1. **Submission Date of Final Report to LISS:** 11/30/2009

2. **EPA Grant Number and Project Title:** LI-97263606, Development of a Long Island Sound-Specific Water Quality Index Using Cluster Analysis and Discriminant Analysis

3. **Grantee Organization and Contact Name:** City College of New York, Pengfei Zhang

4. **Public Summary:**

Two types of water quality indices (WQI) were developed, one based on multivariate cluster analysis and discriminant analysis (Discriminant-WQI) of four water quality parameters (Chlorophyll *a* (Chl-a), total dissolved nitrogen (TDN), total dissolved phosphorus (TDP), and dissolved oxygen (DO)), and the other based on a simple linear combination (Linear-WQI) of the four water quality parameters. Color-coded contour maps of the water quality indices of the entire Sound in every month from 1995 to 2006 were generated to provide an easy overview of the water quality of the Sound over time. While the Discriminant-WQI uses a weighted combination of the four individual water quality parameters and was thought to be superior to the simple combination of the parameters, the Linear-WQI actually provided a more consistent overview of the water quality of the Sound.

The impact of the anthropogenic nitrogen loadings to the Sound on water quality was also examined. Point source and non-point source (NPS) nitrogen loads in the past 26 years from 11 areas along the coast of the Sound were obtained from CTDEP and analyzed. Time series data for total nitrogen load, TDN, TDP, Chl-a, and DO were deseasonalized and the long-term trend and cyclical variations around the trend line were assessed. The cross-correlations between total nitrogen (TN) in LIS water and total nitrogen loadings to the Sound, as well as the cross-correlations among the four individual water quality parameters were examined.

There is no clear trend for the total nitrogen loads from the two major riverine (non-point) sources except for some long-term cyclical variations. In contrast, the total nitrogen loads from areas dominated by sewage treatment plants (STPs) showed a slight decrease trend between 1991 and 1996, a significant drop from 1997 to 1999 (likely as a result of much improved nitrogen removal efficiencies), and a rebound after 2003 probably due to the relaxation of regulations for further facility upgrade.

Chl-a levels at all sampling stations decreased gradually from 1991 to 1999, rebounded quickly around 2000-2002, and then leveled off thereafter. TDN concentrations showed a similar long-term trend. These long-term trends for Chl-a and TDN appeared to reflect the trend of the total nitrogen loads from the major point N sources (major STPs) to the Sound, indicating that the reduction of nitrogen at STPs indeed had a positive impact on LIS water quality. Interestingly, TDP concentrations increased steadily from 1991 to about 2002, dropped in the following year, and then leveled off thereafter. DO levels and minimum annual DO did not show any long-term trends (e.g., improvements). Cross-correlation analysis confirmed that TN concentrations in LIS water would change to a certain degree in response to the total N loadings to the Sound, with a lag time of about 9-10 months.

The maximum cross-correlations between DO and TDN ranges from -0.22 to -0.52 at various stations (average -0.39 for all stations), with the corresponding lag time ranges from 7 months to 10 months (average of 8.6 months). In other words, in most cases DO reaches its lowest level about 8.6 months after the TDN peaks at the same sampling station. The cross-correlations between DO and TDP are almost always higher than the cross-correlations

between DO and TDN for any given sampling station (average -0.521 for all stations), suggesting that TDP may play a more important role on DO concentrations than TDN. Likewise, the cross-correlations between Chl-a and TDP (average 0.27 for all stations) are also higher than the cross-correlations between Chl-a and TDN in most cases (average 0.22 for all stations), further suggesting the importance of TDP on algal blooms.

5. **Project Period:** 9/1/06 – 8/31/09

## 6. Project Description:

The objectives of this project were to develop a Long Island Sound (LIS) specific water quality index that would reflect the trophic status of LIS water, and to examine the impact of human induced nitrogen input on LIS water quality.

The new water quality index was computed using multivariate cluster analysis and discriminant analysis of a set of individual water quality indicators. A numerical water quality index would result, with a value close to 1 indicating good water quality (oligotrophic), a value close to -1 indicating poor water quality (eutrophic), and a slight negative value representing mesotrophic conditions. The new method was compared with the established ASSETS methodology, and was applied to LIS water quality data (1995-2006, at ~17 stations) collected by the Connecticut Department of Environmental Protection (CTDEP). A second water quality index based on the simple linear combination of the same set of individual water quality parameters was also developed. Monthly water quality indices were computed for every station and contour maps of the entire Sound were generated. The relationships between water quality parameters and total nitrogen loads were examined to assess the impact of the anthropogenic nutrient input on LIS water quality. Templates were developed to facilitate the computation of the indices.

Color-coded contour maps (monthly, 1995-2006) of the water quality indices would provide a convenient overview of the water quality conditions of the Sound. Temporal trends in water quality of the entire Sound could be easily examined. The relationships between the anthropogenic nutrient input and LIS water quality would shed light on the effectiveness of the managed nitrogen reduction effort.

## 7. Activities & Accomplishments:

### a. computing water quality indices

Two types of water quality indices (WQI) were developed. The first type was based on multivariate cluster analysis and discriminant analysis of a set of individual water quality parameters (Chlorophyll *a* (Chl-a), total dissolved nitrogen (TDN), total dissolved phosphorus (TDP), and dissolved oxygen (DO)), as originally proposed. Detailed methods of computing this type of water quality index (referred to as Discriminant-WQI) is provided in Section 9 below. The second type of water quality index was computed by a simple linear combination of the four water quality parameters (referred to as Linear-WQI, see Section 9 for details). Templates were developed to facilitate the computation of the indices. Color-coded contour maps of the monthly water indices were generated for the years of 1995 to 2006 (~550 contour maps, listed in Appendix 1). To facilitate the assessment of water quality of the entire Sound, color-coded contour maps of the four individual water quality parameters (Chl-a, TDN, TDP, and DO) for the same period (1995-2006) were also generated (~1100 contour maps, listed in Appendix 2). A huge amount of time was spent to generate these contour maps.

b. examining the impact of human induced nitrogen input on LIS water quality

Point source and non-point source (NPS) nitrogen loads in the past 26 years from 11 areas along the coast of the Sound were obtained from CTDEP and analyzed. Time series data for total nitrogen load, TDN, TDP, Chl-a, and DO were deseasonalized and the long-term trend and cyclical variations around the trend line (Appendix 3) were assessed. Seasonal indices representing the effect of each season on these parameters were also calculated (Appendix 3). The cross-correlations between total nitrogen (TN) in LIS water and total nitrogen loads to the Sound, as well as the cross-correlations among the four individual water quality parameters were examined.

c. training

We have trained colleagues at CTDEP (Matthew Lyman and three other staff members) last fall on the calculation of the water quality index using a template.

8. **Modeling:** N/A.

9. **Summary of Findings:**

a. Water quality indices

The first type of water quality index was developed using cluster analysis and discriminant analysis. Water quality data (April and September 2004, Table 1) from various sampling stations in LIS (Figure 1) were used to illustrate this approach.

The cluster analysis procedure is designed to assemble observations into relatively homogeneous groups (the so-called "clusters") based on the similarities between observations. A commonly used measure of similarity between objects is a standardized *m*-space Euclidean distance, $d_{ij}$, computed as:

$$d_{ij} = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2} \qquad (1)$$

where $x_{ik}$ is the $k^{th}$ variable measured on object *i* and $x_{jk}$ is the $k^{th}$ variable measured on object *j*, and *m* is the number of variables (*1*). In the case of LIS water quality data, the objects are different sampling stations, and the variables are measured parameters (Chl-a, TDN, TDP, and DO). To ensure each variable is weighted equally during the Euclidean distance calculations, each element in the $n \times m$ (*n* is the number of objects, or sampling stations in this case) raw data matrix is standardized by subtracting the column means and dividing by the column standard deviations prior to computing the Euclidean distances (*1*).

The most reasonable strategy to produce clusters from a similarity matrix is the group average method (*2*). This method begins by placing each item into a separate cluster, and then joining clusters based on the average distance between all members of one cluster and all members of the other. This process continues until the desired number of clusters is formed, and the result is displayed as a dendrogram, shown in Figure 2.

Table 1. Water quality data collected during April and September 2004 by CTDEP.

| Station Name | Depth Code | Sampling Date | Chl-a (μg/L) | TDN (mg/L) | TDP (mg/L) | DO (mg/L) | Overall Rating | ***WQI*** |
|---|---|---|---|---|---|---|---|---|
| | | | | April | | | | |
| 09 | B | 4/14/2004 | 6.200 | 0.285 | 0.021 | 10.89 | Fair | **-0.44** |
| 15 | B | 4/12/2004 | 5.900 | 0.086 | 0.056 | 10.81 | Fair | **0.19** |
| **A4** | B | 4/14/2004 | 13.400 | 0.287 | 0.040 | 10.94 | Fair | **-1.41** |
| **B3** | B | 4/14/2004 | 10.600 | 0.207 | 0.025 | 10.80 | Fair | **-0.71** |
| **C1** | B | 4/14/2004 | 11.100 | 0.229 | 0.037 | 10.61 | Fair | **-0.63** |
| C2 | B | 4/14/2004 | 5.100 | 0.165 | 0.037 | 10.80 | Fair | **0.10** |
| D3 | B | 4/14/2004 | 6.800 | 0.201 | 0.020 | 10.55 | Fair | **0.05** |
| E1 | B | 4/12/2004 | 1.800 | 0.108 | 0.049 | 10.56 | Fair | **0.91** |
| F2 | B | 4/12/2004 | 4.800 | 0.096 | 0.032 | 10.64 | Good | **0.48** |
| F3 | B | 4/12/2004 | 1.500 | 0.198 | 0.045 | 10.46 | Fair | **0.83** |
| H2 | B | 4/7/2004 | 2.700 | 0.095 | 0.038 | 10.82 | Good | **0.56** |
| H4 | B | 4/12/2004 | 0.400 | 0.150 | 0.038 | 10.45 | Fair | **1.10** |
| H6 | B | 4/12/2004 | 0.800 | 0.175 | 0.030 | 10.40 | Fair | **1.04** |
| I2 | B | 4/7/2004 | 1.800 | 0.203 | 0.045 | 10.89 | Fair | **0.33** |
| J2 | B | 4/7/2004 | 1.700 | 0.217 | 0.043 | 10.81 | Fair | **0.39** |
| K2 | B | 4/7/2004 | 1.380 | 0.210 | 0.028 | 10.47 | Fair | **0.80** |
| M3 | B | 4/7/2004 | 1.300 | 0.165 | 0.057 | 10.56 | Fair | **0.83** |
| | | | | September | | | | |
| 03 | B | 9/3/2004 | 1.724 | 0.084 | 0.079 | 2.83 | Fair | **-0.44** |
| 04 | B | 9/3/2004 | 3.160 | 0.107 | 0.092 | 3.01 | Fair | **-0.23** |
| 07 | B | 9/3/2004 | 4.265 | | 0.068 | 5.66 | | |
| 09 | B | 9/2/2004 | 3.934 | 0.110 | 0.078 | 4.76 | Fair | **0.46** |
| 15 | B | 9/1/2004 | 3.934 | 0.090 | 0.074 | 6.48 | Fair | **1.05** |
| **A4** | B | 9/2/2004 | 4.376 | 0.235 | 0.150 | 0.98 | Poor | **-1.00** |
| **B3** | B | 9/2/2004 | 3.470 | 0.164 | 0.115 | 1.20 | Poor | **-0.89** |
| **C1** | B | 9/2/2004 | 2.144 | 0.164 | 0.124 | 1.19 | Poor | **-1.10** |
| C2 | B | 9/3/2004 | 1.746 | 0.093 | 0.080 | 2.74 | Fair | **-0.48** |
| D3 | B | 9/3/2004 | 1.856 | 0.072 | 0.082 | 3.19 | Fair | **-0.29** |
| E1 | B | 9/1/2004 | 1.326 | 0.117 | 0.075 | 3.31 | Fair | **-0.40** |
| F2 | B | 9/1/2004 | 2.873 | 0.104 | 0.089 | 2.99 | Fair | **-0.27** |
| F3 | B | 9/1/2004 | 1.414 | 0.122 | 0.082 | 3.68 | Fair | **-0.29** |
| H2 | B | 8/31/2004 | 1.414 | 0.073 | 0.078 | 3.72 | Fair | **-0.18** |
| H4 | B | 9/1/2004 | 3.227 | 0.091 | 0.067 | | | |
| H6 | B | 9/1/2004 | 1.436 | 0.163 | 0.071 | 4.16 | Fair | **-0.20** |
| I2 | B | 8/31/2004 | 3.956 | 0.092 | 0.047 | 6.44 | Good | **1.07** |
| J2 | B | 8/31/2004 | 3.381 | 0.069 | 0.049 | 6.84 | Good | **1.15** |
| K2 | B | 8/31/2004 | 4.022 | 0.059 | 0.042 | 7.03 | Good | **1.33** |
| M3 | B | 8/31/2004 | 2.099 | 0.109 | 0.044 | 7.01 | Fair | **0.95** |

Chl-a: Chlorophyll a; TDN: total dissolved nitrogen; TDP: total dissolved phosphorus; DO: dissolved oxygen; *WQI*: water quality index computed using cluster analysis and discriminant analysis. Color code: red-poor, yellow-fair, and green-good.

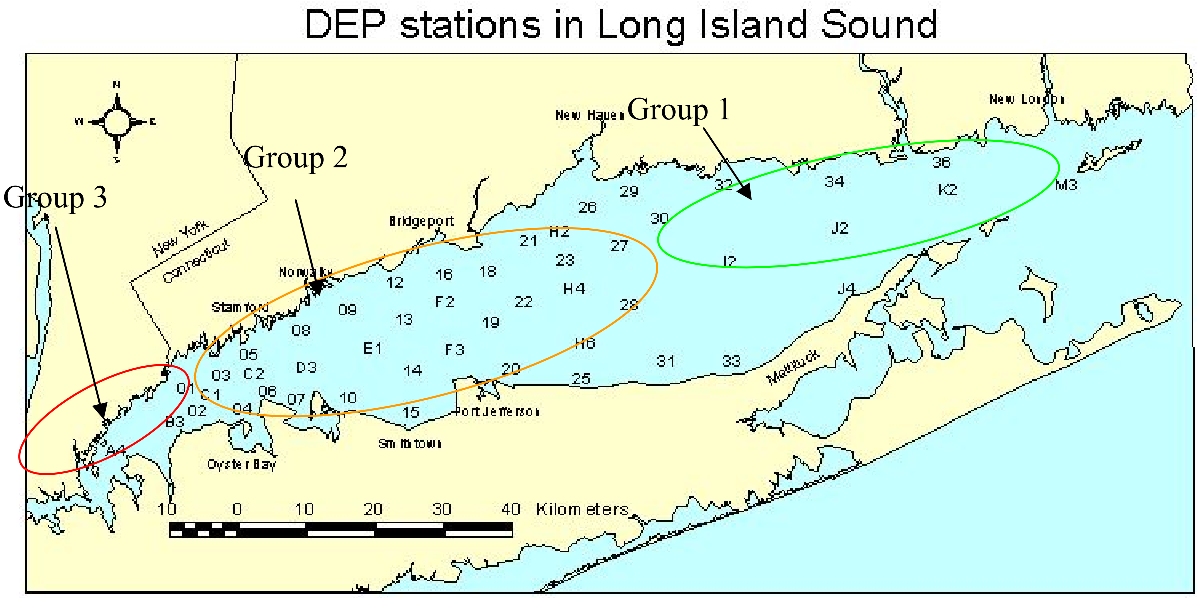


Figure 1. Connecticut DEP sampling stations in Long Island Sound.

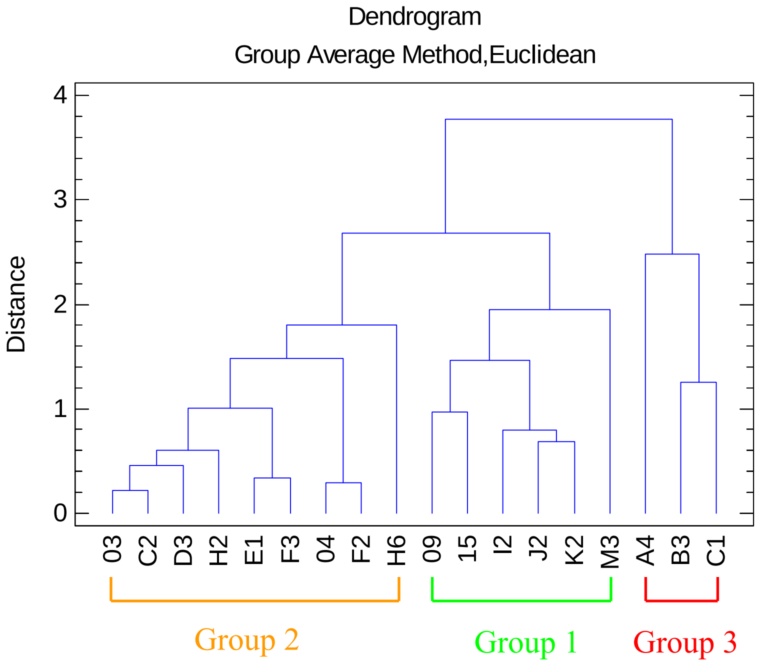


Figure 2. Dendrogram developed using September data in Table 1.

To illustrate the cluster analysis, the September 2004 water quality data (Chl-a, TDN, TDP, and DO, Table 1) from LIS were used to generate the dendrogram (Figure 2). The 18 sampling stations (see Figure 1 for locations) were assembled into 3 groups based on the cluster analysis (Figure 2). Group 1 consists of stations 09, 15, I2, J2, K2, and M3, which have the lowest TDN and TDP, and the highest DO (5-7 mg/L) among all stations. Most of the Group 1 stations are located at the east part of LIS (with lowest anthropogenic nutrient input) and the water is oligotrophic. Group 3 includes stations A4, B3 and C1, which have the highest Chl-a, TDN, and TDP, and the lowest DO (< 2 mg/L). This group is located at the west end of LIS (with highest anthropogenic nutrient input) and the water is clearly eutrophic. Group 2 includes stations 03, C2, D3, H2, E1, F3, 04, F2, and H6, which have moderate TDN, TDP, and DO (3-5 mg/L). These stations are located at the middle part of LIS and the water represents mesotrophic conditions.

Once the observations are assembled to different groups, the discriminant analysis can be applied to distinguish between these groups by constructing the so-called discriminant functions that are linear combinations of the variables. The *j*-th discriminant function, $D_j$, takes the form:

$$D_j = d_{j1}Z_1 + d_{j2}Z_2 + \ldots + d_{jm}Z_m \tag{2}$$

where $Z$'s are the standardized input variables, and $d_j$'s are the discriminant function coefficients that reflect the importance of each variable to the differentiation of the observations (*1*). The discriminant functions are derived in order to maximize the separation of the groups. The numerical values of $D$ are the so-called discriminant scores and can be normalized to yield the new water quality indices, described below.

To illustrate the discriminant analysis, the September 2004 water quality data in Table 1 are first divided into three groups (Group 1: oligotrophic, Group 2: mesotrophic, and Group 3: eutrophic) based on the cluster analysis, and discriminant analysis is then applied. Two discriminant functions are generated by the analysis, with Function 1 describing about 91% of the differences between different groups (in other words, Function 2 can be neglected). Furthermore, the discriminant function coefficients of Function 1 (Table 2) indicate that DO is the most important variable (followed by Chl-a, TDN and TDP) in the differentiation of sampling stations in September. The discriminant score $D_j$ of Function 1 for site *j* in September 2004 is calculated as:

$$D_j = 0.593 \times (Chl\text{-a}_j) + 1.011 \times (DO_j) - 0.279 \times (TDN_j) - 0.060 \times (TDP_j) \tag{3}$$

where $(X_j)$ are the standardized concentrations of the independent variables.

Table 2. Discriminant function coefficients for September 2004 data.

| | Function 1 | Function 2 |
|---|---|---|
| Chl-a | 0.593 | 0.593 |
| DO | 1.011 | 0.275 |
| TDN | -0.279 | 0.484 |
| TDP | -0.060 | 0.404 |

Clearly equation 3 puts more weight on DO on the calculation of the discriminant scores for the September data. Since in late summer primary eutrophication indicators (high nutrients) and symptoms (e.g., high chlorophyll *a*) start to diminish and secondary symptoms (e.g., low DO) are well developed, it is logical to put more emphasis on DO when water quality indices are computed. In contrast, in spring when primary eutrophication symptoms are clear and secondary symptoms are absent, it is necessary to put more weight on chlorophyll *a* and nutrients when water quality indices are computed. The discriminant analysis is capable of doing exactly this. For instance, the discriminant function 1 derived for the April 2004 data (equation 4 below) apparently puts more weight on chlorophyll *a* and nitrogen than on DO:

$$D_j = 1.051 \times (Chl\text{ - a}_j) - 0.691 \times (DO_j) + 0.924 \times (TDN_j) - 0.691 \times (TDP_j) \qquad (4)$$

Such "smart" weighting is indeed the essence of the discriminant analysis, because the process of deriving discriminant functions (maximizing the separation of different groups) is the process of finding the parameters with the highest variations among different groups (e.g., nutrients in early spring and DO in later summer for groups with different trophic conditions).

A plot of the discriminant functions (Figure 3) shows that the three groups (three trophic conditions) are completely separated by the two discriminant functions. Therefore, the centroids of the discriminant scores for different groups (Table 3 and Figure 3) can be used as reference values to indicate the overall water quality (i.e., good quality if a discriminant score is close to 5, or poor quality if the score is close to -5, Table 3).

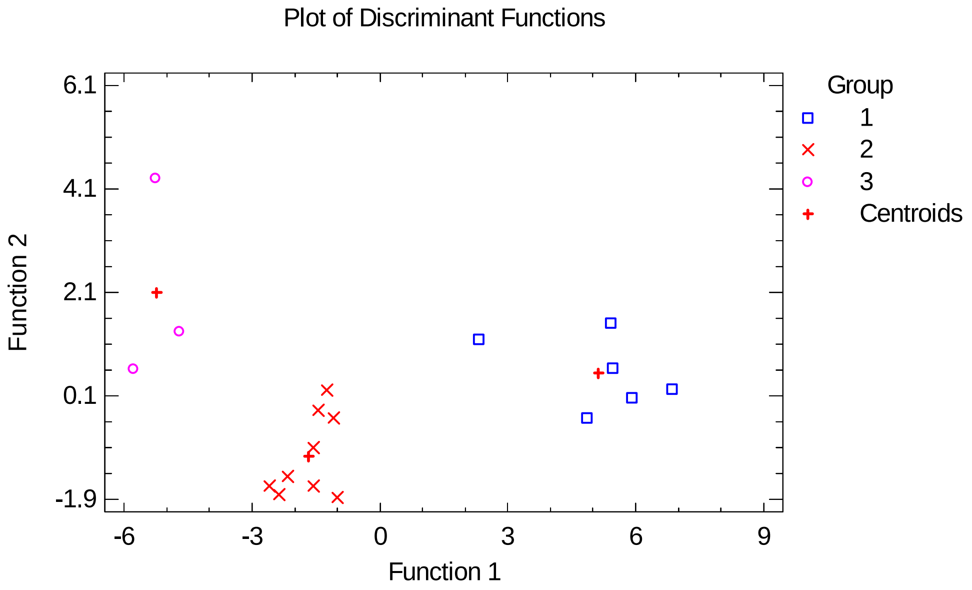


Figure 3. Plot of discriminant functions using September 2004 data.

Table 3. Group centroids of the discriminant scores for September 2004 data.

| Group | Function 1 | Standardized Values |
|---|---|---|
| 1 | 5.135 (a) | 1.00 |
| 2 | -1.671 | -0.31 |
| 3 | -5.256 (b) | -1.00 |

The new water quality index (*WQI*) at a particular site in LIS is defined as the discriminant score *D* standardized via the following formula:

$$WQI = \frac{D-(a+b)/2}{(a-b)/2} \qquad (5)$$

where *a* and *b* are the two centroids having the highest absolute values (e.g., a=5.135 and b=-5.256 for the September 2004 data, Table 3). Such standardization ensures that the group centroids of the discriminant scores range between ~-1 and 1. At each site, a *WQI* value close to 1 indicates good water quality (oligotrophic), a value close to -1 indicates poor water quality (eutrophic), and a slight negative value suggests a mesotrophic condition. The new *WQI* values for April and September 2004 are listed in Table 1.

In essence, the Discriminant-WQI is a water quality index based on a weighted combination of the four individual water quality parameters. The weight of each water quality parameter at a different month may change, and the WQI only tells the relative water quality among the different stations as separated into three groups (i.e., ~-1 and 1 for the two end groups and values in-between for the middle group). There are several potential problems with this type of WQI. First, the discriminant analysis itself does not know which group has the best water quality and which one has the worst; it only tells how different they are. Therefore, we need to assign the correct sign to make sure that a value of -1 represents bad water quality. This potential problem could be fixed by assigning a negative sign to the WQI of A4, a station that usually has very poor water quality. Second, the worst WQI value among the stations is always around ~-1 and the best value is always ~1 for each month and as such comparisons among different months are not possible (in other words, a WQI of -1 at different months may represent different water quality conditions). To overcome this shortcoming, a reference value at station A4 was calculated using a simple linear combination of the four water quality parameters. A numerical score was first assigned to each water quality parameter according to the criteria listed in Table 4, and a reference value was then calculated by simply averaging the four numerical scores. The reference value would potentially range from -2 to 2 but most of the values calculated from the LIS water quality data ranged from 0 to 1. The Discriminant-WQI values at different stations for a particular month were then shifted up or down by a same amount so that the WQI at A4 matches the reference value. The new Discriminant-WQI typically ranges from 0 to 2 after reference correction. The third potential problem with the cluster/discriminant analyses is that an abnormally high or low water quality parameter at any station may throw off the analysis and lead to strange WQI values. The fourth problem with the cluster/discriminant analyses is that the analyses are dependent upon the number of objects (sampling stations in this case) analyzed and missing more than 3 data points at different stations may lead to unreliable WQI values.

Table 4. Criteria for assign a numerical score to each water quality parameter values.

| **Score** | **-2** | **-1.5** | **-1** | **-0.5** | **0** | **0.5** | **1** | **1.5** | **2** |
|---|---|---|---|---|---|---|---|---|---|
| DO (mg/L) | 0-1 | 1-1.5 | 1.5-2 | 2-3 | 3-4 | 4-5 | 5-6 | 6-7 | >7 |
| Chl-a (μg/L) | 0-2 | 2-3 | 3.-5 | 5-10. | 10.-15 | 15-20 | 20-22.5 | 22.5-25 | >25 |
| TDN (mg/L) | >0.7 | 0.6-0.7 | 0.5-0.6 | 0.36-0.5 | 0.23-0.36 | 0.1-0.23 | | | 0-0.1 |
| TDP (mg/L) | >0.13 | 0.09-0.13 | 0.05-0.09 | 0.036-0.05 | 0.023-0.036 | 0.01-0.023 | | | 0-0.01 |

The method of calculating the reference values at A4 was extended to all stations using a spreadsheet, and these values were designated as Linear-WQI, as they are calculated by a simple linear combination of the numerical scores of the individual water quality parameters.

Contour maps of the monthly *WQI* of the entire Sound (Discriminant-WQI and Linear-WQI) were generated (Appendix 1). The two types of WQI maps show some similarities, with the Linear-WQI behaving more consistently than the more complicated Discriminant-WQI.

b. Impact of nitrogen loads on water quality

Point source and non-point source (NPS) nitrogen loads in the past 26 years from 11 areas along the coast of the Sound (see Figure 4 below for locations) were obtained from CTDEP. Of the 11 areas, Areas 2, 4, 8, and 9 supply more than 80% of the total nitrogen to the Sound (Figure 5). Nitrogen input from Areas 2 and 4 is dominated by non-point sources (riverine and coastal NPS, Figure 6); whereas nitrogen input from Areas 8 and 9 is dominated by point sources (sewage treatment plants, STPs, Figure 7). It's worth noting the different nitrogen sources because nitrogen loads from STPs are more manageable than those from non-point sources.

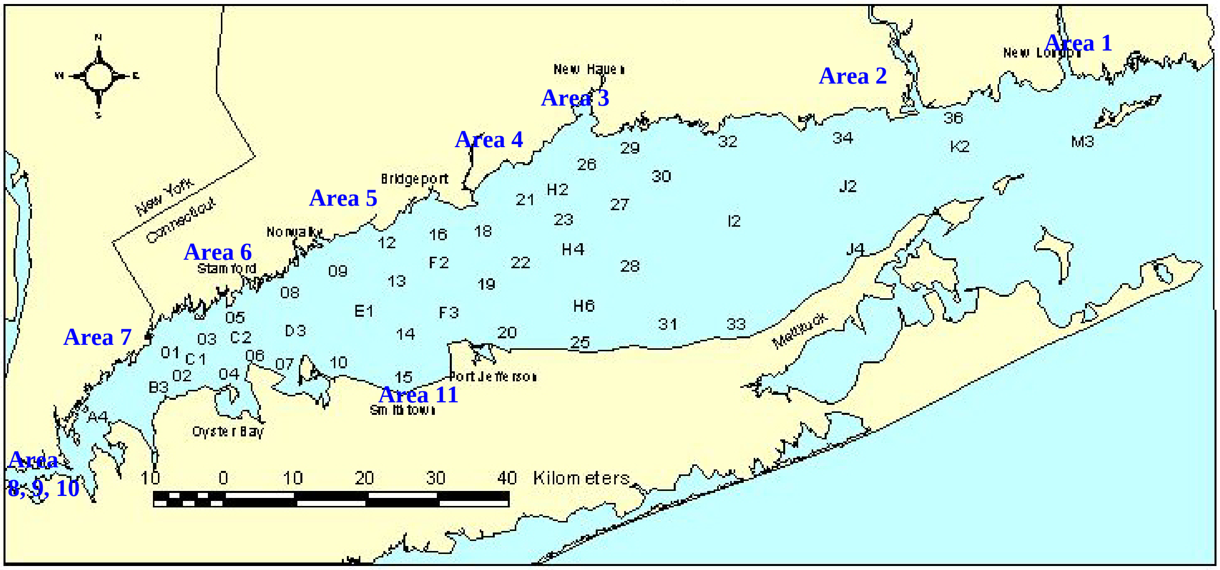


Figure 4. CT DEP sampling stations and areas where total nitrogen loads are available.

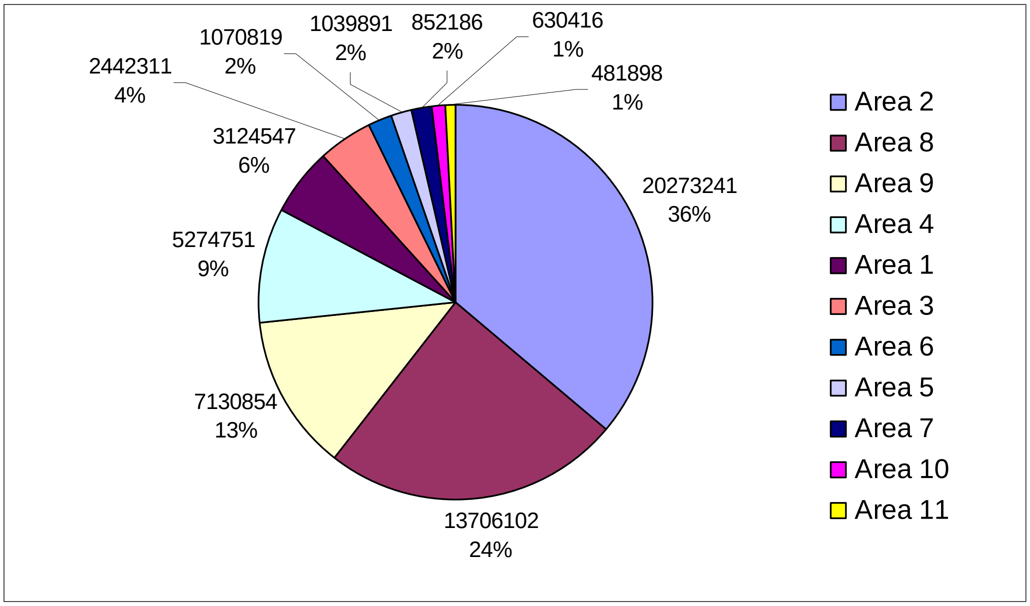


Figure 5. Average annual total nitrogen load (kg) to LIS from 11 monitored areas.

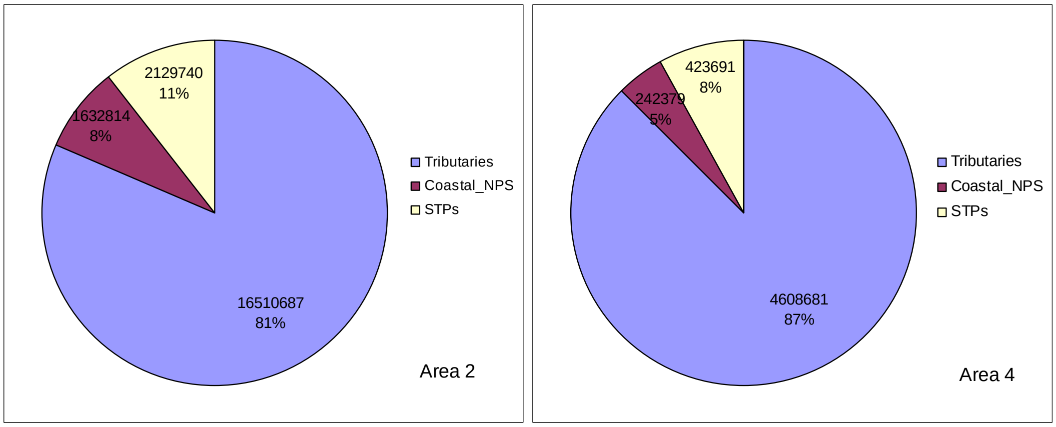


Figure 6. Average annual nitrogen load by category at Areas 2 and 4.

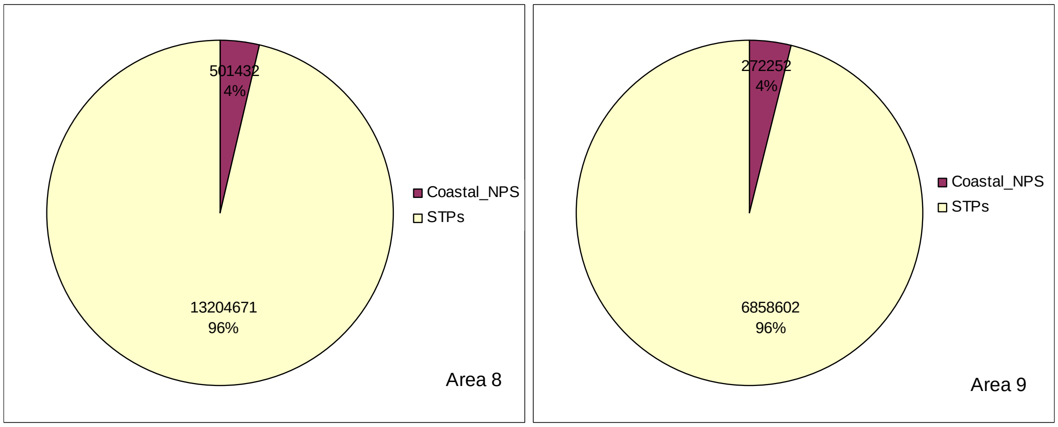


Figure 7. Average annual nitrogen load by category at Areas 8 and 9.

Time series data were deseasonalized (yearly cycles removed) so that the long-term trend and cyclical variations around the trend line (the so called Trend-Cycle Component, see left panel of Figure 8 for an example) can be assessed. In this particular example, it can be seen that bottom TDN at A4 increased from 1994 to 1996, decreased gradually from 1996 to 2000, increased again from 2000 to 2002, and then leveled off after 2002. The trend-cycle component is estimated by smoothing the time series data using a simple moving average with span *k* equal to the length of seasonality *s* (12 mo in this case). Seasonal indices representing the effect of each season were also calculated. For instance, the right panel of Figure 8 shows that there is a seasonal swing of TDN at A4 from 60% (in April) of average to 150% of average (in November) throughout the course of one complete cycle.

Trend-cycle component plot and seasonal indices plot for nitrogen loads at the 11 locations are presented in pages A121-A139 of Appendix 3, whereas the plots for water quality parameters for CTDEP stations are presented in pages A140-A191.

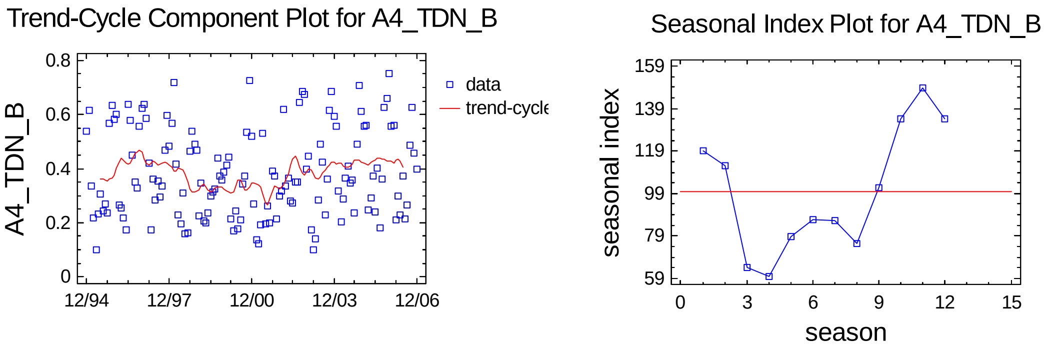


Figure 8. Trend-cycle component plot and seasonal index plot for bottom TDN at A4.

There is no clear trend for the total nitrogen loads from the two major riverine (non-point) sources (Areas 2 and 4) except for some long-term cyclical variations (Figure 9). The nitrogen loads from these places peak around March-April and dip around July-August (Figure 10).

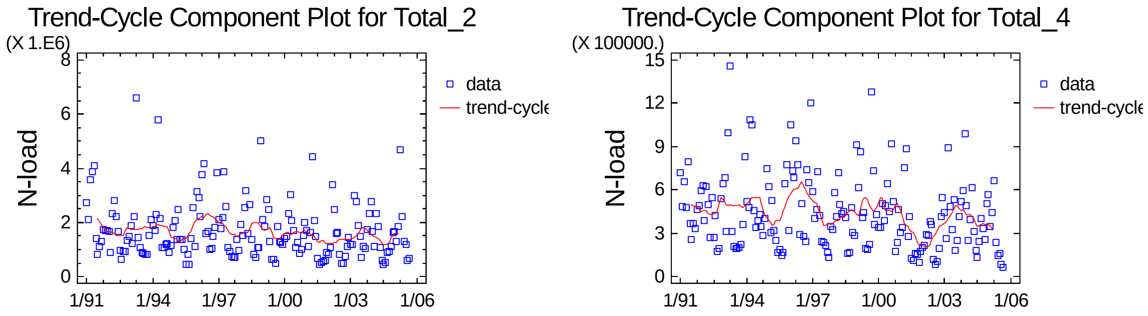


Figure 9. Trend-cycle component plots for Areas 2 and 4.

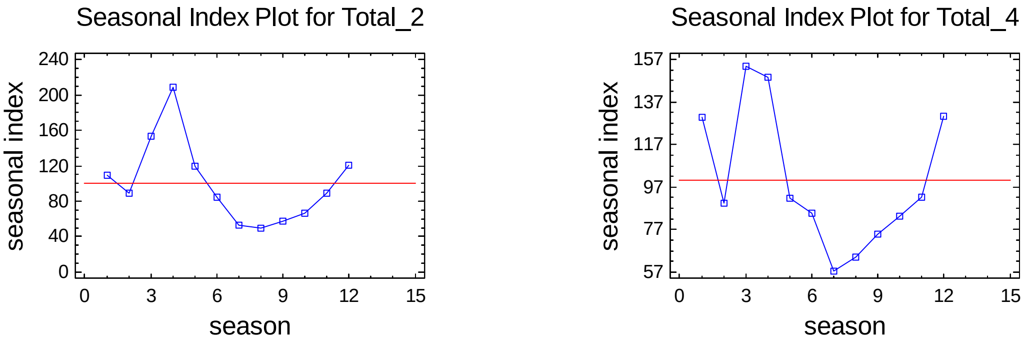


Figure 10. Seasonal index plots for Areas 2 and 4.

In contrast, the total nitrogen loads from area 8 (dominated by STPs) showed a slight decrease trend between 1991 and 1996, a significant drop from 1997 to 1999 (likely as a result of much improved nitrogen removal efficiencies), and a rebound after 2003 probably due to the relaxation of regulations for further facility upgrade (Figure 11, left panel). Total nitrogen loads from area 9 dropped significantly from 1994 to 1995 and slightly from 1995 to 1999, and leveled off thereafter (Figure 11, right panel).

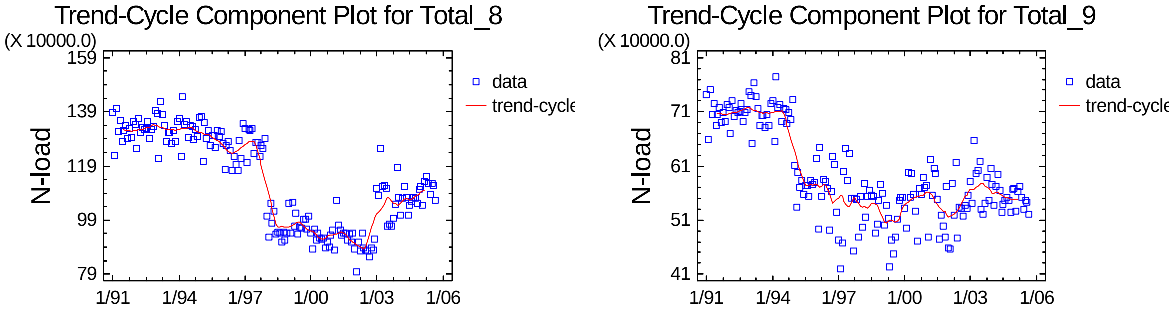
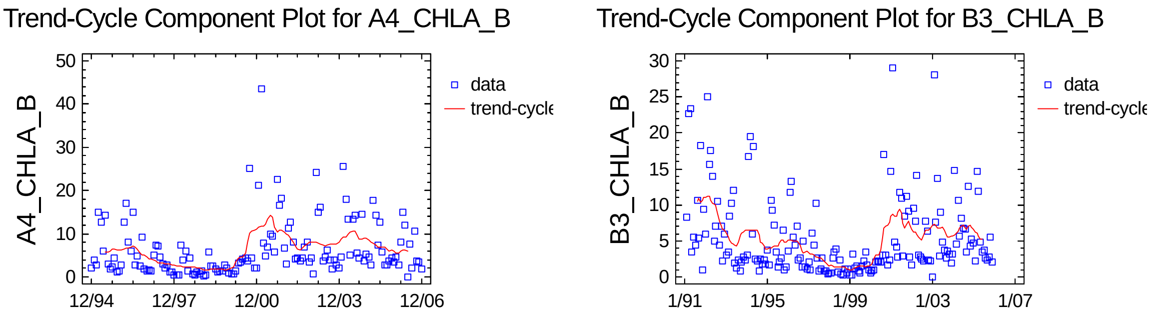


Figure 11. Trend-cycle component plots for total nitrogen loads at areas 8 and 9.

Chl-a levels at all stations decreased gradually from 1991 to 1999, rebounded quickly around 2000-2002, and then leveled off thereafter (see Figure 8 for example; complete list on pages A140 to A152). TDN concentrations showed a similar long-term trend (see pages A153 to A165). These long-term trends for Chl-a and TDN appeared to reflect the trend of the total nitrogen loads from areas 8 and 9, indicating that the reduction of nitrogen at STPs indeed had a positive impact on LIS water quality. Interestingly, TDP concentrations increased steadily from 1991 to about 2002, dropped in the following year, and then leveled off thereafter (see pages A166 to A178). This different trend suggests that the sources for phosphorus and nitrogen might be very different. DO levels did not show any long-term trends (see pages A178 to A191). Minimum annual DO did not improve over the past 15 years either. This lack of improvement in DO levels is intriguing considering the fact that N-loads from major STPs have been reduced significantly. The exact reasons for the lack of DO improvement may be quite complicated and a couple of possible explanations are proposed. First, there might be other important sources of N-loads to LIS (e.g., nitrogen input through groundwater discharge into LIS, or recycling of organic nitrogen from LIS sediment) that are not included in current nitrogen budget. If these uncounted terms are large, then the total nitrogen load may not decrease significantly even though the point source loads are reduced. Second, it is possible that a threshold nitrogen level exists for the DO level to improve, and the reduction in the total nitrogen load has not reached the threshold yet.


Trend-Cycle Component Plot for A4_CHLA_B

Trend-Cycle Component Plot for B3_CHLA_B

Figure 12. Trend-cycle component plots for bottom Chl-a at stations A4 and B3.

To examine the causal relationship between N loadings to the Sound and the total N (TN) in water, the cross-correlations between TN in water at various sampling stations and total N loadings of the entire drainage area as well as areas adjacent to the stations were determined. The cross-correlation at lag *k* measures the strength of the linear relationship between the value of Y (output) at time *t* and the value of X (input) *k* periods earlier. It can be used to determine whether X (e.g., total N loading) would help forecast Y (e.g., TN in water). Cross-correlations between TN and total N loadings averaged 0.23 and 0.27 in surface and bottom waters, respectively, with an average lag time of 9 months for surface water and 10 months for bottom water. In other words, TN concentrations in LIS water would change to some degree in response to the total N loadings to the Sound, with a lag time of about 9-10 months.

Table 5. Cross-correlations between TN in water at various sampling stations and total N loadings from the entire drainage area and areas adjacent to the stations.

| Output | Input | Surface Water | | Bottom Water | |
|---|---|---|---|---|---|
| TN @ | Total N loading from | Lag | Cross-correlation | Lag | Cross-correlation |
| A4 | Entire Area | 8 | 0.282 | 8 | 0.235 |
| A4 | Area_8 | 7 | 0.261 | 11 | 0.338 |
| A4 | Area_8_9_10 | 7 | 0.290 | 11 | 0.341 |
| B3 | Entire Area | 10 | 0.258 | 8 | 0.247 |
| B3 | Area_8_9_10 | 10 | 0.276 | 7 | 0.302 |
| B3 | Area_7 | 10 | 0.296 | 10 | 0.194 |
| C1 | Entire Area | 8 | 0.296 | 10 | 0.218 |
| C1 | Area_8_9_10 | 8 | 0.308 | 9 | 0.358 |
| C1 | Area_7 | 8 | 0.208 | 10 | 0.161 |
| C2 | Entire Area | 8 | 0.246 | 9 | 0.243 |
| C2 | Area_6 | 10 | 0.217 | 10 | 0.175 |
| D3 | Entire Area | 8 | 0.189 | 10 | 0.268 |
| D3 | Area_6 | 10 | 0.107 | 11 | 0.210 |

| | | | | | |
|---|---|---|---|---|---|
| 09 | Entire Area | 10 | 0.183 | 9 | 0.245 |
| 09 | Area_5 | 10 | 0.171 | 10 | 0.217 |
| E1 | Entire Area | 10 | 0.238 | 10 | 0.351 |
| E1 | Area_5 | 10 | 0.185 | 11 | 0.287 |
| 15 | Entire Area | 10 | 0.298 | 8 | 0.357 |
| 15 | Area_11 | 10 | 0.283 | 10 | 0.300 |
| F2 | Entire Area | 10 | 0.271 | 10 | 0.334 |
| F2 | Area_5 | 11 | 0.214 | 10 | 0.299 |
| F3 | Entire Area | 10 | 0.288 | 10 | 0.353 |
| F3 | Area_11 | 10 | 0.235 | 11 | 0.278 |
| H2 | Entire Area | 8 | 0.232 | 10 | 0.314 |
| H2 | Area_3 | 10 | 0.200 | 10 | 0.252 |
| H2 | Area_4 | 12 | 0.159 | 10 | 0.209 |
| H4 | Entire Area | 10 | 0.285 | 10 | 0.370 |
| H4 | Area_3 | 10 | 0.260 | 11 | 0.275 |
| H4 | Area_4 | 10 | 0.207 | 11 | 0.251 |
| H6 | Entire Area | 10 | 0.308 | 10 | 0.347 |
| H6 | Area_3 | 10 | 0.241 | 10 | 0.262 |
| H6 | Area_4 | 10 | 0.230 | 10 | 0.240 |
| I2 | Entire Area | 8 | 0.221 | 10 | 0.245 |
| I2 | Area_2 | 8 | 0.234 | 8 | 0.231 |
| J2 | Entire Area | 11 | 0.224 | 10 | 0.197 |
| I2 | Area_3 | 8 | 0.123 | 11 | 0.183 |
| J2 | Area_2 | 11 | 0.211 | 10 | 0.188 |
| K2 | Entire Area | 3 | 0.088 | 10 | 0.114 |
| K2 | Area_2 | 3 | 0.115 | 10 | 0.102 |
| M3 | Entire Area | 10 | 0.094 | 10 | 0.187 |
| M3 | Area_1 | 8 | 0.138 | 10 | 0.150 |
| Avg. | | 9 | 0.227 | 10 | 0.256 |

<u>c. importance of TDP on DO and Chl-a</u>

Contour maps of Chl-a, TDN, TDP, and DO from Jan. 1995 to Dec. 2006 were created. The criteria and color schemes (Table 6) similar to the ones specified by USEPA (*3*) were used so the quality of the individual indicator could be easily identified. For instance, Figure 13 clearly shows that TDN in bottom water in August 1999 was fair, TDP was poor in most areas, chl-a was good, and DO was poor in west tip of the Sound, fair in mid-part of the Sound, and good in eats-part of the Sound. A complete list of the contour maps is presented in Appendix 2.

Table 6. Criteria for assessing water quality indicators in East/Gulf Coast sites (*3*).

| | Good (green) | Fair (yellow) | Poor (red) |
|---|---|---|---|
| TDN | < 0.1 mg/L | 0.1 - 0.5 mg/L | > 0.5 mg/L |
| TDP | < 0.01 mg/L | 0.01 - 0.05 mg/L | > 0.05 mg/L |
| Chl-a | <5 µg/L | 5 - 20 µg/L | >20 µg/L |
| DO | > 5 mg/L | 2 - 5 mg/L | < 2 mg/L |

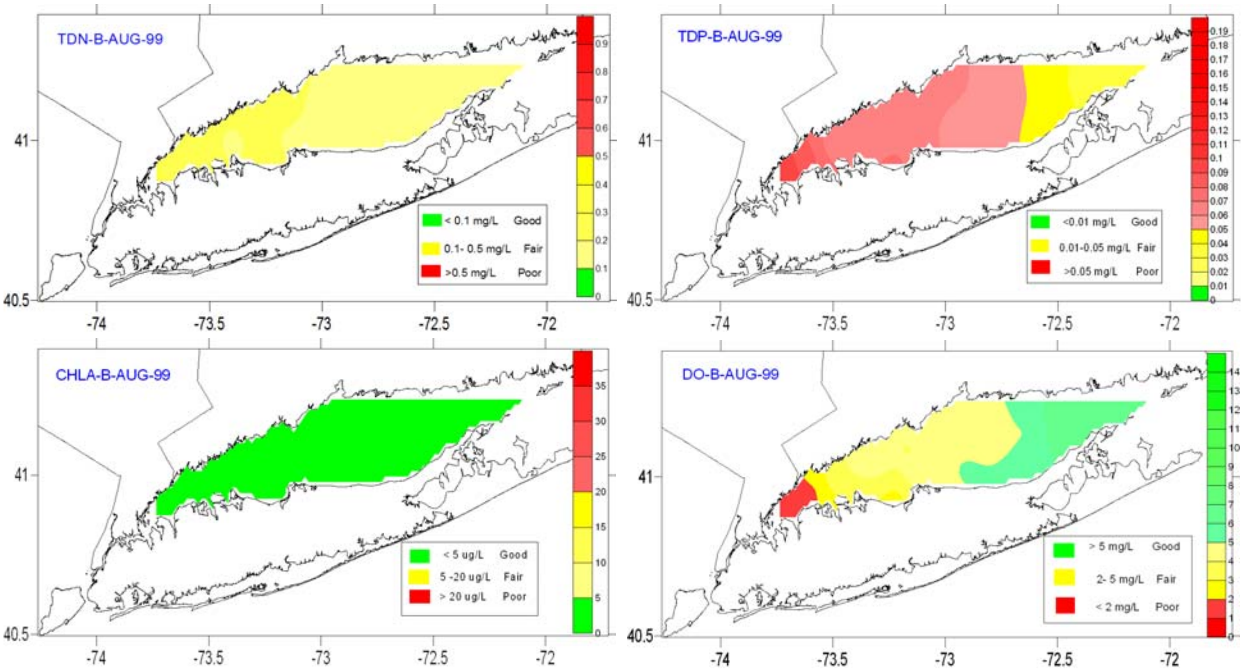


Figure 13. Contour maps of TDN, TDP, Chl-a, and DO in bottom water in August 1999.

The contour maps (pages A-49 to A-84 for surface water and A-85 to A-120 for bottom water) clearly show that TDN is generally fair in both surface and bottom waters in the past decade with some occasional poor rankings in west tip of the Sound during winter months. Chl-a is good to fair in most cases with some poor rankings in early spring and late summer at the west tip of the Sound. DO is generally good in surface water, and exhibits some fair to poor conditions in west LIS during summer months. It is striking to see that TDP is generally poor during cold months (Sept. to Feb.), and only reaches fair conditions in the warmer months.

While N is considered the primary limiting nutrient in LIS and most other coastal marine systems, the absolute concentration levels of N and P also play a crucial role on primary production. For instance, if the concentration of DIP (dissolved inorganic phosphorus) exceeds 5 µg/L and DIN (dissolved inorganic nitrogen) exceeds 300–500 µg/L, neither P nor N may be limiting (*4*). In other words, the limiting factor may depend on the level of nutrient concentrations. Since dissolved organic nitrogen (DON) is utilized as a nitrogen source by many (but not all) phytoplanktonic species, it is important to consider TDN (DIN + DON) rather than DIN alone in primary production (*5*). Since TDN in many of the west LIS sampling stations exceeds 300 µg/L, it is postulated that P plays a very important role on algal blooms. The fact that TDN levels are still high (even after the reduction of N loading from sewage treatment plants, STPs) and TDP levels are increasing over the past 15 years may explain, at least partially, why the DO does not improve.

To test this hypothesis, we examined the cross-correlations between the four water quality indicators (TDN, TDP, Chl-a, and DO) using the time series data from 1991 to 2006. The maximum cross-correlations between the four water quality indicators (TDN, TDP, Chl-a, and DO) at various LIS water sampling stations as well as the corresponding lag time (months) are listed in Table 7. The maximum cross-correlations between DO and TDN ranges from -0.22 to -0.52 at various stations (average -0.39 for all stations), with the corresponding lag time ranges from 7 months to 10 months (average of 8.6 months). In other words, in most cases DO reaches its lowest level about 8.6 months after the TDN peaks at the same sampling station. Negative cross-correlations are used here because a higher TDN level is expected to lead to a lower DO value some time later. The cross-correlations between DO and TDP are almost always higher than the cross-correlations between DO and TDN for any given sampling station (average -0.521 for all stations), suggesting that TDP may play a more important role on DO concentrations than TDN. Likewise, the cross-correlations between Chl-a and TDP (average 0.27 for all stations) are also higher than the cross-correlations between Chl-a and TDN in most cases (average 0.22 for all stations), further suggesting the importance of TDP on algal blooms.

We suggest that the management examine the possibility of P reduction as an alternative means of reducing eutrophication in LIS.

## 10. Conclusions:

Contour maps of WQI and individual water quality parameters would provide the management with a quick overview of the water quality in the entire Sound. These contour maps should be generated as soon as data are available to gain a quick assessment of the water quality in the Sound. The reduction of nitrogen at STPs appeared to have a positive impact on TDN and Chl-a levels in LIS water. TDP levels seemed to have a stronger impact on both DO and chl-a levels than TDN levels, and it is suggested that the management examine (or re-examine) the importance of TDP to the entire water quality of the Sound, and consider the possibility of P reduction as an alternative means of reducing eutrophication in LIS.

## 11. Presentations/Publications/Outreach:

We made three presentations, the first one at the 2007 American Geophysical Union fall meeting, the second one at the 2008 Geological Society of American annual meeting, and the third one at the Long Island Sound Research Conference. One proceeding paper (Long Island Sound Research Conference) was published this year. A master's thesis also is resulted from this project.

## 12. Other Information: None.

## References

1. Davis, J. C., *Statistics and data analysis in geology*. 3rd ed.; John Wiley and Sons: New York, NY, 2002.
2. Sneath, P. H. A.; Sokal, R. R., *Numerical taxonomy: The principles and practice of numerical classification*. W. H. Freem & Co: San Francisco, 1973.
3. USEPA *National Coastal Condition Report II*; EPA-620/R-03/002; Office of Research and Development/Office of Water: Washington, DC 20460, 2004; p 272.
4. Reynolds, C. S., *The Ecology of Freshwater Phytoplankton*. Cambridge University Press: Cambridge, 1984.
5. Capriulo, G. M.; Smith, G.; Troy, R.; Wikfors, G. H.; Pellet, J.; Yarish, C., *Hydrobiologia* **2002,** 475-476, 263-333.

Table 7. Cross-correlations and lag time (months) between DO, TDN, TDP, and Chl-a at various stations. S-surface; B-bottom.

| | DO as output of TDN | | DO as output of TDP | | DO as output of CHLA | | CHLA as output of TDN | | CHLA as output of TDP | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Lag | Crosscorrelation | Lag | Crosscorrelation | Lag | Crosscorrelation | Lag | Crosscorrelation | Lag | Crosscorrelation |
| A4_S | 10 | -0.511 | 10 | -0.581 | 3 | -0.483 | 7 | 0.306 | 7 | 0.464 |
| A4_B | 9 | -0.518 | 10 | -0.591 | 5 | -0.249 | 4 | 0.294 | 4 | 0.197 |
| B3_S | 9 | -0.338 | 10 | -0.513 | 4 | -0.278 | 6 | 0.146 | 7 | 0.324 |
| B3_B | 9 | -0.407 | 10 | -0.584 | 5 | -0.349 | 5 | 0.189 | 6 | 0.181 |
| C1_S | 10 | -0.455 | 10 | -0.491 | 3 | -0.240 | 7 | 0.181 | 7 | 0.357 |
| C1_B | 9 | -0.353 | 9 | -0.582 | 5 | -0.339 | 3 | 0.321 | 4 | 0.361 |
| C2_S | 9 | -0.473 | 9 | -0.576 | 3 | -0.256 | 7 | 0.129 | 8 | 0.335 |
| C2_B | 9 | -0.347 | 9 | -0.539 | 6 | -0.369 | 3 | 0.249 | 2 | 0.408 |
| D3_S | 9 | -0.424 | 10 | -0.565 | 4 | -0.261 | 5 | 0.223 | 7 | 0.261 |
| D3_B | 9 | -0.339 | 9 | -0.452 | 5 | -0.441 | 2 | 0.192 | 4 | 0.295 |
| 09_S | 9 | -0.309 | 10 | -0.465 | 2 | -0.170 | 3 | 0.182 | 7 | 0.152 |
| 09_B | 9 | -0.393 | 9 | -0.593 | 5 | -0.141 | 3 | 0.214 | 3 | 0.250 |
| E1_S | 9 | -0.498 | 10 | -0.620 | 3 | -0.310 | 6 | 0.154 | 8 | 0.272 |
| E1_B | 8 | -0.399 | 9 | -0.417 | 6 | -0.487 | 2 | 0.356 | 4 | 0.264 |
| 15_S | 9 | -0.399 | 9 | -0.538 | 3 | -0.121 | 2 | 0.116 | 8 | 0.259 |
| 15_B | 8 | -0.443 | 9 | -0.552 | 5 | -0.166 | 2 | 0.290 | 3 | 0.220 |
| F2_S | 9 | -0.474 | 9 | -0.576 | 3 | -0.238 | 5 | 0.156 | 6 | 0.184 |
| F2_B | 9 | -0.394 | 9 | -0.567 | 6 | -0.416 | 3 | 0.339 | 3 | 0.384 |
| F3_S | 9 | -0.494 | 9 | -0.507 | 3 | -0.289 | 7 | 0.204 | 8 | 0.216 |
| F3_B | 8 | -0.298 | 9 | -0.609 | 5 | -0.430 | 1 | 0.250 | 3 | 0.305 |
| H2_S | 8 | -0.461 | 10 | -0.532 | 3 | -0.203 | 6 | 0.144 | 9 | 0.228 |
| H2_B | 8 | -0.446 | 10 | -0.526 | 5 | -0.321 | 2 | 0.261 | 3 | 0.209 |
| H4_S | 9 | -0.411 | 9 | -0.612 | 1 | -0.247 | 7 | 0.165 | 8 | 0.276 |
| H4_B | 8 | -0.432 | 9 | -0.555 | 6 | -0.365 | 1 | 0.305 | 3 | 0.317 |
| J2_S | 8 | -0.365 | 10 | -0.454 | 12 | -0.117 | 8 | 0.158 | 9 | 0.261 |
| J2_B | 8 | -0.330 | 9 | -0.467 | 11 | -0.056 | 8 | 0.173 | 3 | 0.176 |
| K2_S | 7 | -0.263 | 9 | -0.380 | 2 | -0.238 | 8 | 0.208 | 9 | 0.365 |
| K2_B | 8 | -0.216 | 10 | -0.377 | 8 | -0.109 | 8 | 0.202 | 3 | 0.254 |
| M3_S | 8 | -0.285 | 10 | -0.395 | 2 | -0.170 | 4 | 0.171 | 9 | 0.271 |
| M3_B | 8 | -0.267 | 10 | -0.404 | 4 | -0.106 | 3 | 0.171 | 3 | 0.141 |
| **Average** | **8.6** | **-0.391** | **9.5** | **-0.521** | **4.6** | **-0.266** | **4.6** | **0.215** | **5.6** | **0.273** |
| Average_S | | -0.411 | | -0.520 | | -0.242 | 5.9 | 0.176 | 7.8 | 0.282 |
| Average_B | | -0.372 | | -0.521 | | -0.290 | 3.3 | 0.254 | 3.4 | 0.264 |