Peer Review Responses for *Assessment of Nutrient Loading and Eutrophication in Barnegat Bay–Little Egg Harbor, NJ in Support of Nutrient Management Planning*

December 17, 2014

Contents

I. Questions Posed to Peer Reviewers & Summary of Peer Reviewer Responses

Questions were posed to the peer reviewers through a document entitled "Peer Review Scope of Work." The Scope of Work identified five key topics for peer reviewers' consideration, followed by more specific examples and more detailed questions for each topic. The key topics were:

- Data Sufficiency to Draw Conclusions / Develop Index;
- Selection of Thresholds to Define Condition;
- Derivation of the Index;
- Use of Index;
- Overall Adequacy of the Report.

The following sections include each question posed and any background context provided to the peer reviewers in bold italics, followed by the unedited peer reviewer responses in separate gray boxes. Additional and summary comments are included in Section II, and literature cited can be found in Section III.

Data Sufficiency to Draw Conclusions / Develop Index: ***There are significant limitations with respect to the data available for this project – examples include a lack of data for some measures across the years considered, limited frequency of data collection (typically quarterly monitoring so that if growing season is selected as stated, there may be only two sampling events), and limited data available for some locations. Given these limitations, can the conclusions drawn and/or the index developed be used for management purposes, developing strategies to target specific levels of nutrients that would be expected to result in support of a healthy ecosystem for this estuary? Are conclusions supported by the data? Specific examples of data limitations are listed below:***

General Data Sufficiency Comments:

a.

Data sufficiency is hard to assess from this report. Data availability information is scattered throughout the report, not organized in a coherent manner. Data gaps, if any, by station and sampling event are not presented. Data are presented in Appendix 3-4 but the data are summarized by year and segment. Because bits and pieces of information are scattered among the various sections of the report, data availability information can be confusing. For example, in the Methods Section, Components Subsection, it is stated that "epiphyte biomass and areal cover measurements were made on seagrass samples collected over a three-year study period (2009-2011). However, Figure 3-2 shows epiphyte biomass data available for 2004-2006 and 2008-2010 (no data for 2007), and Figure 3-50 shows epiphyte results for 1997-2010 (and there are data in 2007). Somewhere else (under Available Data/Data Gaps in Component 3), we learn that epiphyte biomass, and the ratio of epiphyte to seagrass biomass, was estimated backwards to 1997. The estimation methods are not presented. Other examples:

- There appear to be phytoplankton species composition and abundance data for 2009, 2010, and 2011 (Methods Section), but these data are not considered or used.
- Brown tide bloom event monitoring is highlighted in the Methods Section, but this section fails to mention the spatial and temporal patchiness of these collections.
- In the Methods Section of Component 2, we learn that "water samples (N= 72) were collected at 12 transects in 2008 to determine nutrient concentrations", but surely there were nutrient concentration data collected since 1989, mentioned elsewhere.

Overall, however, the Index of Eutrophication does not appear to be limited by the amount of data available for each of its major components, but by the lack of key components that cannot be included in the index simply because they are not monitored, or they are monitored infrequently.

The Index of Eutrophication is limited by the absence of key biotic components. Hard clam abundance would be a valuable component in the index, but data exist only for 2001. Benthic macro-invertebrates are important measures of overall ecological condition in estuaries, but very little data were available except for 2001. The Methods Section states that benthic invertebrates were used in the development of the Index of Eutrophication, but in reality they were not. The EMAP metrics were not included in the index. The index focuses on water quality and seagrass, which are strong components in the index, but biotic components that are often important in assessing ecosystem health, such as phytoplankton community, benthic invertebrate community, and fisheries species, are not included. The problem is that these components are not currently monitored.

b.

There are concerns about data sufficiency centering on the period of record, with the result that the authors turn to literature values and best professional judgment (BPJ) frequently. Although I consider it to be a short period for this kind of comprehensive ecosystem analysis, particularly with the need to tease out cause-effect relationships, the authors make a credible attempt to use the available information from primary and secondary data sources to develop a quantitative framework for management action moving forward.

1. ***Data quality concerns: Were censored values, non-detects, zero values, skewness, outliers handled correctly/adequately?***

a.

Within the constraint above, the monitoring program is adequately framed- sampling frequency, spatial distribution, parameters are appropriate. Data are of high quality and data quality is appropriately assessed and controlled for. The QAPP is standard, well-developed and complete with a data management plan, goals for data quality are well-defined and personnel responsibilities detailed. The field and analytical methods described are appropriate, well-detailed, with time and space scales of sample collection generally good and in conformance with the size and variability of the system. QA/QC procedures are well-defined, the sample handling, sample processing appropriate, and in conformance with standard methods and practice. The analyses performed and conclusions drawn are appropriate, with some caveats as detailed below.

Missing values are handled appropriately. Missing data is properly accounted and handled in the analysis, that is, not treated as zeros. Similarly, a result of zero is appropriately treated in the calculations and included in the analysis as a zero.

Outliers were identified and justification was given for withholding them from the analyses. A brief discussion of the pattern of outliers in the USGS nutrient-versus-%turf assessment was cogent and justifiable. USGS applied the Discordance Test to these data which is acceptable industry standard.

Several instances of censored data occurred in the database for a variety of reasons, such as changes in analytical method applied or readings below detection. The substitution method used by USGS to assign values to these data seem to yield a result that is more useful than a non-detect but I have not used, nor seen, this "rescue" technique used in data QA previously. While the censored data for TN amounted to less than 10%, TP data exhibited an uncomfortably high level of censored data approaching 20%, which could be cause for concern.

Additional parameters that should be considered for inclusion in the monitoring program are described below (5). [*NEIWPCC Note: Refers to Data Sufficiency Question 5.*]

b.

Yes. Despite unavoidable data limitations due to their multi-project, multi-source, multiple sampling design origins, the authors did an excellent job of selecting unbiased and as spatially and temporally representative data as possible.

c.

The USGS study appears to deal well with data availability, compatibility, and usage. Quality-assurance measures have been taken. For censored values (below reporting level), the mean of all detected values less than the reporting level was substituted for the corresponding censored values. Censored data are usually set to one-half of the method detection limit before analysis. As method detection limits can vary among data sets, substitution of censored data with estimates calculated from the dataset of detected values for a particular parameter is a preferred method. As for the main report, the results of the methods for "statistical rigor, robustness, and representativeness" are not presented. There should be a section or appendix presenting at least box plots (the USGS report presents box plots), and documenting how skewness and outliers were treated.

2. ***Are the written conclusions in line with data presented? Should there be any concerns regarding poor statistical correlations?***

a.

While the overall conclusion that BB-LEH is a highly eutrophic estuary appears to be valid, some conclusions are less clear. Take the first finding (Abstract: Findings). "This study confirmed that surface-water concentrations of nutrients (nitrogen and phosphorus) in the BB-LEH estuary are strongly related to land use. Total nitrogen and phosphorus are highest in areas with the highest percentages of urban and agricultural land, and with the lowest percentage of forested and undeveloped land", and in the following order (Component 1: Evaluation of Available Water-quality Data): TN_{north} >TN_{south}>$TN_{central}$, and TP_{north} >TP_{south}>$TP_{central}$. While this may be true for TN, inspection of Figures 15 and 16 (Appendix 1) does not appear to support this conclusion for TP concentration. There is no statistically significant difference between the south and central watershed segments in terms of TP (Fig. 15), and in any case the median TP concentration is higher in the central segment, which is also less developed than the south segment (Fig. 16). Variability in TP concentration in the estuary during the period 1999-2010 is high, and the largest concentrations were observed in the central and south segments, not the north segment (Fig. 2-4). Appendix 1 (pg. 13) states that the estuary has experienced "increases in macroalgal growth, harmful algal blooms, and turbidity, as well as oxygen depletion", but a quick look at Figure 2-5 shows again that total suspended solids are highly variable, and the trend in the graph suggests a decline in mean TSS during the 1989-2010 time series. Same story for chlorophyll (Fig. 2-7).

Regarding the Index of Eutrophication, it is concluded (Abstract: Findings) that index values declined 34% and 36% in the central and south segments, from 73 (actually, 72, pg. 111) and 71 in the 1990s to 48 and 45 in 2010 (do not quite match the values in Figure 3-57, or the final values in Appendix 3-7), "indicating that these segments are currently undergoing eutrophication". However, this assessment fails to mention that about half of the values in the plot in Figure 3-57 (before 1999) do not use TP, while the other half of the values (1999 and after) use TP. A line drawn through 1999 divides the plot into a highly variable cloud of points on the left-hand side of the plot (pre-1999) that is essentially flat (not increasing or decreasing through time) and a much less variable cloud of points on the right-hand side (post 1999) that is probably flat at least for the north and south segments. Eutrophic condition in the north segment is indeed worst, but highlighting changes in scores downwards ("scores in the north segment declined sharply in 2010") conveniently leaves out of the picture the report's own assessment of a significant upward trend in index scores (improving) in the north segment (Fig. 3-57).

Incidentally, I have calculated the average of some of the final scores shown in Figures 3-54c, 3-55c and 3-56c and cannot reproduce index scores shown in Figure 3-57. For example, for the north segment, the WQ index value in 2010 is ~35 (Fig. 3-54c), and the Light Availability index values is ~70 (Fig. 3-55c). The Seagrass Response Index does not apply. A simple average of these two values gives an index value of 52.5 for 2010. In Figure 3-57, the index score for the north segment in 2010 is substantially lower.

b.

The conclusions drawn by the authors are for the most part supported by the data. As discussed below there are instances where the data are not available or are asynchronous with other data, and are in some cases poorly correlated. There are numerous data gaps in a variety of parameters including light, DO and secchi depth and some critical datasets are of notably short periods of record, including TP, seagrass abundance, macroalgae cover and epiphyte cover. Authors' conclusions may not be the only interpretation of the information presented, as many assumptions are made in the analysis. It is important to note gaps in the knowledge base, as the authors do, and identify recommendations so that monitoring, modeling and analysis can be focused on these points as the program advances. This gap analysis is a benefit provided by the study.

c.

The written conclusions are in line with the data presented. I have no substantial concerns regarding poor statistical correlations because real biological data are often more variable and respond less linearly than physical or chemical data. However, in the "Potential Improvements to the Eutrophication Index" section, I have suggested some alternative approaches that should be explored and evaluated before settling on the current index. [*NEIWPCC Note: The Potential Improvements to the Eutrophication section comments from this peer reviewer are included under Section II: Additional and Summary Comments.*]

3. ***The study states that, "The BB-LEH database was analyzed for each segment of the bay, because these segments have been determined to be heterogeneous habitats." If this statement is true, is the determination of one threshold calculation for the entire bay for each indicator appropriate in determining the indicator score for each segment, or should the threshold calculations for indicators be defined separately for each segment in order to determine the indicator score for each segment of the Bay?***

a.

As I understand the "Methods: Determining Thresholds: Rescaling Data" section (on pp 86-87?), I believe single values are appropriate for and applicable to all three segments "based on the numerous literature studies and volume of data that were analyzed." The uncertainty based on global analyses of BB-LEH-like systems is definitely substantially less than uncertainties generated by developing separate thresholds for each segment, using fewer data, and very likely substantially smaller in some cases.

b.

Segmentation of the bay for purposes of data analysis, model development and threshold determination is a widely accepted practice in developing indices and criteria and appropriately invoked here. It allows the system to be analyzed at the sub-estuary landscape level, which is harmonious with most physical and biological processes and with the data available. Authors are using a 3-segment partitioning, which for some analyses is further sub-divided and indexes are calculated for each segment where possible. There is data sufficiency to do this based on nutrient loading, chlorophyll and salinity. The bulk of the analysis for development of and application of eutrophication indexes is based on a three-segment bay: northern, central and southern. An additional partitioning into east-west is made for each segment for a total of six is applied only (apparently) in the case of benthic vegetation due to further compartmentalization based on species and salinity. It is not made clear enough by the authors, however, how the secondary segmentation was derived how and when 3 versus 6 partitions were applied, a full justification for the additional split or a statistical treatment for bounding these areas.

It would seem that the thresholds for indicators should be defined specifically for each segment of the bay. Because the segments are deemed heterogeneous, there would reasonably be different physical, chemical and biological processes acting on the nutrients and on the symptoms of eutrophication. Nutrients are processed in different ways in each segment therefore the calculation of indexes relative to the thresholds should be applied on a segment-by-segment basis.

The bay segmentation was done a priori based on knowledge of the bay. The statistical justification for the segmentation, via ANOVA, was performed a posteriori and found to be significant. In other words, the segmentation process, and the segments chosen, were in fact justified. This process is however, implemented backwards and while the segments were found to be significantly different, statistically, (except for benthic invertebrates), this method may have nonetheless allowed some stations to fall into the “wrong” segment when a priori testing would have sorted it/them into different segments or created more or fewer segments, giving more power to the segmentation scheme. A more rigorous and objective grouping of stations might have been accomplished by first applying a PCA to determine the primary variables influencing patterns seen in the data, followed by cluster analysis to objectively group the stations into appropriate spatial categories based on water quality characteristics, as has been done for other estuaries (e.g. Hardin et al. 1996, Boyer et al. 1997, 2009). If the segmentation were based on initial statistical determination of significant differences, segmentation would increase the resolution of the index calculations, strengthen the underlying rationale and provide better targeting of management actions.

c.

Physicochemical variables and biotic indicators should respond in the same general way to pressures in the estuary regardless of the segment for which the assessment is made. However, if a particular biotic indicator varies with the estuarine gradient, an index developed to assess this indicator should take into account this variability, by providing equitable sampling across the gradient, and by normalizing the indicator or calculating separate thresholds for different regions of the estuary. This is commonly done for phytoplankton and benthic macro-invertebrates because their abundance and species composition typically vary with the salinity gradient. Additionally, benthic invertebrate composition and abundance vary with depth (in water column-stratified estuaries) and sediment type. In BB-LEH, these indicators are not used. Only if biotic indicators were expected to elicit different responses to stressors in different segments, would thresholds have to be defined separately for each segment. In a highly eutrophic system (e.g., north segment) there may be little response of an indicator to improvements in water quality until a threshold is reached. Thus it might be of interest to set lower thresholds in these systems; however these would be management thresholds rather than ecological thresholds.

4. ***Do current USGS studies sufficiently capture (identify and estimate) all substantive N and P loads to the bay? If not, please identify additional sources that should be considered.***

a.

The analysis by USGS from Appendix 1 sufficiently captures a large part of the water budget and watershed nutrient inputs. However, the basis for several simplifying assumptions for groundwater terms in the water budget (e.g. no exchange with adjacent basins, no change in storage in the aquifer and minimal withdrawals, only an average is used) are not completely explicated. I am curious about the potential import of nutrients through the inlet- both from coastal waters loaded by wastewater effluent discharges in the vicinity and by the general coastal water mass that may be subsidizing loading by additional external marine end member sources that need to be accounted in models and management plans. The eastern seaboard is replete with point source discharges creating a broad nutrient-enriched coastal zone and there is a high probability that offshore nutrients can be introduced to BB-LEH from this general elevated milieu of nutrients offshore. Further, enriched BB-LEH waters that are flushed and then reenter the estuary on flood tides represent a pseudo-subsidy that elevates nutrient concentrations and impact by effectively extending residence times on the re-entered water parcels. This represents a complication in calculating loading impacts that are assumed to be minor terms in the budget and further underlines the importance of a sophisticated circulation model. The complete justification for these assumptions is not presented.

Airshed inputs are referenced from earlier studies and are significant but it is not clear how or if they are incorporated into any of the models or analyses. There is certainly no obvious accounting for them in the regressions or the inputs even though Gao et al. (2007) did calculate that atmospheric deposition of N was substantial and significant to the nutrient budget. Perhaps the stepwise adjustment routine described on Pg. 41, App 1 is intended to account for such inputs and other factors where data are fuzzy or incomplete and aggregate to a substantial error term but this is not described. In any case, the delivery of N to the estuary is being calculated with much of the budget unconstrained, making predictive capability of the models uncertain.

A consideration of information that would complement the USGS analysis is a set of potentially important process measurements that may bear additionally on nitrogen budget in the watershed and the estuary. These processes affecting nitrogen delivery and availability include abiotic sequestration and burial or dissimilatory nitrate reduction to ammonium (DNRA) and denitrification, possibly important terms in an estuary where multiple instances of hypoxia have been described by the authors.

b.

The USGS study does appear to capture all substantive N and P loads. Data selection, site selection, and estimation methods are sound and reasonable.

c.

Yes, the comprehensive concentration, load, and yield determinations (USGS: Baker et al. 2014) sufficiently capture N & P loads to the bay. The study is one of the most detailed nutrient spatial estimation efforts worldwide and covers a timespan of more than two decades. Quality assurance efforts such as comparisons of calculated and measured values (USGS Tables 8 & 9, Figure 10) support this conclusion.

However, atmospheric deposition may account for 22% (Wieben and Baker, 2009) to 34% (Hunchak-Kariouk & Nicholson, 2001) of the total nitrogen load (Rutgers Report page 39) and probably should be measured and considered for potential future inclusion.

- If atmospheric deposition was included in the USGS or Rutgers studies it should be clearly stated. I don’t think it was.
- As USGS 2014 p14 states: “an understanding of nutrient cycling from atmospheric and watershed contributions to biotic uptake and degradation and sediment processes is needed” and should continue to be an overall program goal into the future.

5. ***Do the included condition variables include all important parameters of interest regarding the bay’s condition? Is it important or useful to have any estimates of microbial loop or secondary production (e.g., if only for the bay’s herbivores [clams])?***

a.

It seems that there is adequate coverage of the condition variables for this early point in the process of developing indicators and thresholds for BB-LEH. However, there should be a discussion of the data on and importance of benthic chlorophyll. Given the well-established benthic vegetation community there, it is apparent that there should be sufficient light reaching bottom in the shallow community to support potentially significant microphytobenthos mats. These are good indicator variables and also can be key agents of biogeochemical influence, important terms when developing models of functional connections in the system. Benthic infauna are key piece of the picture often used in developing indices of ecosystem health and biotic integrity. There is only one year of benthic invertebrate data in the current study, a secondary dataset from the REMAP study- additional effort should be directed to monitoring of important benthos.

Mercenaria are good representative indicators of a sensitive resident species that is impacted by nutrient enrichment. Indeed they seem to have been- the decline in hard clams is striking and eutrophication would be a clear candidate for causing such a dramatic trend. However, no data are presented concerning other potential factors such as unrelated species shifts, temperature change, disease, predators, invasives, overharvesting. I think eutrophication is a likely dominant reason and the connection between Aureococcus blooms and loss of filter feeders is there, correlatively. The data show the threshold is reasonable based on literature, but again, site-specific data is preferable. There are likely other potential indicators of importance such as Argopecten, the bay scallop. However, there were insufficient periods of record, baseline data and insufficient presence of individuals to make a contribution to the suite of indicators in the IE. Additional years of monitoring data should be conducted to try to bring at least one key indicator higher trophic level into the index.

There is no inventory of or even mention of other higher trophic levels such as fish, larvae or other nekton and these may provide indicator information, especially if the habitat is as impaired as indicated. Such organisms would represent a good integrative indicator of habitat quality over time- and creel census or trawl data may be more readily available for nekton than dredge or grab data for hard clams and scallops. This I find somewhat surprising because the ASSETS model which this study relies upon specifically includes analysis of recreational fishing and cites the wealth of NMFS data on fish in Barnegat Bay, specifically summer flounder.

It is acknowledged that smaller steps must be taken initially in estuarine characterizations- some aspects of this system have been understudied in temporal (period of record), spatial and functional scales and the complexity of incorporating motile higher trophic levels into the monitoring program and the development of the index (with a multitude of variables that determine fish year class, fishing pressure, etc.) extends far beyond nutrient enrichment effects.

Expending additional monitoring effort on labor and cost-intensive variables would provide important insight but would have to be carefully considered. This would be a recommendation for increased monitoring of higher trophic levels, were such an expansion feasible. Microbial loop assessments may enhance understanding of system function and nutrient processing, but are likely of tertiary level importance- such assessments should be of lower priority and deferred until after the initial index is proven, validated and stabilized.

b.

All condition variables for which sufficient spatial and temporal data exist are likely included in this study.

Additional variables could be investigated and collected as a component of the monitoring program suggested above under "Potential future improvements to the index project." [*NEIWPCC Note: The Potential future improvements to the index project section comments from this peer reviewer are included under Section II: Additional and Summary Comments.*]

c.

As noted above, there are important variables that were not included in the Index of Eutrophication, such as phytoplankton community and benthic invertebrate community, but the reason for this is that these components are not monitored. There are no data, or the data are insufficient, to include these components in the index. Benthic invertebrate production can be very important, as it is the base for many fisheries species, but biomass is needed to estimate this production. In essence, the index developed for BB-LEH is more an Index of Eutrophication (nutrient enrichment) than an Index of Ecosystem Health. Seagrass and associated parameters is the only major biotic component of the index.

6. ***Given the methodology used to derive a unit-less score for the index, the index assessment for any given year is opportunistic (limited by the data available for a given year) and not deterministic (informed by data from the full suite of prospective relevant factors). As a result, the importance of setting thresholds against which observations are compared to determine the assessment cannot be overstated. As the value for each threshold is one of the most important elements in determining the outcome of applying the index, it is essential that the threshold values be solidly based in science. In other estuary studies, the causal thresholds (for nutrients N and P) were selected based on modeling the relationship between the causal factor and the response variables in the particular waterbody, which is appropriate because the fate and transport of nutrients will vary given the physical/chemical/biological dynamics unique to that water body. Here, the causal thresholds were selected before that modeled relationship has been determined. Does this limit the study's use for management purposes, developing strategies to target specific levels of N and P that would be expected to result in support of a healthy ecosystem for this estuary?***

a.

As long as appropriate threshold values are selected, picking causal thresholds before the modeled relationship has been determined should not limit the study's use for management purposes. However, providing justification or supportive reasoning would strengthen a case for deviation from widespread standard procedures.

b.

Thresholds should be selected based on biologically relevant relationships between stressors and response measures. Usually a reference data set is examined for relationships and to set the thresholds. Here thresholds were based on habitat requirements for living resources (seagrass, fish), and were based on literature values modified for this particular estuarine system. Setting thresholds based on literature values for similar water bodies is not unusual because linkages between stressors and effects may not have been established. It is stated in this report that analyses of the existing database were conducted and a process was established to help identify thresholds. However the results of these analyses or process are not presented. The threshold selection procedure does not limit use of the study for management purposes, but clearly, more detailed information about the particular method used to set thresholds is needed.

c.

The study is limited in several areas by the availability of specific data, a shortcoming that is acknowledged by the authors. The selection of the thresholds prior to modeling the variables is concerning and represents potential source of error in applying the IE. Absent sufficient data to develop modeling analyses of cause-effect relationships, the authors used what they had access to, and drew informed conclusions based on weight of evidence. As a result, at a minimum, several additional years of validation of the existing framework are warranted until such time as testable models can be implemented to support the conclusions drawn and thresholds established here. Therefore I think the utility for management purposes could be compromised by the broad of range of uncertainty inherent around the stated thresholds. ("Compromised" may be too strong a word but application of the criteria should be done with extra caution until additional data are evaluated.) The thresholds seem to be appropriate and approximate to expected values but additional work is required to verify. In addition to more monitoring information and validation exercises, targeted experimental and process measurements should be done to determine rates of important terms in the nutrient budget as mediated by abiotic and biological processes.

<u>Selection of Thresholds to Define Condition:</u> ***The basis for selecting the thresholds is given as literature, data analysis, best professional judgment (BPJ) or a combination of these factors. Is this a supportable basis for selecting thresholds that would be used to make condition assessments and inform management options designed to effectuate improvement in condition? Specific concerns and questions:***

1. ***Is there sufficient information within the study report to show that there is enough Barnegat Bay data to determine each of the threshold indicator values? Has the report addressed which indicators relied more heavily on BPJ or literature and should be revisited when more Barnegat Bay specific data for that indicator are available? Is the report detailed and transparent enough such that the reader can reproduce all steps taken to get to the conclusions provided?***

a.

The authors model their approach on NOAA's NEEA study using the ASSETS model (Bricker et al. 1999) and expand upon that methodology. This construction is a proven national standard and has been useful in assessing estuaries across many types and regions. The expansion of NEEA's model of five variables to a multi-layered system using 20 variables is ambitious. Applying a numeric scoring system to evaluate components and in aggregating to the final index is an enhancement to the NEEA model that improves precision in following the BB-LEH system over time, but requires more precision in the data to be useful.

The current report contains copious references to primary and secondary data from BB-LEH or the surrounding region from which the work is developed, as well as reference to many literature sources when local data are not available. The authors do a good job of indicating the limitations of the primary data and of justifying the literature sources or BPJ. As detailed elsewhere, the issue is the possible over-reliance on these external information sources where the local monitoring data are wanting. A plan for improved monitoring data collection is recommended and a method for transitioning to a more locally-based analysis is described by the authors. However, the grand accumulation of studies and data from around the mid-Atlantic, east coast and nation, does not enable a clear representation of thresholds that would apply in a particular way to BB-LEH. The studies provide general trends with a lot of scatter and variability and it is premature to set thresholds for one estuary based on these general regional trends. Indeed, in much of the site-specific data for BB-LEH is equally difficult to divine a trend. Figures 3-13 and 3-16 show almost no trend at all in light response factors with increasing nitrogen loading rates.

The report does adequately identify the variables that rely on BPJ or literature values. This is the notion that there is convergence of relationships and processes in estuaries that can be considered common across many estuaries of similar type. There is a passing reference to the ontology of estuarine type and how BB-LEH fits into that spectrum: shallow, confined and so forth. The process employed here is very good at getting within range of a true quantitative characterization, but the data and assumptions may fall short in terms of achieving specific numeric thresholds that accurately identify change points and stable states for this estuary. This will have to be verified in years to come.

b.

The report states that thresholds were defined based on a thorough examination of the literature review, analysis of the assembled database for calibration to BB-LEH, best professional judgment, or a combination of these factors. To this reviewer, it looks like most thresholds were selected based on literature review augmented with best professional judgment. Analyses of the assembled database (e.g., change-point analysis) may have been performed, but the results of these analyses are not presented, except for regression plots between indicators and loadings (Figs. 3-15 through 3-20). It is not clear how much data analysis contributed to threshold selection. The literature reviewed is discussed, but it would have been useful to have a table showing the thresholds that elicit biotic response in each of the studies examined. Also, these studies used different units of measurement. For example, the report mentions Latimer and Rego (2010), who “found that at <50 kg TN loading ha-1 year-1, seagrass extent was variable and likely controlled by other ecosystem factors unrelated to nutrient loading”. This is 5,000 kg TN loading km2/yr, 100 times the lower reference threshold selected for BB-LEH. Relationships between WQ indicators and TN loading (Fig. 3-15) are inconclusive. For example, T and DO vs. TN regression lines are flat, but it looks like there might be a trend if the data were analyzed separately for the central and south segments. Relationships between light variables and TN loading (Fig. 3-16) are also inconclusive. It is impossible to conclude anything about light variables and TN loading for the central and south segments from Figure 3-16, and it looks like the north segment may have relationships of its own. For example, an increase in chlorophyll a at about 8,000 kg TN km2/yr is suggested for the north segment, and the epiphyte/sav ratio appears to decrease at about the same loading. Thus, the final thresholds for TN and TP do not follow from the "above observations and analyses", which is not to say that they are off target, but that they should be based on a more objective analysis. Again, a table showing thresholds from the literature review and those observed from the analysis of the calibration data for each variable, would help in understanding the choice of final thresholds for BB-LEH, and would be more scientifically defensible. In some instances, assessments of literature thresholds are made, but these thresholds are dismissed. For example, in the Maryland coastal bays, biologically relevant thresholds for dissolved oxygen were established so that dissolved oxygen concentrations greater than 6 mg/L met objectives, and concentrations greater than 7 mg/L were better than objectives. For the BB-LEH, concentrations of 7.5 mg/L receive a score of 25, essentially borderline between good and bad. It is not clear why this particular threshold was selected. Nor it is clear why a threshold of 40 µg/L was chosen for TP concentration. Does seagrass biomass really decrease "markedly" at total phosphorus concentrations greater than 40 µg/L (Fig. 3-33)?

Some of the relationships examined are contrary to expectation, so pointing to these relationships to build support for a particular threshold does not make sense. For example, macroalgae percent cover decreased with increasing temperature (at least for the central segment, not clear for the south segment, Fig. 3-22). Is that good? Should high temperatures receive high scores based on this relationship? Another example: the ratio of epiphyte biomass to SAV biomass increases positively with dissolved oxygen (Fig. 3-27). If we want to limit the amount of epiphytes growing on the leaves of SAV, should we act to reduce oxygen concentrations in the estuary?

The threshold scale (values rescaled into scores) is a relative scale (it is possible to have values that produce scores outside this range), and thus it is taken as a reference scale. However, reference conditions have not been established. What should the condition of seagrass beds be in a restored estuary? Uncertainty associated with reference conditions is acknowledged in the report, and it is stated that "assessments were adjusted based on literature values of seagrass biomass". However, it is not clear whether or how the adjustments were made.

c.

The threshold indicator values are based on data and effects from a wider geographic range than just BB-LEH, an approach which is probably less uncertain than relying only on BB-LEH data and effects. The report would benefit from a more concise table of selected threshold values for each indicator. I would guess that, in combination with the SAS code provided in the appendices, a SAS-educated reader wouldn't have much trouble reproducing most or all of the steps.

2. ***Are you aware of any other significant data/studies that are relevant and should be included or referenced in this study and should have been used to help determine the threshold indicator values? Please explain fully.***

a.

I am not aware of any significant data/studies other than the studies already referenced in the report. Nevertheless, it would be expected that a study of this kind would have exhaustively examined the literature relevant to this project. The results of this literature review should have been included in a section of its own, accompanied by tables summarizing the literature examined, the literature thresholds, and (especially relevant to index derivation) the indices and methods of index development employed in other index development efforts.

b.

There are several studies nationwide that have engaged in attempts to establish thresholds, indexes, criteria and relationships between characteristics of the estuary and eutrophication status, showing varying degrees of progress. The present work on BB-LEH is among the more comprehensive of these analyses. All of these works point out that although general trends that can be observed across estuaries of similar a type, each estuary is specific and requires a history of data to develop predictive models to identify critical points with any precision. General guidelines for the kinds of information that are needed to develop site-specific indicators can be found in a USEPA report cited by the authors (Glibert et al. 2010) as well as the NEEA versions (Bricker et al. 1999 and Bricker et al. 2003a, 2003b) and work by Nixon et al. (2001). Several studies undertaken by state management agencies are attempting to proceed with this general approach. These examples include efforts in Mississippi (MS Dept of Env Quality 2004), Florida (FL Dept of Env Quality 2013), California (Sutula et al. 2011), Maine (Bureau of Land and Water Quality 2008) and New Hampshire (NH Dept of Env Services 2009). Each of these reports reflects the local characteristics of the estuary(ies) being assessed as well as general commonalities across estuarine systems. A consensus is forming on a national scale around elements and approaches to setting estuarine thresholds and eutrophication assessment. The current study on BB-LEH is in the forefront of use of these approaches.

c.

I'm not sure that the selected DO thresholds go low enough. The thresholds should cover the entire range of conditions and, based on papers such as Dauer et al. (1992), Diaz & Rosenberg (1995) and Diaz (2001) I would think 1 mg/l, 2 mg/l, and 4 (or 5) mg/l would be more appropriate thresholds.

Derivation of the Index: ***The derivation of the index relies on a Principal Components Analysis (PCA) and a series of manipulations involving raw data values and weighted scores. Is the derivation of the index in the manner indicated supportable? Specific concerns:***

1. ***Determination of index values blends raw scores (comparison of average of raw data to a selected threshold) and weighted scores (square of eigenvector value, considering the factors for which there were data in a given year). Weighted scores simply represent a measure of the variability of the factor, if it is present within a given year. If there are no data, the factor is given no weight. What is the purpose of blending the weighted score with the raw score, and is this a valid approach?***

a.

The utility of blending a weighted score and a raw score is not intuitive. The intention is to incorporate a measure of the variability of the component into the contribution of the component to the overall index. The stated reason for this step is to identify highly variable factors that might be more worthy of monitoring scrutiny, and so should be given more weight. However, the inclusion of the variability as a weighting factor for the score would seem to remove the option by the manager to evaluate priorities based on variability him/herelf because it is folded into the overall score. Without decomposing that score, the manager would not know the relative importance of the factor itself versus its variance in the long term; the index “operator” would know how the scores combined to attain a final score but in transmitting information to others utilizing the scoring system or making decisions, or even to the public, this would then involve reporting three scores for each index component to allow keeping track of how the raw and weighted scores influenced the final score.

b.

Additional justification and exploration of alternatives for the suggested method is required for at least three reasons: (1) the assessment values of raw scores are based on the selected thresholds and this basis is lost or diluted by adding weighted scores (is more variability truly good?); (2) the authors do not show that adding weighted scores is necessary or that it makes a real difference; (3) which means this may be an unnecessary and unwanted complication.

c.

PCA has been used in other studies to select the most important variables for index derivation, and to define reference conditions. For example, PCA was used by Shin and Lam (2001) to select sediment chemistry variables to derive a marine sediment pollution index. Variables with loadings greater than 0.7 were selected and weighted by the loading value in proportion to the eigenvalues obtained from the PCA. In the BB-LEH study, the weights are the square of the loadings (range: 0-1) and weighted scores (range: 0-50) are calculated for each variable by multiplying the raw scores by the weighting. Then, the raw scores and the weighted scores are summed. This is done presumably so that the raw scores (essentially the normalized values of the variables) account for 50% of the index and the weighted scores account for the other 50%

of the index. The final scores for each variable (sum of the raw scores and the weighted scores) are not used. These do not vary from 0 to 100. They would only vary from 0 to 100 if the weighting was 1. To calculate an index for one of the components (e.g., Water Quality) of the Index of Eutrophication, the raw scores are averaged, and the average is multiplied by the sum of the weighted scores of the individual variables. This is in essence a weighted average. The raw average and the weighted average are then summed (range: 0-100).

Variables with high loadings in the PCA have more influence on the index than variables with low loadings. For example, let's say that the raw score for dissolved oxygen is 50 (excellent) and the raw scores for the other three variables of the WQ component each are 10 (bad). The average of these raw scores would be 20, and the WQ index would be 33 (bad). However, if the raw score for TP were 50 (excellent) and each of the raw scores for the other three variables were 10 (bad), the average of the raw scores would still be 20 but now the WQ index would be 56 (moderately good), even though dissolved oxygen in the estuary was low (below 4 mg/L). This is because TP is weighted at 65%. Thus the WQ index is driven mostly by TP and dissolved oxygen does not have much influence on the index. The report does not discuss why this is desirable. If TP were omitted from the WQ index, the index would be driven mostly by temperature and dissolved oxygen, which would be weighted at 61% and 29%, respectively. Years for which TP is included in the index (2000-2010, Table 3-23) should not be compared with years for which TP is not included (1989-1999). The weighting of the variables is different for each block of years and the Index of Eutrophication is influenced by changes in temperature and dissolved oxygen prior to 1999 and by TP after 1999. A change in variability and regression slope through index values after 1999 is noticeable in Figure 3-57.

2. ***The approach taken in using PCA is not standard and no documentation is presented to justify it. Typically, to develop an index using PCA, the scores of the first few principal components would be examined. If the first eigenvalue (score variance) comprises a large amount of the total variability, then the first principal component might be taken as the index. If weighting the index is desired then the first eigenvalue would be used as a weight. In this report, there do not seem to be any attempts to assess the adequacy of using only the first principal component. Is this approach valid? If not, what argument, further analysis, and documentation would justify this approach?***

a.

There are different approaches to weighting an index. The approach taken in a marine sediment pollution index is mentioned above. It appears that for the BB-LEH Index of Eutrophication the weights are the loadings (eigenvectors) of each variable in the PCA, scaled so that the sum of the squares equals 1. The loadings on the first principal component axis were used. Presumably the first principal component accounted for a majority of the variability in the PCA, but PCA results are not shown. A table showing eigenvalues, eigenvectors, and cumulative percentage of variance should be included in a PCA Results Section. If the second or third principal component accounted for a significant proportion of the total variability, then the approach taken would not be justified.

b.

There is not sufficient supporting justification provided for deriving the index in this way. The fact that something has greater variance is not an indication of greater importance as an indicator of eutrophication per se. The same analysis by a manager could be performed on the un-blended weighting factor- variance- without the weighting imposing its effect on the overall condition assessment. This can be remedied by requesting details to be provided by the authors to discuss the rationale for the derivation as performed.

The inclusion of a trend or change from previous years may be a key piece of information. But rather than including the overall variance to weight that factor's score, have the authors considered merely weighting by the per cent change from the previous year to establish whether a component is deteriorating or improving? This end may be satisfied simply separately reporting the trend of the indicator, as is done in other applications of estuarine indicators, as for example in the State of Florida (Doren et al. 2009).

c.

Although PCA is an excellent transformation for reducing dimensionality of data, its use should be justified. In particular, with standardized axes grounded in assessment thresholds, a simpler combination method (the mean of the indicator variable raw scores, for example) may be much simpler and more intuitive to interpret relative to condition assessment. Adding "weighted scores" to calculate a final score dilutes assessment interpretability of the final score. Why is higher variability good?

The authors should better justify the need for PCA and develop a way to select among alternate threshold and analysis options, as discussed in the "Potential Improvements to the Eutrophication Index" section above.

3. ***The approach taken in this report is to use the squared component of the eigenvector as a multiplicative weight for that component of the index. The justification is that this weight would be the variance of the component. Is this claim correct? If the variables had been standardized to a variance of 1, then there would be some basis for this, although correlations between variables would also have to be considered. The SAS* code in the appendices shows that no variance standardization was done during the PCA analysis and it did not appear to have been done before that. Should the use of multiplicative weighting not be justified, as well as this particular weighting method? Do these concerns affect the validity of the index's derivation, and what can be done to address them?*** [*NEIWPCC Note: Statistical Analysis System, a software suite developed by SAS Institute for advanced data analytics.]

a.
Some of the reasoning underlying this section is a bit opaque. The squaring of the eigenvalue as a multiplicative weighting term relates to the variance. I believe that additional information should be provided to the authors to buttress their support for this treatment. There is no discussion as to why the departure from standard practice is warranted. There is too much need to reference the SAS code in an appendix to find answers rather than having the entire rationale and step-by-step method clearly explained, with supporting documentation, in the body of the text. Multiplicative weighting per se is not inherently "blurring"- the main question is what the weighting is composed of. Here, that weighting factor is a derivation of variability that, though potentially useful, is also insufficiently justified.

b.
Yes, indicators that show higher spatial and temporal variability have higher loading values, thus higher weights. These indicators have a larger influence in the index. The variables used in the Index of Eutrophication have been standardized to a common scale (0 to 50), so no prior normalization is needed to carry out the PCA. The weights are used to influence the more important variables in the dataset (i.e., those that are spatially and temporally variable), as illustrated above.

c.
As stated above, the authors should better justify adding the multiplicative weight to the threshold-normalized raw index score. The addition probably dilutes the ability of the index to interpret condition. Setting some expectation or criterion for evaluation of index performance that is related to eutrophication ecological responses would be a welcome enhancement.

4. ***The sole justification for combining the weighted and raw indices is that it integrates the multiple indicators and their variability. The advantage of this approach is not obvious and requires some justification and documentation. Would combining the two indices serve to blur any useful measure, or instead improve it? Do these concerns affect the validity of the index measures?***

a.
I am skeptical that the variance of the component should be as important in the derivation of the index as to merit position as a weighting factor in half of the score. Weight should be given to those factors which may have more amplified or multiplicative effect in the system, or are where the system is poised close to "tipping points," etc. The authors need to expound on their reasons for targeting factors with greater variance as being of greater importance in the index than factors of lower variance. It would not have been difficult to present an alternative analysis using the first principal component as the index and then provide a comparison of results to their technique with to legitimize the non-standard approach taken. The end result is that the power of the raw index may indeed be reduced by the "blurring" effect. This is not to say that it is a definitive diminishment of the index, but the process of combining essentially two different

measures of each variable may serve to lessen the efficacy of each. Without doing a systematic analysis of each variable and the effect of the derivation of the index, the extent of this effect is not known, but the authors may have done such an analysis and it would serve the user well to see it.

b.

According to Table 3-23 (weightings used to calculate the weighted scores for indicators), the final index uses 3 components: Water Quality, Light Availability, and Seagrass. HABs and benthic invertebrates are not used in the final index (despite some claims elsewhere in the report) because existing data are not deemed to be adequate for inclusion in the index. The final index score is a simple average of the component scores. In the end, the reliability of the index depends on the data that feed the index. These data were averages (or medians) of various collections taken over the course of the growing season at many locations within a segment. There appear to be differences among sampling events for some indicators. Considerable effort is given to describe these differences, especially for the seagrass indicators. Given the large amount of inter-annual variability observed in the scores of some indicators (Figs. 3-49 and 3-50), and the seasonal trends observed for some of the seagrass indicators (Tables 4-2 and 4-3), it is not clear how representative of conditions is the averaged indicator score that goes into the index. If the spatial or seasonal variability of the indicators is larger than the inter-annual variability, the final index score won't be very useful for tracking changes over time. What is needed is a measure of the confidence of the index, and whether the final index score calculated for one year is within the level of error of index scores calculated for previous years.

c.

I'm not going to repeat myself (again).

<u>Use of Index</u>: Objective 5 of this study is "To generate an Index of Eutrophication as a tool to evaluate future conditions using water quality and biotic indicators to assess eutrophication, eutrophic impacts, and overall ecosystem health of the BB-LEH Estuary..."

1. ***Does the study report provide enough information on how one can use the Index of Eutrophication to evaluate future conditions using newly acquired water quality data? Is the report detailed and transparent enough such that the reader can reproduce all sets taken to get to the conclusions provided?***

a.

I believe that the report is complete enough, if not fully transparent, to allow a reader to both reproduce the index and assemble data needed to incorporate additional years of measurements. The section of the report describing the method is somewhat convoluted and it is not obvious how simple the derivation would be without going through the exercise of repeating all steps to create and populate the IE. But the road map and step by step information is complete. In fact it is replete with the nuts and bolts information. It needs more supporting information. The Index of Eutrophication is a complex assessment tool that requires a sophisticated monitoring program and many years of data to back it up. If the tool is to be applied in the near term, possibly as a management program is ramped up, it may be advisable to use a subset of parameters that are more tested as robust ecosystem indicators.

b.

I would guess that, in combination with the SAS code provided in the appendices, a SAS-educated reader wouldn't have much trouble reproducing most or all of the steps and results.

c.

Surprisingly, the Index of Eutrophication is not calculated for 2011, the year reserved for index validation. The reasons why the index was not calculated for 2011 are unknown. The whole validation section is an evaluation of the data collected in 2011. Was 2011 a typical year? Validation would have been more useful for years that differ from the norm (e.g., wet vs. dry years), to see if the index behaves according to expectations. The 'validation' appears to be against the NEEA assessment (nothing to do with 2011). The NEEA assessment documented that BB-LEH had "high overall eutrophic condition" in 2007, whereas the Index of Eutrophication was "moderate" in that year. We cannot conclude much from this comparison. In any case "highly eutrophic" and "moderate" is quite in disagreement. What the index does in 1999 (pg. 123, second paragraph) is irrelevant for this comparison. The Index of Eutrophication quantifies status and trends in the estuary relative to watershed pressures. One aspect of the validation could have been a comparison between changes in TN and TP loadings and changes in the status of water quality and biotic indicators for the validation year. However, TN and TP loadings changed very little across years in the last eight years, so it would have been difficult to assess a change in the index relative to a change in nutrient loadings. Perhaps more than one year (years differing in nutrient loadings) should have been set aside for validation. Figure 3-58 does not provide a good validation because the north segment is so different from the other two segments. Relationships between the index and the nutrient loadings should have been examined separately for the central and south segments, so that the response of the index to changes in nutrient loading in the good to moderate range can be assessed. A measure of the sensitivity of the index to changes in watershed pressures, and 95% confidence limits are needed.

The report should include a section on how to apply the index in future years using newly acquired data; what to do if data are unavailable; and more importantly what should be done if new data are collected to monitor other components of ecosystem health, such as hard clam abundance or benthic invertebrate condition. Appendix 3-7 may be the only guide that is needed to calculate the index in future years, but this needs to be stated.

2. ***In your opinion, what are the weakest and the strongest aspects of the Eutrophication Index and the Threshold determinations? Please make suggestions on how the weakest parts can be strengthened.***

a.

The variety and quality of the data that were used to derive the Index of Eutrophication is impressive and certainly one of the strengths of this study. Data documentation could have been much better, but that's a problem with how the report is organized (discussed below). The index does not include components of ecosystem health (e.g., benthic invertebrates, fisheries species) other than seagrasses. The seagrass data are excellent, even if no goals for what the areal extent of seagrass beds in BB-LEH should be have been established (e.g., compare to Chesapeake Bay goals). The approach to deriving the index sounds reasonable, even though important questions remain to be answered (see above). The weakest element is probably not being able to place confidence in the values of the index, and to determine how sensitive the index is to changes in watershed pressures and how much is just random noise. The individual station data could be used to calculate error terms for the assessment.

b.

The potentially weakest aspects of the Eutrophication Index are (1) addition of the "weighted score," and (2) use of PCA to combine previously "lightly" reduced (mean or median) data. They may unintentionally weaken the assessment ability of the index, and the latter may also be more complicated than desirable or necessary.

c.

One of the weaknesses of the analysis is the lack of a detailed understanding of residence time and circulation. Development of a hydrodynamic/water quality model should be initiated to provide better fate and transport tracking, as well as bay circulation information and residence time calculations under a variety of conditions (wet years, dry years). Airshed modeling or wet and dry deposition data should be acquired over several years to allow users to set general bounds on this potentially important term to distinguish watershed-level controllable variables versus inputs that are not controllable by managing the watershed.

As for strengths, the authors have done a commendable job of vetting and analyzing the data that they have, organizing and assembling it into a coherent whole and developing a credible analysis. I think the Index of Eutrophication is quite useful as starting point, follows established ecological (if not all statistical) concepts has shown initial validation. The report, while serving as a management tool in itself, also represents a good inventory of scientific knowledge about the bay and serves well as a gap analysis and template for guiding future monitoring priorities. The first-steps taken in this project will serve as an excellent foundation for further testing and strengthening a robust index that will well-serve management of the estuary.

3. ***Are there any elements missing from the Eutrophication Index which you think need to be included or which would strengthen the tool? Please explain fully.***

a.

A regularly conducted and well-designed benthic monitoring program with assessments by a well-constructed benthic index would probably improve the eutrophication index. The Chesapeake Bay B-IBI has proved capable of identifying sites declining due to eutrophication induced abundance and biomass increases. The most common criticism of indices such as the NCA-stimulated AMBI (Gillett et al., in Press) is response to eutrophication rather than toxic pollutants.

b.

As mentioned above, phytoplankton, benthic invertebrates, and fisheries species components (hard clams or other species) should be evaluated for inclusion in the Index of Eutrophication if data become available. Phytoplankton communities are important components of estuarine food webs. Because of their short life cycles, phytoplankton species can rapidly integrate the effects of nutrient and sediment loading and the effects of grazing from higher trophic levels, so they provide a holistic assessment of estuarine health (bottom up and top down controls). Phytoplankton measures could include shifts in species composition in addition to abundance of harmful species. Benthic macro-invertebrates integrate conditions over time and respond to different sources of stress (e.g., nutrient enrichment, low dissolved oxygen, toxic pollution) in predictable ways. They are sensitive indicators of ecosystem change. These types of indicators would augment the biotic response component of the index, now assessed only by seagrass indicators. Although seagrasses acutely respond to human-induced pressures in the estuary, especially changes in water clarity, they are also sensitive to global changes, such as changes in water temperature and carbon dioxide in the atmosphere. These are external influences that cannot be managed by controlling sources of eutrophication within the estuary. Thus a more robust, corroborative suite of biotic indicators would provide a more accurate assessment of BB-LEH.

c.

In terms of missing components, I believe that several relationships need to be fleshed out- how seagrass responds to epiphyte colonization, to turbidity, to different components of turbidity, especially chlorophyll a, but also FDOM and inorganic particulates. I think that process-oriented measurements should be made- denitrification, DNRA and N sequestration and burial, for example- to establish order of magnitude bounds on the importance of those processes in removing N from the system. Bioassays should be done to develop phytoplankton dose-response curves for additions of species and N and P. The datasets do not extend far back enough to view the system in a "reference condition," nor apparently is there a similar estuary type in pristine condition from the biogeographic province that can be used. Beyond the aforementioned assessment of higher trophic species as potential indicators, and benthic mats and infauna as also mentioned previously, I do not think that additional variables are required for consideration for incorporation into the IE.

4. ***The Estuaries and Coasts article Mind the Data Gap: Identifying and Assessing Drivers of Changing Eutrophication Condition (Fertig, et al.) identifies grouping the variables into three major categories, one of which is seagrass, to develop an index of eutrophication; however, there are no seagrass data available for the first 15 of the 25 years of data used to develop the index. Thus, can we be confident in using and applying this index?***

a.

Indices should not be endpoints in estuarine assessments, but tools that allow us to summarize complex data into one number that can help guide management actions. Only recently we have become aware of the value of long-term monitoring. Monitoring is expensive, often opportunistic, and perfect data collections usually do not exist. There is nothing we can do about the lack of seagrass data for the first 15 years of the data series analyzed for this study. Does this mean that we should abandon all efforts to develop an Index of Eutrophication that includes a seagrass component? Seagrasses are very important in many ways to ecosystem integrity in coastal lagoons. Seagrasses are just one component of the index, and currently the only component of biotic response. The seagrass response index contributes 1/3 to the average of the final index. The individual components of the final index and their patterns can and should be examined separately for insight into the reasons for determining condition. One element that I think has not been well examined in this report is how individual components influence the final index, especially over the years for which data are unavailable for a particular component. Figure 3-57 should be discussed in this context.

b.

We cannot be as confident in the seagrass component of the IE as in other parts of it. The dataset for this highly variable ecosystem component is small. Seagrasses are known to migrate around within their home range, Ruppia particularly, expand and contract and have seasonal cyclical increases/decreases in biomass in response to a variety of factors, not necessarily associated with eutrophication. This is borne out in the data which show high variability and lack of clear trends in the few years of data availability. Authors admit that the remote sensing validation exercise revealed some patterns of seagrass unexpected increase and other patterns of decrease that were explainable by changes in sediment geomorphology due to shifting sand bedloads. Simply stated, more years of data are needed to create a more robust seagrass component of the index. This represents a good start.

c.

We have to choose between (1) using the best and most complete data available for today and the future, with 15 years of >10 year old potentially inaccurate assessments, or (2) throwing out important data and having inaccurate assessments forever because we knew less 25 years ago than we know now. The choice is obvious and yes, we can be confident using and applying such an index.

5. ***Does the approach used here "validate" the developed eutrophication index?***

a.
Validation of the IE is claimed by assessing the quarantined 2011 dataset and by comparing to a NEEA analysis of the same data. The index construction and conclusions do hold up when evaluated in this validation exercise but one year of validation data do not a fully validated model make. Further years of data acquisition and validation should be required before fully embracing either the approach or the specific of the index. These results are promising and I believe will be justified in terms of concept and implementation of protocol. The numerical ranges and thresholds may need to be tweaked and additional variables should be considered for incorporation into the index as detailed above.

b.
A validated, unambiguous approach to evaluating eutrophic condition in the estuary is essential. I don't think this study has yet adequately assessed validation.

c.
The approach used here validates that the proposed eutrophication index behaves in the same way with independent 2011 data as it does with the roughly 20 years of data used to develop it. A more comprehensive approach would validate the assessments based on observed effects and assessment categories across a wide range of conditions.

Overall Adequacy of the Report

1. ***Is the organization of the document appropriate and does it present the material in a clear and concise manner? Please explain fully.***

a.
In general, yes. See the "Editorial Improvements" section above for a few potential improvements. [*NEIWPCC Note: The Editorial Improvements section comments from this peer reviewer are included under Section II: Additional and Summary Comments.*]

b.
I find the report to be comprehensive, generally complete, well-written, mostly transparent, mostly well-supported and with much good and useful content. Yet it is also uneven in the presentation of some analyses, based on assumptions not fully explained and has some organizational problems that make it less- readable, as described below. In terms of adequacy of addressing the goals, the work is a very positive step toward understanding enrichment impacts and defining thresholds and indicating steps needed for remediation in BB-LEH.

c.

The report is not a careful assemblage of material. Beyond the general sections (Introduction and Components 1-4), subsections are often redundant and material is scattered throughout. The text is extremely repetitive. The same statements are repeated over and over throughout the document, sometimes verbatim. For example, the last paragraph of pg. 121 (Component 4: Epiphytes) is repeated word by word in the last paragraph of pg. 124 (Component 4: Conclusions). Subsection headings often do not flow logically from previous headings. Sampling design, sampling methods, and index methods are given piecemeal over various subsections. The text is cluttered by numbers to the point that connections and key points are difficult to make. The methods for indicators (temperature, etc.) and the methods for index components (WQ, etc.) are comingled, and this creates confusion. There are numerous errors, mostly related to calling the wrong figure or table in the text. The USGS report (Appendix 1) by comparison is a carefully crafted document. The following suggestions may improve the BB-LEH report:

a) Make statements once and then recap key findings in a summary section. It will reduce substantially the size of the report and make it more concise.

b) Put key assessments and methods in one place. Document methods with a logical flow, so that sampling design (number of stations, years), parameters, and index component information is complete and can be fully understood at every step.

c) Provide numbers and statistical results either in the text or in the tables, but not both. Make key points in the text without going through every single number that is shown in a table. Make sure that numbers in text, tables, and figures agree.

2. *Are the stated objectives adequately met? Please explain fully.*

[NEIWPCC Note: The seven key objectives of this study are listed below for reference.

1. *To document the influence of human altered land use on past and present nutrient export from the BB-LEH Watershed to the BB-LEH Estuary using physical and chemical watershed data and land-use patterns, and spatially explicit models.*
2. *To determine if nutrient loading quantified by subwatershed and biotic response is stable or is temporally and spatially variable.*
3. *To quantify baseflow, runoff, and total nutrient loads and to determine the relative importance of turf area coverage.*
4. *To determine estuarine biotic responses to the loading of nutrients across a gradient of upland watershed development and associated estuarine nitrogen loading, and to identify key biotic responses across a variety of estuarine organisms by examining shifts in phytoplankton, benthic macroalgae, seagrass, epiphytes, benthic invertebrates, and shellfish structure and function. Each of these parameters will be examined and assessed for statistical validity and inclusion in the index development for the 1989 to 2010 period.*
5. *To generate an Index of Eutrophication as a tool to evaluate future conditions using water quality and biotic indicators to assess eutrophication, eutrophic impacts, and overall ecosystem health of the BB-LEH Estuary and to develop threshold levels of biotic decline and numeric loading criteria that can support an effective nutrient management plan.*
6. *To apply a conceptual model of eutrophication and determine if ecosystem structure and function have been altered in the BB-LEH Estuary.*
7. *To document the current biotic and seagrass habitat conditions of the BB-LEH estuary at the end of the investigation using the most recent biotic data collected (2011) and index methods developed from data collected through 2010.]*

a.
The Objectives are stated in the Introduction Section. Seven objectives are listed. Objectives 1-3 are met. Objective 4 is partially met. The objective to identify key biotic responses across a variety of estuarine organisms is met for benthic macroalgae, seagrass, and epiphytes; partially met for shifts in phytoplankton (Aureococcus anophagefferens, only species examined) and shellfish structure and function (paucity of data; one assessment comparing 1986-87 to 2001); and not met for benthic invertebrates (problems with EMAP data; data collected by NCA not summarized). Objective 5 (to develop an Index of Eutrophication) is largely met, although significant questions remain as to the performance, sensitivity, and application of the index (see above). Not sure about Objective 6. The conceptual model was applied, but whether a determination has been made about ecosystem function is not clear from the observations presented in the report. Finally, Objective 7 is partially met. Current biotic and seagrass habitat conditions are documented for 2011, but index methods were not applied to 2011 or validated.

b.
Yes, with no reservations.

c.
The study moves quite far toward supporting the goals of identifying key factors causing eutrophication, placing quantitative bounds on environmental condition and pointing to concrete ways to ameliorate problems and restore the estuary. This work is obviously the first step in a long process, but it is a worthy step. Additional validation of the index and further refinement is warranted. Although the scope of this study is specific to BB-LEH, application to other estuaries in the New Jersey coastal zone (and beyond) is possible, with modification, which would make this work even more valuable. As a conceptual and methodological framework for constructing an index for this type of shallow, slowly flushed estuary, it is a useful model. But the wider applicability of the index in direct use for evaluation of other estuaries is currently beyond the scope of this study, as the authors themselves note.

The work of both NOAA’s NEEA estuary assessment program and the EPA’s national nutrient criteria standards needs to be supplemented by work such as this, as the national level view encompasses such a large dynamic range. The application of nationally-based index thresholds or criteria to specific estuaries would be too generalized without the local level guidance provided here. These studies “on the ground” in local and regional areas are an essential part of the larger goal to develop estuarine nutrient standards, indicators and criteria for the nation.

3. Do the results from the study support the authors’ conclusions and recommendations?

a.

Yes. However, there are a few alternatives, including some index simplifying alternatives, which should be explored before settling on the current index. They are discussed in the “Potential Improvements to the Eutrophication Index” section.

b.

The state of the estuary is documented and synthesized in the Summary and Conclusions Section of Component 3 (Index Development). These conclusions are further discussed and augmented with references from other studies conducted in the BB-LEH estuary in Component 5 (Synthesis and Management Recommendations). The main conclusion of the study, that “once loading increases beyond 2000 kg TN/km2/yr or 100 kg TP kg/km2/yr eutrophication condition reaches a new, lower steady state of poor condition”, is made from observations in the north segment (Fig. 3-58), because only the north segment exhibited such high nutrient loading. The north segment received some of the lowest scores of the Index of Eutrophication, but some years scored within the moderate condition range. A decline in the condition of the north segment with increasing nutrient loadings beyond 2000 TN kg/km2/yr has not been proven, only that lower scores overall occur there. We do not know what the condition of the north segment was historically. Also, not all the indicators for the north segment were bad. The north segment had moderate to excellent light availability and moderate WQ (Figs. 3-54 and 3-55), although percent surface light, TN, and TP were generally bad. The seagrass condition index was not applied to the north segment. Therefore, this main conclusion is tenuous and should be taken cautiously.

c.

The authors clearly have excellent knowledge of the BB-LEH estuary and have observed in anecdotal and data trends the decline in several measures of ecosystem quality. This attempt to develop a quantitative, though somewhat unorthodox, index is positive, overall. Additional justification is required as to the reasons for some departures from common methodology. Specifically the incorporation of the eigenvalue as a variability term in the index itself raises questions. Modeling analysis is required to evaluate the utility of this derivation versus more standard methods of developing indices based on PCA. High variability in parameters, per se, may not represent a reason to give greater weight to a particular characteristic of the estuary. A cleaner presentation to management may be offered by separating the two components of the index (the weighted and the raw scores) to allow judgments about priorities and management action based on variability separately from the impact of the indicator.

Finally, the results of the study do support the authors’ conclusions and recommendations. The authors are to be commended for doing a thorough and innovative job with an imperfect supply of tools and data (as ecological data always are). I believe the authors have a good sense of the ecological trends in the system and make an accurate assessment of the current and future pressures applied, specifically regarding sources of enrichment occurring due to land use changes in the watershed. Specifics of biogeochemical processing within the estuary are the details which require further study.

II. Additional and Summary Comments

a.

The authors are to be commended for credibly accomplishing a very large, multifaceted, and complicated data selection, compilation, quality assurance, and analysis effort. The work is an appropriate first step that is more than sufficient to justify strong and immediate nutrient management actions. However, additional work would probably be useful to justify and possibly modify (a) one or two indicator thresholds and (b) index formulation. The present work should be considered an excellent first step based on available (asymmetric and incomplete) data and an opportunity to design a monitoring program that collects symmetric and complete data. After five or ten years of monitoring, these data can be used to test assumptions and revise thresholds, models, and TMDLs.

<u>Major Strengths</u>

The report is an excellent, successful, very large, multifaceted, and complicated data selection, compilation, quality assurance, and data analysis effort.

- The number of indicators, sources, datasets, variables, and data reviewed and the quantity of data accumulated for many indicators is very impressive.
- The steps taken to deal with incomplete, missing, and censored data and avoid spatial and temporal bias were also extensive, successful and impressive.
- Uniform axes grounded in assessment-related threshold values were created for about 20 indicator variables, potentially simplifying integrating indicator values into an integrated eutrophication index.

<u>Potential improvements to threshold selection documentation</u>

A few simple enhancements will make threshold selection and the data rescaling process much easier to understand.

- **Threshold selection and application principles.** The threshold section badly needs an opening section presenting an overview of threshold selection principles to help readers understand the basic principles of what was done to indicator variables. The present text overloads the reader with detail and obscures the common and basic principles. As best as I can determine, for each selected indicator, the process involved:
    - Pick three thresholds demarcating four impact levels at approximately equal intervals of ecosystem impact between the minimum (strongest impact, 0) and maximum (least impact, 50) for each indicator.

    - Impact intervals between thresholds were (approximately?) equal, but indicator measurement intervals were not. For example in Table 3-11, as with all indicator thresholds, raw score intervals were 12.5. Temperature intervals were uniformly 4°C while dissolved oxygen intervals varied from 4.0 to 1.0.
    - Raw indicator scores of 0.0 indicate maximum ecosystem impact and raw scores of 50.0 indicate minimum ecosystem impact.
    - The raw threshold score-indicator measurement relationship was used to construct an indicator and estuary segment-specific rescaling equation, which was subsequently used to convert indicator measurements to raw indicator scores.
- **Threshold table scores**. The threshold tables should include a decimal place on the threshold score to make it easier to understand that thresholds were picked at equal 12.5 score intervals. It is much easier to get this with scores of 12.5, 25.0, 37.5, and 50.0 than with 13, 25, 38, and 50.
- **Comprehensive threshold table.** A single comprehensive table presenting thresholds selected for all indicator variables should be included up front. The other tables for individual indicators with literature values and text estimating the quantity of BPJ included can follow.

Potential improvements to the eutrophication index

Establishing criteria for "success", eliminating addition of "weighted scores", combining raw scores without PCA, adjusting evaluations of data bias and adequacy, and including a better performing benthic index formulation are steps that can potentially improve the eutrophication index.

- **Establish success criteria.** The authors imply that eutrophication index performance is acceptable because it meets expectations, such as low eutrophication index values in the northern segment then in the center or south. It would be useful to explicitly state these expectations, including any necessary spatial and temporal specifications. The expectations can then be used to evaluate and compare the performance of alternate index formulations and choices.
- **Eliminate "weighted scores" addition.** Provision of additional justification for the selected index construction steps and exploration of alternatives is required for at least three reasons: (1) the assessment values of raw scores are based on the selected thresholds and this basis may be lost or diluted by adding weighted scores (more variability may not necessarily be good); (2) the authors do not show that adding weighted scores is necessary or that it makes a real difference; (3) which means this may be an unnecessary and unwanted complication. The alternative approaches should be evaluated and the better approach selected, based on expectations established in the previous section.
- **Combine raw scores simply (without PCA).** Having converted all the indicator measurements into standardized indicator scores on uniform indicator axes, combination of indicator scores in assessment accurate ways may be as simple as calculating mean values across indicators. Using PCA for axis combination may be an unnecessary complication. The

alternative methods should be evaluated based on the expectations established above, and the best performer should be selected. The authors should provide justification if the feel strongly that PCA is the preferred combination method.

- **Revise evaluations of data bias and adequacy with more emphasis on environmental reality and less emphasis on statistics.** The authors expressed serious concerns about using dissolved oxygen as an indicator due to bias due to potential sampling time of day and "index period" or seasonal bias citing a string of statistical references. However, Summers et al. (1997) found that results of ambient bottom DO measurements on vessel sampling visits were remarkably consistent with continuous in situ DO logs. The take home message is that environmental measurements may perform better than expected based on statistical assumptions.
- **Improve benthic monitoring and index formulation.** A regularly conducted and well-designed benthic monitoring program with assessments by a well-constructed benthic index would probably improve the eutrophication index. Benthic index approaches not included in this study have been used to detect eutrophication. For example, the Chesapeake Bay B-IBI has proved capable of identifying sites declining due to eutrophication induced abundance and biomass increases. Also, a common criticism of indices such as the NCA-stimulated AMBI (Gillett et al., in Press) is response to eutrophication rather than toxic pollutants, which is exactly what the present effort requires. The benthic measures considered for this project (EMAP Benthic Index, Gleason's Diversity, and "top-3 abundance") are unlikely to respond to eutrophication.

<u>Potential improvements to the eutrophication index</u>

The present work should be considered an excellent first step based on available (asymmetric and incomplete) data and an opportunity to design a monitoring program that collects symmetric and complete data. After five or ten years of monitoring, these data can be used to test assumptions and revise thresholds, models, and TMDLs. Atmospheric deposition should also be measured and evaluated for inclusion in the index.

- **Recommend monitoring program design.** The recommended management actions should include a long-term monitoring program that includes specifications for (1) the variables to be monitored and (2) the spatial and temporal sampling designs to be used to collect monitoring data. The design should try to solve the problems encountered when acquiring, compiling, and analyzing existing data for the current project. After five or ten years of monitoring, analysis of the entire body of data including (1) loads, concentrations, and yields, and (2) water column, biota and sediment eutrophication symptoms can be used to test testable assumptions and "tweak" parameters identified in the present study such as thresholds, models, and TMDL's. Presently, assumptions and parameters are based on (asymmetric and incomplete) that were available for the present study. Analysis of data collected using optimal designs can improve selections and strengthen confidence in assumptions.

- **Evaluate atmospheric deposition for monitoring and inclusion in the index.** Atmospheric deposition may account for 22% (Wieben and Baker, 2009) to 34% (Hunchak-Kariouk & Nicholson, 2001) of the total BB-LEH nitrogen load (Rutgers Report page 39) and probably should be measured and considered for potential future inclusion. As USGS 2014 page 14 states: "an understanding of nutrient cycling from atmospheric and watershed contributions to biotic uptake and degradation and sediment processes is needed" and should continue to be an overall program goal into the future.

Editorial Improvements

As with any efforts of this magnitude, there were minor editorial improvements that would make reading and understanding the report easier for the uninitiated.

- **USGS Concentrations, Loads, and Yields Report**
  - I couldn't find a definition of "yield" or descriptions of how yields were calculated in the report. Inclusion of the explanation in the Rutgers Report, Nutrient Loading Analysis Chapter, Base-Flow Yields on a HUC-14 Scale section, that yields are load values normalized by unit area would be very helpful to many readers.
- **Rutgers Assessment of Nutrient Loading and Eutrophication Report**
  - Pages 1-96 are not numbered, making difficult specific references such as the one above.
  - Several figures and tables have very small text and require a magnifying glass in order to get the message. Many look as if they were copied or scanned from other publications. It would be helpful if they were all legible.
  - It would be helpful if the Component 3 Introduction "Building on the NEEA Report" is written more like a methods section. The way it is written now is confusing and it is difficult to distinguish what was done here from what NEAA did. I recommend presenting what was done first, and NEEA rationale and comparisons later.
  - The three or four pages on REMAP and NCA benthic data in the "Evaluation Process For Inclusion of Secondary Data into the Index of Eutrophication" seems rather long and unfocused, especially since the authors consider them of limited value because (1) of the existence of only one year of spatially widespread data, and (2) perceived habitat gradient issues.
  - There are occasional misplaced page breaks and carriage returns which are confusing.

b.

Research Questions

The project can be seen as attempting to answer four key questions:

- **Is Barnegat Bay-Little Egg Harbor impacted/impaired?**
- **Is nutrient enrichment causing or contributing to the environmental degradation?**
- **What is the relationship between nutrient input and environmental degradation?**

- **What can be recommended to halt or reverse the proximate causes of the enrichment impact and the system-level cause of environmental degradation?**

The report successfully addresses each of these questions, to varying degrees. These answers require a number of tasks such as the parameterization of watershed models to calculate hydrologic flow and nutrient loading; the coordination of multiple agencies and datasets; integration of multiple components including watershed modeling and data collection, remote sensing and in situ monitoring; and the development, application and validation of an Index of Eutrophication (IE). The generally good success in addressing these tasks is to some degree limited by the lack of available data. There is need for continued and expanded long term monitoring, for additional analysis and empirical experimental data on ecosystem response to nutrient enrichment specific to this system. The goals would be strongly aided by a hydrodynamic model to describe fate and transport of nutrients in the estuary, to understand circulation regimes, and to derive flushing and residence times. Understanding the sources of nutrients requires additional examination of the groundwater input component, internal regenerative sources and fluxes at the system boundaries, particularly the marine end member component.

<u>Assessment of the Scientific Basis of the Report</u>
The authors organize the analysis into five components that culminate in synthesis and management recommendations. Presented here are summary comments on the adequacy of the report in addressing each of those components:

- **Component 1- Watershed Nutrient Loading**

The USGS report on watershed contributions to nutrient loading is based on sound protocols. The period of record of the dataset is about the minimum required to observe long-term trends and tease out inter-annual variability; the analysis meets the standard required to derive appropriate relationships. The USGS analysis of turf land use impact on N and P loading is cogent and well executed, indicating that developed land with turf is much more likely to export higher loads of TN and TP through the watershed to the estuary.

The report states on pg 44 (Component 1) that initial water quality data were collected from 1970-2011 USGS yet the appendix states that the period of record was 1980-2011. Am I misreading the POR of the data?

I have questions about the process for determining segmentation of the estuary. Often such divisions are based on an analysis of multiple factors to determine spatial differentiation based on multivariate analysis. I do not see that kind of analysis presented in support of the segmentation. In this case, the segments (North, Central, South) are done a priori, based apparently on professional judgment and on familiarity with the nutrient and other data. Sediment type and some other factors are mentioned. However, no evidence is presented to demonstrate the statistical basis for segmentation. This could undermine spatial clustering relationships if the divisions are locked into a spatial framework that precludes the discovery of similarities based on a different parsing of the data. Nonetheless, the evidence presented does support the existing segmentation utilized in the analysis and proceeds based on the assumption that the segmentation is justified. Given that assumption, the north basin is significantly higher in TN and TP, followed by the south and central, which is better flushed through a tidal inlet.

I am not convinced of conclusions based on residence time. There seems to be very little known, relatively, about the hydrodynamics of this important estuary. For this reason, water residence time cannot be used in the Index. Yet the authors make several inferences and draw outright conclusions throughout the document based on residence time calculations. There is no existing 2D or 3D circulation model of the bay. The study referenced for residence time calculations (Guo et al. 2004) is not found in the Literature Cited and so cannot be checked, although there are two other similar documents there. (Correction- Guo et al. 2004 is just misplaced in the bibliography, concatenated with another reference). The Guo 2004 study, which investigates circulation, is based on current and surface level measurements at a few locations in the estuary during parts of three seasons in a single year (1995). This small amount of data can supply a sketch of the relative magnitude of seasonal residence time at a point in time but hardly provides a multi-year picture of annual and seasonal average residence times that can be relied upon for making management decisions. The lack of a hydrodynamic model is not the fault of the authors, but caution is advised in utilizing a limited amount of information on which to base analyses of the fate of nutrients.

- **Component 2: Estuarine Biotic Responses**

Component 2 is the complex heart of the analysis; the job of characterizing estuarine biotic response is daunting; it is a weaker part of the analysis than the water budget and loading calculations simply because there is often insufficient data to do the complete job- 7 years of SAV data, 3 years of epiphyte data, 1 year of benthic invertebrate data, for example. The chlorophyll record is an acceptable baseline but many other biotic variables are under-sampled for the kind of analysis it is being asked to support, including USGS's own assessment of water quality data within the estuary.

Data collection from 2004-2011 is a very short period of record to apply ecological data for determination of thresholds and biotic responses of seagrasses. Although some variables such as DO, nutrients and chlorophyll have longer records, the "biotic response" analysis focused only on the years 2004-2011 and used only selected key biotic indicators to determine response- SAV, epiphytes and shellfish. Another seagrass study done by Lathrop and Haag (2011) in BB-LEH showed an increase in total area of SAV of 138 ha between 2003 and 2009.

As for nutrients, which enjoy a relatively long period of record, the data seem to show a fairly stable pattern of average N concentrations since 1989, with some spikes in peaks through the years but no dramatic increase in the overall mean (Figure 2-3). Notwithstanding, it appears that the concentrations have increased slightly over time, although the main driver of this pattern is the most recent 2 out of 3 years in the dataset where loads were 4 - 8% above long term means and initial loadings 2 decades ago. This increase in 2009 and 2010 is associated with an increase in the precipitation levels in those years to the highest in the data record. Mean P concentrations have actually halved in the past decade (Figure 2-5).

Chlorophyll a over the period of record has declined by over half in the south segment since the initial years of measurement (1997-1999) and remained steady and fairly low in all segments until present. Although peak annual chlorophylls in the 15-30 µg/L range correspond to moderate eutrophy, mean levels of 5 or below are not considered eutrophic.

- **Component 3: Index of Eutrophication Development**

The authors were deliberative and conservative as to the datasets that met the level of integrity for inclusion into the index. This simply points to the need for longer term data on which to base many of the relationships in the index and many of the conclusions. The idea of building on NEEA is good, and this method harmonizes the analysis and local results with the national effort. Evaluation of Datasets good,

Caution in advised in assessing the seagrass change maps- comparing RS SAV change 2003 and 2009 data, to 2004-2010 ground data for validation seems fraught with chances for errors. In the analysis of 2011 data for above below ground relationships, A/BG dynamics are notoriously complex with storage, allocation, translocation, light, nutrient sufficiency/surplus affecting the ratio as the plant uses different strategies to optimize the thermodynamics of energy balance, light and space requirements. Not every reduction in belowground is necessarily a “loss”

Conversely, reduction in density, shoot count, biomass (but not necessarily blade length or width) of SAV usually is a “loss” but some of these losses are driven by temperature, herbivory, disease, or simply interannual variability, rather than precisely reaction to nutrient enrichment. Indeed, in the data for many years is not a trend.

Many of the relationships between environmental and ecological variables and TN loading (Figures 3-15 and 3-16) are full of scatter and essentially flat over a wide range of loading rates. This should instill caution in utilizing regression analysis to derive dose-response relationships.

The SAV response variables do not always behave as expected in much of the analysis increasing uncertainty about the biotic response calculations in the IE. Shoot density of eelgrass increases with a tripling of chlorophyll concentration (Figure 3-37) and doubling of TSS (Figure 3-38) and almost every metric for SAV increases positively with increasing macroalgae cover (Figure 3-40).

The development, testing and analysis of the IE is impressive and results clearly shows that this index is useful as a tool for evaluating estuarine eutrophication in BB-LEH. Figure 3-58 shows a clear (and expected) relationship between N loading and the EI. However, because N concentration is related to N loading and is also a component of the EI, part of this relationship is circular and needs to be parsed. The discussion of how the factors important in eutrophication change over time is interesting and valuable use of the IE.

The decline in SAV in recent years and with increasing watershed pressure factors is clear though could be for a number of reasons for this as earlier stated and many of the underlying relationships do not support eutrophication as the mechanism. The fact that light availability is increasing nearly system-wide is a conundrum since the general model for rooted vascular decline is a response to severely reduced light climate. The SAV degradation could be related to other loading-related causes like nitrogen inhibition/toxicity but the BB-LEH data show N loading and especially concentration to be quite low to have this effect.

- **Component 4: Validation Dataset (2011) for Eutrophication Assessment**

The validation section accomplishes some things and fails to accomplish others. A comparison to the NEEA assessment shows that the IE developed specifically for BB-LEH and scored specifically for 2007 data is in concordance with NEEA's result in 2007. This is fairly expected since the IE is built on the NEEA framework using some of the same variables. Additional examination of this result would be interesting- how would application of the NEEA formula to all other years used in the IE analysis compare? This could give insight as to how much an improvement the new formulation actually is. Analyzing the differences and similarities may show weaknesses in the IE methodology (or strengths) and reveal some new aspects of the relationships between estuarine stress and biotic response.

The other validation exercise used a quarantined benthic vegetation dataset from BB-LEH in 2011 and ran the IE algorithm using it. The resulting scores were compared to other years' scores. They were found to be similar to recent years' scores for BB-LEH benthic data. The fact that the IE did not "blow up" when encountered a new dataset is a form of validation I suppose. I expected to have seen some comparison of the IE to N loading and other pressures variables a la Figure 3-58. The validation exercise seems only to go halfway toward assessing and validating the behavior of the model.

- **Component 5: Synthesis and Management Recommendations**

The Synthesis and Management Recommendations are quite spare. In terms of synthesis, the authors offer six almost random figures, one a fairly standard conceptual model of eutrophication, one of TN concentrations (which is not a component of the IE), two maps of sampling stations for nutrients and DO, and two related to sea nettles. There is none of SAV, phytoplankton, macroalgae or epiphytes. More synthetic figures such as light climate, competitive interactions or nutrient enrichment pathways are not presented.

Table 5-1 is a very important summary of the changes in land use over time, emphasizing the loss of almost every land type at the expense of urban land, which increased many fold over the period of record. The summary in the text is a bit confusing because the data discussed for specific year intervals do not match up with the table referred to (Table 5-1). The urban data begin in 1972, which is not in the table. A per cent land use change is discussed between 1995 and 2006, while the data presented in the table are for 1995 to 2002 then skips to 2007. While the trend in land use is clear and is supportive of the authors' concerns, tightening this up would enable more straightforward evaluation and more confidence in the interpretation.

Table 5-2 of essentially raw DO data is not a synthesis.

There is an entire section of Part 5 that references tables in Part 2 requiring a lot of jumping back and forth to see what is going on.

The discussion of hypoxia occupies a large part of this section and again, is not synthetic. It is a general assessment that the use of DO as a sole criterion as too variable and unreliable and difficult to measure with accuracy. This all may well be correct. But the reader is left wondering about the synthesis of information and what the authors would recommend, both in terms of a more acceptable monitoring program, and of management recommendations, rather than the lengthy presentation of shortcomings of DO as a target variable.

Similarly, the sea nettle section seems out of place since the sea nettle problem is not necessarily linked to eutrophication and is not part of the IE, (even though it may be indirectly). But no evidence presented for that even indirect link, through shifts in food web dynamics for example.

The cause of an increasing sea nettle population might just as well be importation from the marine end member, increasing water temperature or changes in salinity. The complex forces that determine the dynamics of the sea nettle population will certainly be disruptive to both food web ecology and to human use of the estuary but there is not much to be recommended locally other than removal of hardened shorelines, if that.

The management applications discussion essentially boils down to the following recommendations spread over two different sections:

- Improved storm water management, specifically reducing stormwater discharge directly to streams
- Implementation of best practices for favorable land use, fertilizer application, turf reduction
- Seeking ways to minimize/reduce impervious surfaces
- Establishment of a TMDL for the estuary
- Additional monitoring throughout the estuary
- Establishing nutrient criteria based on cause-effect relationships
- Other remedial actions
- Development of a coordinated, holistic management plan

I believe that all of these recommendations are important and achievable. Perhaps chief among them for sharpening a management plan is the recommendation for a more robust monitoring program and study of dose-response, cause-effect relationships. This grows from the recognition that the estuary responds in proprietary and unique ways to nutrient inputs and requires site-specific data to develop the models. This point is implicit in much of the discussion; however, here is it explicitly stated: only with improved understanding of BB-LEH hydrodynamics, and knowledge of complex site-specific processing through ecological modeling, can fate and transport of nutrient inputs be determined and effective thresholds for nutrient loading be developed.

- **CONCLUSIONS**

Given that the authors of the study are clearly concerned about the ecological decline of the estuary and have a long history and large body of work there, it is clear that symptoms of degradation have been observed and are real and management action is required. Nor is there much doubt that a large part of the cause is nutrient inputs from the watershed. Questions arise concerning specific quantitative links between cause and effect, and attempts to develop recommendations for management strategies based on them. The magnitude of a response to a particular stress, or the reduction of that stress, which can involve a certain amount of hysteresis, carries a large uncertainty. The question in this instance, and for that matter in all coastal zone management is, is the uncertainty within the bounds that reasonable assuredness can be ascribed to a particular management action, or set of management actions that will move the state of the system toward a desired result.

Within the report the authors identify seven key objectives of the study. Each goal is enumerated below and is followed by comments, in italics, on the relative success in achieving each:

1) To document the influence of human altered land use on past and present nutrient export from the BB-LEH Watershed to the BB-LEH Estuary using physical and chemical watershed data and land-use patterns, and spatially explicit models. *This was accomplished. Land use patterns were analyzed and evaluated for contribution to loading to the estuary providing a critical basis underlying the entire study and development of the IE.*

2) To determine if nutrient loading quantified by subwatershed and biotic response is stable or is temporally and spatially variable. *It is not clear what the importance of nutrient loading stability or biotic response stability, temporally or spatially is and how it was useful in analysis or index development.*

3) To quantify baseflow, runoff, and total nutrient loads and to determine the relative importance of turf area coverage. *This goal to develop water budgets and nutrient loading estimates for the watershed was executed well.*

4) To determine estuarine biotic responses to the loading of nutrients across a gradient of upland watershed development and associated estuarine nitrogen loading, and to identify key biotic responses across a variety of estuarine organisms by examining shifts in phytoplankton, benthic macroalgae, seagrass, epiphytes, benthic invertebrates, and shellfish structure and function. Each of these parameters will be examined and assessed for statistical validity and inclusion in the index development for the 1989 to 2010 period. *This goal represents a major portion of the reason for this study and was partially accomplished. The information presented and interpreted provides an excellent comprehensive ecosystem picture of the BB-LEH system over time. The analysis of statistical validity of each dataset was impressive and thorough. However, additional process work needs to be done to understand "key biotic responses."*

5) To generate an Index of Eutrophication as a tool to evaluate future conditions using water quality and biotic indicators to assess eutrophication, eutrophic impacts, and overall ecosystem health of the BB-LEH Estuary and to develop threshold levels of biotic decline and numeric loading criteria that can support an effective nutrient management plan. *An index was successfully developed and tested. The framework for development and the methodology outlined were successful in creating a working Index of Eutrophication. Some issues regarding assumptions and statistical treatment remain to be better explained and more robust datasets to be gathered. Additional validation years are warranted.*

6) To apply a conceptual model of eutrophication and determine if ecosystem structure and function have been altered in the BB-LEH Estuary. *The conceptual model was in the background of the plan for assessing ecosystem structure and impairment but frankly the model was not used in an explicit way as a guiding outline, to test hypotheses or to reveal data/knowledge gaps. Nor was it effectively used as a communication tool in explaining the IE or the overall analysis. The model could be showcased and used more effectively.*

7) To document the current biotic and seagrass habitat conditions of the BB-LEH estuary at the end of the investigation using the most recent biotic data collected (2011) and index methods developed from data collected through 2010. *This goal was accomplished.*

The project generally accomplishes its goals. It is hampered by lack of specific data and a short period of record in which all datasets do not match up, temporally. The monitoring program design is appropriate, though focused on the lower level variables, as necessary in a nascent program. Expansion of the monitoring is warranted. The data are a concern because of the low frequency of collection, missing data and lack of a long history of baseline data. There is no historical period or specific location that might be utilized as a of a reference condition for this estuary. The understanding of management endpoints is also uncertain. If there is not a pristine "benchmark" or a reference condition or even a set of designated uses defined as the goals, the recommendations are going to be vague. The Index of Eutrophication is well-developed within the constraints of the data available, based on sound science, and generally supported by the authors' justification, although with caveats. Two weaknesses of the IE are the lack of long-term data on which to base several components of the Index, and over-reliance on literature values for thresholds. The use of BPJ and literature is certainly the best path when local data are not available, but the uncertainty imported expands the confidence limits on any relationships such that practical use of the Index should be deferred until additional testing and validation is completed. The interannual variability of the system creates a large noise:signal ratio.

The wider applicability of the IE is questionable, as it is specifically designed for the BB-LEH system, and the thresholds and targets selected for this kind of estuary. However, as a design protocol for developing IE for other systems, selecting appropriate variables, addressing specific weighting, the algorithm used, the validation technique, this is an excellent model. I am here speaking of a process, rather than explicit use of this algorithm applied to other estuaries. Each estuarine system should be evaluated on its own, the model fitted to its own data, validated and evaluated autonomously, as I believe the authors themselves state. I am concerned that application of the model at this point in anything other than testing/validation mode even to its own estuary, when many of the components are based on only a few years of very variable data, may be premature. Additional testing and development with more years of data will strengthen confidence in this approach and specific parameters.

Therefore, in basing management decisions on the outcomes of models, managers must be wary of going forward too soon with new protocols, when making decisions affecting many stakeholders. If a model overpromises and does not function as advertised or as quickly as advertised, it can undermine the entire effort. Having said that, it is important not to allow the perfect to be the enemy of the good. Clearly, the estuary is in need of a management plan in science-based framework. The authors have made the best with what they have in terms of data availability. If there is a bit of overreach, that can be understood, and the value of best professional judgment and best available data is real. All estuary nutrient management programs are based on this leap of faith to some degree. The development of a TMDL for this estuary should be a short-term goal. The continuation of a monitoring program, with the enhancements stipulated in the report which exposes the shortcomings of the data is certainly obligatory.

Part 3- Detailed Assessment of Overall Report Organization

This section presents an editorial review of the documents to discuss issues and questions that concerning the scientific reasoning and development of the index.

Overall comments: Although there is a well-developed and detailed Table of Contents, much of the document has no page numbers! It greatly adds to the difficulty of reviewing it.

Many of the figure captions are inadequate. The figures are difficult enough to match with text since they are gathered at the back of the report but a number of them cannot stand alone on the information provided in the caption. Some of these include sampling period, location (Fig 5-5), frequency, averaging scheme (e.g. Fig 5-4). Many of the most important figures supporting the authors' conclusions are extremely small and/or of low resolution (Fig 4-2; Table 3-6) making it difficult to decipher what point is being made or supported.

- **Comments on Key Findings Section**

The Key Findings section is a means to establish the main points of the analysis. It is an all-important section as it will be the most-read portion of the document, and for many people the only portion read. The section could be made much more cogent with an organizational scheme that groups like findings together, such as gathering the descriptive, stressor and response items in separate groups, possibly in a hierarchical framework. Many of the bullets presented are confusing and jump around from topic to topic. Although a Key Findings section is not meant to be detailed, some level of context must be provided for the points to make sense and capture the reader's interest and attention. When conclusions are drawn and no period of record stated it is difficult to gauge the importance of the point. Note that for many of the issues mentioned below, as regards the effectiveness of the Key Findings, the points are very well addressed later in the main body of the report. The point is to optimize the Key Findings section to reflect the quality to be found deeper in the report as the most effective means of communicating findings to many readers. Since bullets are not numbered it is difficult to reference them. These comments go in order but do not address each bullet.

- In the land use description is the 10% impervious a subset of the 25% urban? Of the 34% developed or are these three additive categories (This question is answered deeper into the report, but that should be made clear in the key findings, particularly since additional reference to these categories is made within the Key Findings).
- The study confirms that the nutrients are strongly related or correlated to land use? How was the link made?
- What is the basis for the turf/non-turf relation to N loading? At least a mention of the test or model or evidence for it could be presented here.
- This whole series of statements about land use is confusing. Bullets #4 and 6 could be better organized to state conclusions about land use and nutrient loading. The statement about turf implies that turf = development (is that suburbs?) which drives more N and P loading than non-turf urban. But developed is sometimes used interchangeably with urban, sometimes not (see bullet #1 versus bullet #4 "urban development" versus "loads and yields are generally higher in areas with more development") so it is difficult to follow the ranking of least- to most-loading land uses.
- The bullet #7 beginning "Baseflow contributes more than 80%..." also suffers from this lack of definition of developed versus undeveloped.
- In the bullet regarding low dissolved oxygen, BB-LEH was hypoxic 82 times in 22 years. Without a spatio-temporal context, it is difficult to grasp the severity of this phenomenon. Does it mean that there were 82 instances of estuary-wide hypoxia? Or hypoxia was detected at 82 stations over 22 years? To illustrate in extremis, this could represent a single

station in the bay going hypoxic on 82 occasions while the rest of the bay was well-oxygenated. This is a significant point and the bullet statement needs to be clarified here.

- What is meant by spatio-temporal alignment of data collection? How will this alignment in the monitoring aid in the analysis?
- “Initial rapid declines.” This sounds like rapid temporally (to new readers). This really means there is great sensitivity in the Index to low levels of loading, or to a large response to small increases in loading. The wording should be changed to reflect the correct meaning. Does this refer to a component of IE or the whole index? “Shifting of dominant factors” does not convey enough meaning and does not indicate how that would influence the IE.
- According to the description, there seems to be a lot of bouncing up and down of the IE over time, between good and bad years, making any trend toward increasing eutrophication difficult to discern.
- There is a whole series of bullets describing the IE results that is difficult to follow. This lack of clarity comes from reporting results of the IE before the reader knows how the IE works. The phrase “total nutrient loadings” is misleading. Authors must mean the IE score for loadings. Some of this becomes clearer in the main report but it detracts from the initial impression given in the Key Findings:
    - The IE score indicates that low loading rates are found in the northern bay, but it variously has the lowest (or highest) nutrient inputs; the 2009 score for the north was highest (best) during the period of record (but declined in 2010) and its lowest (worst) was in 1991 near the beginning of the record.
    - The statement that the “north has already undergone eutrophication” is supported by what? It sounded as though nutrient loading is lowest in this segment.
    - There are some seemingly contradictory findings- “total nutrient loadings in the north were very low (IE=7)”- this is directly contradictory to an earlier statement “increases in total nitrogen (concentrations and loads) is more intense in the north.” Please clarify.
- “Favorable temperatures”- favorable for algae growth or for environmental improvement? Favorable on the IE scale?

IE- outputs for 3 areas- geographic or conceptual areas? (e.g. nutrients, light biology) Impt because separate components are discussed farther down.
Pg 15 contains a confusing discussion of IE values, increases and declines. Lowest IE was in 1991 in north? Doesn’t this contradict the statement that the north has been getting steadily worse?

- **Executive Summary**

Some statements in the Executive Summary are made without sufficient supporting information- macroalgae cover of what? Specific stations? The whole bay bottom? Over what time period.

Changes in land use have been shown- where- in BB-LEH? Reference? This discussion is very general and not focused on BBLEH. Much reads like a textbook but is not coherently linked to BB-LEH issues and data.

The lengthy definition of eutrophication quoted here and elsewhere throughout the document is awkward, even if expressing an authoritative definition. It is a long laundry list that becomes a long run-on sentence. If it is to be effective, this passage needs to be parsed into digestible portions.

Development does not necessarily lead to eutrophication with proper planning and infrastructure. This is being argued in this document is acknowledged, but this is too sweeping a generalization and lacking supporting documentation. The authors are generally good at keeping a clinical, objective position but occasionally an advocacy tone creeps into the narrative. It should be restrained.

P16 In Para 2, is the focus on BB-LEH or on "the estuary" in general. The shifting tone makes it confusing as to which the authors are referring to. For example "can lead to" indicates speculative position that refers broadly to a general characteristic of eutrophication of estuaries, not to a finding of this study and specific to BB-LEH. It actually emphasizes that it hasn't led to these conditions in BB-LEH or the authors would have stated so. In fact, the tone of the passage gives the sense that it has not actually occurred in BB-LEH.

Concerning Dissolved Oxygen as the lynchpin of a management monitoring program, I totally agree with the authors on their point- clearly a program monitoring ecological health of coastal waters based only on DO is underspecified and even inappropriate.

Para 3 is out of place.
P 19 Para 1 Fix: which the Assessment of Estuarine Trophic Status (ASSETS) Model.

P 19 Para 1 Fix: (rather than 5 used in NEEA).

Data are analyzed separately for each segment of the bay, because they have been determined to be heterogeneous habitats. This requires caution- the scheme for segmentation is not fully explained nor is the bisection of each segment (totaling six ever fully justified quantitatively in the analysis.

While recognizing the threshold as an important guiding metric, a data-based metric is only applicable to a particular system after the response has occurred in the system. Moreover, quantitative criteria can be established only when developed from robust long-term data have been gathered from the target system.

The authors themselves note that some variables are both stressors and response variables-epiphytes is one of these dual variables and its use as an indicator is confounded by its own response to phytoplankton in the water column and resultant light attenuation. Therefore, the more phytoplankton respond to nutrient enrichment, the less valuable epiphytes become as an indicator. They would tend to asymptote at an intermediate level.

Point of Organization: the discussion jumps from SAV to macroalgae to HABs to benthic inverts back to SAV.

P15 para 2 is the nutrient index for central and southern portions referring to loading or concentration? This is a very important distinction and one that is argued in the endeavor of developing site specific nutrient criteria.

Pg 15 The sea nettle paragraph- The occurrence of sea nettle blooms in the north segment has posed a hazard to human use of some waters in the estuary.- this perhaps is not primarily tied to eutrophication. It is in fact linked by the authors explicitly to bulkheading of the shoreline along the BB-LEH estuary. I don’t believe the authors of this document are intending to take a position on hardening of shorelines as a management recommendation since this is about nutrient loading.

Eutrophication condition was worst in the north segment despite modest improvements, in contrast to stages and trends in the south and central segments. This is indicated as being a response to land use (development) but have improvements in the south and central water treatment been taken into account? Or declines?

P22 para 1 is very confusing- the discussion of consistency as a rationale for adding the raw and weighted scores. It is not clear at this point how this works, nor the reason for adding a weighted score (which contains information about the raw score) to the original raw score. One hopes that the main report clears up the confusion.

Also, the term “in flux” is vague. A timescale needs to be introduced by which to measure change in the IE- what constitutes change? What level of change becomes a significant trend?

Long residence times are mentioned- how has this been calculated-not via modeling but by estimation yet the details of the data this estimation is based upon, method, the assumptions and the period of record are not detailed. This point is encountered throughout the document and is in fact a critical issue. Estuaries can be overloaded with nutrients every day of the year but with sufficient throughput could theoretically suffer few ill effects. Apparently BB-LEH is not one of these lucky estuaries, as stated repeatedly, but how and where are the calculations showing this? Obviously there is no physical model but there must have been studies on the basis of which to justify the statements regarding flushing? Is the stated residence time really that long compared to other estuaries? (not really) many are of a similar magnitude or greater.

Also it is well known that residence times vary throughout the year with changing hydrology, spring flood, etc. Therefore, is the residence time being stated the maximum residence time? There may be periods during the year of much greater flushing rates and lower residence time. Is this an annual average residence time? A growing season residence time (when authors state the loading is highest)? Lack of a hydrodynamic circulation model limits evaluation of residence time, water budgets etc. problematic.

P23 I am not sure what the penultimate paragraph is doing here. There seem to be references to research from other estuaries quite randomly without explanation as to how they compare to BB-LEH or why they are referenced.

- **Comments on the Main Body of the Report**

Problem Statement:
Pg 31 and Figure 1-2 I am a bit bewildered by this figure as there is no explanation as to how the contours were arrived at. Is this straight interpolation via kriging? The high TN concentrations in the north bay, showing what is I guess a 20 mean, seems based on a handful of stations a small subset of them at the >60 concentration, showing colors that do not seem to add up to the red and yellow regions depicted in the polygons. It is not clear how the maps were derived.

Pg 31 Definition of eutrophication is long and awkward. All of it is valid but the sentence would have greater impact broken into two.

Pg 34 “Seagrass now covers 5200ha”- this statement has no context without a coverage from previous year(s). Is that a lot? A little? The area of the bay is never given in the site description but from Figure 1-1 it looks to average 50km x 5km or about 25,000ha (excluding Great Bay. The SAV cover is a little more than 20% of that. What would it look like in the past?

Pg 38. Ah! Here is the surface area of the bay.

The list of degradation indicators in degraded condition, including: epiphytes, algal blooms, macroalgae expansion, are compelling and indicative of eutrophic conditions. A stronger monitoring program examining these variables over the long term is warranted.

Pg 68 “Seitzinger et al. (2001) showed that benthic algal dynamics can significantly influence sediment-water nutrient fluxes in the estuary, particularly ammonium from sediments which may sustain system eutrophy.” This statement is not followed up- how does algae affect sediment fluxes? And by what mechanism?

Pg 68 There is a lot of information able macroalgae dynamics from other studies but not from BB-LEH. This reduces the significance of the role it may play in the IE.

Pg 70 No measurement, apparently, of dissolved organic N in BB-LEH to support statements about brown tide blooms. Relationship to high salinity makes a connection to watershed inputs more difficult to make.

Pg 73 Problems with fonts on this page and elsewhere

Pg 75, 76 Note there is repetition of this passage:
“Substantial effort during this project went into identifying, assembling, characterizing, analyzing, and evaluating available datasets and databases. Qualitatively, these were examined for availability, completeness, and representativeness. Quantitatively, these were examined through a variety of methods for statistical rigor, robustness, and representativeness. The goal of these efforts was to determine the suitability of including variables within these datasets and databases as indicators for inclusion in the Index of Eutrophication, as specified in the project QAPP (page 60).”

Pg 79 The selection, a priori, of the variables conforms generally to the set of variables detailed in the EPA nutrient criteria program and in the NOAA NEEA eutrophication assessments.

Pg 84 I am not convinced that all of the seagrass indicators are appropriate. A spatial extent (area % cover), and aboveground biomass parameter are appropriate. Shoot density is somewhat redundant with biomass. Belowground biomass might vary orthogonally to aboveground biomass for a number of reasons not related to eutrophication. And blade length may not be related to eutrophication at all.

Pg 95-96 **THIS:** *Defined thresholds for Harmful Algal Blooms are listed in* **and THIS:** Table 3 - 17The rescaling equation that is generated from these thresholds is listed in Table 3 - 4. **Need repair.**

Pg 97 This passage is repetitive of an earlier section (pg 86): “Thresholds are determined and defined through examination of: (a) the literature, (b) analysis of available data for BB-LEH, (c) Best Professional Judgment, and (d) some combination of a-c. Raw scores range from 0 (degraded condition) to 50 (excellent condition) and are evenly weighted between indicators within the component index. Thus, for example, the raw score for each of the four Water Quality indicators contributes 12.5% of the score for the Water Quality Index (25% * 50% = 12.5%).”

It is not clear how thresholds are ultimately determined: eyeballing a collection of literature relationships? Combining that with available BB-LEH relationships?

Pg 100 Sensitivity Analysis section: wouldn’t different variables have different temporal scales on which to measure variability? Authors settle on Multi-year for water quality. Is this analysis done for each variable and what is the result?

Pg 122 Expressing epiphyte biomass per unit area of bottom seems incorrect. The load per leaf area seems more relevant to seagrass response.

- **Comments on Figures and Tables**

Some figures are very difficult to read and interpret. Others are poorly keyed, referenced or titled.

In Fig 1-4 the animal feeding sites are hard to detect against the similar-colored background.

Fig 1-5 What are the red dots? Water quality, I assume by comparing with Fig 1-8.

Fig 1-8 The seagrass map is very convoluted and difficult to discern a point.
In general, it is difficult to read many of the legends- very fuzzy resolution troubles many of the figures in the report.

A prolonged period of low DO in Figure 2-2 from 92-96.

Fig 2-3: the TN shows an interesting pattern where large increases in maximum loading in the north from 96-03 coincided with the recovery of the DO profile.

Fig 2-4 TP: There are lots of missing data before 1998.

Fig 2-6: Secchi actually seems to be improving since the early 2000's. How do the authors explain this and account for it in their conclusion of increasing eutrophication and turbidity trends?
Fig 2-7: Chlorophyll a, the all-important variable, the lynchpin in the eutrophication argument, shows a fairly steady mean concentration at low levels around 5 for all sectors, although the maxima are elevated periodically, into the 30s. But the variance would be much lower and showing the maxima does not give a good feel for how persistent the elevated periods are, given these are annual values. These levels are below levels found in many estuaries that are considered eutrophic.

For macroalgae figures, why is the north segment excluded? These values for macroalgae, while not insignificant are hardly overwhelming. Therefore, the authors' assertions that they are levels that are crowding outcompeting the SAV need rely on other measurements and additional data need to be presented to support this case.

Fig 3-9 show evidence of a cause effect relationship but the data are from two other estuaries.

Fig 3-13 gives yield per km2 which is not the unit of the previous supporting studies.

Fig 3-33 shows as much benefit to SAV as harm by increasing N and especially P.

Fig 3-58 This is an impressive figure.

Fig 4-2 is a mess. It is difficult to see it, impossible to read it and difficult to understand what point the authors intend to make.

Table 4-2 Table 4 - 2 Characteristics of submerged aquatic vegetation (SAV) by sampling period in the BB-LEH Estuary during 2011. Should indicate the N for each kind of sample.

c.

All comments from the third reviewer were directed to specific questions from the Peer Review Scope of Work.

III. Literature Cited

a.

Shin, P.K.S. and W.K.C. Lam. 2001. Development of a marine sediment pollution index. Environmental Pollution 113: 281-291.

b.

Boyer, J.N., J.W. Fourqurean, and R.D. Jones. 1997. Spatial characterization of water quality in Florida Bay and Whitewater Bay by multivariate analysis: Zones of similar influence. Estuaries 20(4): 743–758.

Boyer, J.N., C.R. Kelble, P.B. Ortner, and D.T. Rudnick. 2009. Phytoplankton bloom status: Chlorophyll a biomass as an indicator of water quality condition in the southern estuaries of Florida, USA. Ecological Indicators 9S: S56–S67.

Bricker, S.B., C.G. Clement, D.E. Pirhalla, S.P. Orlando, and D.R.G. Farrow. 1999. National Estuarine Eutrophication Assessment; Effects of Nutrient Enrichment in the Nation's Estuaries. NOAA, National Ocean Service, Special Projects Office and the National Centers fro Coastal Ocean Science. Silver Spring, MS: 71 pp.

Bricker, S.B., J.G. Ferreira, T. Simas. 2003a. An Integrated Methodology for Assessment of Estuarine Trophic Status. Ecological Modelling. 169:39-60.

Bricker, S., G. Matlock, J. Snider, A. Mason, M. Alber, W. Boynton, D. Brock, G. Brush, D. Chestnut, U. Claussen, W. Dennison, E. Dettmann, D. Dunn, J. Ferreira, D. Flemer, P. Fong, J. Fourqurean, J. Hameedi, D. Hernandez, D. Hoover, D. Johnston, S. Jones, K. Kamer, R. Kelty, D. Keeley, R. Langan, J. Latimer, D. Lipton, R. Magnien, T. Malone, G. Morrison, J. Newton, J. Pennock, N. Rabalais, D. Scheurer, J. Sharp, D. Smith, S. Smith, P. Tester, R. Thom, D. Trueblood, R. Van Dolah. 2003b. National Estuarine Eutrophication Assessment Update: Workshop summary and recommendations for development of a long-term monitoring and assessment program. Proceedings of a workshop September 4-5, 2002, Patuxent Wildlife Research Refuge, Laurel, Maryland. NOAA, National Ocean Service, National Centers for Coastal Ocean Science. Silver Spring, MD: 19 pp.

Bureau of Land and Water Quality (Maine) 2008. Development of Nutrient Criteria for Maine's Coastal Waters. 31 pp.

Doren, R.F., J.C. Trexler, A.D. Gottlieb, M. Harwell. 2009. Ecological Indicators for System-wide Assessment of the Greater Everglades Ecosystem Restoration Program. Ecological Indicators 9(6): S2-S16.

Florida Department of Environmental Quality. 2013. Status of Efforts to Establish Numeric Interpretations of the Narrative Nutrient Criterion for Florida Estuaries and Current Nutrient Conditions of Unimpaired Waters. 68 pp.

Glibert, P., C. J. Madden, W. R. Boynton, 2010. A framework for developing estuarine nutrient criteria. In: P. Glibert, C. J. Madden, J. Sharp, R. Smith, E. Dettmann, N. Detenbeck, J. Kurtz, J. Latimer, J. Lehrter, W. Nelson [eds.]. Development of nutrient criteria for the nation's estuaries: State of the science. USEPA. Office of Water. Report of the National Nutrient Criteria Expert Panel.

Hardin, M. M., H. A. Rines, C. E. Rose, and M. B. Abel. 1996. Spatial heterogeneity of macrophytes and selected invertebrates in Narragansett Bay (RI, USA) as revealed by principal component and cluster analyses. Estuarine, Coastal and Shelf Science 42:123-134.

Lathrop, R. G. and S. M. Haag. 2011. Assessment of seagrass status in the Barnegat Bay-Little Egg Harbor Estuary system: 2003-2009. Technical Report, Center of Remote Sensing and Spatial Analysis, Rutgers University, New Brunswick, New Jersey. 56 pp

Mississippi Department of Environmental Quality Office of Pollution Control. 2004. Mississippi's Plan for Nutrient Criteria Development. 23 pp.

Nixon, S. W., B. Buckley, S. Granger, and J. Bintz. 2001. Responses of very shallow marine ecosystems to nutrient enrichment. Human Ecological Risk Assessment, 7:1457-1481.

New Hampshire Dept. of Environmental Services. 2009. Numeric Nutrient Criteria for the Great Bay Estuary. 120 pp.

Sutula M. 2011. Review of Indicators for Development of Nutrient Numeric Endpoints in California Estuaries. Southern California Coastal Water Research Project Technical Report No. 646. December 2011. 76 pp.

c.

Dauer, D.M., A.J. Rodi and J.A. Ranasinghe. 1992. Effects of low dissolved-oxygen events on the macrobenthos of the lower Chesapeake Bay. Estuaries 15:384-391.

Diaz, R.J. 2001. Overview of hypoxia around the world. Journal of Environmental Quality 30:275-281.

Diaz, R.J. and R. Rosenberg. 1995. Marine benthic hypoxia: A review of its ecological effects and the behavioural responses of benthic macrofauna. Oceanography and Marine Biology: An Annual Review 33:245-303.

Gillett, D.J., S.B. Weisberg, T. Grayson, A. Hamilton, V. Hansen, E.W. Leppo, M.C. Pelletier, A. Borja, D.B. Cadien, D.M. Dauer, R.J. Diaz, M. Dutch, J.L. Hyland, M. Kellogg, P. Larsen, J. Levinton, R. Llanso, L.L. Lovell, P. Montagna, D. Pasko, C.A. Phillips, C. Rakocinski, J.A. Ranasinghe, D.M. Sanger, H. Teixeira, R.F. Van Dolah, R.G. Velarde and K.I. Welch. In Press. Effect of ecological group classification schemes on performance of the AMBI benthic index in US coastal waters. Ecological Indicators.

Summers, J.K., S.B. Weisberg, A.F. Holland, J.Y. Kou, V.D. Engle, D.L. Breitburg and R.J. Diaz. 1997. Characterizing dissolved oxygen conditions in estuarine environments. Environmental Monitoring and Assessment 45:319-328.