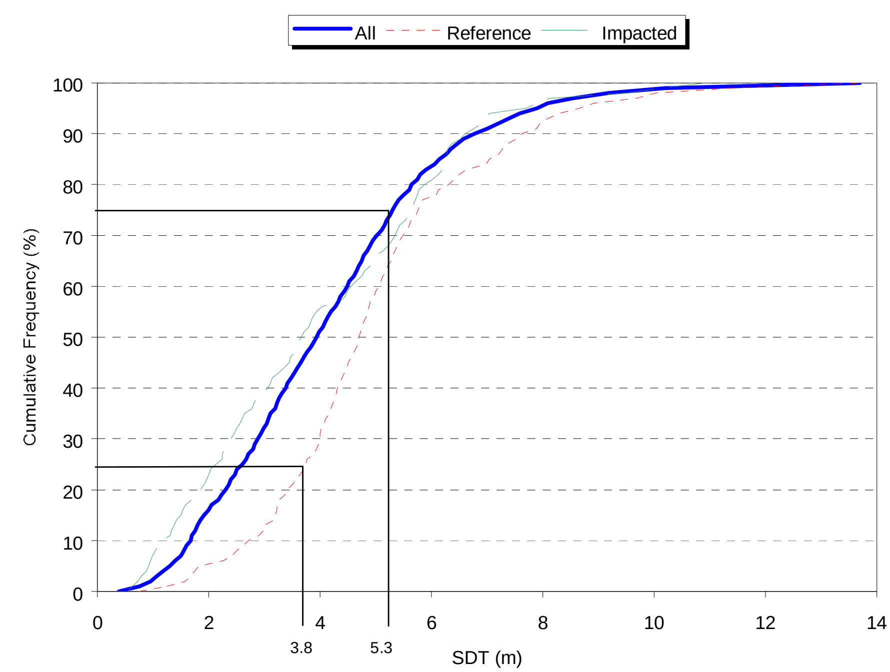


Figure 6-5 Comparison of cumulative frequency distribution of Secchi Disk Transparency measurements for reference, impacted and all lakes populations.

Figures 6-2 to 6-5 show differences in the cumulative frequency distributions of the reference and general populations for all four selected trophic parameters. As expected even greater differences are seen between the reference and impacted lake populations. It can be seen that that for TP, chl *a* and TN, the 25th percentile of the general lake population is lower than the 75th percentile of the reference lakes. In the case of SDT depth, the distinction between impacted and reference lake populations is less pronounced. This is most likely due to the strong influence of lake depth has on the measured SDT. To test this, the shallow ($\leq$5 m mean depth) and deep (>5 m mean depth) lakes were segregated. Figure 6-6 shows the distributions broken down in shallow and deep lakes with an expected separation of these two groups of distributions.

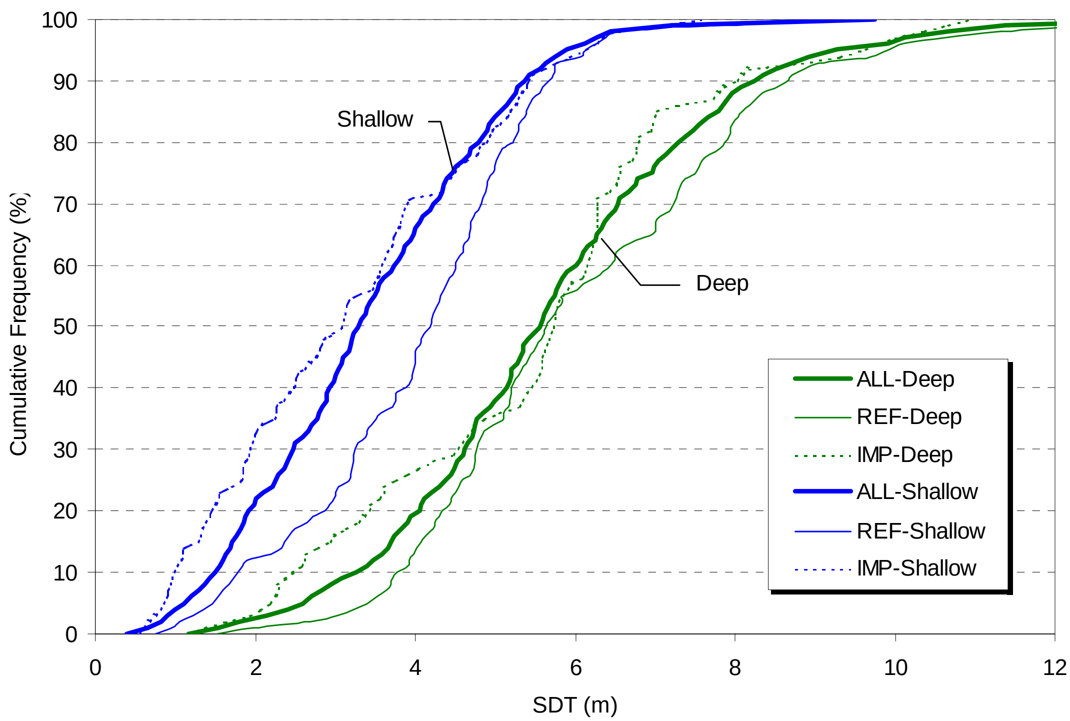


Figure 6-6 Distribution of SDT measurements by lake depth category (shallow and deep), for reference, impacted and all lakes populations.

6.3.1 Nutrient Ecoregions

Based on the analysis conducted in Section 5.3 (see Table 5-6), significant statistical differences occur between ecoregions for many of the trophic parameters. Figures 6-7 to 6-10 show the comparison of the distribution of trophic parameters measurements for the four New England nutrient ecoregions. Note that this data is similar to that displayed in Figures 5-6 to 5-9, but has been plotted to allow easier comparison between ecoregions.

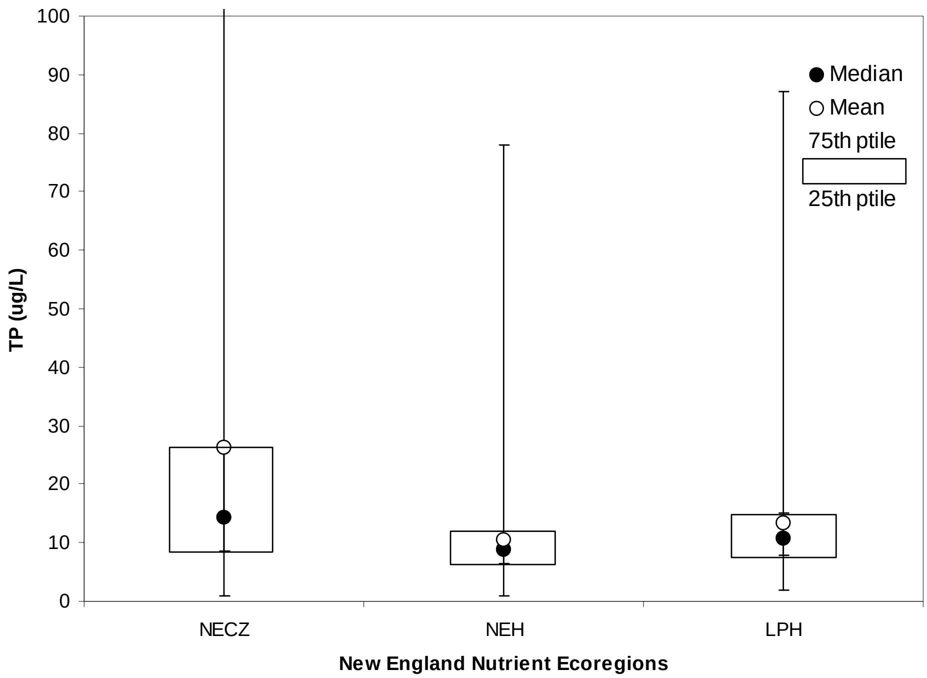


Figure 6-7 Comparison of distribution of Total Phosphorus measurements for New England nutrient ecoregions.

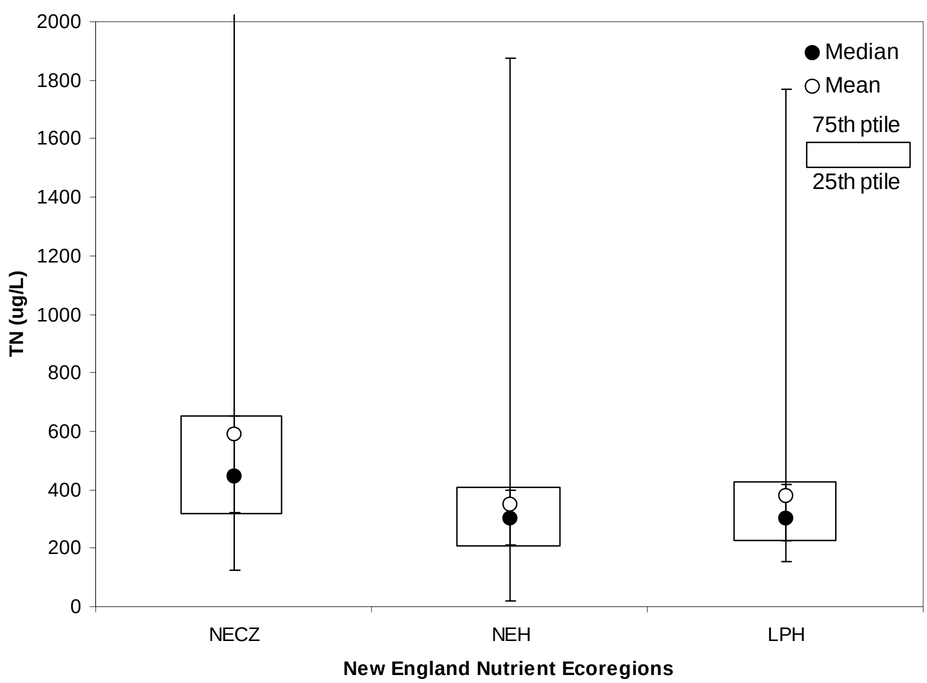


Figure 6-8 Comparison of distribution of Total Nitrogen measurements for New England nutrient ecoregions.

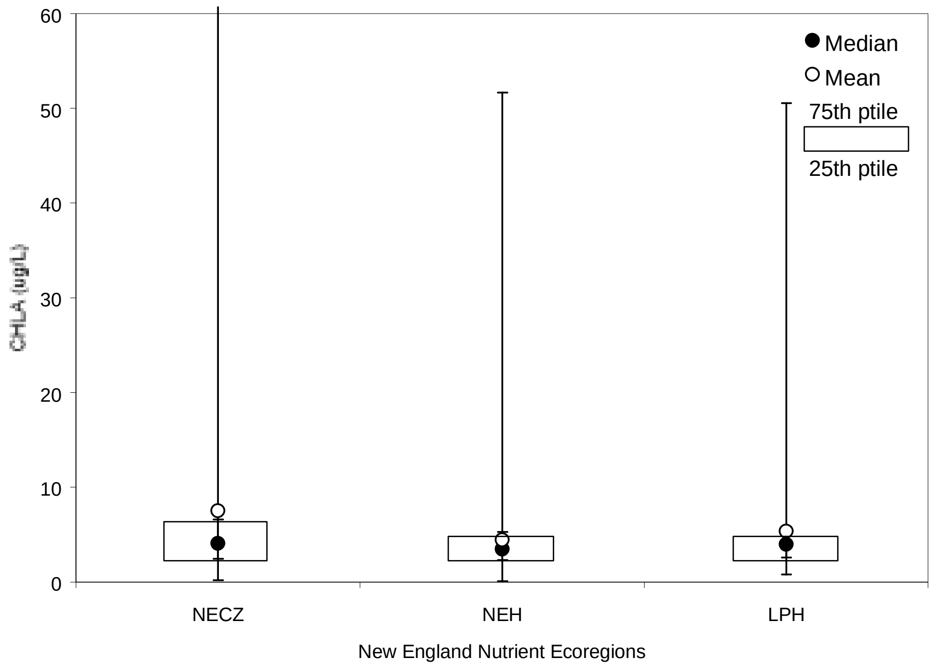


Figure 6-9 Comparison of distribution of Chlorophyll-a measurements for New England nutrient ecoregions.

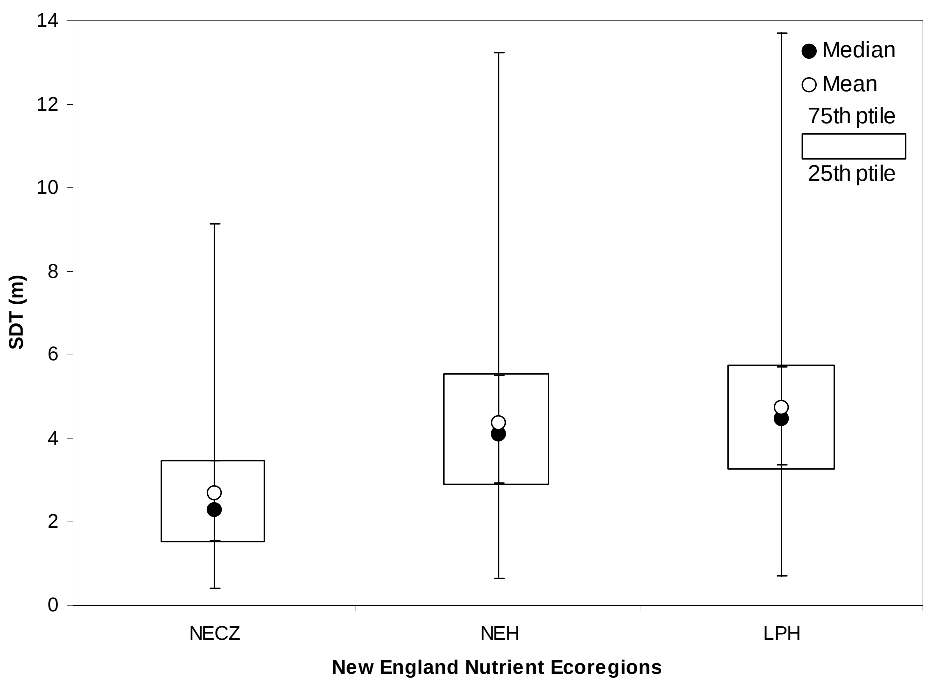


Figure 6-10 Comparison of distribution of Secchi Disk Transparency measurements New England nutrient ecoregions.

Figures 6-7 to 6-10 generally show slight difference between the New England Highland (NEH) and Laurentian Plains and Hills (LPH) ecoregions. The New England Coastal Zone (NECZ) however, consistently shows a distribution characterized by lower water quality (i.e. higher concentrations of TP, TN and chl *a*, and lower SDT).

Using the method suggested by the Lake and Reservoir *Technical Guidance Manual*, comparison was made between the general population and reference lakes for each of the nutrient ecoregions as a means to derive preliminary draft nutrient criteria. Tables 6-3 to 6-6 summarizes the statistical distributions for general and reference lakes in each of New England's nutrient ecoregions for TP, TN, chl *a*, and SDT, respectively. The 75th percentile (most impacted quartile) of reference lakes, and 25th percentile (least impacted quartile) of all assessed lakes are highlighted in the tables for TP, TN and chl *a*. In the case of SDT, the 25th percentile of reference lakes and the 75th percentile of all assessed lakes are highlighted, since for that parameter, water quality is inversely proportional to the measured secchi depth. These ecoregional-specific sets of values were used to generate draft nutrient values for TP and TN.

Table 6-3 Comparison of Total Phosphorus distributions (ug/l) in New England nutrient ecoregion for all assessed lakes and reference lake populations.

	All lakes and reservoirs			Reference lakes		
	NECZ	NEH	LPH	NECZ	NEH	LPH
Arithmetic Mean	26.38	10.62	13.36	11.32	8.52	9.64
Median	14.35	8.78	10.81	8.47	7.13	9.00
25th percentile	8.59	6.50	7.93	7.22	6.00	7.00
75th percentile	26.00	11.97	15.07	11.22	10.00	11.66
St. Dev.	41.27	8.01	9.93	8.09	6.24	3.65
Min	0.95	1.00	2.00	4.80	1.00	2.00
Max	376.25	77.94	87.12	46.63	67.00	18.33
No. Waterbodies	219	425	284	37	135	113

Table 6-4 Comparison of Total Nitrogen (ug/l) distributions in New England nutrient ecoregion for all assessed lakes and reference lakes populations.

	All lakes and reservoirs			Reference lakes		
	NECZ	NEH	LPH	NECZ	NEH	LPH
Arithmetic Mean	588.27	349.66	378.76	369.09	289.82	362.71
Median	447.83	300.00	300.00	303.43	269.50	236.00
25th percentile	322.62	213.00	225.00	270.84	200.50	202.50
75th percentile	650.63	400.00	415.00	444.49	364.23	299.25
St. Dev.	468.85	227.00	301.31	162.33	125.81	411.61
Min	122.92	20.00	153.00	208.56	50.00	153.00
Max	3796.77	1877.39	1770.00	1000.00	602.46	1770.00
No. Waterbodies	177	113	29	32	34	14

Table 6-5 Comparison of Chlorophyll-a (ug/l) distributions in New England nutrient ecoregion for all assessed lakes and reference lakes populations.

	All lakes and reservoirs			Reference lakes		
	NECZ	NEH	LPH	NECZ	NEH	LPH
Arithmetic Mean	7.52	4.46	5.36	2.25	3.23	3.69
Median	4.08	3.47	3.95	1.90	2.72	3.30
25th percentile	2.45	2.31	2.62	1.34	2.00	2.17
75th percentile	6.60	5.30	5.80	2.63	3.83	4.81
St. Dev.	14.56	4.41	5.24	1.41	1.97	2.02
Min	0.20	0.10	0.80	0.60	0.20	0.80
Max	172.25	51.65	50.57	7.53	10.59	11.70
No. Waterbodies	221	420	270	37	130	110

Table 6-6 Comparison of Secchi Disk Transparency (m) distributions in New England nutrient ecoregion for all assessed lakes and reference lakes populations.

SDT (m)	All lakes and reservoirs			Reference lakes		
	NECZ	NEH	LPH	NECZ	NEH	LPH
Arithmetic Mean	2.70	4.36	4.73	3.87	4.96	5.25
Median	2.27	4.09	4.48	4.28	4.81	4.72
25th percentile	1.55	2.92	3.35	2.70	3.71	3.99
75th percentile	3.46	5.49	5.72	4.85	5.83	5.96
St. Dev.	1.57	1.97	2.08	1.45	2.03	2.07
Min	0.39	0.63	0.71	1.21	0.75	1.00
Max	9.14	13.24	13.70	7.50	13.20	13.70
No. Observations	221	491	340	37	161	143

Table 6-6 presents the range of ecoregional preliminary draft nutrient criteria for TP and TN indicated by the two suggested percentiles. The ranges of values shown in Table 6-6 show considerable overlap between ecoregions. For all three ecoregions, the 75th percentile for the reference lakes exceeded the 25th percentile of all lakes for both TP and TN. This is similar to the case study presented in the Lake and Reservoir *Technical Guidance Manual*, using Minnesota Data.

Table 6-7 Ecoregional Preliminary Draft Nutrient Criteria (TP, TN) derived by Statistical Method

Ecoregions	Phosphorus All (25th)	Phosphorus Ref (75th)	Nitrogen All (25th)	Nitrogen Ref (75th)
Laurentian Plain and Hills	7.9 ug/l	11.7 ug/l	225 ug/l	299 ug/l
New England Highland	6.5 ug/l	10.0 ug/l`	213 ug/l	364 ug/l
New England Coastal Zone	8.6 ug/l	11.2 ug/l	322 ug/l	444 ug/l

If the midpoint of the ranges are considered, the ecoregions rank for TP - NEH (8.3 ug/l), LPH (9.8 ug/l), NECZ (9.9 ug/l) and for TN - LPH (262 ug/l), NEH (289 ug/l), NECZ (383 ug/l). For phosphorus these midpoint values would be considered near the nutrient concentrations (approximately 10 ug/l) where literatures values suggest a potential shift from oligotrophic to mesotrophic conditions may occur (Horne and Goldman, 1994). Since many of the designated uses that are of concern to lake managers are likely to be protected in lakes/ponds by nutrient concentrations resulting in mesotrophic conditions, the Statistical Method provides a very conservative approach to derivation of preliminary draft nutrient criteria. This conservatism would result in a highly protective approach and nutrient concentrations that may not be achievable in some lakes. On the other hand, this approach could be easily modified by selecting different population percentiles as the critical values, perhaps for nutrient concentrations that result in more pronounced eutrophic conditions.

6.3.2 Mean Depth

As part of the investigation of nutrient criteria, additional classifications were considered to better refine the criteria. As shown in Figure 6-6, subdivision of lakes, ponds and reservoirs in the Nutrient Database into shallow (< 5 m mean depth), and deep (>5 m mean depth) lakes provides some differentiation of SDT values. This concept was explored for TP and chl *a* (Figures 6-11 and 6-12). These figures illustrate the trend to higher TP and chl *a* in shallower lakes. This was further assessed within one ecoregion – the NECZ. Figures 6-13 to 6-15, respectively, show the influence of lake depth on distributions of TP, TN and chl *a* measurements for lakes within the NECZ ecoregion. The trend to higher nutrient concentrations for shallower lakes is indicated in Figures 6-13 to 6-14 and shows greater difference within the ecoregion than for New England. In contrast, the chl *a* distributions of shallow and deep lakes (Figure 6-15) are similar for low concentrations and separate only at the higher concentrations.

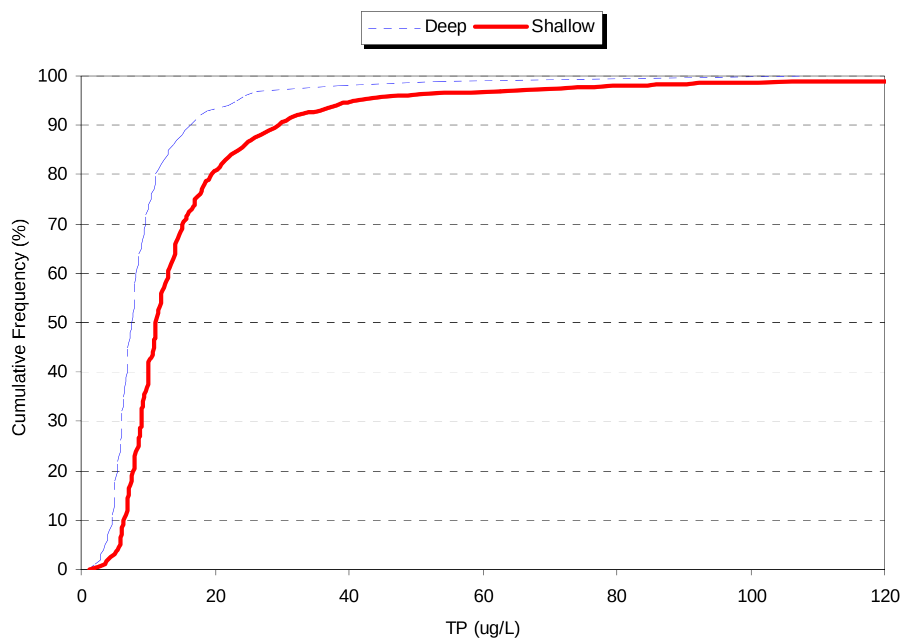


Figure 6-11 Cumulative frequency distribution of Total Phosphorus measurements in deep and shallow lakes of New England.

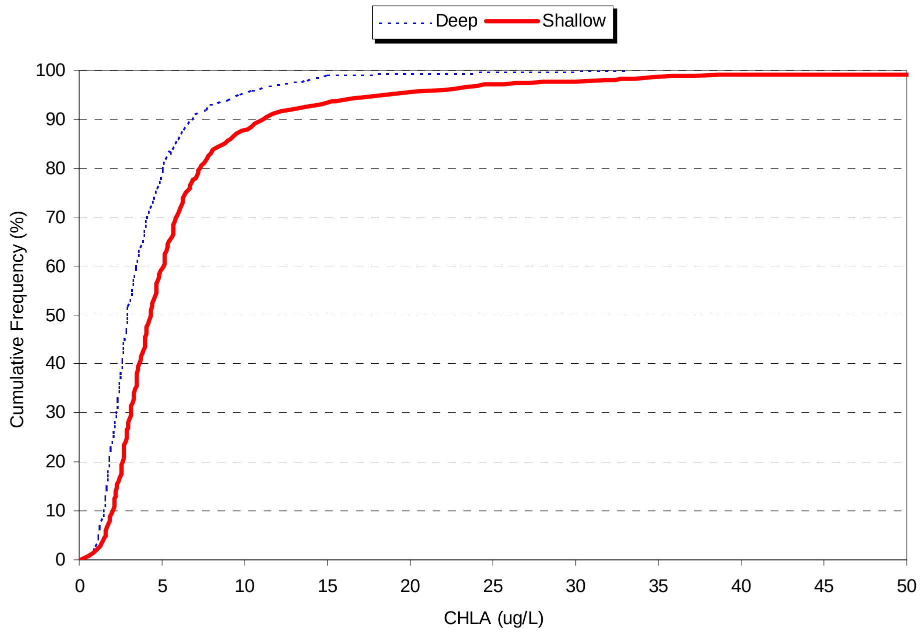


Figure 6-12 Cumulative frequency distribution of Chlorophyll-a measurements in deep and shallow lakes of New England.

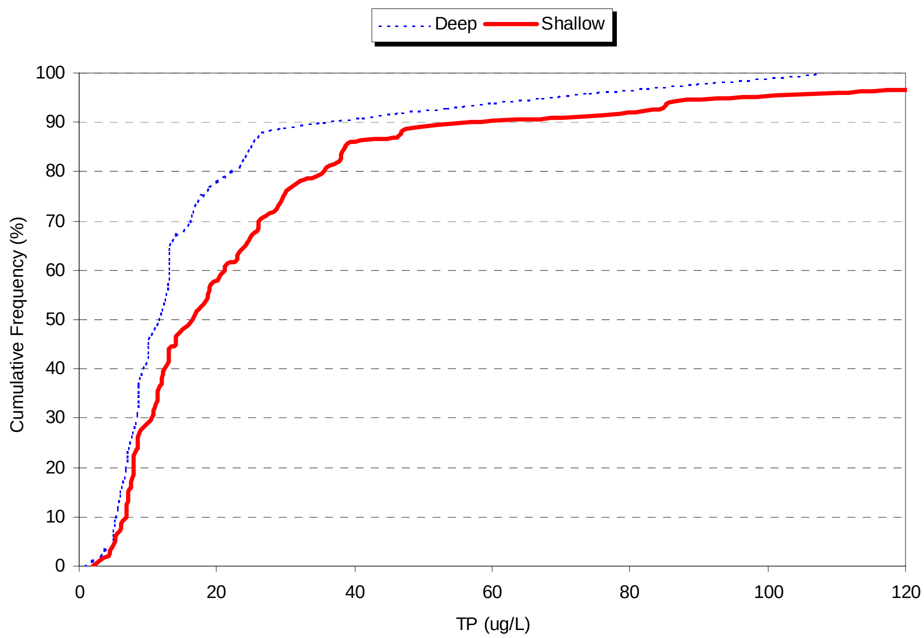


Figure 6-13 Cumulative frequency distribution of Total Phosphorus measurements in deep and shallow lakes of the New England Coastal Zone ecoregion.

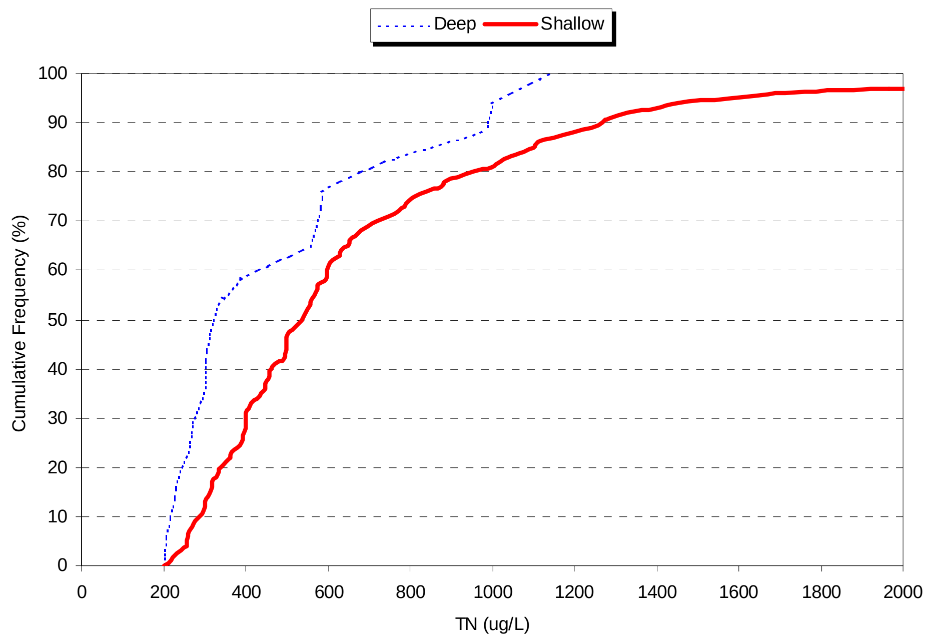


Figure 6-14 Cumulative frequency distribution of Total Nitrogen measurements in deep and shallow lakes of the New England Coastal Zone ecoregion.

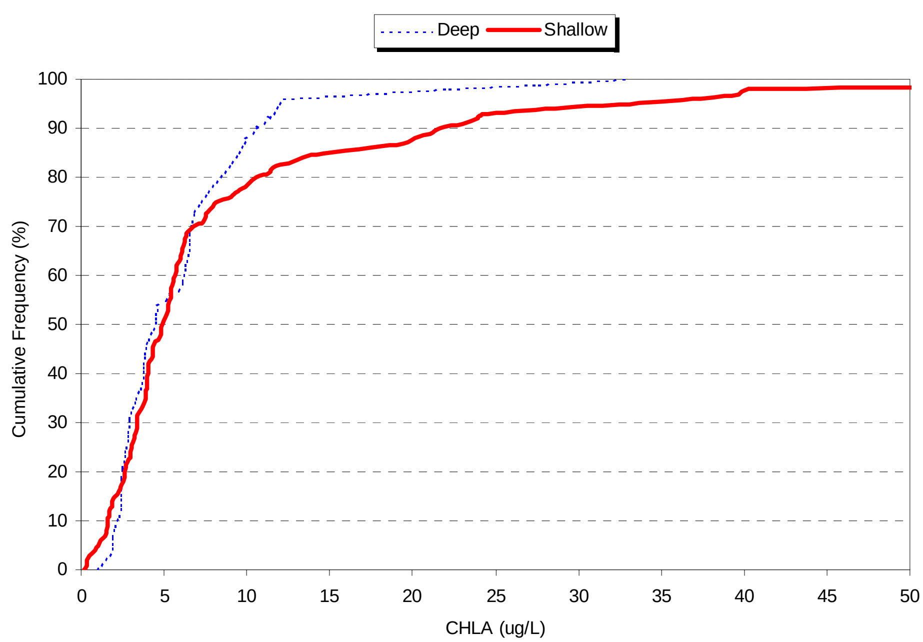


Figure 6-15 Cumulative frequency distribution of Chlorophyll-a measurements in deep and shallow lakes of the New England Coastal Zone ecoregion.

These analyses indicate that separation of ecoregional lakes/ponds into further sub-divisions such as depth should be more thoroughly explored during the development of ecoregional nutrient criteria.

6.4 Designated Uses

Another method of deriving preliminary draft nutrient criteria is through consideration of the important regulatory-protected designated uses of a waterbody. The Designated Use Method does not rely on the statistical distribution of trophic parameters for reference and general waterbody populations. For this method, the scientific literature and BPJ are used to identify potential nutrient concentrations associated with overall water quality shifts that may result in a loss or impairment of a particular function. It should be noted that the threshold values used are those typically associated with shifts in trophic state and are not directly linked with impairment of designated uses. Potential impacts to designated uses were inferred from trophic state shifts. The Lake and Reservoir *Technical Guidance Manual* (U.S. EPA, 1999a) provides a range of hypothetical designated uses ranging from those requiring high water quality (drinking water supply, outstanding resource waters) to moderate water quality (contact recreation, aquatic life protection) to low water quality (boating, flood control, irrigation). Along with these uses, the *Technical Guidance Manual* proposes a set of TP criteria, expressed as TSI (TP) values (see Table 7-2 in *Technical Guidance Manual*) as estimates of shifts of lake attributes or conditions.

These criteria were adapted for use in New England by using the range of designated uses described in waterbodies by the state experts for assessment reference and impacted conditions (Section 6.2). Identified designated uses for water classes are varied among the New England states but generally include pubic water supply, protection and propagation of aquatic life, contact recreation (i.e., swimming), non-contact recreation (boating), and Irrigation and other agricultural uses. Most lakes/ponds in New England are Class A (all uses including drinking water) or Class B (all uses except drinking water) or their equivalents.

To evaluate the potential ability of New England lakes and ponds to fully support their designated uses, a set of estimated TSI (TP) criteria were selected. The following criteria and associated TP water concentration were selected:

- TSI (TP) value of < 30 (equivalent to < 6 ug/l TP) – waterbodies less than this value are expected to be oligotrophic and were considered to be highly likely to support all designated uses, including public drinking supply;
- TSI (TP) value from 30 to < 50 (equivalent to 6 - 24 ug/l TP) – waterbodies in this category would be considered mesotrophic and highly likely to support protection and propagation of aquatic life (Note: drinking water supply may or may not be present), contact recreation (i.e., swimming), non-contact recreation (boating); and other non-water quality dependent uses (irrigation, flood control);

- TSI (TP) value from 50 to < 70 (equivalent to 24 - 96 ug/l TP) – waterbodies in this category would be considered eutrophic and likely to support protection and propagation of warmwater fisheries only, non-contact recreation (boating); and other uses. Lakes in this category would be considered potentially impacted; and
- TSI (TP) value from ≥ 70 (equivalent to > 96 ug/l TP) – waterbodies in this category would be considered highly eutrophic and likely to support only pollution-tolerant fisheries, non-contact recreation (boating); and other uses. Lakes in this category would be considered heavily impacted.

Based on potential impairment of designated uses, lakes/ponds could be judged acceptable or unacceptable. For example, it could be proposed that TSI (TP) values of <50 would indicate acceptable conditions, values of 50 to <70 may be presumed as potentially unacceptable - with a lake-specific investigation required for an exact determination, and values of ≥70 considered unacceptable (unless waterbody is restricted to non-contact recreation uses).

This scenario is displayed in Figure 6-16 using all lakes/ponds in the New England Database. It can be seen that most of the lakes/ponds fall under the acceptable criterion of TSI (TP) = 50; only about 12% exceeded the criteria and were potentially "unacceptable". For comparison, the TSI (TP) equivalent to the approximate preliminary draft nutrient criteria of 10 ug/l TP (approximate mean value of LPH, NEH, and NECZ midpoints, as determined by percentiles; Table 6-6) is indicated. It can be seen that this value is well below that needed to support designated uses and that many lakes exceed this value.

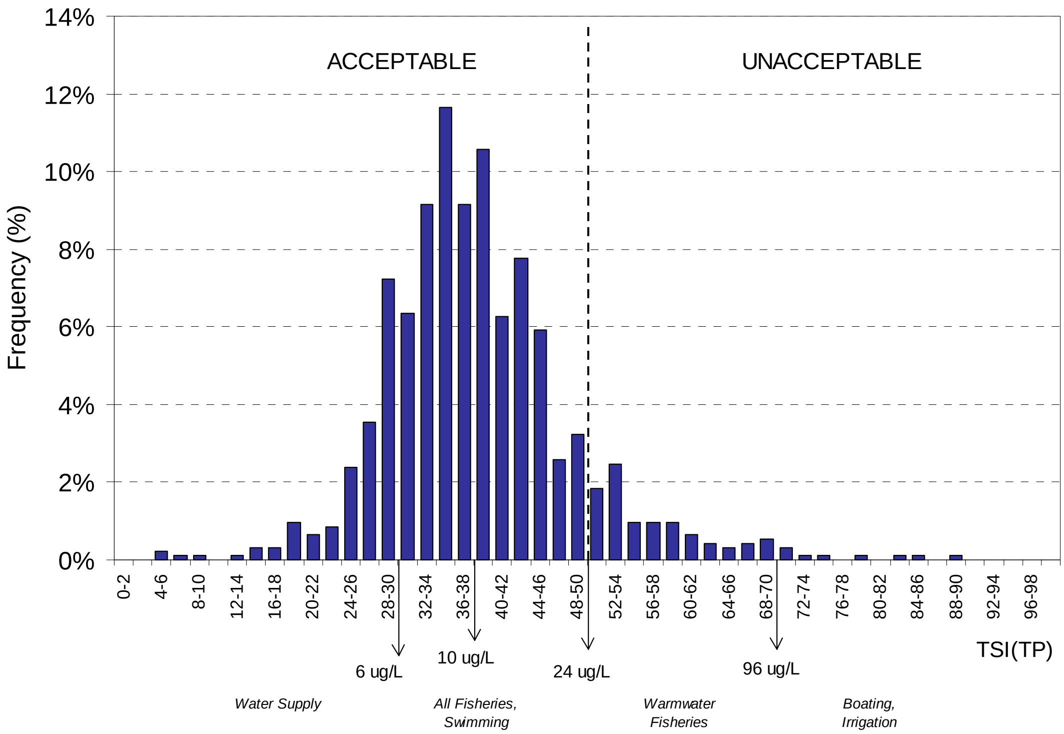


Figure 6-16 Distribution of Carlson TSI(TP) for lakes and reservoirs in New England.

This analysis was carried further using the three ecoregions (Figure 6-17 to 6-19), using the same framework, but simply inserting the percentile-derived ecoregional draft criteria in place of the general New England value (as the midpoint of TP values from Table 6-7). In all cases, the majority of lakes in the three ecoregions meet the acceptable TP criterion, while the percentile-derived criteria are conservative with regard to protection of the most water-quality dependent designated uses.

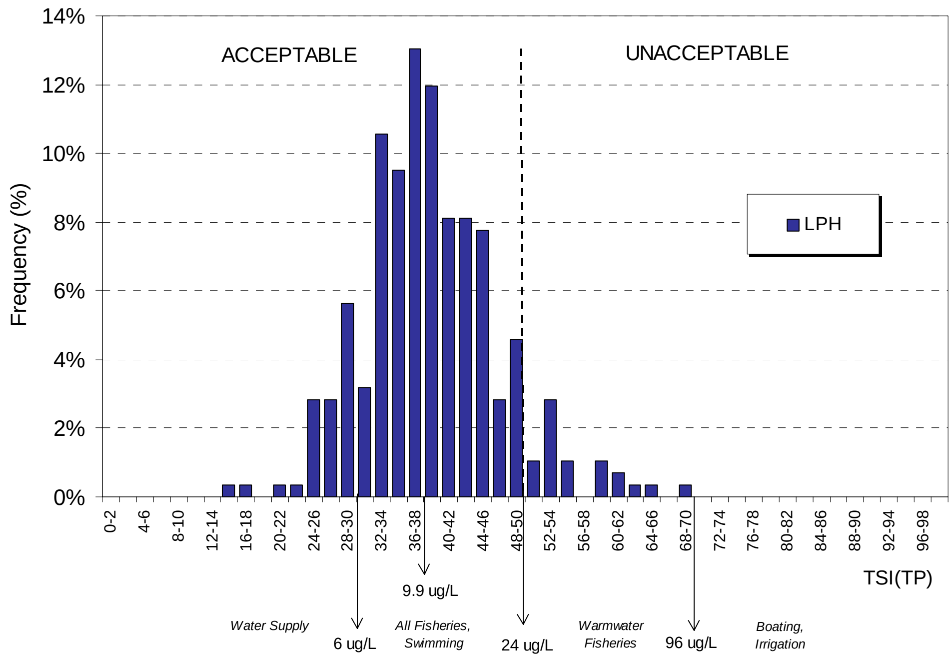


Figure 6-17 Distribution of Carlson TSI(TP) for lakes and reservoirs in the Laurentian Plains and Hills ecoregion.

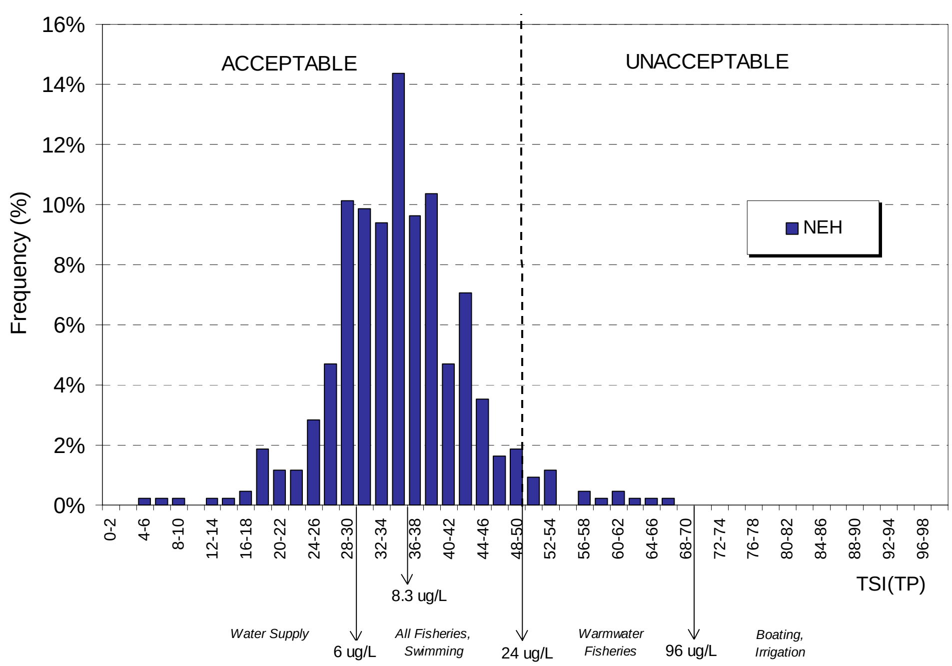


Figure 6-18 Distribution of Carlson TSI(TP) for lakes and reservoirs in New England Highlands ecoregion

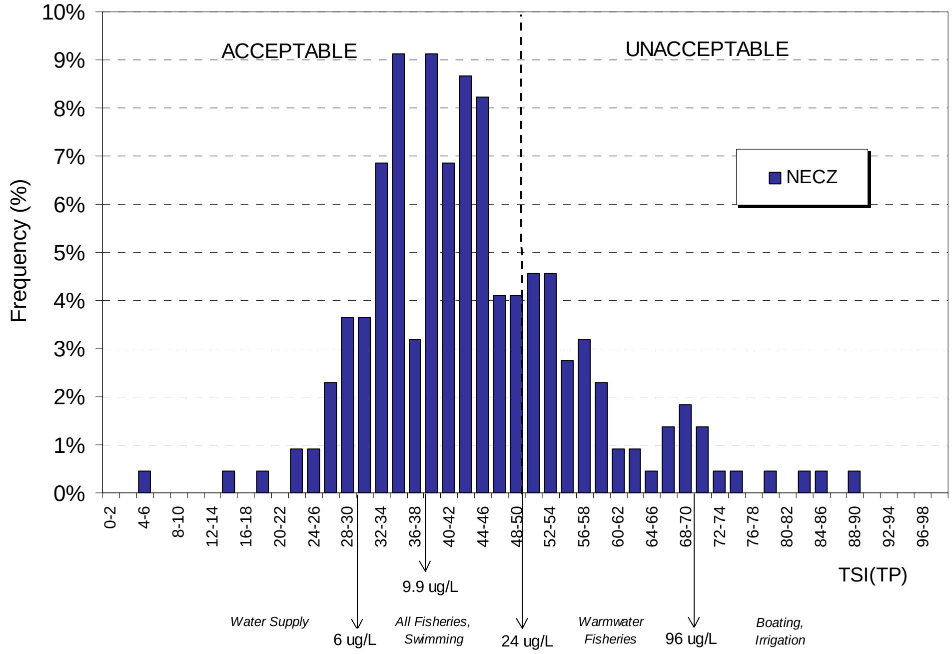


Figure 6-19 Distribution of Carlson TSI(TP) for lakes and reservoirs in New England Coastal Zone ecoregion

Finally, the distribution of TSI (TP) values for all New England reference and impacted lakes is shown in Figure 6-20. This figure indicates the less than substantial difference in water quality between these two sets of waterbodies. The reference lake population is virtually all within “acceptable” conditions, while the impacted lakes contribute a much greater number of “unacceptable” waterbodies. Yet, it can be seen that many impacted lakes have water quality conditions similar to those found in the reference lakes. This overlap probably indicates the potential influence of macrophytes in determining lake status, but may also be indicative of some amount of natural variation seen within regional lake datasets.

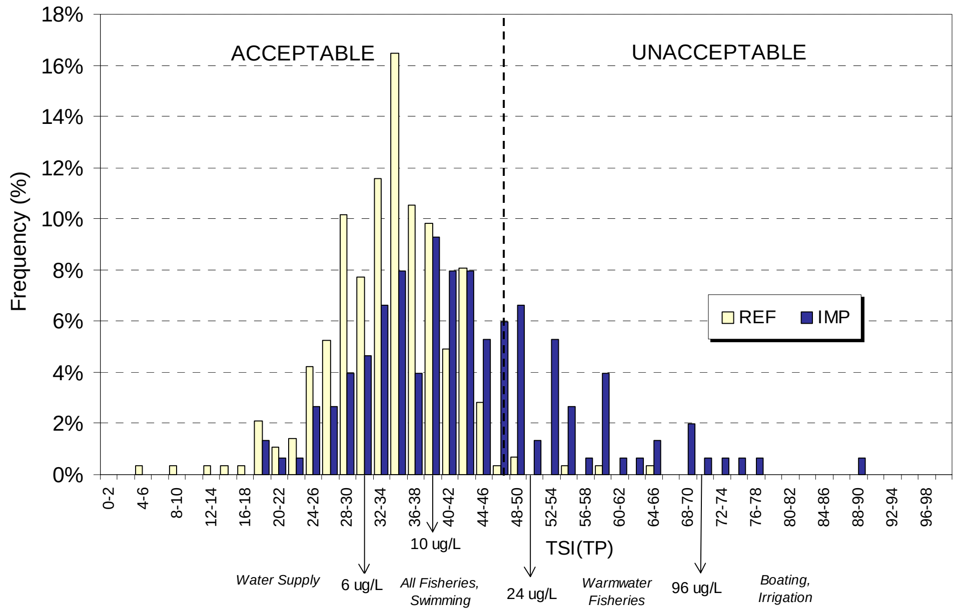


Figure 6-20 Comparison of the distribution of Carlson TSI(TP) for Reference and Impacted waterbodies in New England.

6.5 Summary of Draft Preliminary Nutrient Criteria Development

Draft preliminary nutrient criteria were developed for L/P/R waterbodies in three New England ecoregions using two approaches. The Statistical Method used the statistical distribution and identified percentiles of reference waterbodies (75^{th}) and all waterbodies (25^{th}) within an ecoregion and took the midpoint between as a draft criterion. For all New England waterbodies within the three ecoregions of interest, the nutrient criterion for TP derived by this method was approximately 9.3 ug/l, while that for TN was 311 ug/l. A second approach, the Designated Use Method, applied the statistical distributions of ecoregion L/P/R/ waterbodies to well-accepted literature values for trophic states along with inferred effects on protected designated uses. This approach does not directly determine a nutrient criterion, but did indicate that the criterion developed by the Statistical Method was a conservative estimate of the nutrient concentration required to protect most designated uses. Further work is needed on the linkage between nutrient concentrations and effects on designated uses on New England L/P/R waterbodies to produce a true designated use derived nutrient criterion.

7 Outstanding Issues

During the course of the construction of the Nutrient Database and the development of preliminary draft nutrient criteria, numerous issues were identified regarding the procedures, protocols or assumptions used. Many of these issues were successfully addressed during the development process, in discussion with the Regional NAT in meetings, or through communication with individual NAT members or state experts. In some cases, issues were identified that were not fully resolved and which may need further investigation as the regional nutrient criteria are developed. In many cases, these issues concern alternative procedures or assumptions that reflect different approaches used by New England states to collect or analyze data.

These issues may be broadly categorized into two areas of concern. The first area deals with concerns regarding the selection of data (location, season, depth, and type) to be included in the nutrient database (Section 7.1). The second area of concern is the nature of the draft preliminary nutrient criteria and their potential application (Section 7.2).

7.1 Selection of Data to be Included in the Nutrient Database

Several issues were identified with the selection of waterbodies, nutrient data, and other parameters that were incorporated into the New England Nutrient Database. These issues are discussed further below.

7.1.1 Issues associated with the selection of Lakes and Ponds

One of the issues raised by reviewers of the draft document is that the nutrient data were not collected in an unbiased manner (i.e., by a statistical random selection) and how this could lead to potential bias in the database and resulting draft nutrient criteria. The concern is that the nutrient database relied on existing data that were largely collected by state agencies that had different reasons for selecting lakes to sample and used different levels of effort to collect the data. In addition, due to availability of nutrient data and supporting metadata, the final nutrient database contains more lakes in Maine, Vermont, and New Hampshire while it appears that lakes and ponds in Massachusetts, Connecticut, and Rhode Island may be underrepresented. It has been suggested that this apparent biased selection of lakes from the north could lead to a bias in the distribution of data in the ecoregions.

Some reviewers commented on the potential inherent difference between the typical northern New England and southern New England lakes and ponds with regard to basin origin. For example, many of the lakes in northern New England are probably deep lakes of glacial origin, while many lakes in southern New England are simply small streams which were dammed to form shallow impoundments to provide hydropower.

On a nationwide scale, there appears to be a north to south trend of increasing eutrophication. The results of the “Great American Secchi Dip-In” survey of lakes (reported in Lakeline 17(2): 33) showed northern states like Maine, Vermont, and New Hampshire reporting SDT values > 4.5 meters, while Massachusetts was about 2.2, with SDT depths declining further south into New Jersey, Georgia and Florida. This suggests there may be a trend in the data to more eutrophic conditions along a north-south gradient even within an ecoregion.

These questions regarding the source and amount of data have been considered several times during the database development. While the authors acknowledge that the observations above are probably valid, it should be recalled that the nutrient data collection has never attempted to conduct a comprehensive collection of data on the entire spectrum of New England lakes and ponds. As noted earlier in the Technical Memorandum (ENSR, 1999), it was recognized at a very early stage that there were several limitations to development of the database due to:

- an uneven amount of nutrient and trophic-state related data available between the six New England States;
- the variable measured parameters in the databases provided by States, Tribes, federal agencies, and the academic community;
- the heterogeneous quality of the data, in terms of sampling effort, amount of supporting *metadata*, analytical precision, and analytical accuracy; and
- the uneven regional coverage of waterbodies, with a likely overrepresentation of smaller lakes and those with recognized water quality problems.

As noted in the Technical Memorandum, the primary technical focus became the development of a nutrient database that was sufficient to support preliminary development of draft criteria. ENSR used target goals as a means to identify the appropriate number of waterbody categories to establish reference and population sets of waterbodies to be acquired. These target numbers were used as the basis of evaluating data sufficiency and data gaps.

Some consideration was given to increasing the number of lakes and ponds selected for the New England Coastal Zone during the Data Gaps analysis (see Section 4.5). The incorporation of additional hard copy data increased the representation of Southern New England lakes and ponds in the Database to some extent.

However, we agree that the final Nutrient Database is subject to some degree of bias due to the factors identified above. For example, identification of reference lakes and ponds indicates a higher proportion of non-impacted waterbodies in the two predominantly northern ecoregions (i.e., LPH, NEU) as compared to the NECZ. This may reflect the reduced watershed populations and anthropogenic inputs found in more rural areas or basic differences in state lake programs and/or available electronic data files (e.g., emphasis on monitoring vs. studying "problem" lakes).

In short, it can be conceded that the final Nutrient Database does not provide a comprehensive, non-biased selection of lakes and ponds from New England. At the same time, the Nutrient Database contains a "best-available" compilation of waterbodies and associated nutrient data on which to base preliminary decision-making for draft nutrient criteria. However, it is appropriate that the final Nutrient Database be further explored to identify potential bias to allow more useful interpretations of proposed application of the data in preliminary draft nutrient criteria. Such investigations would also be useful in directing future data collection efforts.

7.1.2 Issues associated with use of Summer Index Period

One of the fundamental assumptions used in development of the Nutrient Database was the restriction of data to that gathered during the period July to September. This period is consistent with the timing of seasonal biological responses to eutrophication (i.e., nuisance algal blooms) and with the sampling of most lake water quality monitoring (especially volunteer monitoring groups) used to generate the Nutrient Database. However, this period may not correspond with the timing of maximum nutrient concentrations, typically associated with the vernal (spring) overturn in most lakes/ponds. Spring phosphorus concentrations are available in some state databases (e.g. VT conducts systematic, statewide lake surveys with spring phosphorus sampling), but these data were not considered as they fall out of the Index Period. This has led to a loss of potentially valuable information in some cases. For example, VTDEC had spring phosphorus data on approximately 60 candidate reference lakes, but use of the Index Period to screen data led to the exclusion of all but three of these reference lakes. While the summer is typically the period of greatest biological response, phosphorus concentrations in the summer water column are not always well correlated to response levels (i.e., chl *a*) due to heterogeneity and distribution of algal blooms over time, depth, and space and/or uptake by rooted macrophytes. Further, it has also been suggested that examination of nutrient ratios (i.e., N: P ratios) should also be considered when evaluating potential response to phosphorus input. This increases the uncertainty in identifying a TP level associated with a predictable biological response

The central concern regarding this issue is the time lag between the supply of nutrients (spring) and measurement of the biological response (summer) they produce. A summer index period provides a consistent, if somewhat arbitrary, approach for inclusion of data into the Database that would be lost if each state provided values obtained under different monitoring approaches. It may also be argued that more urban lakes receiving stormwater may be receiving irregular pulses of nutrients that may be less seasonally predictable. Calculation of summer phosphorus concentrations from spring phosphorus may be feasible (e.g, Rohm, 1995). Alternatively, expanding the Index Period to earlier in the year (e.g. before stratification) would incorporate spring nutrient concentrations. Each of these alternatives has tradeoffs with regard to the final data included in the Nutrient Database. The issue of seasonal-restricted observations should be further discussed and alternatives more fully explored before developing final reference conditions and regional nutrient criteria.

7.1.3 Depth-Integrated Sampling

The issue of depth-integrated sampling is similar to the concerns discussed above for the Seasonal Index Period. Some agencies use depth-integrated sampling (i.e., sampling of the entire water column in epilimnion and/or photic zone as opposed to sampling at discrete depths). This sampling method can provide a more comprehensive and accurate evaluation of nutrients and production in the upper waters than a restricted number of depth-discrete samples. In addition, this method will potentially sample water from depths greater than 5 m. Restriction of samples in the Nutrient Database to the $\leq$ 5 m zone may reduce information about the lower portions of the photic zone and bias the Database with regard to chl a concentrations (e.g., surface bloom concentrations more represented).

As with the spring phosphorus issue, the differences between sampling efforts by various state agencies result in the need for a common approach for data included in the Nutrient Database. Hence, the assumption of samples collected in the 0 to 5 m range was used. However, relaxation of this criterion to expand the Database by inclusion of epilimnetic samples and/or photic zone samples could be considered. Determination of the depth of the epilimnion from temperature profiles can be made from lake and date-specific records, but such determinations would be very labor-intensive and should be conducted only if initial analyses indicate that such an effort is justified. Alternatively, the use of the reported SDT depth to estimate an approximate photic zone limit on a particular sampling date may be considered. Since the majority of the lakes/ponds data are typically surface or surface and bottom samples, it is not clear that additional analyses would be merited.

7.1.4 Macrophytes

The reporting of and relative importance given to data regarding the distribution and abundance of macrophytes differs between the states. Typically, macrophyte distributions and abundance are more heavily weighed by CT, MA, and RI state agencies in their determination of trophic state and impacted status. VT provides information regarding the identification of macrophyte species, while ME and NH do not formally consider macrophyte growth in state classification, preferring to focus on water column nutrient concentrations for trophic classification. This division between states' approaches appears related to the shallower depth found in NECZ waterbodies relative to the other ecoregions and/or greater abundance of organic substrate available for colonization. Modification of trophic status by macrophyte abundance seems warranted, but does not easily fit into a conventional nutrient - phytoplankton prospective, because macrophyte abundance is less easily linked to water column nutrient conditions (i.e., in some cases nutrient-poor lakes can sponsor extremely luxuriant macrophyte growth). Equally important, there is no general consensus regarding the levels of macrophyte coverage (>50%, >75%?) or abundance associated with impairment of designated uses. These issues need to be further discussed and a more satisfactory method proposed to integrate macrophyte community factors into the development of regional nutrient criteria.

7.2 Nature of Regional Nutrient Criteria

The nature of the preliminary draft nutrient criteria generated in Section 6.0 will need to be more fully considered. Use of a percentile approach provides a feasible means of establishing a numeric criteria but has many implications that need to be further considered. By definition, establishing a set percentile as the target concentration for criteria means that a certain percentage of waterbodies would be out of compliance automatically. While it can be assumed that at least a portion of the waterbodies are seriously impacted by elevated nutrients, many waterbodies may be considered in violation, even though many of these lakes may fully support their designated uses (e.g., waterbodies with high nutrients that do not exhibit signs of eutrophication). [Note: it has been suggested that as a "reality check" any proposed nutrient criterion be evaluated as to, if it were applied, how many reference lakes would be targeted for management and how many impacted lakes would fail to be identified].

Some reviewers questioned the utility of a single criterion as a lake management tool and suggested that a range of values be developed. As noted earlier, many of the designated uses contained in water quality standards are protected by a wide range of nutrient conditions. This is supported by the distribution of lake TP concentration relative to those expected to lead to loss of impairment of designated use function (Figures 6-16 to 6-19). Moreover, there is an underlying assumption that an overall reduction in nutrients and shifting of waterbodies to more oligotrophic conditions is desirable in all cases. While this is generally true for most impacted lakes, it does not consider the need or utility for a range of differing lake trophic states to provide a wider range of recreational and ecological function. It can also be seen that low to moderate nutrient conditions is no guarantee that a lake will not be considered impacted (Figure 6-20).

In addition, it is well known that there are regional differences in opinion about what constitutes acceptable water quality where eutrophication is concerned. Unlike other pollutants such as toxic heavy metals (which are considered harmful even in trace amounts), nutrients are natural and, in many cases, desirable, depending on the designated use of the waterbody. However, the designated use and the public perception of acceptable or not-acceptable water quality condition changes from region to region (Heiskary and Walker, 1988). For example, a lake with 24 ppb phosphorus and a SDT of 2 meters may be considered a good bass lake in southern New England, while such a lake in Maine could be considered heavily impacted and potentially unacceptable in a cold water fisheries region. The regional ecosystems do not fully account for the temperature, fisheries and local public opinion changes that influence the designated uses within an ecoregion. Thus, even though the EPA recognized regional differences rather than setting national criteria, even the regional criteria may not be scaled to a fine enough degree to account for these "subregional" differences.

While this discussion is by no means intended as a full exploration of the implications of decision-making regarding development of regional nutrient criteria, it does indicate the need for considerable dialogue and consensus among the state and federal agencies responsible for deriving the regional nutrient criteria.

8 Summary

Water quality data, lake characteristics, and watershed information were collected from over 7,000 waterbodies in New England as part of the "Collection and Evaluation of Ambient Nutrient Data" Project (Section 1.0). The primary source of information was from electronic data files obtained from various state and federal agencies, Tribal nations, and academic institutions (Section 2.0). A relational database was designed and implemented in Microsoft Access97® to accumulate and manipulate the extensive amount of available electronic data (Section 3.0).

Review of the initial data collection for four key trophic parameters (TP, TN, chl *a*, SDT) indicated that these were not available for a majority of the waterbodies (Section 4.0). Subsequently, the data were refined to focus on waterbodies with good trophic parameter data and further supplemented with additional hardcopy data to address data gaps and produce the final New England Nutrient Database. The final Nutrient Database, with 1,155 lakes, ponds and reservoirs represented, was used to develop preliminary draft regional nutrient criteria for lakes/ponds in the three EPA Level 3 non-aggregated ecoregions (LPH, NEH, NECZ). Further evaluation of the river and stream data collected from the data sources noted above was deferred to a future phase of this project.

The Nutrient Database was analyzed and its general characteristics described (Section 5.0). The distribution of the ecoregion-specific trophic parameters was tested and statistical differences detected between many of the trophic parameters. The trophic status (oligotrophic, mesotrophic, eutrophic) of lakes within the ecoregions were determined and compared to previous work with generally good agreement.

Two methods (Statistical Method, Designated Use Method) were used to develop preliminary draft regional nutrient criteria (Section 6.0). Region Nutrient Assessment Team members and state experts identified reference and impacted waterbodies. Using the recommended percentile for reference (lower quartile) and all lakes (upper quartile), a range of possible nutrient criteria was generated for the three ecoregions. The composite midpoint of the range for the three ecoregions was approximately 10 ug/l TP. A second set of nutrient criteria were generated using literature values and recommended TSI(TP) criteria from the Lakes and Reservoirs *Technical Guidance Manual* (U.S. EPA, 1999). Here, the emphasis was prevention of loss of designated use and a value of 24 ug/l TP was selected to distinguish between acceptable and non-acceptable conditions. Comparison of the results generated by the two methods suggests that the Statistical Method would be overly conservative as a means of protecting key designated uses. Outstanding issues were identified that must be further discussed and consensus reached during the development of a regional approach to nutrient criteria (Section 7.0).

9 References

Canavan, R.W., and P.A. Silver. 1994. Chemical and Physical Properties of Connecticut Lakes With Emphasis on Regional Chemistry. Lake and Reservoir Management, Vol. 10: 175-188.

Connecticut Department of Environmental Protection. 1991. Trophic Classifications of 49 Connecticut Lakes. Bureau of Waste Management. Hartford, CT.

Connecticut Department of Environmental Protection. 1998. Trophic Classifications of 12 Connecticut Lakes. Bureau of Waste Management. Hartford, CT.

ENSR Consulting and Engineering, 1999a. *Technical Memorandum for Developing Target Goal for Waterbodies*, Acton, MA.

ENSR Consulting and Engineering, 1999b. *Data Distribution Report*, Acton, MA

Gibson, G. 1999. *Draft Nutrient Data Sequencing*, Memorandum, U.S. EPA, Washington, D.C.

Heiskary, S.A. and W.W. Walker, Jr. 1988. *Development Phosphorus Criteria for Minnesota Lakes.* Lake and Reservoir Management 4:1-9.

Heiskary, S.A., Wilson, C.B., and D.P. Larsen. 1987. *Analysis of Regional Patterns in Lake Water Quality: Using Ecoregions for Lake Management in Minnesota*. Lake and Reservoir Management 3:337-344.

Lakeline. 1995. Great American Dip In. Vol. 17(2) 33. NALMS.

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.

Mattson, M., Godfrey P.J., Walk, M.I., Kerr, P.A., and O.T. Zajicek. 1992. Regional Chemistry of Lakes in Massachusetts. Water Resources Bull. 28: 1045-1056.

Omernick, J.M. 1987. Ecoregions of the conterminous United States. Map (scale 1:7,500,000). Annals of the Association of American Geographers. 77(1):116-125.

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

United States Environmental Protection Agency. 1998a. *Lake and Reservoir Bioassessment and Biocriteria. Technical Guidance Manual*. Office of Water. Washington, D.C. EPA Document EPA 841-B-98-007.

United States Environmental Protection Agency. 1999a. *Nutrient Criteria Technical Guidance Manual. Lake and Reservoirs. Technical Guidance Manual*. *Review Draft*. Office of Water. Office of Science and Technology. Washington, D.C.

United States Environmental Protection Agency. 1999b. *Nutrient Criteria Technical Guidance Manual. Rivers and Streams. Technical Guidance Manual*. *Review Draft*. Office of Water. Office of Science and Technology. Washington, D.C.

United States Environmental Protection Agency. 1999c. Personal communication between Jim Fritz (US EPA, Region 1) and Isabelle Morin (ENSR) regarding electronic land use files.

APPENDIX A

ENSR-NEIWPCC STATISTICAL ANALYSIS RESULTS

TECHNOLOGY PLANNING AND MANAGEMENT CORPORATION (TPMC)

To:	David Mitchell
Cc:	Jeff Rosen, Isabelle Morin; Jose Sobrinho
From:	Kristyn Stevens
Date:	August 31, 2001
Subject:	***ENSR – NEIWPCC Statistical Analysis Results***

INTRODUCTION/BACKGROUND

This memorandum summarizes the results and conclusions by TPMC in an effort to support ENSR in providing a technical basis and support for the United States Environmental Protection Agency, New England ("EPA") development of nutrient criteria for lakes, ponds, and reservoirs ("L/P/R") in New England. The statistical analyses described by this memo indicate support for development of sub-classifications of L/P/R that may justify development of categorized nutrient criteria, dependent on the physical and chemical characteristics.

Ecoregions were used as a basis for some data segregations prior to analysis. Three of the four major New England Non-aggregated Level 3 ecoregions were included in these analyses, including: North Eastern Coastal Zone (NECZ); Laurentian Plains and Hills (LPH); and North Eastern Highland (NEH). Variables included in the analyses were morphometric, chemical, and land use categories. Morphometric variables included mean depth, maximum depth, area, Osgood Index (defined as mean depth / square root of the surface area), and latitude (coordinates of estimated lake center). Land use categories included residential, commercial, forested, agricultural, recreational, shrubland, wetland, and barren. Chemical variables included total nitrogen (TN), total phosphorus (TP), secchi disk transparency (SDT), color, chlorophyll *a* (chl a), alkalinity, and pH.

All statistics were performed on the New England Nutrient Database provided by ENSR. This database consisted of the geometric means of trophic parameters for over 1000 waterbodies during the summer index period. The database was produced by averaging duplicate samples taken at a particular station, date, time and depth. Then, the geometric mean was calculated for each waterbody based on the "averaged" samples collected during the summer index period, in the upper 5 meters of the water column. This resulted in a unique value for each waterbody.

ENSR proposed the following two questions which characterized the statistical investigation:

1. *Using reference lake data only, what variables create the greatest "separation" (i.e., clustering in groups along morphological or water quality-based axes) for the trophic state indicators (TP, SDT, chl a TN) ?* As *a priori* assumptions, we expected mean lake depth and alkalinity or color within an ecoregion to be important.

2. *What factors allow the greatest "separation" of reference and impacted lakes in terms of trophic variables ?*

Based on discussion between TPMC, ENSR, and EPA on September 28th, 2000, it was decided to break this analysis into a phased approach, due to budgetary and time constraints. The first recommended step was to identify whether or not there were distinct categories of reference lakes. If distinct categories could be discerned then each category must be considered separately since the between group variability may possibly mask detectable differences between reference and impaired lakes.

METHODS AND RESULTS

Cluster analyses, correlations, and prinicpal component analysis (PCA) were selected as the most appropriate methods to properly classify impaired, general, and reference conditions in each ecoregion. Since these suggested analyses are predicated on a normal distribution, exploratory analyses and tests for normality were conducted to test this assumption. These methods are described below.

Exploratory Analyses

Descriptive Statistics

Three waterbody assessment classes were identified from the New England Nutrient Database. These were reference waterbodies (identified by States through watershed analyses or best professional judgment (BPJ)), impaired (identified by presence on Clean Water Act Section 303(d) lists or state-specific evaluations), and test lakes. Test lakes refer to waterbodies in the New England Nutrient Database that were not identified as reference nor listed on the State's 303(d) lists. Waterbodies identified as "Test2" (subset of Maine lakes with additional comments), and "None" (no comments from States) were put in Test category for statistical analyses.

Preliminary and exploratory analyses of the data were performed for the overall dataset including all assessment types, by assessment type (reference, test, and impaired) and by a combination of EPA ecoregion and assessment type. Sample size, mean, standard deviation, minimum and maximum values were calculated for all numeric variables surveyed in this study. These results are presented in Tables 1 through 5.

Table 1 provides a summary of selected parameters of the New England Nutrient Database used as the master data source for all subsequent statistical analyses. Table 2 provides summary statistics of the entire dataset broken down into the three assessment categories (reference, impaired, test). Tables 3 through5 provide statistical summaries of each assessment category, broken down by ecoregion (i.e., NECZ, LPH, and NEH).

TPMC noted that the majority of the lakes included in the reference assessment class were from Maine. This is likely to introduce some bias in the dataset, both geographically and from a north/south gradient perspective. It may also introduce bias from any specific differences between Maine and the other states in hydrology.

Testing for Normality of Variables and Optimal Data Transformation Analyses

Parametric statistical analyses are often more powerful than non-parametric analyses. However, this is not true when the assumptions of parametric statistical analyses are violated. One of the most critical assumptions that characterize parametric analyses is that a sample taken from a population should be normally distributed. In the case that this assumption fails, data transformations can often result in data that do meet this assumption. Data transformation analyses typically achieve or improve normality, as well as simplify the structure of the model and stabilize the variance.

Tests were conducted to determine if the variables were normally distributed. The Shapiro-Wilk test was used to test the normality of samples with fewer than 2000 observations and the Kolmogorov test for larger sample sizes. None of the numeric variables examined had normal distributions at the 95% confidence level ($p < .05$). Since the majority of data from this study showed a highly skewed and non-normal distribution with p-values of less than 0.01 for most variables, data transformation (power analyses from Box and Cox (1964) analyses were employed to determine what transformation would result in approximate normal distributions.

Two transformations of the morphological and chemical data were attempted - the square root and the log10 transformation. Few of the variables examined had normal or near normal distributions at the 95% confidence levels after the recommended data transformations were performed. These variables included lake area, Osgood Index, color, and chlorophyll *a* for the log 10 transformation and SDT for the square root transformation.

Transformation analyses were also performed on the proportional data (e.g., land use) using the (variable) +1 to account for zeros in the data set. Since the Arc sine inverse of the data is used for proportional data, this transformation was also applied (Krebs, 1989). All measurements were $p<0.01$ which meant that the transformed values were still not normal, most likely due to zeros. None of the transformations helped to reach normality, therefore non-parametric tests were used.

Verification of Classifications through Wilcoxon Rank Sum Tests

The next set of analyses was performed to evaluate whether the classifications made by the State experts in this study defined a statistically distinguishable waterbody population. Table 6 presents the results of non-parametric statistical tests between the impaired and the reference data for each variable measured for both the impaired and reference waterbodies. Although t-tests are somewhat tolerant of a lack of normality, because of the data's extensive departure from normality, non-parametric tests (e.g., Wilcoxon Rank Sum Test) were used instead.

In general, the tests indicated that there were significant statistical differences between most parameters for the impaired lakes versus the reference lakes. The exceptions to this were lake area, Osgood index and color. All other parameters showed statistically significant difference between the impaired and the reference lakes. This suggests that the State experts identified a statistically distinct set of lakes and that there are, in fact, differences between the impaired and the reference lakes, which can be measured statistically with a sample size on the order of a few hundred lakes.

Note: Cluster analyses were also originally proposed to determine whether the data support the impaired versus reference categories over the entire data set. This would present a quantitative perspective on whether or not the a priori classifications are reasonable. Since the Wilcoxon Rank Sum tests indicated that the classifications done by experts are supported by the data, this step was not necessary.

Based on these statistical differences between impaired and reference lakes, the parameters for the reference lakes were summarized by ecoregions. Table 7 shows the results of non-parametric Wilcoxon Rank Sum tests for each variable from the reference lakes when a pair-wise comparison is done between ecoregions. This table indicates that from the perspective of morphometrics (lake area and max depth), ecoregion NECZ is significantly different from both LPH and NEH. However, it is important to note (as shown in Table 3) that the sample sizes for NECZ are much smaller than they are for LPH, which in general is slightly smaller than NEH.

The differences in the sample sizes may account to some degree for some of the consistent differences that we have observed between NECZ, LPH, and NEH.

When looking at the land use categories, the ecoregions are different in all of the following land use classifications: forested, recreational, and wetlands. We do see some differences in residential, shrubland, and commercial land use percentages between NECZ and LPH and NECZ and NEH, while LPH and NEH are the same. To some degree this may be a function of the smaller sample size for NECZ. In addition, NECZ is different from LPH and NEH for secchi depth, color, and chlorophyll *a*.

These results indicate that the ecoregions are sufficiently different and there are consistent, observable, and detectable differences that can be identified. This would suggest that the ecoregion designation is a pertinent one and is likely to be appropriate for use in development of water quality criteria and indicators.

Correlation analyses

To establish a set of parameters that can most effectively be used to separate classes of reference lakes from one another with minimum redundancy, correlation analyses (Pearson or Spearman) were performed to identify parameters that co-vary strongly. If two variables were thought to be correlated, when one changed the other did so in a related manner. The correlation coefficient is a number ranging between –1 (perfect negative correlation) and 1 (perfect positive correlation) that acts as a measure of association between the two variables of interest. Correlations were run on morphometric, land use, and chemistry variables. For purposes of this analysis, a correlation coefficient >0.50 was considered indicative of a strong correlative relationship. Normality tests were used to identify whether Spearman (non-parametric) or Pearson (parametric) correlation methods should be used.

Due to the high non-normality of the environmental variables, the non-parametric Spearman Rank correlations were used. In addition, for the few variables that were found to be normally distributed after transformation (e.g., lake area, Osgood index, color, chl a, and SDT), Pearson correlations were also conducted. These results concurred with the Spearman correlations. Correlation analyses for all data in the New England Nutrient Database (regardless of assessment type) are presented in Table 8. Note that all shaded values in the correlation tables represent coefficients approximately 0.5 or above.

Few highly significant correlations were observed for the overall data including pH and alkalinity, lake mean depth and lake mean depth (which would be expected). However, several environmental factors displayed correlation coefficients approximately at or above 0.50. This included positive relationships between chl *a* and TP, color and TP, lake area with both mean and maximum ("max") depth, SDT and mean and max depth, TN and TP, as well as color and TN. This also included inverse relationships between color and SDT, TN and mean depth, TP and max depth, Osgood Index and lake area, SDT and TN, as well as SDT and TP.

Correlation analyses results for the reference data only are shown in Table 9. Results for the reference data were similar to that of the overall data with few exceptions. The inverse correlation with TP and max depth was not evident in the reference data nor was the inverse relationship between both color and chl *a* with TP. In addition the relationship between lake area and mean depth was also not evident.

Correlations were also run by ecoregion for NECZ (presented in Table 10) and the LPH and NEH ecoregions combined (presented in Table 11). An inverse relationship between SDT and chl *a*, as well as color and max depth was evident in the NECZ. Interestingly, relationships between mean or max depth were not strongly correlated with lake area. In addition, Osgood Index and lake area were not strongly correlated. The relationships between TN and mean depth and TP and max depth were not as strong as they were in the overall dataset.

Results for the LPH and NEH ecoregions combined were very similar to that of the overall dataset with few exceptions. An inverse relationship between chl *a* and SDT was evident in this dataset but it was not as strong in the overall dataset. In addition, an inverse relationship was evident between TN and max depth but was not evident between TP and max depth.

In summary, few correlations were observed between parameters across the major three data categories (morphometric, land use and physiochemical). The main correlations among parameters (across the three data categories) were TN and TP correlated with mean and max depth respectively. In addition, SDT was correlated with both lake mean and max depth and color was correlated with max depth.

Cluster Analyses (Reference Data Only)

Cluster analyses were conducted to determine if clearly identifiable groups could be identified within the reference lakes. Both mean depth and surface area were removed when running a cluster analysis with Osgood Index (see definition earlier) included in the morphological variables.

Preliminary cluster analyses were run on the morphometrics for the reference lakes to see if general classes could be distinguished based on size, depth etc. The cluster analysis separated out into size classes dominated by the lake area. In order to reduce the overwhelming impact of surface area, a square root transformation was performed on the raw value of lake surface area and the cluster analysis was repeated. It was expected that this would increase the influence of the max and mean depths into consideration for the clustering. However, lake surface area continued to dominate the clusters.

Lake surface area clusters were also analyzed to see if they were related to land use, water quality and ecoregions. No apparent trends were observed. To verify that no trends exist with regard to nutrients, we also plotted TN and TP concentrations against lake surface areas and did not observe any apparent relationships (Figures 1 and 2). Based on these results, TPMC concluded that there are no apparent clusters of lakes that would justify consideration of separate criteria based on the morphometrics of the lakes.

Principal Components Analysis (PCA)

PCA reduces the number of parameters required to define and delineate water conditions by combining those parameters into a linear combination that accounts for the greatest amount of variance in the data. This typically provides a clear grouping of parameters that co-vary, resulting in regressions that best define the relationships in nature. A correlation matrix was used to identify the relevant parameters. The analysis results in equations that can be used to linearly combine multiple co-varying parameters, resulting in a new set of parameters that carry the maximum variability for the original parameters while eliminating covarying data.

PCA was conducted on the overall data as well as the overall dataset split into two geographic regions (i.e., NECZ vs. LPH + NEH), and by assessment type. Parameters included in the analysis consisted of the morphometric and chemistry variables only. [Note: all land use variables were excluded from the PCA].

Specifically PCA was conducted on seven dataset-parameter combinations:

1) All data (all parameters, excluding TN and TP)
2) All data (all parameters excluding TN, TP, mean depth and surface area)
3) NECZ (all parameters excluding TN, TP, mean depth and surface area)
4) LPH and NEH combined (all parameters excluding TN, TP, mean depth and surface area)
5) Reference waterbodies (all parameters excluding TN, TP, mean depth and surface area)
6) Test waterbodies (all parameters excluding TN, TP, mean depth and surface area)
7) Impaired waterbodies (all parameters excluding TN, TP, mean depth and surface area)

We initially began with 1,125 waterbody records. However, PCA needs a data point for each field in each record. Due to the low sample size (22 to 44 records), which resulted from using all of the data we removed TN from the analyses (since it only had a limited number of values). Once TN was removed; the sample size was approximately half the actual number of records. We also removed surface area and mean depth from the majority of the analyses since Osgood index is an integrated measure of both surface area and mean depth. Once these analyses were completed we also tried a set of analyses without TP to see if that would change any of the results. Since much is known about TN and TP in lake systems, we did not want to mask other important factors. The results for the analyses without TN and TP are provided below. See Appendix A and Tables A1-A7 for the PCA results that exclude only TN.

TPMC identified the most influential parameters with respect to discriminating/classifying the trophic indicator data. Since the goal of PCA is to identify a new set of reduced variables (principal components) to account for the variance of the dataset we also included variance information. PCA can be described using eigenvalues and eigenvectors. The eigenvalue indicates the amount of variance explained by an eigenvector (or principal component) out of the total variance. The way to determine the number of eigenvectors is to follow some general guidance or stopping rules. The two criteria that were used in the PCA were:

1. The percentage of variance – the specification that factors are to be extracted until some percentage of the total variance has been explained (our experience with environmental data is to use 85%).
2. "Kaiser's stopping rule" –extract (or retain) only eigenvector with eigenvalues of at least 1.

Since this study involves many moderately correlated variables and a sufficient sample size, a large number of factors would be retained if we use the 85% rule. This led us to reduce the percentage of variance criterion to about 70% coupled with the Kaiser stopping rule.

PCA Results

Results from the PCA have been divided by assessment type/ecoregion and also by the inclusion of parameters. Results from each of the seven analyses are presented in Tables 12 through 18, respectively. Each table includes information on the eigenvalues, the cumulative percentage of the total variance contributed by each principal component, and the individual principal components (X_1-X_n) (or eigenvectors).

<u>PCA #1 - All data</u> (all parameters, excluding TN and TP) – Table 12

The eigenvalue for a principal component indicates the percentage of the variance that the Principal component accounts for. In PCA #1. The first principal component accounts for (2.459/9) *100% = 27%, the second for (1.647/9) or 17%, the third for (1.166/9) or 13% and the fourth for (1.113) or 12%. Cumulatively, the second component accounts for approximately 45%, the third for 58% the fourth for 70% and the fifth for 81%, making the last five components account for the remainder of the variance (approximately 19%)

Principal Component 1 (PC1) was dominated by SDT and mean depth with inverse values for chl *a* and color. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by lake area with an inverse value for Osgood index.

PCA #2 - All (all parameters excluding TN, TP, mean depth and surface area) - Table 13

PC1 was dominated by chl a and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #3 - NECZ (all parameters excluding TN, TP, mean depth and surface area) – Table 14

PC1 was dominated by chl a and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with inverse values for SDT and color. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #4 - LPH and NEH combined (all parameters excluding TN, TP, mean depth and surface area) – Table 15

PC1 was dominated by chl *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #5 - Reference (all parameters excluding TN, TP, mean depth and surface area) – Table 16

PC1 was dominated by SDT with inverse values for chl *a* and color. PC2 was dominated by pH and alkalinity with an inverse value for SDT. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #6 - Test (all parameters excluding TN, TP, mean depth and surface area) – Table 17

PC1 was dominated by chl *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for color. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #7 - Impaired (all parameters excluding TN, TP, mean depth and surface area) – Table 18

PC1 was dominated by chl *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

In general, for the seven dataset-parameter combinations, PCA has given us a fairly consistent set of principal components. Principal Component 1 is dominated by chl *a* and color exhibiting positive factors while SDT has a negative factor. Principal Component 2 is dominated by pH and alkalinity with positive factors and color with a negative factor. Principal Component 3 is dominated by Osgood Index with a positive factor and pH with a negative factor. Principal Component 4 is dominated by max depth with a positive factor and Osgood index with a negative factor. The overall assessment (excluding TN and TP only) and the reference PCA display similar components, except Principal Component 1 is dominated by SDT (+ factor) and both chl *a* and color (- factors). Principal Component 2 for the reference lakes only is also dominated by alkalinity (+) and pH (+) and SDT (-).

The results of this PCA suggest that there are consistent parameters that seem to influence the characteristics of lakes and ponds. These same parameters emerge as influential in the principal component analyses regardless of how the data are aggregated and classified. This consistency suggests that these parameters contain the signals of the responses of lakes to their natural and anthropogenic influences. These parameters also probably contain the signals of other less influential parameters that are correlated with the parameters that most influence the principal components.

Finally, it should be noted that the subset of reference lake results indicate the same dominant parameters in each of the principal components as the impacted and test lakes, but with opposite signs from those analyses. This supports the contention that these parameters are fundamentally indicative of the lakes' conditions. More specifically, the factors' reverse sign indicates that when the data are aggregated, the same parameters that dominate for impacted lakes (with positive sign) are negative sign for reference (on unimpacted) lakes. This inverse relationship indicates that if parameters are fundamentally connected to the condition of impaired lakes (i.e., show high factors), we would expect those same parameters to show very low number for reference lakes.

CONCLUSIONS

On the basis of the statistical analyses conducted, the following conclusions were reached. However, it should be noted that there may be underlying uncertainty associated with these conclusions due to different sample size among ecoregions and parameters.

- The classifications of assessment type (reference versus impacted) by experts are supported by the data as seen in the descriptive statistics and the Wilcoxon Rank Sum tests.
- The NECZ ecoregion is significantly different from the LPH and the NEH ecoregions for morphometric parameters. This supports the conclusion that the ecoregions do provide statistical segregation for many important lake parameters
- Based on the results of the cluster analyses, there are no apparent clusters of lakes that would justify consideration of separate criteria based on the morphometrics of the lakes.
- PCA provides tentative identification of the four most influential principal components. Principal Components 1 and 2 are a function of water quality. Principal component 3 is a function of morphometrics with inverse relationships to water quality. Principal Component 4 is a function of morphometrics alone.
- PCA identified clear but not strong principal components. Each component cannot be easily and strongly named and described. This may be attributed to the fact that most of the variables are moderately correlated with one another in some manner which may account for the smaller percentages represented by each component.

RECOMMENDATIONS

Results from the PCA show promise that the parameters measured can be recombined into principal components that account for the majority of the variability observed in all the parameters measured. This would reduce the complexity of further exploratory analyses and might indicate that only the dominant parameters need to be measured in the future in order to properly classify lakes as impacted or not impacted. The best way to determine the efficacy of the dominant parameters would be to calculate the principal components for each observation and then use these data in a step wise discriminant analysis. This would establish if discriminant functions could be developed based on the principal components that can properly classify lakes with a high enough classification efficiency.

The approach would be as follows:

1. Subset out the assessment type = test. These records will not be included in the discriminant analysis.

2. Calculate the principal components for each record using the factors calculated in this study. Keep only the categorization parameters (assessment class, the ecoregion, the sampling date, etc.) and the principal components.

3. Partition the remaining data set into two equal subsets by randomly selecting equal numbers of reference and impaired lakes from the entire data set.

4. Use one set of the data (we'll call it the training set since it will be used to train the discriminant analysis to identify reference and impaired) for the stepwise discriminant analysis and the second set of data for the validation of the discriminant function.

5. Predefine what an acceptable rate of correct classification of lakes will be. This rate will be considered following step 8 below.

6. With the first set, run a stepwise discriminant analysis to identify the principal components that best segregate between impaired and reference. Define stopping rules similar to the ones employed for the Principal Component Analysis.

7. Once the discriminating principal components are identified, run a regular discriminant analysis. This will output a discriminant function (based on the principal components) and a decision rule indicating what values of the discriminant function classify into the reference category and which values classify into the impaired category.

8. Using the outputs from step 6 above, the second set of data (the validation set) should be used to calculate the appropriate discriminant scores and the resulting classifications. These classifications should be compared to the a-priori classification in the attribute assessment class, and a percent correct classification calculated.

9. If the percent correct classification exceeds the percentage set in step 5 above then you have a useful disciminant function, which you can apply and use in resource management. If not, then you can try to redo your stepwise discriminant function or consider if the lower classification efficiency is marginally acceptable. The other alternative is that the data still contain too much variability to develop a strict mathematical approach to identifying lakes at risk.

10. More complicated assessments of the efficacy of classification can be performed by separately considering the percent correct classification for the known impaired systems and the known reference systems. This gets quite complicated but can be used to determine if a particular discriminant function classifies "protectively". An example would be if the impaired lakes are properly classified 80% of the time and the reference lakes are only properly classified 50% of the time then the approach would be protective. This means that you might be requiring monitoring or reduction of inputs in some lakes that don't really need mitigation. This is still classifying protectively in that 80% of the impaired lakes will have action taken (20% will not) and 50% of the lakes that do not need any action will have actions taken that improve the environment. If the approach needs to balance cost with protection then the percentages of misclassification of reference lakes need to be looked at very carefully.

11. If the determination of classification efficiencies are acceptable based on the criteria in either step 9 or 10 above then the discriminant function development should be repeated, but for this iteration no principal components should be calculated. Instead those parameters which emerged as dominant in the principal components should be used directly with no additional transformation or calculations. If the classification efficiency is still acceptable then only the dominant parameters need to be measured. The discriminant function developed for the parameters is applied and the result gives you a classification (impaired or reference) with a confidence interval that can be used to make management decisions.

REFERENCES

Box, G.E.P., and Cox, D.R. 1964. An Analysis of Transformations. Journal of the Royal Statistical Society., Ser. B 26:211-243.

Kaiser, H.F. 1960. The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Krebs, C.J. 1989. Chapter 12 – Survival rates In: Ecological Methodology pp. 446-451.

Table 1. Descriptive Statistics for Overall Dataset (all Assessment Types)

Variable	Statistics				
	N	Min	Max	Mean	Std Dev
Lake Mean Depth (m)	996	0.2	42.7	5.2	4.09
Lake Max Depth (m)	963	0.5	5468.0	18.8	176.15
Lake Area (ha)	1073	0.40	30876.00	327.92	1348.608
Osgood Index	992	0.578	134.397	6.181	6.9661
Land Use Residential (%)	1029	0	0.6452	0.0300	0.0800
Land Use Commercial (%)	1029	0	0.2187	0.0113	0.0248
Land Use Forested (%)	1029	0	0.9986	0.8179	0.1525
Land Use Agricultural (%)	1029	0	0.6375	0.0612	0.0743
Land Use Recreational (%)	1029	0	0.1939	0.0076	0.0205
Land Use Shrubland (%)	1029	0	0.6044	0.0034	0.0316
Land Use Wetland (%)	1029	0	0.4097	0.0539	0.0454
Land Use Barren (%)	1029	0	0.2794	0.0119	0.0274
Total Nitrogen (μg/L)	321	20.00	3796.77	484.38	400.385
Total Phosphorus (μg/L)	937	0.95	376.25	15.16	22.284
Secchi Disk Transparency (m)	1111	0.39	13.92	4.21	2.106
Color (color units)	724	1.00	315.00	26.89	23.901
Chlorophyll *a (μg/L)*	926	0.10	172.25	5.48	8.369
Alkalinity (mg/L as $CaCO_3$)	786	0.11	1897.53	36.74	129.675
pH (S.U.)	710	4.27	8.94	6.81	0.545

Notes:
Includes REF, TEST, TEST2, NONE, and IMP

REF = waterbodies identified as reference by State experts
IMP = waterbodies listed on a State 303(d) list
TEST = waterbodies identified as neither reference nor impacted
TEST2 = Maine waterbodies identified as neither reference nor impacted
NONE = waterbodies with none comments from States, assumed to be neither impacted nor reference

Table 2. Descriptive Statistics by Assessment Types

Variable	Reference					Test					Impaired				
	N	Min	Max	Mean	Std Dev	N	Min	Max	Mean	Std Dev	N	Min	Max	Mean	Std Dev
Lake Mean Depth (m)	353	0.2	24.7	5.9	3.98	493	0.2	42.7	4.7	4.21	150	0.8	33.3	5.2	3.71
Lake Max Depth (m)	330	0.5	67.7	15.6	10.55	485	0.5	5468.0	22.5	248.01	148	1.0	107.0	13.4	13.88
Lake Area (ha)	375	0.80	18043.40	426.35	1257.431	539	0.40	30876.00	247.73	1478.872	159	2.00	11949.00	367.60	1050.494
Osgood Index	353	0.578	28.520	5.900	4.5490	489	0.657	134.397	6.537	8.6104	150	0.707	45.255	5.681	5.4951
Land Use Residential (%)	383	0	0.4336	0.0112	0.0426	492	0	0.6452	0.0343	0.0808	154	0	0.6452	0.0632	0.1245
Land Use Commercial (%)	383	0	0.1045	0.0042	0.0130	492	0	0.2187	0.0140	0.0265	154	0	0.2062	0.0202	0.0352
Land Use Forested (%)	383	0	0.9986	0.8833	0.1054	492	0	0.9967	0.7922	0.1492	154	0	0.9915	0.7374	0.1941
Land Use Agricultural (%)	383	0	0.3601	0.0305	0.0515	492	0	0.5524	0.0744	0.0724	154	0	0.6375	0.0955	0.0975
Land Use Recreational (%)	383	0	0.1939	0.0038	0.0147	492	0	0.1837	0.0089	0.0221	154	0	0.1661	0.0126	0.0253
Land Use Shrubland (%)	383	0	0.0532	0.0027	0.0062	492	0	0.6044	0.0048	0.0453	154	0	0.0272	0.0009	0.0033
Land Use Wetland (%)	383	0	0.4097	0.0438	0.0424	492	0	0.3063	0.0613	0.0473	154	0	0.3106	0.0555	0.0420
Land Use Barren (%)	383	0	0.2794	0.0176	0.0351	492	0	0.2577	0.0090	0.0226	154	0	0.1170	0.0070	0.0147
Total Nitrogen (µg/L)	80	50.00	1770.00	334.28	215.213	202	20.00	3175.55	494.92	364.395	39	198.00	3796.77	737.68	661.398
Total Phosphorus (µg/L)	287	1.00	67.00	9.32	5.697	495	0.95	290.25	15.85	21.440	155	3.32	376.25	23.78	36.653
Secchi Disk Transparency (m)	386	0.75	13.92	5.07	2.069	557	0.39	13.29	3.69	1.898	168	0.55	10.92	3.95	2.216
Color (color units)	247	1.00	129.61	25.36	18.299	350	2.00	315.00	28.94	28.852	127	5.00	89.15	24.22	17.173
Chlorophyll *a* (µg/L)	283	0.20	11.70	3.30	1.993	494	0.10	51.65	5.39	5.795	149	1.50	172.25	9.90	17.015
Alkalinity (mg/L as $CaCO_3$)	253	0.11	408.64	23.60	53.848	390	0.20	1897.53	48.61	175.540	143	1.83	357.64	27.63	51.195
pH (S.U.)	213	4.60	8.36	6.74	0.482	348	4.27	8.94	6.77	0.594	149	5.37	8.90	7.02	0.455

Notes:
Test - includes TEST, TEST2, and NONE
TEST = waterbodies identified as neither reference nor impacted
TEST2 = Maine waterbodies identified as neither reference nor impacted
NONE = waterbodies with none comments from States, assumed to be neither impacted nor reference

Table 3. Descriptive Statistics by Ecoregion and Assessment Type - Reference Data Only

Variable	North Eastern Coastal Zone						Laurentian Plains and Hills						North Eastern Highland					
	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev
Lake Mean Depth (m)	15	1.1	16.3	4.6	5.1	3.46	152	0.9	24.2	4.5	5.6	3.63	186	0.2	24.7	5.1	6.2	4.28
Lake Max Depth (m)	29	2.7	19.7	7.0	8.1	4.55	139	3.4	67.7	12.8	15.9	10.79	162	0.5	54.9	13.7	16.6	10.63
Lake Area (ha)	34	3.00	1673.40	38.82	139.15	337.825	152	4.00	7405.00	164.50	515.80	1056.203	189	0.80	18043.40	77.58	406.08	1485.800
Osgood	15	1.518	17.898	6.225	7.336	5.1194	152	0.578	28.520	3.765	5.097	4.3251	186	0.742	25.425	5.243	6.439	4.5970
Land Use Residential (%)	31	0	0.4336	0.0408	0.0895	0.1206	162	0	0.0725	0.0003	0.0039	0.0087	190	0	0.1284	0.0001	0.0047	0.0145
Land Use Commercial (%)	31	0	0.0923	0.0093	0.0210	0.0273	162	0	0.0489	0.0003	0.0022	0.0056	190	0	0.1045	0.0003	0.0031	0.0120
Land Use Forested (%)	31	0	0.9441	0.8152	0.7548	0.1837	162	0	0.9986	0.8923	0.8782	0.0949	190	1	0.9974	0.9252	0.9085	0.0777
Land Use Agricultural (%)	31	0	0.1724	0.0257	0.0323	0.0365	162	0	0.3377	0.0013	0.0351	0.0579	190	0	0.3601	0.0016	0.0263	0.0475
Land Use Recreational (%)	31	0	0.1939	0.0072	0.0280	0.0420	162	0	0.0260	0.0001	0.0020	0.0038	190	0	0.0758	0.0000	0.0014	0.0066
Land Use Shrubland (%)	31	0	0.0069	0.0000	0.0006	0.0017	162	0	0.0331	0.0003	0.0025	0.0058	190	0	0.0532	0.0001	0.0032	0.0069
Land Use Wetland (%)	31	0	0.1348	0.0460	0.0632	0.0290	162	0	0.4097	0.0475	0.0569	0.0543	190	0	0.1327	0.0248	0.0296	0.0241
Land Use Barren (%)	31	0	0.1062	0.0014	0.0103	0.0240	162	0	0.2794	0.0039	0.0160	0.0356	190	0	0.2756	0.0044	0.0202	0.0361
Total Nitrogen (µg/L)	32	208.56	1000.00	303.43	369.09	162.331	14	153.00	1770.00	236.00	362.71	411.610	34	50.00	602.46	269.50	289.82	125.809
Total Phosphorus (µg/L)	37	4.80	46.63	8.52	11.33	8.085	115	2.00	18.33	9.00	9.62	3.645	135	1.00	67.00	7.13	8.52	6.211
Secchi Disk Transparency (m)	38	1.21	7.50	4.30	3.93	1.483	160	1.00	13.92	4.84	5.32	2.068	188	0.75	13.20	4.89	5.09	2.102
Color (color units)	10	3.74	35.52	8.50	11.54	9.678	114	4.47	129.61	22.18	28.21	19.838	123	1.00	99.00	18.34	23.84	16.680
Chlorphyll *a* (µg/L)	37	0.60	7.53	1.90	2.25	1.407	115	0.80	11.70	3.13	3.72	2.062	131	0.20	10.59	2.70	3.23	1.967
Alkalinity (mg/L as $CaCO_3$)	13	2.00	402.82	9.00	44.59	108.780	117	2.45	408.64	8.73	21.08	44.290	123	0.11	367.56	7.74	23.78	53.918
pH (S.U.)	11	6.00	8.30	6.70	6.73	0.712	111	4.60	8.36	6.81	6.78	0.458	91	5.30	8.00	6.74	6.69	0.479

Table 4. Descriptive Statistics by Ecoregion and Assessment Type - Test Data Only

Variable	North Eastern Coastal Zone						Laurentian Plains and Hills						North Eastern Highland					
	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev
Lake Mean Depth (m)	102	0.7	14.6	2.7	3.2	2.34	119	0.6	18.0	3.9	4.9	3.66	272	0.2	42.7	3.7	5.2	4.82
Lake Max Depth (m)	110	0.8	24.1	4.5	6.6	5.23	108	1.8	5468.0	9.8	63.0	525.05	267	0.5	93.9	9.1	12.8	12.08
Lake Area (ha)	134	1.00	469.72	28.30	52.19	69.279	119	1.00	7257.60	84.00	362.58	920.577	286	0.40	30876.00	48.77	291.56	1936.500
Osgood	102	1.293	66.000	4.526	6.378	7.4065	119	0.657	71.000	3.939	6.027	8.4298	268	0.667	134.397	5.198	6.824	9.1181
Land Use Residential (%)	96	0	0.6452	0.0676	0.1248	0.1478	121	0	0.0844	0.0034	0.0085	0.0148	275	0	0.1201	0.0063	0.0140	0.0213
Land Use Commercial (%)	96	0	0.2187	0.0205	0.0364	0.0438	121	0	0.0483	0.0008	0.0051	0.0101	275	0	0.1524	0.0036	0.0101	0.0179
Land Use Forested (%)	96	0	0.9591	0.6941	0.6431	0.1931	121	0	0.9943	0.8402	0.8203	0.1165	275	0	0.9967	0.8515	0.8318	0.1064
Land Use Agricultural (%)	96	0	0.3317	0.0528	0.0636	0.0646	121	0	0.5524	0.0725	0.0785	0.0811	275	0	0.4073	0.0614	0.0764	0.0709
Land Use Recreational (%)	96	0	0.1837	0.0206	0.0345	0.0383	121	0	0.0563	0.0016	0.0045	0.0082	275	0	0.0683	0.0001	0.0019	0.0074
Land Use Shrubland (%)	96	0	0.0119	0.0000	0.0007	0.0024	121	0	0.6044	0.0000	0.0120	0.0751	275	0	0.5664	0.0000	0.0030	0.0343
Land Use Wetland (%)	96	0	0.2329	0.0828	0.0878	0.0433	121	0	0.3063	0.0458	0.0601	0.0553	275	0	0.2889	0.0415	0.0527	0.0413
Land Use Barren (%)	96	0	0.1810	0.0023	0.0084	0.0207	121	0	0.1170	0.0022	0.0076	0.0157	275	0	0.2577	0.0038	0.0099	0.0256
Total Nitrogen (μg/L)	116	122.92	3175.55	452.04	579.67	412.029	11	211.00	654.00	333.00	387.91	136.392	75	20.00	1877.39	312.33	379.52	260.513
Total Phosphorus (μg/L)	144	0.95	290.25	15.80	24.48	35.720	103	3.00	87.12	10.94	14.02	10.932	248	1.26	77.94	9.53	11.60	8.267
Secchi Disk Transparency (m)	145	0.39	9.14	2.15	2.50	1.407	122	0.71	11.36	4.32	4.47	2.035	290	0.63	13.29	3.67	3.96	1.781
Color (color units)	62	4.00	315.00	24.15	37.20	46.322	96	5.00	155.00	23.00	32.65	28.099	192	2.00	150.00	20.10	24.43	19.838
Chlorphyll *a (μg/L)*	144	0.20	40.50	4.07	6.27	7.133	98	1.60	50.57	3.98	5.51	6.005	252	0.10	51.65	3.95	4.84	4.720
Alkalinity (mg/L as $CaCO_3$)	93	1.00	1680.98	11.00	82.67	255.374	99	1.90	332.71	10.00	23.72	48.155	198	0.20	1897.53	7.08	45.06	168.420
pH (S.U.)	77	4.50	8.94	6.67	6.72	0.733	97	5.47	8.38	6.93	6.93	0.471	174	4.27	8.80	6.80	6.71	0.574

Table 5. Descriptive Statistics by Ecoregion and Assessment Type - Impaired Data Only

Variable	North Eastern Coastal Zone						Laurentian Plains and Hills						North Eastern Highland					
	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev	N	Min	Max	Median	Mean	Std Dev
Lake Mean Depth (m)	29	0.8	12.5	2.8	3.7	2.79	79	0.8	15.0	4.6	5.1	2.73	42	1.9	33.3	5.3	6.4	5.25
Lake Max Depth (m)	33	1.0	22.9	7.3	7.9	5.66	75	2.7	51.8	10.1	13.4	10.22	40	3.0	107.0	12.7	18.1	21.31
Lake Area (ha)	37	3.80	524.00	23.60	80.93	129.299	79	2.00	2716.00	141.00	383.12	564.823	43	10.10	11949.00	156.00	585.76	1850.140
Osgood	29	1.473	16.101	4.872	5.832	3.8989	79	0.707	45.255	3.615	5.859	6.6820	42	1.303	21.083	4.129	5.242	3.7601
Land Use Residential (%)	28	0	0.6452	0.1466	0.2386	0.2104	83	0	0.2152	0.0105	0.0202	0.0318	43	0	0.1173	0.0233	0.0319	0.0299
Land Use Commercial (%)	28	0	0.2062	0.0359	0.0623	0.0603	83	0	0.1110	0.0045	0.0109	0.0173	43	0	0.0556	0.0071	0.0108	0.0123
Land Use Forested (%)	28	0	0.8982	0.5765	0.5267	0.2578	83	0	0.9915	0.8110	0.7821	0.1304	43	0	0.9805	0.8222	0.7882	0.1586
Land Use Agricultural (%)	28	0	0.1628	0.0310	0.0498	0.0504	83	0	0.6375	0.0952	0.1205	0.1123	43	0	0.4618	0.0607	0.0769	0.0734
Land Use Recreational (%)	28	0	0.1661	0.0170	0.0374	0.0467	83	0	0.0896	0.0049	0.0088	0.0136	43	0	0.0316	0.0016	0.0036	0.0063
Land Use Shrubland (%)	28	0	0.0206	0.0000	0.0013	0.0046	83	0	0.0125	0.0000	0.0009	0.0023	43	0	0.0272	0.0000	0.0007	0.0041
Land Use Wetland (%)	28	0	0.1017	0.0748	0.0704	0.0206	83	0	0.2724	0.0409	0.0488	0.0405	43	0	0.3106	0.0508	0.0586	0.0521
Land Use Barren (%)	28	0	0.1170	0.0042	0.0133	0.0274	83	0	0.0414	0.0026	0.0053	0.0076	43	0	0.0655	0.0018	0.0061	0.0127
Total Nitrogen (μg/L)	31	200.00	3796.77	629.16	836.66	705.800	4	200.00	708.00	365.50	409.75	219.988	4	198.00	400.00	298.00	298.50	96.341
Total Phosphorus (μg/L)	41	5.47	376.25	25.29	46.25	63.591	71	4.47	72.78	14.00	18.17	12.874	43	3.32	53.57	8.00	11.63	10.064
Secchi Disk Transparency (m)	42	0.55	6.63	1.66	2.28	1.697	83	0.80	9.05	3.81	4.16	1.896	43	1.09	10.92	5.41	5.18	2.296
Color (color units)	17	8.00	68.47	21.00	29.34	20.796	70	6.32	70.87	21.57	24.81	15.032	40	5.00	89.15	13.77	21.03	18.791
Chlorphyll *a* (μg/L)	41	2.40	172.25	6.46	16.68	29.451	67	1.60	35.61	5.41	8.28	7.460	41	1.50	39.38	4.19	5.77	6.560
Alkalinity (mg/L as $CaCO_3$)	29	4.24	357.64	18.12	64.93	97.043	73	2.00	150.58	15.49	23.05	26.783	41	1.83	39.52	8.49	9.40	5.762
pH (S.U.)	27	5.37	8.90	7.10	7.08	0.634	81	5.70	8.25	7.08	7.08	0.434	41	6.08	7.39	6.91	6.85	0.296

Table 6. Non-parametric comparison (Wilcoxon test) between Reference and Impaired data

Variable	P-value	Significant? Yes/No
Lake mean depth (m)	0.032	Yes
Lake max depth (m)	<0.010	Yes
Lake area (ha)	0.431	No
Osgood Index	0.217	No
Land Use Residential (%)	<0.010	Yes
Land Use Commercial (%)	<0.010	Yes
Land Use Forest (%)	<0.010	Yes
Land Use Agricultural (%)	<0.010	Yes
Land Use Recreational (%)	<0.010	Yes
Land Use Shrubland (%)	<0.010	Yes
Land Use Wetland (%)	<0.010	Yes
Land Use Barren (%)	0.043	Yes
Total Nitrogen (μg/L)	<0.010	Yes
Total Phosphorus (μg/L)	<0.010	Yes
SDT (m)	<0.010	Yes
Color (color units)	0.401	No
Chlorophyll *a* (μg/L)	<0.010	Yes
Alkalinity (mg/L as $CaCO_3$)	<0.010	Yes
pH (S.U)	<0.010	Yes

Table 7 - Non-parametric comparison (Wilcoxon test) between EPA Ecoregions (reference data only); Ecoregion 1= NECZ, Ecoregion 2 = LPH, Ecoregion 3 = NEH

Variable	P-values	Significant? Yes/No
Lake mean depth	0.679 for Ecoregions 1 and 2 0.295 for Ecoregions 1 and 3 0.174 for Ecoregions 2 and 3	No
Lake max depth	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.324 for Ecoregions 2 and 3	Yes No
Lake area	<0.010 for all three comparisons	Yes
Osgood	0.073 for Ecoregions 1and 2 0.535 for Ecoregions 1 and 3 <0.010 for Ecoregions 2 and 3	No No Yes
L. Use Residential	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.125 for Ecoregions 2 and 3	Yes No
L. Use Commercial	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.271 for Ecoregions 2 and 3	Yes No
L. Use Forest	<0.010 for all three comparisons	Yes
L. Use Agricultural	0.100 for Ecoregions 1 and 2 0.053 for Ecoregions 1 and 3 0.947 for Ecoregions 2 and 3	No
L. Use Recreational	<0.010 for all three comparisons	Yes
L. Use Shrubland	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.895 for Ecoregions 2 and 3	Yes No
L.Use Wetland	0.039 for Ecoregions 1 and 2 <0.010 for Ecoregions 1 and 3, and 2 and 3	Yes
L. Use Barren	0.082 for Ecoregions 1 and 2 0.020 for Ecoregions 1 and 3 0.227 for Ecoregions 2 and 3	No Yes No
Total Nitrogen	0.023 for Ecoregions 1 and 2 0.026 for Ecoregions 1 and 3 0.626 for Ecoregions 2 and 3	Yes No
Total Phosphorus	0.944 for Eparegions 1 and 2 <0.010 for Eparegions 1 and 3, and 2 and 3	No Yes
Secchi depth	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.525 for Ecoregions 2 and 3	Yes No
Color	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.065 for Ecoregions 2 and 3	Yes No
Chlorophyll *a*	<0.010 for Ecoregions 1 and 2, and 1 and 3 0.055 for Ecoregions and 2 and 3	Yes No
Alkalinity	0.858 for Ecoregions 1 and 2 0.584 for Ecoregions 1 and 3 0.659 for Ecoregions 2 and 3	No
pH	0.287 for Ecoregions 1 and 2 0.627 for Ecoregions 1 and 3 0.239 for Ecoregions 2 and 3	No

Table 8. Spearman Correlations for the overall dataset including all assessment types (REF, IMP, TEST, TEST2, and NONE)

Parameter	Mean Depth	Max Depth	Lake Area	OI	Latit	LU_Res	LU_Com	LU_For	LU_Agr	LU_Rec	LU_Shr	LU_Wet	LU_Bar	TN	TP	SDT	Color	Chl a	Alk	pH
Lake Mean Depth (m)	1.000	0.898	0.543	0.344	0.264	-0.129	-0.113	0.179	-0.117	-0.147	0.050	-0.214	0.050	-0.574	-0.480	0.662	-0.381	-0.303	-0.050	0.137
Lake Max Depth (m)		1.000	0.590	0.193	0.402	-0.230	-0.200	0.249	-0.191	-0.205	0.120	-0.242	0.078	-0.475	-0.501	0.668	-0.323	-0.273	-0.050	0.099
Lake Area (ha)			1.000	-0.541	0.397	-0.244	-0.224	0.183	-0.206	-0.183	0.124	-0.091	0.069	-0.337	-0.268	0.403	-0.109	-0.201	-0.009	0.118
Osgood Index (OI)				1.000	-0.169	0.126	0.113	-0.016	0.156	0.040	-0.075	-0.129	-0.016	-0.253	-0.196	0.218	-0.257	-0.059	-0.061	0.021
Latitude (dec. degree)					1.000	-0.713	-0.601	0.457	-0.344	-0.460	0.340	-0.407	0.202	-0.347	-0.146	0.347	0.166	-0.033	0.129	0.110
L.U. Residential (%)						1.000	0.845	-0.677	0.526	0.663	-0.349	0.398	-0.089	0.307	0.201	-0.225	-0.099	0.167	-0.014	0.122
L.U. Commercial (%)							1.000	-0.670	0.506	0.562	-0.245	0.352	0.052	0.260	0.189	-0.231	-0.062	0.142	0.039	0.144
LU Forested (%)								1.000	-0.571	-0.537	0.170	-0.570	-0.088	-0.433	-0.352	0.328	-0.103	-0.252	-0.196	-0.272
L.U. Agricultural (%)									1.000	0.303	-0.351	0.128	-0.204	0.116	0.206	-0.143	-0.040	0.244	0.134	0.339
L.U. Recreational (%)										1.000	-0.176	0.342	-0.044	0.336	0.308	-0.222	-0.002	0.143	0.134	0.161
L.U. Shrubland (%)											1.000	-0.073	0.575	-0.077	-0.026	0.061	0.126	-0.040	0.005	-0.130
L.U. Wetland (%)												1.000	0.040	0.353	0.188	-0.275	0.166	0.117	-0.125	-0.114
L.U. Barren (%)													1.000	0.048	0.013	-0.010	0.095	-0.011	0.069	-0.019
T. Nitrogen (µg/L)														1.000	0.643	-0.690	0.612	0.414	-0.073	0.117
T. Phosphorus (µg/L)															1.000	-0.668	0.514	0.581	0.280	0.164
SDT (m)																1.000	-0.583	-0.496	-0.246	0.023
Color (color units)																	1.000	0.439	0.047	-0.081
Chlorophyll *a* (µg/L)																		1.000	0.118	0.128
Alkalinity ($CaCO_3$ mg/L)																			1.000	0.689
pH (S.U.)																				1.000

Table 9. Spearman Correlations for Reference Data

Parameter	Mean Depth	Max Depth	Lake Area	OI	LU_Res	LU_Com	LU_For	LU_Agr	LU_Rec	LU_Shr	LU_Wet	LU_Bar	TN	TP	SDT	Color	Chl a	Alk	pH
Lake Mean Depth (m)	NA	0.895	0.489	0.290	-0.084	-0.069	0.165	-0.114	-0.139	-0.052	-0.172	0.006	-0.551	-0.469	0.582	-0.327	-0.392	0.014	0.085
Lake Max Depth (m)		NA	0.622	0.062	-0.245	-0.192	0.265	-0.223	-0.267	-0.007	-0.194	0.029	-0.471	-0.420	0.562	-0.247	-0.264	0.005	0.088
Lake Area (ha)			NA	-0.629	-0.374	-0.312	0.208	-0.382	-0.297	0.111	-0.079	0.106	-0.355	-0.218	0.256	0.000	-0.148	0.110	0.101
Osgood Index (OI)				NA	0.308	0.234	-0.086	0.308	0.199	-0.098	-0.063	-0.091	-0.267	-0.163	0.211	-0.287	-0.067	-0.124	-0.018
L.U. Residential (%)					NA	0.771	-0.539	0.688	0.674	-0.353	0.420	-0.303	0.073	0.042	-0.042	-0.196	-0.010	-0.265	-0.008
L.U. Commercial (%)						NA	-0.517	0.590	0.578	-0.257	0.343	-0.123	0.118	0.017	-0.059	-0.134	-0.069	-0.106	0.050
LU Forested (%)							NA	-0.502	-0.499	-0.013	-0.671	-0.153	-0.296	-0.164	0.211	-0.072	-0.126	-0.028	-0.159
L.U. Agricultural (%)								NA	0.498	-0.329	0.289	-0.343	0.095	0.107	-0.070	-0.100	0.208	-0.101	0.157
L.U. Recreational (%)									NA	-0.133	0.338	-0.091	0.081	0.207	-0.147	-0.098	0.039	-0.073	0.127
L.U. Shrubland (%)										NA	-0.023	0.652	0.043	0.048	-0.065	0.170	0.090	0.012	-0.068
L.U. Wetland (%)											NA	0.014	0.281	0.091	-0.151	0.106	0.031	-0.230	-0.089
L.U. Barren (%)												NA	0.150	0.014	-0.090	0.133	-0.025	0.052	-0.042
T. Nitrogen (µg/L)													NA	0.541	-0.597	0.672	0.180	-0.216	-0.033
T. Phosphorus (µg/L)														NA	-0.528	0.442	0.471	0.055	0.055
SDT (m)															NA	-0.593	-0.407	-0.255	0.000
Color (color units)																NA	0.394	0.088	-0.003
Chlorophyll *a* (µg/L)																	NA	0.002	-0.029
Alkalinity (mg/L as $CaCO_3$)																		NA	0.625
pH (S.U.)																			NA

Table 10. Spearman Correlations for the North Eastern Coastal Zone (NECZ) dataset

Parameter	Mean Depth	Max Depth	Lake Area	OI	Latit	LU_Res	LU_Com	LU_For	LU_Agr	LU_Rec	LU_Shr	LU_Wet	LU_Bar	TN	TP	SDT	Color	Chl a	Alk	pH
Lake Mean Depth (m)	1.000	0.823	0.452	0.562	0.249	0.076	-0.064	0.173	-0.071	-0.318	-0.222	-0.027	-0.156	-0.410	-0.373	0.596	-0.447	-0.236	-0.138	0.182
Lake Max Depth (m)		1.000	0.345	0.334	0.130	-0.040	-0.078	0.165	-0.045	-0.256	0.060	-0.026	-0.171	-0.335	-0.420	0.534	-0.505	-0.134	-0.126	0.122
Lake Area (ha)			1.000	-0.430	0.069	-0.111	-0.193	0.168	0.093	-0.232	-0.049	0.130	-0.156	-0.175	-0.226	0.262	-0.230	-0.153	-0.162	0.053
Osgood Index (OI)				1.000	0.128	0.218	0.129	-0.096	-0.112	-0.017	-0.168	-0.130	0.094	-0.242	-0.137	0.288	-0.172	-0.032	0.026	0.132
Latitude (dec. degree)					1.000	-0.002	-0.050	0.204	0.292	-0.628	0.063	-0.203	0.133	-0.001	0.048	0.221	0.247	0.244	-0.299	-0.070
L.U. Residential (%)						1.000	0.783	-0.812	-0.108	0.313	-0.162	0.014	0.373	0.222	0.243	-0.176	-0.076	0.160	0.345	0.375
L.U. Commercial (%)							1.000	-0.820	-0.073	0.399	-0.081	0.149	0.461	0.133	0.241	-0.182	0.055	0.165	0.429	0.455
LU Forested (%)								1.000	-0.015	-0.549	0.117	-0.333	-0.432	-0.319	-0.401	0.339	-0.042	-0.235	-0.499	-0.351
L.U. Agricultural (%)									1.000	-0.371	0.198	-0.038	0.308	0.106	0.044	-0.109	0.085	0.171	0.053	0.187
L.U. Recreational (%)										1.000	-0.139	0.214	-0.012	0.134	0.330	-0.301	-0.098	0.016	0.632	0.079
L.U. Shrubland (%)											1.000	-0.038	0.240	0.015	0.051	-0.050	0.111	0.245	0.115	-0.089
L.U. Wetland (%)												1.000	0.128	0.124	0.038	-0.042	0.104	-0.064	-0.252	-0.204
L.U. Barren (%)													1.000	0.080	0.097	-0.126	0.232	0.171	0.245	0.493
T. Nitrogen (µg/L)														1.000	0.571	-0.612	0.650	0.444	-0.038	0.159
T. Phosphorus (µg/L)															1.000	-0.602	0.547	0.584	0.390	0.347
SDT (m)																1.000	-0.572	-0.540	-0.320	-0.080
Color (color units)																	1.000	0.490	-0.164	-0.020
Chlorophyll *a* (µg/L)																		1.000	0.304	0.320
Alkalinity (mg/L as $CaCO_3$)																			1.000	0.668
pH (S.U.)																				1.000

Table 11. Spearman Correlations for the combined Laurentian Plains and Hills (LPH) and North Eastern Highland (NEH) dataset

Parameter	Mean Depth	Max Depth	Lake Area	OI	Latit	LU_Res	LU_Com	LU_For	LU_Agr	LU_Rec	LU_Shr	LU_Wet	LU_Bar	TN	TP	SDT	Color	Chl a	Alk	pH
Lake Mean Depth (m)	1.000	0.904	0.519	0.330	0.173	-0.093	-0.079	0.148	-0.115	-0.102	0.063	-0.193	0.068	-0.625	-0.484	0.648	-0.368	-0.310	-0.014	0.119
Lake Max Depth (m)		1.000	0.578	0.190	0.281	-0.175	-0.151	0.206	-0.216	-0.122	0.098	-0.198	0.093	-0.517	-0.484	0.640	-0.298	-0.334	-0.019	0.086
Lake Area (ha)			1.000	-0.575	0.342	-0.182	-0.164	0.116	-0.251	-0.086	0.123	-0.041	0.089	-0.371	-0.208	0.337	-0.078	-0.202	0.056	0.125
Osgood Index (OI)				1.000	-0.218	0.126	0.113	-0.005	0.176	0.036	-0.069	-0.136	-0.028	-0.285	-0.227	0.238	-0.271	-0.070	-0.088	-0.008
Latitude (dec. degree)					1.000	-0.636	-0.525	0.329	-0.475	-0.241	0.358	-0.299	0.245	-0.086	0.078	0.110	0.222	-0.034	0.339	0.154
L.U. Residential (%)						1.000	0.817	-0.599	0.678	0.569	-0.350	0.320	-0.139	-0.064	0.088	-0.067	-0.132	0.172	-0.102	0.148
L.U. Commercial (%)							1.000	-0.587	0.625	0.458	-0.231	0.271	0.010	-0.028	0.090	-0.106	-0.090	0.131	-0.052	0.145
LU Forested (%)								1.000	-0.691	-0.429	0.138	-0.524	-0.063	-0.246	-0.276	0.213	-0.111	-0.254	-0.138	-0.303
L.U. Agricultural (%)									1.000	0.406	-0.414	0.176	-0.275	0.077	0.254	-0.151	-0.060	0.263	0.134	0.355
L.U. Recreational (%)										1.000	-0.136	0.236	-0.041	0.082	0.213	-0.060	0.008	0.164	0.056	0.189
L.U. Shrubland (%)											1.000	-0.052	0.626	0.042	-0.019	0.031	0.125	-0.085	-0.001	-0.134
L.U. Wetland (%)												1.000	0.032	0.215	0.124	-0.197	0.169	0.127	-0.153	-0.085
L.U. Barren (%)													1.000	0.120	0.000	0.001	0.072	-0.055	0.024	-0.059
T. Nitrogen (μg/L)														1.000	0.596	-0.703	0.649	0.388	0.053	0.037
T. Phosphorus (μg/L)															1.000	-0.651	0.517	0.588	0.236	0.144
SDT (m)																1.000	-0.600	-0.511	-0.198	0.013
Color (color units)																	1.000	0.435	0.080	-0.078
Chlorophyll *a* (μg/L)																		1.000	0.059	0.086
Alkalinity (mg/L as $CaCO_3$)																			1.000	0.704
pH (S.U.)																				1.000

Table 12. The eigenvalues and eigenvectors of the correlation matrix for the Overall dataset, all numeric variables except TN & TP

			Eigenvector, coefficient of								
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	LKMNDEP	LKAREA	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.459	0.273	0.033	0.033	0.470	0.229	0.571	-0.430	-0.419	-0.194	-0.019
PRIN 2	1.647	0.456	-0.046	0.130	0.240	0.149	0.034	-0.160	0.251	0.624	0.650
PRIN 3	1.166	0.586	0.409	0.766	0.299	0.029	-0.015	0.238	0.192	-0.006	-0.250
PRIN 4	1.113	0.709	-0.156	-0.265	0.261	0.784	-0.099	0.388	0.197	-0.079	-0.134
PRIN 5	0.964	0.817	0.893	-0.356	-0.146	0.172	-0.055	-0.075	-0.055	0.052	0.104
PRIN 6	0.568	0.880	0.034	-0.160	0.130	-0.130	0.126	-0.383	0.801	-0.365	-0.043
PRIN 7	0.460	0.931	0.080	-0.354	0.416	-0.473	0.361	0.573	0.044	0.095	0.046
PRIN 8	0.421	0.978	0.026	0.138	0.073	-0.016	-0.197	0.201	-0.120	-0.648	0.679
PRIN 9	0.202	1.000	-0.007	0.171	-0.590	0.197	0.689	0.247	0.160	-0.046	0.145

N= 472

Table 13. The eigenvalues and eigenvectors of the correlation matrix for the Overall dataset, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.077	0.297	-0.044	0.050	-0.567	0.467	0.545	0.362	0.168
PRIN 2	1.545	0.517	-0.047	0.039	0.225	-0.424	0.015	0.549	0.681
PRIN 3	1.097	0.674	0.661	0.734	0.066	-0.019	0.046	0.083	-0.098
PRIN 4	0.919	0.806	0.746	-0.646	-0.106	-0.010	-0.019	0.001	0.117
PRIN 5	0.561	0.886	-0.003	0.040	-0.101	0.487	-0.785	0.362	0.060
PRIN 6	0.427	0.947	-0.008	-0.192	0.410	0.029	0.194	0.619	-0.610
PRIN 7	0.373	1.000	0.036	-0.022	0.659	0.603	0.214	-0.214	0.329

N= 472

Table 14. The eigenvalues and eigenvectors of the correlation matrix for NECZ ecoregion, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.683	0.383	-0.435	-0.016	-0.554	0.441	0.451	0.325	0.025
PRIN 2	1.451	0.591	0.146	-0.134	0.106	-0.315	0.148	0.536	0.736
PRIN 3	1.081	0.745	0.169	0.917	-0.149	-0.134	0.228	-0.130	0.146
PRIN 4	0.723	0.848	0.651	-0.228	0.007	0.467	0.400	-0.332	0.189
PRIN 5	0.464	0.914	0.334	0.196	-0.175	0.474	-0.649	0.416	0.025
PRIN 6	0.411	0.973	0.370	0.019	0.146	-0.173	0.353	0.540	-0.629
PRIN 7	0.188	1.000	-0.297	0.225	0.780	0.465	0.129	0.128	0.068

N= 44

Table 15. The eigenvalues and eigenvectors of the correlation matrix for LPH+NEH ecoregions, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.053	0.293	-0.044	0.042	-0.554	0.440	0.540	0.394	0.220
PRIN 2	1.543	0.514	-0.034	0.107	0.271	-0.458	-0.010	0.528	0.652
PRIN 3	1.129	0.675	0.633	0.744	0.092	0.060	0.038	0.090	-0.157
PRIN 4	0.912	0.805	0.770	-0.594	-0.144	-0.094	-0.004	-0.028	0.154
PRIN 5	0.583	0.889	0.013	0.001	-0.132	0.483	-0.801	0.296	0.142
PRIN 6	0.408	0.947	0.010	-0.267	0.354	0.050	0.109	0.657	-0.598
PRIN 7	0.371	1.000	0.053	-0.097	0.669	0.590	0.230	-0.195	0.317

N= 428

Table 16. The eigenvalues and eigenvectors of the correlation matrix for Reference data only, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.322	0.332	0.470	0.046	0.569	-0.492	-0.421	0.076	0.169
PRIN 2	1.531	0.550	-0.041	-0.096	-0.166	-0.046	0.170	0.687	0.677
PRIN 3	1.131	0.712	0.148	0.876	0.088	-0.045	0.421	0.130	-0.086
PRIN 4	0.727	0.816	0.732	-0.114	0.000	0.669	0.037	0.053	0.011
PRIN 5	0.580	0.899	0.166	-0.318	0.194	-0.246	0.722	-0.434	0.255
PRIN 6	0.427	0.960	-0.042	0.329	-0.202	0.153	-0.304	-0.560	0.648
PRIN 7	0.282	1.000	-0.437	0.012	0.750	0.472	0.039	0.040	0.141

N= 141

Table 17. The eigenvalues and eigenvectors of the correlation matrix for Test data only, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.041	0.292	-0.044	0.089	-0.571	0.508	0.547	0.325	0.027
PRIN 2	1.586	0.518	-0.115	-0.156	0.189	-0.321	0.132	0.582	0.683
PRIN 3	1.103	0.676	0.661	0.705	0.116	-0.130	0.147	0.047	0.110
PRIN 4	0.909	0.806	0.734	-0.634	-0.111	0.066	-0.107	0.160	-0.074
PRIN 5	0.545	0.884	0.019	0.175	-0.130	0.529	-0.695	0.034	0.433
PRIN 6	0.438	0.946	-0.065	0.088	0.535	0.329	-0.097	0.598	-0.476
PRIN 7	0.377	1.000	0.071	-0.174	0.555	0.479	0.396	-0.411	0.317

N= 236

Table 18. The eigenvalues and eigenvectors of the correlation matrix for Impaired data only, all numeric variables except surface area, mean depth, TN, and TP

			Eigenvector, coefficient of						
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.795	0.399	-0.366	0.014	-0.553	0.438	0.463	0.316	0.230
PRIN 2	1.651	0.635	0.275	0.409	0.098	-0.284	-0.087	0.585	0.563
PRIN 3	1.002	0.778	-0.119	0.816	0.015	0.242	-0.123	0.056	-0.492
PRIN 4	0.735	0.883	0.785	-0.010	0.113	0.390	0.452	-0.004	-0.125
PRIN 5	0.428	0.944	-0.127	0.297	0.054	-0.537	0.687	-0.361	0.047
PRIN 6	0.238	0.978	-0.225	-0.267	0.433	-0.146	0.284	0.623	-0.448
PRIN 7	0.151	1.000	-0.306	0.081	0.693	0.455	0.080	-0.192	0.411

N= 95

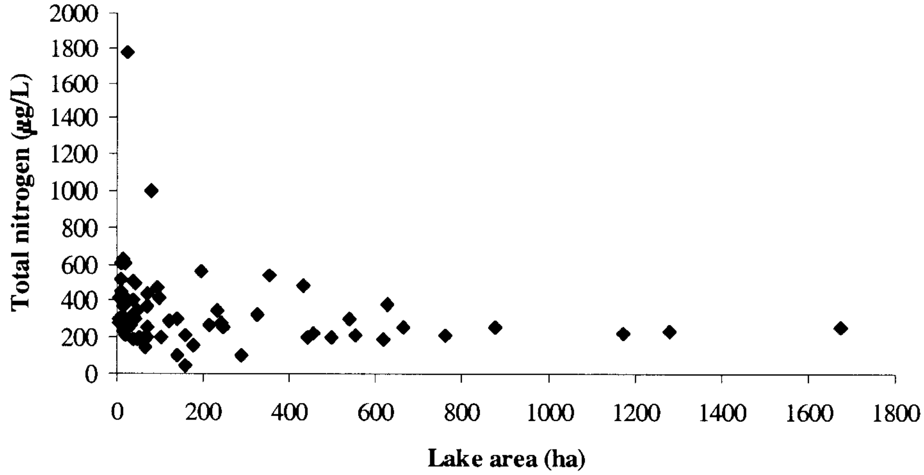


Figure 1. Scatter plot showing the relationship between Total Nitrogen (y- axis) and Lake area (x-axis) for the reference lakes. Note that some datapoints were beyond the scale on the x-axis

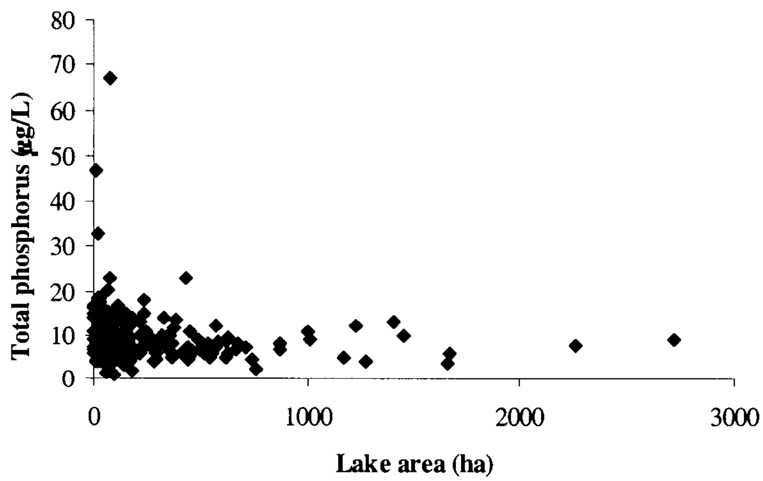


Figure 2. Scatter plot showing the relationship between Total Phosphorus (y- axis) and Lake area (x-axis) for the reference lakes. Note that some datapoints were beyond the scale on the x-axis

Appendix A

PCA was run seven ways:

1) Overall data (all parameters, excluding TN)
2) Overall (all parameters excluding TN, mean depth and surface area)
3) NECZ (all parameters excluding TN, mean depth and surface area)
4) LPH and NEH combined (all parameters excluding TN, mean depth and surface area)
5) Reference (all parameters excluding TN, mean depth and surface area)
6) Test (all parameters excluding TN, mean depth and surface area)
7) Impaired (all parameters excluding TN, mean depth and surface area)

PCA Results

Results from the PCA have been divided by assessment type/ecoregion and also by the inclusion of parameters. Each analysis is presented in Tables A1 through A7. Each table includes information on the eigenvalues, the cumulative percentage of the total variance contributed by each principal component, and the individual principal components (X_1-X_n) (or eigenvectors).

<u>PCA #A1 - Overall data</u> (all parameters, excluding TN) - Table A1

The eigenvalue for a principal component indicates the percentage of the variance that the Principal component accounts for. The first principal component accounts for (2.861/10) *100% = 28%, the second for (1.703/10) or 17%, the third and fourth for 11% each, and the fifth for about 9%. Cumulatively, the second component accounts for 45%, the third for 58% the fourth for 68% and the fifth for 78%, making the last five components account for the remainder of the variance (approximately 22%)

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by lake area with an inverse value for Osgood index.

<u>PCA #A2 - Overall</u> (all parameters excluding TN, mean depth and surface area) – Table A2

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

<u>PCA #A3 - NECZ</u> (all parameters excluding TN, mean depth and surface area) – Table A3

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with inverse values for alkalinity, color, and SDT. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #A4 - LPH and NEH combined (all parameters excluding TN, mean depth and surface area) – Table A4

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #A5 - Reference (all parameters excluding TN, mean depth and surface area) – Table A5

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #A6 - Test (all parameters excluding TN, mean depth and surface area) – Table A6

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for color. PC4 was dominated by max depth with an inverse value for Osgood index.

PCA #A7 - Impaired (all parameters excluding TN, mean depth and surface area) – Table A7

PC1 was dominated by TP, chlorophyll *a* and color with an inverse value for SDT. PC2 was dominated by pH and alkalinity with an inverse value for color. PC3 was dominated by Osgood Index with an inverse value for pH. PC4 was dominated by max depth with an inverse value for Osgood index.

Table A1. The eigenvalues and eigenvectors of the correlation matrix for the Overall dataset, all numeric variables except TN

			Eigenvector, coefficient of									
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	LKMNDEP	LKAREA	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.861	0.286	-0.029	-0.019	-0.374	-0.164	0.428	-0.501	0.382	0.441	0.230	0.065
PRIN 2	1.703	0.456	-0.031	0.133	0.353	0.225	0.184	0.152	-0.220	0.196	0.551	0.599
PRIN 3	1.184	0.575	0.309	0.539	0.405	0.347	0.176	0.024	0.323	0.216	-0.138	-0.359
PRIN 4	1.139	0.689	-0.312	-0.600	0.118	0.668	0.133	-0.020	0.194	0.063	-0.137	-0.062
PRIN 5	0.964	0.785	0.893	-0.357	-0.147	0.168	0.003	-0.053	-0.078	-0.056	0.051	0.104
PRIN 6	0.692	0.854	0.033	-0.208	0.114	-0.283	0.529	0.338	-0.420	0.407	-0.319	-0.156
PRIN 7	0.473	0.902	0.049	-0.278	0.277	-0.375	0.219	0.376	0.533	-0.317	0.336	-0.130
PRIN 8	0.435	0.945	0.064	-0.073	0.280	-0.227	-0.200	-0.041	0.363	0.185	-0.550	0.591
PRIN 9	0.349	0.980	-0.022	0.232	-0.179	0.136	0.599	-0.041	-0.009	-0.604	-0.292	0.291
PRIN 10	0.199	1.000	-0.006	0.157	-0.580	0.191	-0.087	0.681	0.254	0.218	-0.028	0.136

N= 471

Table A2. The eigenvalues and eigenvectors of the correlation matrix for the Overall dataset, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.574	0.322	-0.033	0.027	0.491	-0.474	0.391	0.510	0.314	0.142
PRIN 2	1.545	0.515	-0.046	0.039	0.008	0.233	-0.431	0.010	0.544	0.679
PRIN 3	1.097	0.652	0.658	0.738	-0.015	0.056	-0.011	0.050	0.089	-0.095
PRIN 4	0.922	0.767	0.749	-0.646	0.073	-0.061	-0.047	-0.021	-0.025	0.098
PRIN 5	0.679	0.852	-0.034	-0.016	0.551	0.410	-0.453	0.437	-0.300	-0.202
PRIN 6	0.465	0.910	-0.009	-0.106	0.337	0.464	0.311	-0.351	0.539	-0.387
PRIN 7	0.388	0.959	0.028	0.095	0.237	0.370	0.513	-0.160	-0.449	0.554
PRIN 8	0.329	1.000	0.025	-0.127	-0.528	0.442	0.307	0.630	0.124	-0.050

N=471

Table A3. The eigenvalues and eigenvectors of the correlation matrix for NECZ ecoregion, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	3.140	0.393	-0.369	-0.027	0.425	-0.475	0.366	0.463	0.322	0.063
PRIN 2	1.518	0.582	0.242	-0.123	0.223	0.223	-0.392	0.123	0.435	0.684
PRIN 3	1.083	0.718	0.187	0.912	0.053	-0.127	-0.135	0.234	-0.169	0.109
PRIN 4	0.813	0.819	0.491	-0.240	0.476	0.251	0.136	0.416	-0.437	-0.173
PRIN 5	0.644	0.900	0.458	-0.064	-0.387	-0.258	0.614	-0.056	-0.053	0.432
PRIN 6	0.431	0.954	0.474	0.148	0.088	0.044	0.135	-0.153	0.669	-0.505
PRIN 7	0.223	0.981	-0.238	0.251	0.405	0.502	0.487	-0.433	-0.018	0.202
PRIN 8	0.149	1.000	-0.196	0.073	-0.468	0.571	0.208	0.573	0.192	-0.057

N = 44

Table A4. The eigenvalues and eigenvectors of the correlation matrix for LPH+NEH ecoregions, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.797	0.350	-0.029	0.025	0.542	-0.454	0.360	0.485	0.324	0.170
PRIN 2	1.544	0.543	-0.035	0.108	-0.006	0.263	-0.452	-0.004	0.533	0.655
PRIN 3	1.129	0.684	0.633	0.743	0.002	0.093	0.058	0.037	0.088	-0.158
PRIN 4	0.913	0.798	0.770	-0.596	0.030	-0.122	-0.113	-0.010	-0.041	0.145
PRIN 5	0.622	0.876	-0.022	-0.018	0.254	0.291	-0.521	0.660	-0.314	-0.213
PRIN 6	0.420	0.928	0.015	-0.266	0.177	0.471	0.138	-0.059	0.603	-0.539
PRIN 7	0.374	0.975	0.050	-0.052	0.085	0.626	0.573	0.131	-0.299	0.400
PRIN 8	0.202	1.000	0.013	-0.076	-0.776	-0.053	0.175	0.555	0.222	-0.030

N= 427

Table A5. The eigenvalues and eigenvectors of the correlation matrix for Reference data only, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.841	0.355	-0.406	-0.043	0.473	-0.513	0.423	0.399	0.007	-0.090
PRIN 2	1.572	0.552	0.056	-0.072	0.147	-0.052	-0.139	0.098	0.686	0.683
PRIN 3	1.134	0.693	0.171	0.876	0.051	0.117	-0.064	0.400	0.110	-0.105
PRIN 4	0.728	0.784	0.747	-0.123	0.050	0.021	0.649	0.028	0.044	0.000
PRIN 5	0.584	0.857	0.194	-0.335	0.147	0.244	-0.307	0.678	-0.420	0.197
PRIN 6	0.484	0.918	0.153	-0.177	0.619	0.244	-0.283	-0.110	0.348	-0.537
PRIN 7	0.380	0.965	0.135	0.257	0.567	-0.106	-0.095	-0.444	-0.466	0.398
PRIN 8	0.278	1.000	-0.406	0.035	0.152	0.768	0.440	-0.030	-0.009	0.166

N=141

Table A6. The eigenvalues and eigenvectors of the correlation matrix for Test data only, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	2.495	0.312	-0.037	0.053	0.486	-0.463	0.404	0.513	0.336	0.075
PRIN 2	1.601	0.512	-0.101	-0.168	0.081	0.278	-0.396	0.055	0.525	0.665
PRIN 3	1.104	0.650	0.664	0.702	0.025	0.128	-0.138	0.139	0.035	0.100
PRIN 4	0.911	0.764	0.731	-0.642	0.068	-0.077	0.040	-0.120	0.132	-0.097
PRIN 5	0.649	0.845	-0.057	-0.120	0.593	0.395	-0.418	0.326	-0.209	-0.388
PRIN 6	0.463	0.903	-0.021	0.109	0.411	0.503	0.558	-0.483	0.139	0.027
PRIN 7	0.425	0.956	-0.075	0.081	-0.311	0.173	-0.022	0.080	0.703	-0.600
PRIN 8	0.353	1.000	0.057	-0.179	-0.368	0.497	0.414	0.595	-0.188	0.147

N=235

Table A7. The eigenvalues and eigenvectors of the correlation matrix for Impaired data only, all numeric variables except surface area, mean depth, and TN

			Eigenvector, coefficient of							
		Cumulative Percentage	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
Component	Eigenvalue	of Total Variance	LKMXDEP	OSGOOD	TP	SDT	COLOR	CHLA	ALK	PH
PRIN 1	3.564	0.445	-0.299	-0.008	0.482	-0.478	0.391	0.441	0.260	0.187
PRIN 2	1.655	0.652	0.247	0.413	-0.043	0.067	-0.261	-0.071	0.601	0.573
PRIN 3	1.010	0.779	-0.203	0.801	-0.092	-0.039	0.257	-0.136	0.072	-0.467
PRIN 4	0.800	0.879	0.781	0.108	0.269	0.230	0.251	0.382	-0.050	-0.206
PRIN 5	0.455	0.935	0.225	-0.299	-0.226	-0.133	0.644	-0.516	0.331	0.000
PRIN 6	0.255	0.967	-0.227	-0.269	0.261	0.457	-0.157	0.020	0.610	-0.450
PRIN 7	0.151	0.986	-0.310	0.074	-0.040	0.695	0.456	0.117	-0.176	0.402
PRIN 8	0.110	1.000	0.032	0.096	0.751	0.062	-0.052	-0.596	-0.227	0.108

N=95

APPENDIX B

EVALUATION OF "OUTLIER" LAKES IN NEW ENGLAND NUTRIENT DATABASE

ENSR INTERNATIONAL

Waterbody Name	Comments
AMHERST LAKE, VT	SDT is low even though lake is relatively deep
BABCOCK POND, CT	SDT low due to shallownesss of lake - most probably hit bottom
BEACHDALE POND, CT	CHLA is low. However only 1 measurement.
BRADBURY LAKE, ME	
BRYANT POND, ME	CHLA at limit of meso-eutrophic.
COVENTRY RESERVOIR, RI	SDT at limit of meso-eutrophic. Relatively shallow reservoir.
CRANBERRY PO, ME	SDT low due to shallownesss of lake - most probably hit bottom
DOG POND, CT	SDT low due to shallownesss of lake - most probably hit bottom. TP value elevated while CHLA is low
EAST CARRY POND, ME	
ECHO LAKE (PLYMOUTH), VT	SDT at limit of meso-eutrophic.
FLINT POND (SHREWSBURY), MA	
FOREST LAKE (CHESHIRE), NH	
FORSTER POND, CT	
FOURTH MACHI, ME	
GARDNER LAKE, CT	
HALLS POND, ME	
HARRISVILLE POND, NH	
HEALD POND (SOMERSET), ME	
HODGDON POND, ME	
HOWELL POND, CT	SDT low due to shallownesss of lake - most probably hit bottom. TP value elevated while CHLA is low
ISLAND POND, VT	
JEWETT POND (OXFORD), ME	
LAKE QUASSAPAUG, CT	
LAKE QUINSIGAMOND, MA	
LAKE WILLOUGHBY, VT	
LANTERN HILL POND, CT	
MANSFIELD LAKE, MA	
MESSERSCHMIDTS POND, CT	
MONO POND, CT	
PARTRIDGE LAKE, NH	
PICKEREL LAKE, CT	SDT low due to shallownesss of lake - most probably hit bottom. TP value elevated while CHLA is low
PLUNKETT POND, ME	
POST POND, NH	
ROUND POND (FRANKLIN-E), ME	SDT low due to shallownesss of lake - most probably hit bottom. TP value elevated while CHLA is low
ROUND POND, ME	
SHOWELL POND, NH	
TRIPP POND, ME	
TWIN ISLAND POND, ME	SDT low due to shallownesss of lake - most probably hit bottom. TP value elevated while CHLA is low
WENTWORTH POND, ME	
WOOD CREEK POND, CT	SDT low due to shallownesss of lake - most probably hit bottom. TP value elevated while CHLA is low

Table B-1. List of NE Nutrient Database Outlier Lakes and Additional Lakes with Mixed Trophic Indicators.

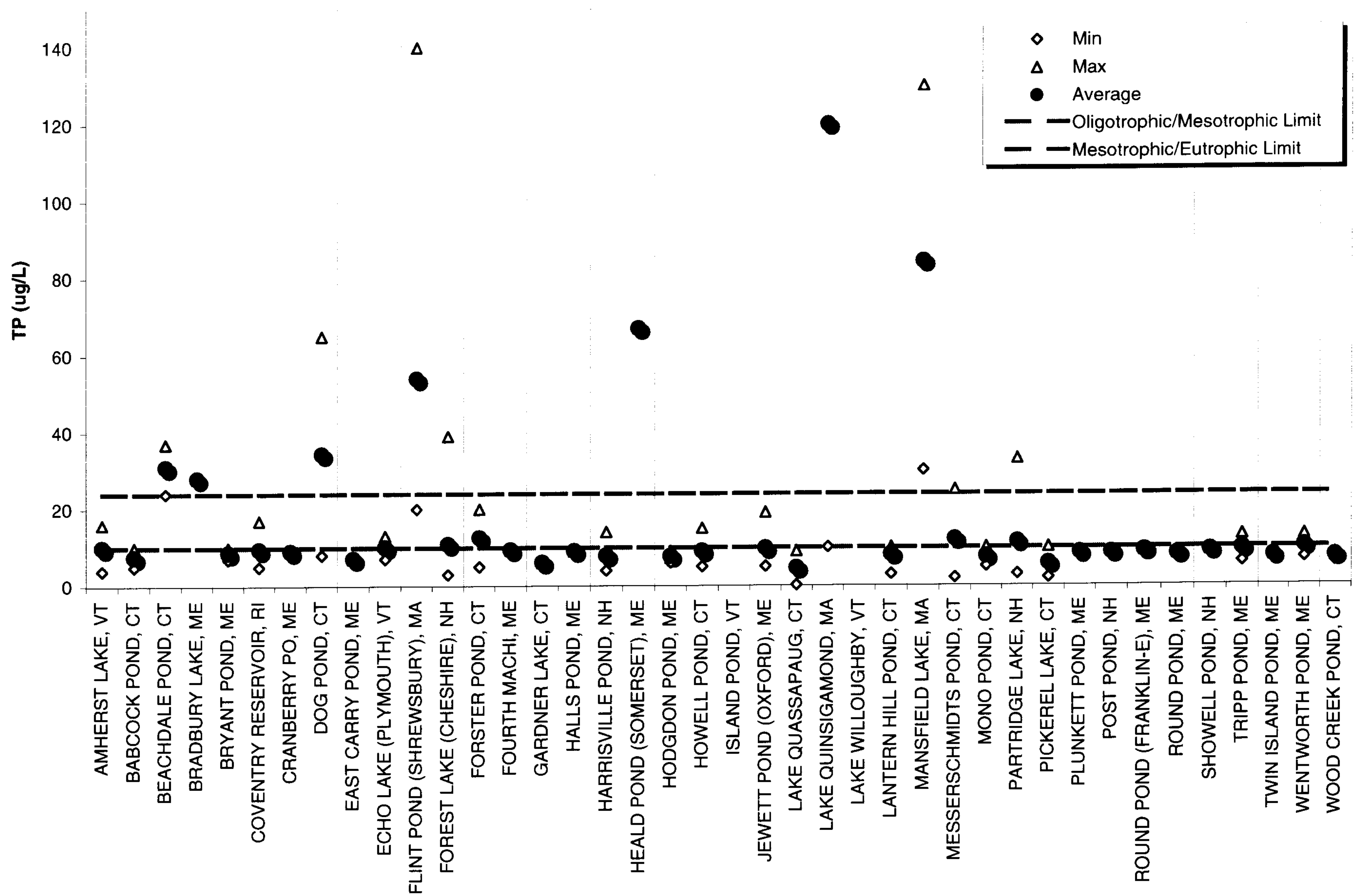


Figure B-1. **Statistical Ranges of Outlier and Additional Lakes with Mixed Trophic Indicators – TP.**

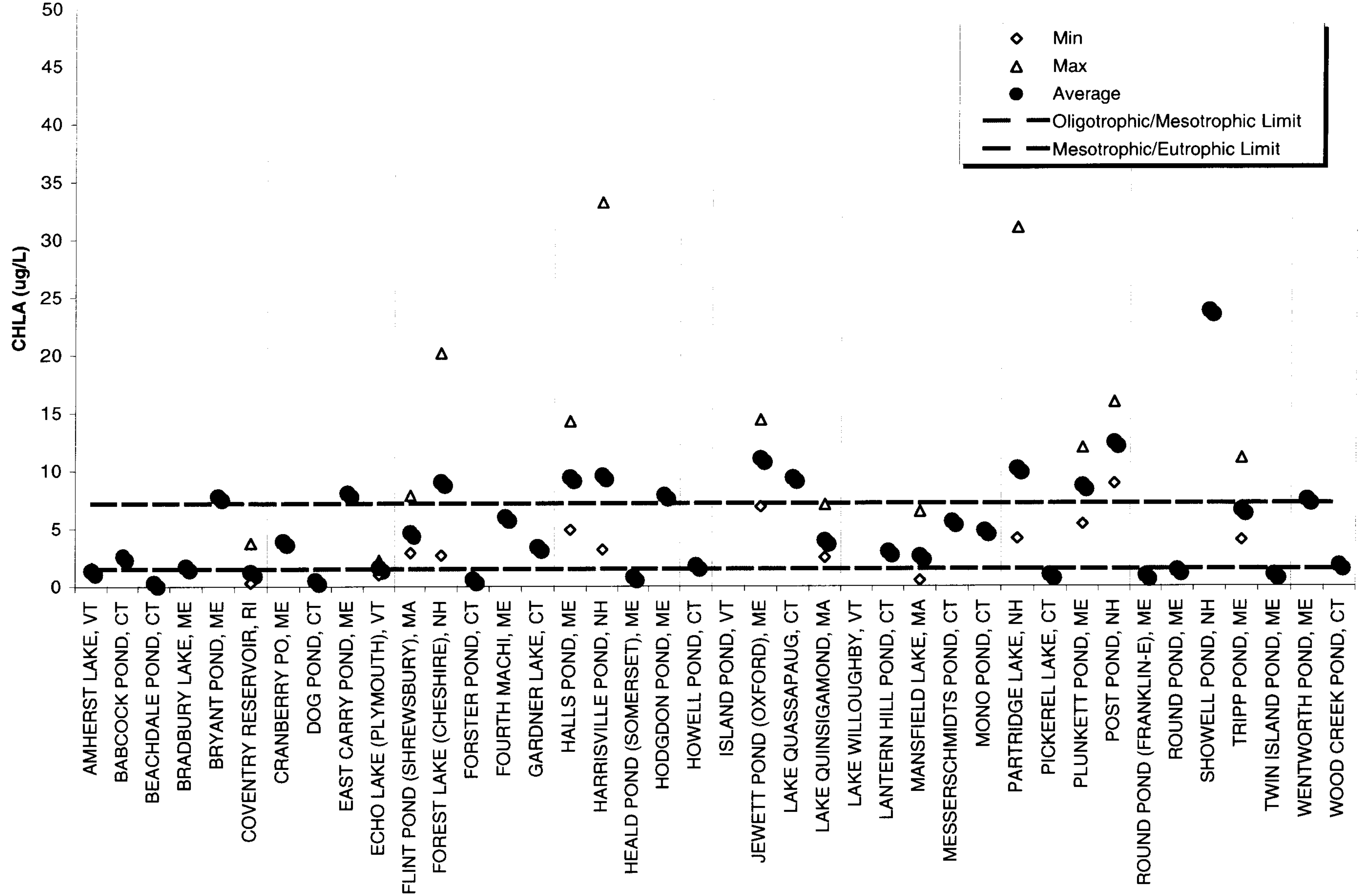


Figure B-2. **Statistical Ranges of Outlier and Additional Lakes with Mixed Trophic Indicators – Chl a.**

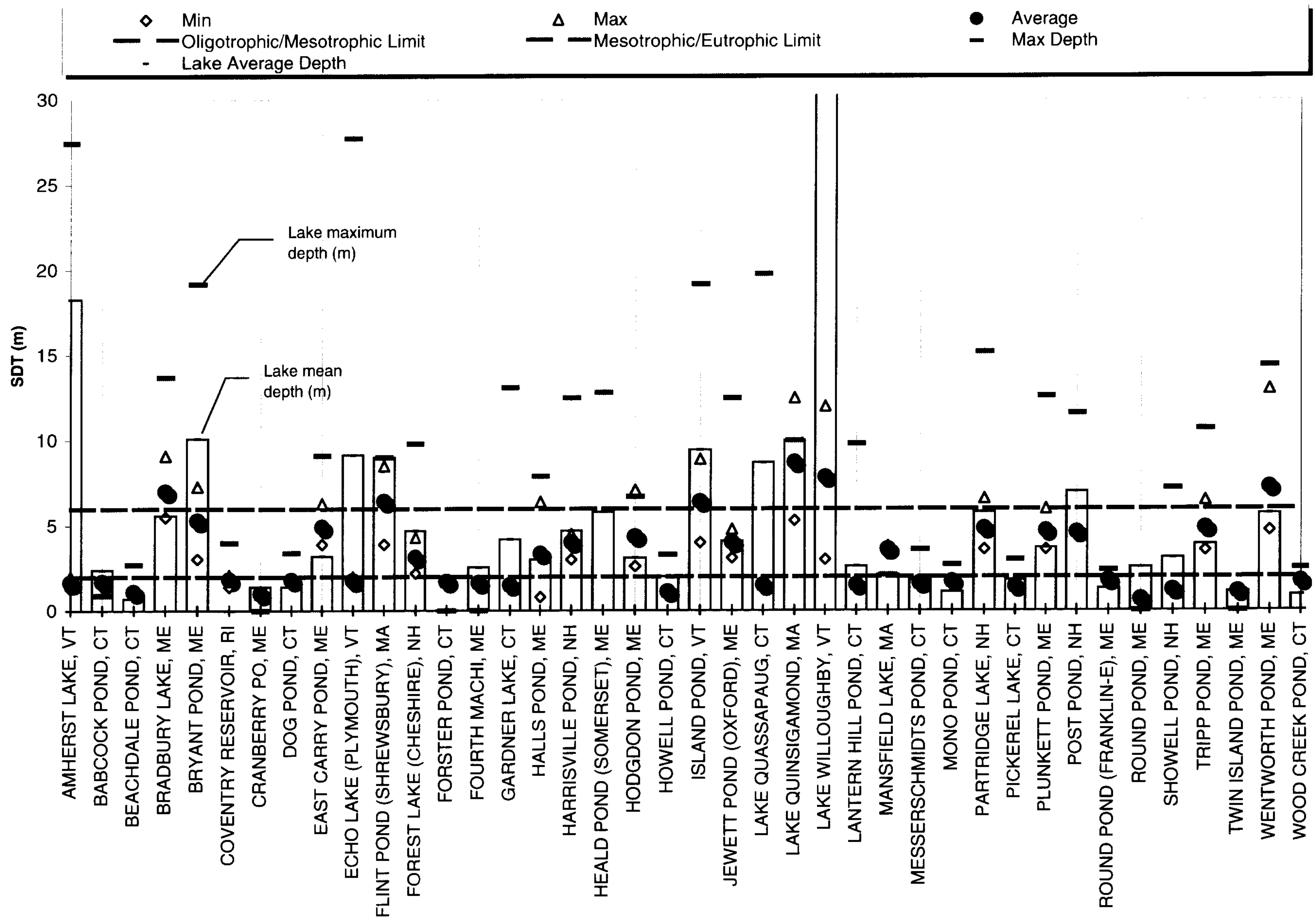


Figure B-3. Statistical Ranges of Outlier and Additional Lakes with Mixed Trophic Indicators – SDT.

APPENDIX C

DRINKING WATER IMPAIRMENT - NUTRIENT LEVELS

ENSR INTERNATIONAL

ENSR

Memorandum

To:	Matt Liebman, Beth Card	**Date:**	September 10, 2000
From:	David F. Mitchell	**File:**	4933001-510
RE:	Drinking Water Impairment - Nutrient Levels	**CC:**	File Isabelle Morin

This memorandum reviews information gathered by ENSR regarding the potential for determining nutrient levels (i.e., nitrogen, phosphorus) that are likely to be protective of a sensitive designated use – public drinking water supplies. This information is part of an overall attempt to provide a closer linkage between nutrient levels and impairment of specific designated uses, as requested by the RTAG following review and comment on the Final Report: *"Collection and Evaluation of Ambient Nutrient Data for Lakes, Ponds, and Reservoirs in New England."*

ENSR contacted a number of drinking water facilities in New England to evaluate information used by these facilities regarding at what threshold levels algal community composition or numbers could potential impair drinking water production and/or quality. Two thresholds were considered – the first was connected with aesthetic taste and odor problems and the second levels was associated with filtration difficulties due to clogging of the filters by algal biomass. It was hoped that problematic algal cell counts or concentrations could be correlated with waterbody chlorophyll a concentrations and/or phosphorus concentrations.

Discussions were had with Ted Kenney, Deputy Executive Director, New England Water Works Association; Larry Pistrang, MWRA; and Bob Hoyt, Water Quality Director, City of Worcester Filtration Works. This is not a definitive list of contacts, but provided some insight as to the likely availability of information from similar facilities around New England. From these contacts it was established that:

- chlorophyll a is not routinely measured at intakes of drinking water facilities, since it is only indirectly related to critical levels of algal species that drive drinking water quality treatment decisions (see below);
- total phosphorus was monitored in tributaries of some reservoirs as part of watershed assessment, but the method detection limits (ND = < 100 ug/L) are not sufficient sensitive to distinguish trophic states;
- cell counts or functional algal standard units (ASU) were regularly collected. These were the based for decision-making regarding treatment to control algal populations;
- an alternative means of measuring biomass is via chemical doses (chlorination) required to oxidize levels of total organic material as a pre-disinfectant. The higher the phytoplankton bloom, the greater the chlorination dose used. Records of dose rates are indirectly indicative of seasonal patterns of runoff and bloom formation.

At Wachusett Reservoir, weekly (or more frequent during bloom conditions) sampling at six depths are used to determine exactly what algal genera are present and how their populations are changing. Relating historic data to complaint information in a somewhat arbitrary manner has led to the development of some "thresholds". Treatment of the reservoir with copper sulfate sometimes occurs if one or more of these threshold is exceeded.

Synura was indicated as the most troublesome genus for the MWRA Wachusett facility, primarily because it causes taste and odor problems at very low concentrations. Treatment normally was conducted when *Synura* concentrations exceed 20 ASU/mL (2 colonies per mL). The diatom *Asterionella*, on the other hand, while also reported in the literature as a taste and odor genus, could be present in concentrations of 1000-1500 ASU/mL (1200-1800 cells per mL) without causing problems. Some of the treatment thresholds reported by MWRA

Synura: treat at 20 ASU (2 colonies per mL); [average colony - 12 ASU; colony range 3-50 ASU]

Anabaena: treat at 50 ASU (250 cells per mL); [during exponential growth phase, average cell - 0.2 ASU]

Asterionella: treat at 1000 ASU (1200 cells per mL); [usually not done, average cell - 0.8 ASU]

Dinobryon: treat at 300 ASU (150 cells per mL); [average cell + lorica - 2 ASU]

Uroglena: treat at 300-400 ASU (3 colonies per mL); [average colony size - 120 ASU, range 6-350 ASU]

It should be obvious from this comparison that any use of a simple chlorophyll number is not likely to be appropriate since both the species composition and relative biomass are critical. Accordingly, drinking water facilities are unlikely to be able to directly correlate chlorophyll a concentrations with direct impairment.

The preliminary results of this survey were discussed with NEIWPCC and EPA during the July 15, 2000 project status meeting. The overall focus of the Nutrient Criteria project is on potential impairment as a L/P/R waterbody shifts from mesotrophic to eutrophic conditions; whereas drinking water operational difficulties are associated with less measurable levels of eutrophication that may be undistinguishable with a reference waterbody. Accordingly, it was agreed not to pursue this area of investigation further at this point.

APPENDIX D

COMPARISON OF NUTRIENT DATABASE WITH ROHM AND GRIFFIN SUB-ECOREGIONS

ENSR INTERNATIONAL

ENSR

Memorandum

To:	Matt Liebman (EPA), Beth Card (NEIWPCC)	**Date:**	September 8, 2000
From:	Isabelle Morin, Dave Mitchell (ENSR)	**Project:**	4933-001
RE:	Comparison of Nutrient Database data with Rohm and Griffith sub-ecoregions	**CC:**	

Introduction

Ecoregions have been proposed by various researchers as a way to define geographical areas that have relatively similar characteristics and observed nutrient levels within their waterbodies. Omernik (1987) divided the six New England states into four such geographical areas: the Atlantic Coastal Pine Barrens, Atlantic Coastal Zone, New England Highlands, and Laurentian Plains and Hills. The data collected for the Nutrient Database were evaluated to determine whether these ecoregions were good indicators of observed nutrient levels. The results of the evaluation were presented in the Data Summary Report (ENSR, 2000). The evaluation showed significant difference between the "lower coastal" and "inland" regions both in terms of the nutrient concentrations typically observed and the characteristics of the lakes (depth, color, alkalinity, etc.), with shallower more colored lakes in southern Massachusetts, Rhode Island and Connecticut showing higher nutrient concentrations than deeper, clearer lakes of northern Maine, New Hampshire and Vermont.

Additional sub-divisions have been proposed by Rohm (1995) and Griffith (XX) for New England and Massachusetts, respectively. The present memorandum discusses the findings of a comparison ranges of total phosphorus concentration reported in the literature for these sub-ecoregions with those found in the Nutrient Database..

Rohm Sub-Ecoregions of New England

Rohm (1995) subdivided Northeastern United States into 61 sub-ecoregions. The sub-ecoregions were defined based on analysis of the soil types, land uses, and measured Total Phosphorus (TP) concentration for a population of 2,893 lakes larger than 1 hectare. The sub-ecoregions cover 8 phosphorus classes defined by range of TP concentrations.

ArcView, a Geographic Information System (GIS) software was used to combine the spatial coverage of Rohm's sub-ecoregions with the coordinates of the lakes, ponds and reservoirs in the Nutrient Database, and assign to each waterbody in the database its appropriate Rohm sub-ecoregion.

The corresponding phosphorus class was then determined from a table of the sub-ecoregion characteristics. Note that no sub-ecoregion of Class 1 (TP < 5ug/L) was found in the six New England state considered.

The sub-ecoregions located within the six EPA Region 1 states are listed in Table 1. The table provides, for each sub-ecoregion, its identifier, the states where it is located, Rohm phosphorus class, number of waterbodies from the Nutrient Database, and general comments on soil, land use or lake characteristics.

Table 1: Rohm Sub-Ecoregion in New England States.

Region ID	States	Phosphorus Class	Waterbodies in Nutrient Database	Comments
5801	ME	7	5	Land use dominated by agriculture, moderate population density. Enriched soils
5802	ME	3	13	Relatively low values. Lakes typically small in size, and mostly shallow.
5803	ME	4	12	Wide range of phosphorus values. Low topographic relief. Extensive agriculture.
5804	ME	3	73	Diversity of land use. Moderate population density.
5805	ME, NH	2	147	TP low but wide range of value.
5806	ME, NH	2	76	Phosphorus values consistently low.
5808	ME	2	66	
5809	ME	3	10	Most lakes are small and shallow.
5810	NH, VT	3	5	Relatively low TP
5811	NH, VT	6	5	High TP probably due to urbanization and agriculture
5812		2	31	Deep lakes with very low TP. Steep-sided low mountains.
5813	NH, MA	4	31	Bi-modal distribution of phosphorus.
5815	VT, NH, MA, CT	3	41	Most lakes small but deep.
5816	VT	2	14	High fraction of lakes with TP<10 ug/L. Most lakes are small but deep.
5829	VT, MA	3	3	
5901	ME	3	17	Lakes typically shallow. Predominantly hilly with low relief.
5902	ME, NH	5	123	TP highly variable.
5903	ME	4	27	Larger and deeper lakes.
5904	MA, NH, ME	7	34	
5905	MA, NH	3	5	
5906	MA	8	8	
5908	MA, RI	6	9	
5909	MA, RI	8	19	Glacial deposition surface. High degree of development
5910	MA, CT, RI	3	96	TP values range widely. Range attributed to amount of watershed development and lake origin and depth. Most lakes are small and shallow with many human-made.
5911	NH, VT, MA, CT	8	8	
5912	CT	5	7	Fairly high TP values. Natural lakes are uncommon.
5913		2	13	
6001	VT	6	6	
6003	VT	4	3	

TP Measurements in Rohm Sub-Ecoregions

The distribution of TP measurements from the Nutrient Database is shown in Figure 1. The TP measurements are grouped by the Rohm phosphorus class (on the x-axis) corresponding to the sub-ecoregion where the waterbody is located. Range of TP concentration as defined by Rohm for each phosphorus class are given in parenthesis.

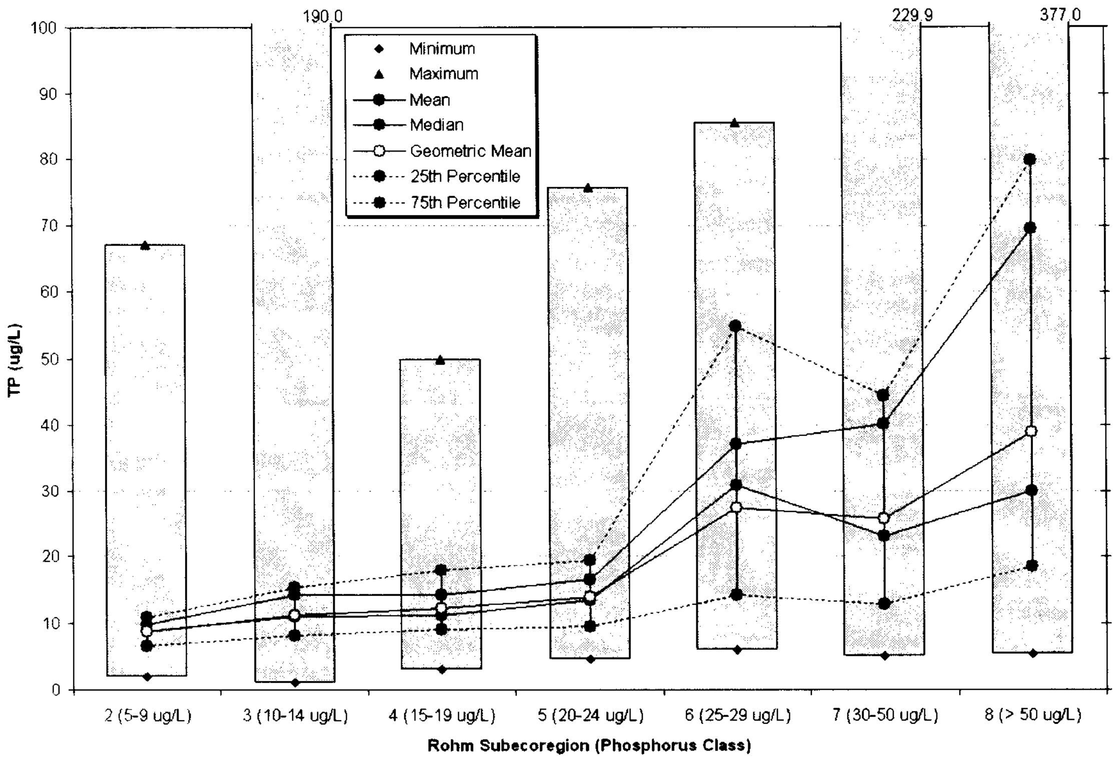


Figure 1: Comparison of phosphorus classes defined by Rohm with distribution of TP measurements from the Nutrient Database.

Overall, we can note a good agreement between the indicators of central tendencies calculated from the Nutrient Database waterbodies and the phosphorus classes defined by Rohm for the sub-ecoregions were the waterbodies are located. The maximum and indicators of central tendencies of TP in the Nutrient Database vary from class to class, with a tendency for mean TP value to increase as expected from class 2 through class 8. Table 2 presents a comparison of the arithmetic and geometric means of the TP measurements in the Nutrient Database with ranges defined for each sub-ecoregion phosphorus class.

Table 2: Comparison of geometric mean of Nutrient Database TP measurements with Rohm phosphorus classes.

Class	Rohm phosphorus range (ug/L)	Arithmetic/geometric mean of TP in Nutrient Database
2	5-9	8.7 / 9.8
3	10-14	11.2 / 14.2
4	15-19	12.2 / 14.2
5	20-24	14.0 / 16.5
6	25-29	27.4 / 37.1
7	30-50	25.6 / 40.2
8	> 50	39.0 / 69.5

TP measurements from the Nutrient Database compare favorably with expected phosphorus range, with the expection of phosphorus classes 4 and 5 where the measurements are lower than expected. For the other classes, either the arithmetic or the geometric mean, or both, are within the range expected from Rohm's definition of the sub-ecoregions.

Figure 2 shows the distribution of TP measurements for individual sub-ecoregions. On that figure, the box represents the range of TP as defined by Rohm for the various sub-ecoregions (phosphorus class).

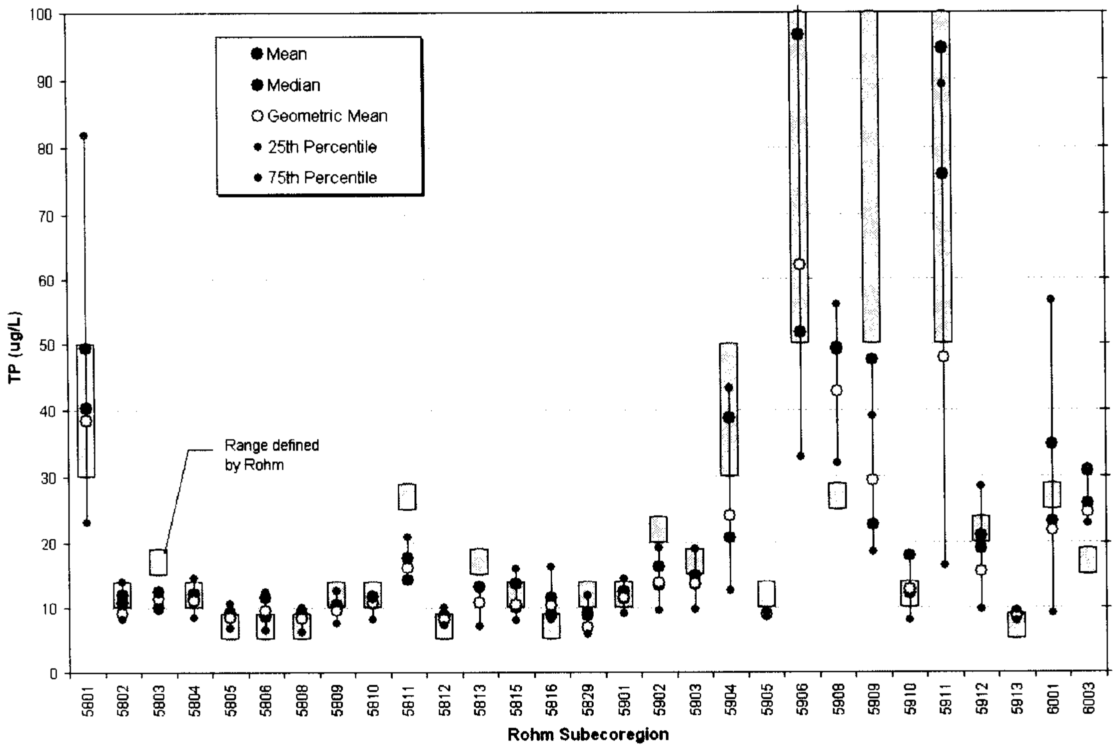


Figure 2: Comparison of phosphorus data from Nutrients Database with range defined by Rohm for sub-ecoregions of New England.

Griffith Sub-Ecoregions of Massachusetts

Griffith has defined 5 sub-ecoregions in Massachusetts. The sub-ecoregions were sub-divided based on observed TP measurements, land use, and soil types. Coordinates of the 60 lakes, ponds, and reservoirs in Massachusetts were used to assign the corresponding Griffith sub-ecoregion.

TP Measurements in Griffith Sub-Ecoregions

Figure 3 shows the distribution and statistical indicators of TP measurements in Massachusetts lakes, ponds and reservoirs for each of the 5 Griffith sub-ecoregion class. The figure shows noticeable differences between the Griffith sub-ecoregions with regards to the range and central tendency of TP measured. Although the population of waterbodies in the Nutrient Database is relatively small for Massachusetts (only 60 with known geographical coordinates), the figure shows a general trend towards increased TP concentration as the sub-ecoregion increases. A copy of the original paper by Griffith could not be located, and the expected phosphorus concentration range could therefore not be used for comparison.

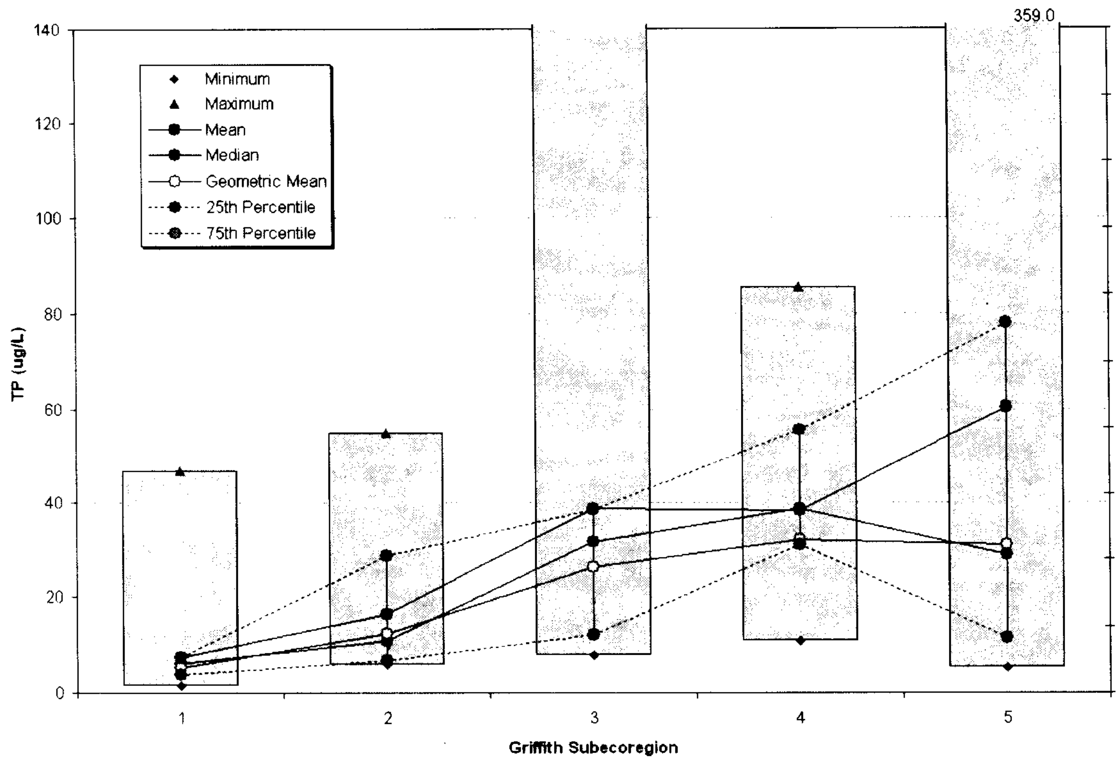


Figure 3: Comparison of phosphorus classes defined by Griffith with distribution of TP measurements in Massachusetts waterbodies from the Nutrient Database.

References

ENSR (2000). Data Synthesis Report, Interim Final, report submitted to the New England Interstate Pollution Prevention Commission, EPA Region 1 and the Nutrient Assessment Team, April, 87 pp.

Griffith (XX).

Omernik, J.M. (1987). Ecoregions of the Conterminous United States. Annals of the Association of American Geographers. 77(1), pp. 118-125.

Rohm, C.M., Omernik, J.M., and Kiilsgaard, C.W. (1995). Regional Patters of Total Phosphorus in Lakes of the Northeastern United States, Lakes and Reservoir Management, 11(1), pp. 1-14.