# Summary Report

## Technical Review of Select Memorandums Supporting the Development of Nitrogen Endpoints for Three Long Island Sound Watershed Groupings: 23 Embayments, 3 Large Riverine Systems, and Western Long Island Sound Open Water

**Prepared for:**

U. S. Environmental Protection Agency Region 1
U.S. EPA Contract Number 68HE0118A0001
Order Number 68HE0118F0006

**Prepared by:**

HydroAnalysis, Inc.

On behalf of:

PARS Environmental, Inc.

Comprehensive Environmental, Inc.

**January 29, 2019**

# Table of Contents

## 1. Introduction

Long Island Sound (LIS or "Sound") suffers from periods of low dissolved oxygen (DO) that have led to adverse ecological effects. Concentrations of DO greater than 5 mg/L are considered protective of aquatic life in Long Island. During the summer, DO concentrations in the bottom waters of the Sound often fall below 3 mg/L, an occurrence referred to as hypoxia. Excess loading of nitrogen is the primary cause of hypoxia in the Sound. In addition to the adverse effects to aquatic life, excess nitrogen can also produce algal blooms, decrease water clarity, and limit the growth of submerged aquatic vegetation (Long Island Sound Study, 2018).

In 2016, the U.S. Environmental Protection Agency (USEPA) Region 1 contracted with Tetra Tech to provide technical support with the development of nitrogen endpoints for Long Island Sound, and the calculation of nitrogen load allocations for the LIS watershed. The development of nitrogen endpoints the Sound focused on three categories of waterbodies: 1) 23 embayments; 2) three large riverine systems (Connecticut, Housatonic, and Thames Rivers); and 3) Open water in Western LIS. The project, entitled *Application of Technical Approach for Establishing Nitrogen Thresholds and Allowable Loads for Three LIS Watershed Groupings: Embayments, Large Riverine Systems and Western LIS Point Source Discharges to Open Water,* was completed in March 2018. In order to ensure that the work was conducted using scientifically-sound methodologies consistent with professional and relevant scientific practices, USEPA commissioned an independent technical review of the following technical memorandums (hereinafter, "technical memorandums" or "memorandums") from the project:

1. Summary of Hydrodynamic Analysis (Subtask E Memorandum) (USEPA, 2018a).
2. Summary of Empirical Modeling & Nitrogen Endpoints (Subtask F/G Memorandum) (USEPA, 2018b).

The Hydrodynamic Analysis subtask (Subtask E; USEPA, 2018a) used output from the System Wide Eutrophication Model (SWEM) and other sources to accomplish two key objectives: 1) Define the areas of influence for the Connecticut, Housatonic, and Thames Rivers (i.e., "regions within which water from the rivers exerts a predominant effect on water quality condition"), and calculate their estimated nitrogen loading contributions to select LIS embayments and throughout all of Long Island Sound; and 2) Calculate the relative mixing between open water in LIS and individual embayments.

The results of the Hydrodynamic Analysis subtask (Subtask E; USEPA, 2018a) were used to support the Empirical Modeling & Nitrogen Endpoints subtask (Subtasks F/G; USEPA, 2018b). The objective of the Empirical Modeling & Nitrogen Endpoints subtask was to develop nitrogen endpoints for each of the selected embayments that are protective of seagrass and that prevent adverse effects related to macroalgae and DO. The results from both analyses (Subtask E and Subtask F/G) are going to be used to support the calculation of nitrogen load allocations for the LIS watershed, and to estimate source specific load reductions to meet the nitrogen endpoints (Subtask H).

The goal of the *Empirical Modeling & Nitrogen Endpoints* subtask (Subtasks F/G; USEPA, 2018b) was to develop nitrogen endpoints for the watersheds selected for the study (see Figure F-1 in USEPA 2018b). The candidate endpoints for total nitrogen were developed using the following three empirical approaches (also referred to as "lines of evidence" in the memorandums) (USEPA, 2018b):

1. Scientific Literature Analysis
    a. Identify literature-based nitrogen endpoints (loads and concentrations) from similar estuaries associated with the protection of key assessment/response variables for LIS (e.g., seagrass, aquatic life).
2. Stressor-Response Analysis
    a. Develop nitrogen endpoints using existing water quality data from LIS to establish empirical statistical models of the relationship between chlorophyll *a* and total nitrogen.
    b. Develop chlorophyll *a* endpoints using empirical statistical models of the relationship between key assessment/response variables (seagrass and aquatic life), light availability (Secchi depth or light attention), and DO, as a function of chlorophyll *a*.
3. Distribution-Based Approach
    a. Develop nitrogen endpoint concentrations using the 25$^{th}$ percentile of total nitrogen concentration distributions for LIS embayments and open water stations.

The following is a summary of the final total nitrogen endpoints selected for each of the above empirical approaches (USEPA, 2018b):

1. <u>Scientific Literature Analysis</u>: Median total nitrogen from literature-based values protective of seagrass
    a. Embayments: Range of 0.30–0.50 mg/L; median of 0.39 mg/L, rounded to 0.40 mg/L
    b. Open Water: Range of 0.30–0.60 mg/L; median of 0.41 mg/L, rounded to 0.40 mg/L
2. <u>Stressor-Response Analysis</u>: Mean total nitrogen associated with chlorophyll *a* endpoints
    a. Embayments: Range of 0.06 mg/L–2.52 mg/L
    b. Open Water: Not applicable
3. <u>Distribution-Based Approach</u>: 25$^{th}$ percentile of total nitrogen observed in LIS embayments and open water stations.
    a. Embayments: 0.27 mg/L
    b. Open Water: 0.24 mg/L

# 2. Technical Review Process and Review Team

## 2.1. Technical Review Process

HydroAnalysis, Inc. (under USEPA Contract No. 68HE0118A0001 with PARS Environmental and Comprehensive Environmental, Inc.) was commissioned by USEPA to coordinate and manage an independent technical review (hereinafter, "technical review") of two selected technical memorandums from the LIS nitrogen endpoints project (see Section 1). HydroAnalysis' responsibilities included identification and selection of technical reviewers (hereinafter, "technical reviewers", "reviewers", "Review Team", or "Technical Review Team"), coordination of the technical review, production of a summary report for the technical review, and development and delivery of a webinar to inform stakeholders of the outcomes of the review.

HydroAnalysis was given directive authority by USEPA for planning, coordinating, and managing all aspects of the technical review. The USEPA remained independent from the technical review, and did not play a role in the selection of technical reviewers or in the production of the summary report. The USEPA was given an opportunity to review the draft report prior to final publication, and ask for clarification on Review Team responses, if needed. Clarification was not needed.

HydroAnalysis assembled a group of four technical reviewers with expertise in the areas of estuarine water quality (e.g., eutrophication), estuarine ecology and biology (e.g., biological response indicators), and estuarine hydrodynamic and water quality modeling. The reviewer selection process included a screening for independence and conflict of interest. All four reviewers were asked a series of questions concerning potential conflict of interest, and signed forms certifying that they had no conflicts of interest related to the technical review. In addition to considerations of expertise, experience, and conflicts of interest, selection was also based on the reviewer's availability to complete the technical review during the timeframe allotted for the review.

The four technical reviewers were charged with performing an independent review of the two selected technical memorandums from the LIS nitrogen endpoints project, and given specific questions to respond to (see Section 2.3). Each technical reviewer submitted written responses to the review questions directly to HydroAnalysis. The technical reviewers did not communicate with one another during the review process. The reviewers also did not communicate with USEPA or with Tetra Tech during the review process or during the development of this summary report.

HydroAnalysis reviewed the Review Team responses, and coordinated closely with the reviewers to obtain clarification on responses as needed, and to obtain agreement for recommended edits to address major grammatical or spelling errors. None of the edits modified, interpreted, or enlarged upon the technical reviewer's responses. The reviewers were given an opportunity to review the draft report, and provide clarification or corrections, if needed. The responses of the Review Team as provided in this report (see Section 4) represent the individual opinions and assessments of each of the technical reviewers.

## 2.2. Technical Review Team

Brief descriptions of the experience and areas of expertise for each of the technical reviewers are provided below.

**Victor J. Bierman, Jr., Ph.D., BCEEM**

Dr. Victor Bierman is a Senior Scientist Emeritus at LimnoTech with 45 years of experience in the development and application of water quality models for eutrophication and the transport and fate of toxic chemicals, leading to his publication of over 100 technical papers and reports. He is a former USEPA National Expert in Environmental Exposure Assessment, and a former Associate Professor in the Department of Civil Engineering at the University of Notre Dame. He is also a Board Certified Environmental Engineering Member (by Eminence) of the American Academy of Environmental Engineers and Scientists. Dr. Bierman conducts research and development on projects for federal, state and regional government clients. He also provides scientific peer review, litigation support, and expert testimony on a variety of environmental issues for government agencies, and industrial, regulatory and private clients. Dr. Bierman is a leading expert in the assessment and solution of problems related to nutrients, DO, nuisance algal blooms, nitrogen fixation, exotic species, and ecosystem processes. He has conducted studies in watersheds, lakes, major rivers, estuaries, coastal marine systems, the Great Lakes, and at USEPA Superfund sites. Key accomplishments by Dr. Bierman related to the topic of this review include service as Panel Chair for a scientific peer review of the Massachusetts Estuary Project (MEP) linked watershed-embayment model for protection of eelgrass and aquatic life, service as a consultant to the USEPA Science Advisory Board for peer review of draft technical guidance on using stressor-response models to derive numeric nutrient criteria, and service on a scientific peer review panel for numeric nutrient criteria for protection of eelgrass in the Great Bay Estuary, New Hampshire.

**Mark J. Brush, Ph.D.**

Dr. Mark Brush is an Associate Professor of Marine Science at the Virginia Institute of Marine Science (VIMS) in Gloucester Point, VA, part of the College of William and Mary. Dr. Brush received his B.S. in Biological Sciences from Cornell University in 1995 and his Ph.D. in Biological Oceanography from the University of Rhode Island in 2002, and has been at VIMS since 2002 as a postdoctoral fellow, research scientist, and faculty member. His research program focuses on the ecology of coastal marine ecosystems such as estuaries and lagoons, through field- and lab-based ecological investigations, synthesis of water quality monitoring data, and interdisciplinary ecosystem simulation modeling. Recent projects have focused on modeling the response of coastal systems to nutrient enrichment and climate change, with a focus on water quality and ecosystem function, quantifying coastal ecosystem metabolism and watershed nutrient loading, and development of living resource models of shellfish, fish, and submerged vegetation to quantify their role in ecosystem function and their response to nutrient loading and climate change. A key aspect of Brush's research involves development of reduced complexity, readily accessible modeling tools that can be delivered online for direct use by other researchers, managers, and educators. He has participated in technical reviews of the Long Island Sound Study Systemwide Eutrophication Model and hypoxia models for the northern Gulf of Mexico. Brush teaches courses in interdisciplinary coastal field research, estuarine ecology, and ecosystem modeling. He recently served as President of the Atlantic Estuarine Research Society and is currently a Member-at-Large for the Coastal and Estuarine Research Federation. He has served on the editorial boards of the Journal of Sea Research, Biogeochemistry, and Estuaries and Coasts.

**Anthony Janicki, Ph.D.**

Dr. Anthony Janicki is president of Janicki Environmental and has worked for more than 25 years dealing with water quality and quantity concerns in Florida. Much of his work had been for local governments, water management districts, and national estuary programs. His most recognized efforts have been for the Tampa Bay Nitrogen Management Consortium and developing numeric nutrient criteria for southwest Florida estuaries and in the examination of water quality problems in the Indian River Lagoon, Caloosahatchee River, and Loxahatchee River. In each of these projects the focus has been on the development of endpoints for ambient water quality, nutrient loading, and seagrass health. The State of Florida, with approval by the Environmental Protection Agency, has adopted numeric nutrient criteria for the Gulf Coast estuaries as a result of his work. He also has been instrumental in the development of environmental flows endpoints for four of the state's water management districts.

**Dubravko Justic, Ph.D.**

Dr. Dubravko Justic is the Texaco Distinguished Professor in the Department of Oceanography and Coastal Sciences at Louisiana State University (LSU). Previously, he was the Eric L. Abraham Distinguished Professor in Louisiana Environmental Studies and Director of LSU's Coastal Ecology Institute. Dr. Justic has 35 years of experience in the development and application of hydrodynamic and biogeochemical models for coastal eutrophication, hypoxia, and potential impacts of climate change on coastal ecosystems. He has extensively studied low oxygen zones in the northern Adriatic Sea and northern Gulf of Mexico and has employed various types of numerical simulation models to describe controls of environmental factors on hypoxia and predict the consequences of management actions. He is presently working on characterizing connectivity among wetland, estuarine, and shelf ecosystems in the northern Gulf of Mexico and evaluating tradeoffs associated with different Mississippi River management alternatives.

## 2.3. Technical Review Questions

Each technical reviewer was charged with evaluating the scientific and technical merits of the following two technical memorandums developed under the LIS nitrogen endpoints project:

1. Summary of Hydrodynamic Analysis (Subtask E Memorandum) (USEPA, 2018a).
2. Summary of Empirical Modeling & Nitrogen Endpoints (Subtask F/G Memorandum) (USEPA, 2018b).

Additional reports and model files cited or referred to in the two technical memorandums were also provided to the reviewers for reference as needed during their review.

Each reviewer was asked to consider and answer the specific questions listed below while keeping in mind USEPA's objective for the technical review (i.e., to ensure that the work was conducted using scientifically-sound methodologies consistent with professional and relevant scientific practices).

**Review Topic 1: Hydrodynamic Analysis (Subtask E Memorandum)**

1. Comment on the overall organization, clarity, and general effectiveness of the memorandum. Is it clear what was done, why it was done, and what was learned? If not, state deficiencies and provide recommendations or suggestions on how the deficiencies might be resolved or improved (e.g., re-organization of the memorandum).

2. Comment on the overall technical quality of the memorandum. Are the assumptions used in applying the New York Harbor Observing Prediction System (NYHOPS) model for "particle tracking," embayment contribution modeling, area of influence estimation, and salinity modeling reasonable? Is employment of the NYHOPS model for "particle tracking," embayment contribution modeling, area of influence estimation, and salinity modeling consistent with relevant existing and emerging scientific practices? Are the results reasonable, and are the conclusions justified and adequately qualified where necessary? Are the results consistent with sound ecological science? Do the embayment mixing values seem realistic and hydrologically valid?

3. Is it ecologically valid to assume that total nitrogen (TN) is conservative (i.e., that it is not being removed from the system in significant amounts) for the purposes of this modeling effort?

**Review Topic 2: Empirical Modeling and Nitrogen Endpoints (Subtask F/G Memorandum)**

1. Comment on the overall organization, clarity, and general effectiveness of the memorandum. Is it clear what was done, why it was done, and what was learned? If not, state deficiencies and provide recommendations or suggestions on how the deficiencies might be resolved or improved (e.g., re-organization of the memorandum).

2. Are the TN endpoints and targets laid out in an understandable way in the *Subtask G. Nitrogen Endpoints* section of the memorandum? Are the graphs showing the hierarchical model easily understandable?

3. Comment specifically on the methods used to recommend TN endpoints. Are the methods used to identify recommended TN endpoints and ranges scientifically valid and laid out in a clear way? Are the TN endpoint values reasonable for protection of the region? Are the assumptions clearly presented? What are the minimum data requirements for applying the methods to establish TN endpoints applicable to individual embayment whether for purposes of protecting Long Island Sound or the embayment itself? What considerations should be given to application of the methods to non-homogenous embayments to ensure that the TN endpoints are protective of all portions of the embayment?

4. Is it reasonable to group the western and eastern narrows together for modeling and endpoint development purposes?

5. Is it reasonable to use eelgrass protection as an endpoint in both embayments and shallow open water (i.e., in the Western and Eastern Narrows)? Is the rationale for using eelgrass protection as an endpoint, both in embayments and shallow open water, well-articulated?

6. Does the model and do the data used depict a reasonable snapshot of current condition in the Sound? Could such a model be adapted to consider future conditions (i.e., higher temperatures and sea level rise)?

7. Is the rationale for use of in-water TN concentration (as opposed to other nitrogen endpoints such as watershed TN loading) in the stressor-response modeling well-explained and documented? Are there additional considerations that should be taken into account when relating nitrogen endpoints to response variables such as chlorophyll *a* and DO?

8. Comment on the approach used for the Literature Review Analysis (LRA) Line of Evidence Method. Is this approach consistent with professional and relevant existing and/or emerging scientific practice? Is the outcome reasonable? Are the literature values selected reflective of protective values for the geographic area? Is the rationale for inclusion or exclusion of values from certain geographic areas justified and valid (i.e., Great Bay, Chesapeake Bay, etc.)? Would application of values from excluded geographic areas (i.e., Great Bay, Chesapeake Bay, etc.) be scientifically appropriate? Is the use of the MassBays reports for the literature review justified given the similar geographic location and hydrological features to Long Island Sound? Is the rationale for this decision apparent in the memorandum? Is the exclusion of Chesapeake Bay literature justified based on geographic location and hydrological features compared to Long Island Sound? Is the rationale for this decision apparent in the memorandum?

9. In your opinion, is it scientifically valid to eliminate TN values from the LRA Line of Evidence Method that are in excess of values known to cause severe degradation and to cap recommended TN endpoint values at levels known to be protective? In your opinion, is the chosen cut-off value of 0.8 mg/L TN and above an appropriate cap value for this purpose? Note: using a degradation cut-off threshold of 0.8 mg/L TN and above resulted in a maximum literature value of 0.6 mg/L TN (i.e., the next highest value below 0.8 mg/L TN).

10. Comment on the Stressor-Response Modeling (SRM) Line of Evidence Method. Comment specifically on the method used to construct the hierarchical models, their execution, and outputs.

    a. Are the selected target light attenuation values reasonable and consistent with accepted ecological science for the Long Island Sound and Southern New England regions? Do tannin-colored waters (e.g., Pawcatuck River) impact the light extinction coefficients?

    b. Comment on the quantile regression model used for chlorophyll *a* versus the light attenuation coefficient, $K_d$. Is the use of this technique sound and is it an adequate model for the goal of setting chlorophyll *a* endpoints? Are the selected chlorophyll *a* endpoints scientifically valid for the LIS?

    c. Is the use of a hierarchical model appropriate for this kind of analysis? Is adequate justification provided in the memorandum for the use of this methodology? Are the statistical methods used in the hierarchical models clearly explained and technically valid? Is the goodness of fit of each modeled relationship adequately presented and interpreted? Should acceptable significance values or quality standards be made explicit? Are the nitrogen concentration endpoints developed in this model ecologically reasonable? Would they be considered protective of eelgrass in the region? Is it appropriate to show the modeled TN concentrations for two chlorophyll *a* levels (when applicable) in a single embayment?

    d. Is it reasonable to include the lower Connecticut River with the 23 priority embayments for modeling purposes? Is this inclusion ecologically and hydrologically sound? Is it reasonable to model a TN endpoint for the Connecticut River based on a hierarchical model built on water quality observations from the 23 priority embayments?

    e. The outputs of the hierarchical model were often above 0.5 mg/L or below 0.2 mg/L. Is it regionally, ecologically, and scientifically credible to assume TN values above 0.49 mg/L are not protective of eelgrass and concentrations below 0.2 mg/L are below the background concentration for the region? Is it appropriate to give the unaltered output of the model a caveat explaining this purportedly realistic/protective range? Is it regionally, ecologically, and scientifically valid to assume TN values above 0.49 mg/L are not protective of eelgrass and concentrations below 0.2 mg/L are below the background concentration for the region?

f. Is the use of chlorophyll *a* corrected rather than chlorophyll *a* measurement adequately explained and justified? Are the methods used to collect chlorophyll *a* data appropriately assessed and interpreted as similarly indicative of phytoplankton biomass (e.g., considering whether measurements represent similar corrections for dead biomass that does not contribute to life processes for production or respiration) when using chlorophyll *a* for stressor-response relationships? How should dead biomass be treated?

11. Comment on the approach used for the Distribution-based Approach (DbA) Line of Evidence Method. Is this approach scientifically valid? Is the outcome reasonable? Is the rationale behind this approach clear? Are the TN values reflective of protective values for the Long Island Sound's geographic area?

12. Many estuaries and embayments on the central and eastern regions of Long Island Sound currently have TN and chlorophyll *a* concentrations that are near the levels recommended (chlorophyll *a* of 3-10 mg $L^{-1}$ and TN of 0.3 to 0.5 mg $L^{-1}$) by the Literature Review Analysis (LRA), Stressor-Response Modeling (SRM), and Distribution-based Approach (DbA) approach used in the analysis (examples include G1 Pawcatuck River, CT and RI, G2 Stonington Harbor, CT, G5 Mystic Harbor, CT, G6 Niantic Bay, CT, G9 Northport Centerport Harbor, NY, G10 Port Jefferson Harbor, NY, G11 Nissequogue River, NY, G12 Stony Brook Harbor, NY and G13 Mt. Sinai Harbor, NY). Despite TN and chlorophyll *a* near the target threshold values, ecosystem function and aquatic life support are still impaired in many of these systems as evidenced by reduced DO, macroalgal blooms, harmful algae blooms (HAB) (e.g., annual HAB shellfish closures in Northport Harbor system), reduced benthic infauna abundance and diversity, and declining eelgrass abundance. In light of these facts, are the recommended chlorophyll *a* and TN targets justified as being protective of aquatic life? Is it adequately documented that water column TN and chlorophyll *a* targets are protective of aquatic life in embayments dominated by macroalgae?

## 3. Overview of Major Findings and Recommendations

This section contains an overview of the major findings and recommendations provided by the technical review team. This overview does not represent an exhaustive list of all of the findings and recommendations from the technical review. Readers are strongly encouraged to review the full technical review responses in Section 4. Note also that this summary does not represent USEPA positions or interpretations on the reviewer responses.

**Review Topic 1: Hydrodynamic Analysis (Subtask E Memorandum)**

1) The reviewers noted numerous instances in the technical memorandum where information was unclear, incomplete, or confusing. Some reviewers stated that the lack of clarity may have limited their ability to fully assess the technical quality of some of the work. The individual reviewer responses in Section 4 provide detailed feedback on specific instances where additional detail is needed to better communicate a concept or to justify a decision.

2) The reviewers agreed that the use of the NYHOPS model for particle tracking, embayment contribution modeling, area of influence estimation, and salinity modeling is consistent with existing scientific practices and a reasonable approach given the constraints in time and resources available to support the analysis. However, the reviewers noted numerous limitations in the use of the NYHOPS model, and provide extensive feedback on potential ways to strengthen the analysis and/or improve upon the explanation and justification in the report (see Section 4). The following are some of the key items that were noted as needing additional analysis, explanation, and/or justification:

    a. Spatial scale - Include more detail on the spatial scale (e.g., grid cells) used to represent the individual embayments in the model.

    b. Temporal scale - Reviewers expressed concern that focusing the model analysis on a limited temporal scale (i.e., July - September) may underestimate the annual total nitrogen loading and the far-field influence of large rivers. The reviewers recommend including more detail and rationale for the use of a limited (summertime only) temporal scale in the analysis.

    c. Dilution threshold - Include more detail and rationale for selection of 40% as the dilution threshold used to identify the areas of riverine influence.

3) The reviewers expressed concern with the assumption made in the modeling analysis that total nitrogen behaves as a conservative substance in the embayments, and do not feel that the assumption is justified in the technical memorandum. One reviewer stated that the assumption is not generally valid. In addition to providing more detail and justification for this assumption, the reviewers recommend inclusion of analyses to validate the assumption that total nitrogen behaves as a conservative substance in the embayments.

**Review Topic 2: Empirical Modeling and Nitrogen Endpoints (Subtask F/G Memorandum)**

1) The reviewers noted numerous instances in the technical memorandum where information was unclear, incomplete, or confusing. Some reviewers stated that the lack of clarity may have limited their ability to fully assess the technical quality of some of the work. The individual reviewer responses in Section 4 provide detailed feedback on specific instances where additional detail is needed to better communicate a concept or to justify a decision.

2) One reviewer noted inconsistency and confusion in the definition and use of important terms in the technical memorandum (i.e., endpoint, response variable, and assessment endpoint). The *LIS Literature Review* Memorandum is noted as accurately defining these terms consistent with USEPA technical guidance documents (i.e., TN concentration is a primary causal variable, chlorophyll *a*, $K_d$, and DO are primary response variables, and eelgrass and aquatic life are assessment endpoints). However, the terms are not used accurately in the Subtask F/G Memorandum (e.g., chlorophyll *a* is referred to as an endpoint, but it is a response variable).

3) The reviewers agreed that the use of a multiple lines of evidence approach (i.e., literature review analysis, stressor-response analysis, and distribution-based approach) to establish total nitrogen endpoints is scientifically-sound, consistent with existing approaches, and makes the best use of all available data and information.

4) The reviewers noted that the technical memorandum is lacking critical justification for grouping the western and eastern narrows together for modeling and endpoint development. The reviewers recommend inclusion of additional detail and justification for this decision, including use of data visuals to support the rationale for this decision.

5) The reviewers agreed that the use of eelgrass protection as an endpoint for the embayments is scientifically-sound and appropriate. However, some of the reviewers expressed concern and uncertainty about the validity of eelgrass protection as an endpoint for the western and eastern narrows (i.e., open water), and that additional detail and justification for this decision is needed in the technical memorandum.

6) The reviewers noted that the technical memorandum does not include discussion on the rationale for use of total nitrogen concentrations in the stressor-response analysis. While the reviewers agreed that use of total nitrogen concentrations is scientifically-sound and consistent with approaches in other estuarine systems, some of the reviewers noted that additional analysis may be warranted to evaluate and consider the relationship between estuarine chlorophyll *a* and watershed total nitrogen loading.

7) The reviewers agreed that the use of a literature review analysis approach for establishing total nitrogen endpoints is a scientifically-valid method and a good first (screening level) step. However, reviewers noted that estuarine total nitrogen concentrations are highly site-specific, and cautioned against assuming that total nitrogen concentrations from other systems can be directly translated to LIS.

8) The reviewers agreed that the use of a stressor-response modeling approach for establishing total nitrogen endpoints is a scientifically-valid and rigorous method. However, the reviewers expressed numerous concerns with how the method was applied to LIS, identified flaws in some of the assumptions made in the analysis, and question the scientific validity of the final results. For example, one of the reviewers noted a fundamental flaw in the conceptual model that $K_d$ is assumed to depend only on chlorophyll *a* concentrations (in addition to this not being correct, it is not supported by observed data for LIS and other estuaries and bays). The reviewers recommend that the methodological issues that they noticed (see Section 4) be addressed prior to acceptance of the total nitrogen endpoints derived using the stressor-response modeling method. The reviewers also noted that discussion of the stressor-response modeling approach and presentation of results in the technical memorandum is incomplete and lacks citations.

9) The reviewers agreed that the use of a distribution-based approach for establishing total nitrogen endpoints is a scientifically-valid method and consistent with USEPA technical guidance (USEPA, 2001). One reviewer noted that the analysis is limited due to the lack of inclusion of LIS data on the assessment endpoints (i.e., eelgrass and aquatic life) in the analysis; and that the analysis could be strengthened through analysis of data from other estuarine systems.

# 4. Technical Reviewer Responses

This section contains the original responses written by each of the technical reviewers. Statements represent the individual view of each technical reviewer; none of the statements represent analyses by or positions of USEPA.

## 4.1. Review Topic 1: Hydrodynamic Analysis (Subtask E Memorandum)

1. Comment on the overall organization, clarity, and general effectiveness of the memorandum. Is it clear what was done, why it was done, and what was learned? If not, state deficiencies and provide recommendations or suggestions on how the deficiencies might be resolved or improved (e.g., re-organization of the memorandum).

***Dr. Bierman's Response***

Given its purpose, the overall organization, clarity, and general effectiveness of the memorandum are adequate. Overall, it is clear what was done, why it was done, and what was learned.

***Dr. Brush's Response***

The memo is well-organized with a structure that was easy to follow, and I think overall it is effective at communicating what was done at a fairly high level and what was found. It ends with a very strong conclusions section. However there were a number of places where I felt greater explanation was warranted, or where the text was too confusing to fully understand what was done. I have detailed most of those in my response to Question 2 below, because they relate to my ability to assess the technical quality of the work.

The one item not covered in my response to Question #2 below is as follows: While not specifically relevant to Subtask E, not enough information was presented to understand Equation 1, and a proper citation was not provided. There was no derivation or discussion of the equation to allow for adequate review of its appropriateness for % reduction calculations.

***Dr. Janicki's Response***

The overall organization is good.

The clarity of presented information is adequate but could use improvement: Why is the SWEM mentioned in the Introduction, then never again? Is it necessary only because the Scope for this Task specified this model? If so, providing that logic would be helpful as well as an explanation for not including that discussion in the model selection section. (Page E-1).

The Introduction would benefit from additional text clarifying the Percent Reduction equation. Specifically, how are $C_w$ and $S_{LIS}$ defined? (Page E-1).

The addition of maps defining the extent of the embayments would be appropriate, so that the regions contributing to "...average salinities inside and outside the selected embayments..." could be defined. (Page E-8).

The identification of areas of influence is confusing. The maps show isopleths of percent river water contribution, but the captions for Figures E-4 through E-6 identify these lines as depicting "percent dilution". (Pages E-11 - E-13).

Tables E-4 through E-6 define dilution factors for each embayment. The algorithm used to calculate the dilution factors is not presented.

***Dr. Justic's Response***

The memorandum is well organized and well written.

2. Comment on the overall technical quality of the memorandum. Are the assumptions used in applying the NYHOPS model for "particle tracking," embayment contribution modeling, area of influence estimation, and salinity modeling reasonable? Is employment of the NYHOPS model for "particle tracking," embayment contribution modeling, area of influence estimation, and salinity modeling consistent with relevant existing and emerging scientific practices? Are the results reasonable, and are the conclusions justified and adequately qualified where necessary? Are the results consistent with sound ecological science? Do the embayment mixing values seem realistic and hydrologically valid?

***Dr. Bierman's Response***

Given its purpose, the overall technical quality of the memorandum is adequate. The assumptions used in applying the NYHOPS model are reasonable and consistent with relevant existing and emerging scientific practices. The results for salinity dilution analyses and dilution factors for the Connecticut, Housatonic, and Thames Rivers for selected embayments are reasonable.

More detail could be provided on model spatial scale for the individual embayments. In Table E-2 it is stated that the standard model grid is 500 x 500 meters, but that small embayments may not be fully resolved. It would be useful to provide a table that contains the number and spatial scale of model grid cells in each of the selected embayments.

The issue of model spatial scale is important because the spatial scales relevant to eelgrass in the individual embayments are much smaller than that of the standard NYHOPS model grid. For example, the LIS-wide Eelgrass Habitat Suitability Index (EHSI) model (Vaudrey et al., 2013) is based on grid cells that are 30.48 x 30.48 meters. The EHSI sub-models used for six selected study sites are based on grid cells that are 7.62 x 7.62 meters.

It is stated on Page E-17 that the lack of vertical mixing and diffusive exchange among grid cells affected the estimates of river water movement throughout the LIS. The reason for this is not clear.

Did the NYHOPS model not represent vertical mixing and diffusive exchange, or was information on these processes not included in the NYHOPS model outputs?

***Dr. Brush's Response***

I will divide my comments into those related to choice of hydrodynamic model, salinity modeling, and particle tracking (and associated area of influence calculations). I am unclear which aspect of the memo "embayment contribution modeling" refers to, but my review assesses all aspects of the document. As noted above, I thought the Conclusion section was a nice summary of the results with appropriate caveats and qualifications.

<u>Hydrodynamic Model Selection</u>

The case for using the NYHOPS model was solid and convincing. The model appears to be state of the art and entirely in line with existing practice. There are some limitations but I agree with the assessment that the best available model and approach were used given the available time and resources. However there are a few key details that should be addressed to fully evaluate the approach:

- Table E-2 indicates that the NYHOPS developer "recently completed a successful effort to validate the model performance for flow, temperature, and salinity in LIS." It is difficult to know how much confidence to place in modeled salinity and currents without seeing those validation results.

- Table E-2 also notes that "Small embayments may not be fully resolved on a 500-m grid," and it is not clear from Figure E-1 if all of the selected embayments contain grid cells. If the latter was the case, salinity could not be computed for those embayments. The text additionally states that different embayments have different numbers of grid cells (which could lead to biased salinity estimates), and that this likely led to overestimation of dilution of the landward, freshwater end-members. That said, I agree that this is the best approach with available resources. However the issue of embayments without any grid cells should be addressed.

- It is unclear if the individual embayments in the NYHOPS model were forced with reasonable estimates of local freshwater discharge. That is likely critical to an accurate simulation of salinity within the embayments and an accurate estimate of dilution with LIS water. If embayment discharge was not included, the computed dilution rates could be biased high. This should be clarified.

Salinity Modeling

As noted below, I have some concerns about the assumption of conservative mixing of TN. Putting that aside, I followed the calculations in this section and found it to be an elegant approach for estimating mixing. The approach is consistent with empirical estimates often employed by estuarine ecologists to compute bulk parameters like flushing time. As an aside, a hydrodynamic model is being used that computes velocity and therefore volume exchange, so I wondered why dilution was not computed directly, or via a particle tracking approach. Regardless, I find the approach entirely valid.

(Note: I did have one question about equations 6 and 7. While these are not required to compute the dilution rates, it seems that implementing these equations requires one to know the value of $C_o$ for each embayment. This will not affect the results of the subtask, but since these equations were given, I wondered if it was possible to implement them with available data.)

I was initially concerned about limiting the analysis to July-September, but I appreciated the sensitivity analysis using a broader temporal window, and agreed with the justification for using the more restricted time period. However a bit more justification on why this is a “critical time period” would be helpful.

It was unclear to me why salinity was computed only over the top five vertical layers of the model. I assume this was related to depth but this choice was not justified. Since the model uses sigma (terrain-following) coordinates, this approach could be averaging salinity inside and outside of the embayments over different depth ranges. I assume all grid cells were averaged within each embayment, although this is not stated. The report also does not state how many cells outside of each embayment were used to compute external salinity.

In conclusion, I find the approach elegant and appropriate given available resources, and in line with existing scientific practice. I also find the estimated dilution factors reasonable given the small volume of these embayments and what I imagine must be small freshwater inputs and strong tidal mixing with LIS. However, there are a few issues that should be addressed to increase confidence in the computed values.

Particle Tracking

Particle tracking is a commonly used, state of the art approach for tracking river plumes and dilution of point source inputs, and the overall approach used with the NYHOPS output appears rigorous. Particles were released only in the upper six vertical layers, because “only those layers had significant net lateral particle movement in the model.” This may be a problem because model estimates of particle transport would potentially be overestimated; i.e., if particles were released uniformly throughout the water column, those in lower layers would not travel nearly as far as those in surface layers. That said, model results show practically no influence of the rivers on the embayments, so addressing this issue would not likely affect the results in a significant way.

A potentially more important issue is that the calculations did not allow for vertical exchange of particles. While I agree that the best approach was used given resource constraints, the lack of vertical exchange may be problematic in a system characterized by two-layer estuarine circulation.

I found it very unclear how the actual percent dilution of particles was computed (pp. E-7 to E-8). Particles were released every four hours over the entire growing season. It was not clear to me how resulting concentrations were averaged or integrated over each release, the entire season, each year, and the six depth layers. It was also unclear how the original release point cell concentrations were computed for the same reasons. I also wondered if movement of particles among sigma layers and cells with varying dimensions led to artificial concentration or dilution of particles in these calculations.

The choice of 40% river water / 60% dilution seems to be an arbitrary choice, even with the justification provided on page E-10. An alternative approach might be to express influence across the continuum of dilutions. (Note: I think there are typos on p. E-10 in that 'river water' was intended instead of 'percent dilution' in the three places this term appears. This is also the case in the legends for Figures E-4 to E-6. I also note that the justification on p. E-10 about spring runoff seems out of place since the model analysis is limited to July-September.)

The release points on Figures E-5 and E-6 seem too far up-river (≥ 2 miles) to represent discharge at the river mouth. I was also concerned about the increasing contours on Figure E-4 as one moves up-river, and the up-river pattern in Figure E-6 where the percentages decrease to 50, then increase to 90, and then decrease again to 80. These patterns were not explained, but I think they pose concerns for the model output. Perhaps the grid resolution is not sufficient in the rivers for this analysis? (Admittedly the focus of the analysis is down-river from the release point, so this may not be an issue.)

Finally, the calculation of river dilution factors (p. E-13, Table E-4) was not explained so I am unable to evaluate these findings. That said, the values are extremely low which is what I would expect. However, the underlying assumption is that concentrations originating from these rivers are "conserved and superposable" (p. E-16). I recognize that far-field impacts of river discharges is a common management question in LIS, but given that nutrients do not behave conservatively within estuaries, particularly over the large spatial scales between these river mouths and most embayments, I am not sure how much can be gathered from these types of calculations.

Overall, I think that there are a number of uncertainties in the particle tracking approach. I again agree that the best approach was used given the constraints. However, I think a number of the issues identified above should be addressed and more clearly explained before accepting the particle tracking results.

***Dr. Janicki's Response***

The assumptions are clearly stated and reasonable. Use of the NYHOPS model is appropriate and results are reasonable and consistent with normal practices. The conclusions are justified and qualified appropriately. The embayment mixing values appear reasonable and valid, but are limited by the temporal and spatial resolution of the model used to derive them.

***Dr. Justic's Response***

The NYHOPS model appears well suited for the analysis of hydrodynamics and salinity in LIS. However, it is unclear why only the July-September period was used in the analysis. While it is true that this period coincides with the summer growing season, the freshwater inflows are minimal at this time, which has profound effect on the estimates of dilution as well as the estimates of the areas of influence of the Connecticut, Housatonic, and Thames rivers. One could argue that taking into account only the summer growing season underestimates the far-field influence of these rivers.

Further, the calculated dilution ratios are close to 1 for most embayments, as there are virtually no salinity differences between embayments and open areas during July-September (Table E-3, page E-9). From an ecological point of view, it would be valuable to estimate the areas of influence during the spring runoff period, when the regions of riverine influence could be substantially larger compared to the July-September period. This is important because in shallow coastal systems, such as LIS embayments, a considerable portion of external nutrient loading can be taken up and stored in biota and sediments over weeks and months, and subsequently recycled to fuel pelagic and/or benthic food webs.

No clear justification was provided as to why the 40% dilution threshold was used to identify the areas of riverine influence (page E-10): "*A value of more than 40 percent dilution was used to define the area of influence because there is error around this value as a function of changes in flow*". Further, the attempt to extend the area of influence estimates for the growing season to higher flow periods does not appear to have a sound scientific basis: "*As such, 40 percent based on the growing season period would encompass an area more likely to include more than 50 percent dilution during higher flow periods (e.g., during spring runoff).*"

The fact that a coupled hydrodynamic-biogeochemical model was not available for this study resulted in several assumptions that are not well supported by the available data, such as the assumed conservative behavior of TN (see my response to Question 3 below). Another unsubstantiated statement refers to the fate and residence times of riverborne nutrients (page E-17): "*Because of significant tidal flushing of water, nutrient loads from winter are likely retained only into the late summer primarily through storage in sediment and biota (dissolved nutrients will be flushed out).*" I cannot comment on whether a coupled hydrodynamic-biogeochemical model could have been implemented to aid in this study, but these uncertainties are important and if they remain unaddressed, can lead to later challenges.

3. Is it ecologically valid to assume that total nitrogen (TN) is conservative (i.e., that it is not being removed from the system in significant amounts) for the purposes of this modeling effort?

***Dr. Bierman's Response***

Nitrogen was not part of the hydrodynamic modeling effort described in Subtask E. The memorandum states that results from this effort for dilution of salinity will be used as a proxy for nitrogen dilution under the assumption that nitrogen within embayments is approximately conservative. This assumption is not generally valid, especially within embayments and nearshore areas. Not only will there be some settling and volatilization losses, as the memorandum states on Page E-1, but there will also be gains due to sediment diagenesis and the resulting sediment-water diffusion of nitrogen. These processes are complex and can vary in both space and time, especially between embayments and open water areas.

As an example, in their linked watershed-embayment modeling study of the Pleasant Bay System, Massachusetts, Howes et al. (2006) investigated sediment-water exchanges of nitrogen. Howes et al. (2006) Figure IV-20 (below) is a conceptual diagram showing seasonal variation in sediment nitrogen flux. During summer (i.e., the primary period of interest identified in Subtask E), sediment-water nitrogen flux is at maximum values. Howes et al. (2006) Table VI-2 (below) contains total nitrogen loads for individual sub-embayments. The loads for net benthic flux were based on site-specific measurements during the summer period. For most of the individual sub-embayments, and for the system as a whole, net benthic flux of nitrogen to the water column was larger than external nitrogen loads from the watershed itself. For the Pleasant Bay System, the assumption that nitrogen within sub-embayments is approximately conservative is violated by greater than a factor of two.

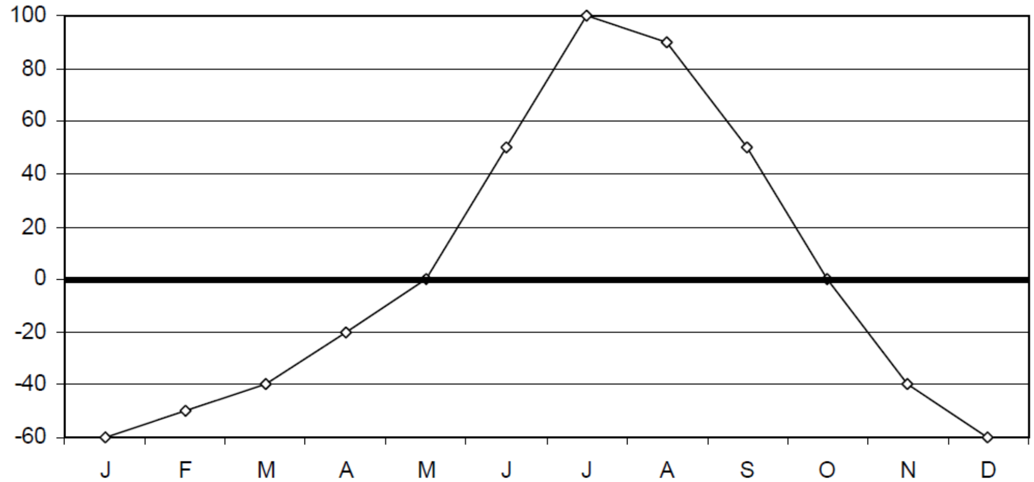


Figure IV-20. Conceptual diagram showing the seasonal variation in sediment N flux, with maximum positive flux (sediment output) occurring in the summer months, and maximum negative flux (sediment up-take) during the winter months.

Table VI-2. Sub-embayment and surface water loads used for total nitrogen modeling of the Pleasant Bay system, with total watershed N loads, atmospheric N loads, and benthic flux. These loads represent **present loading conditions** for the listed sub-embayments.

| sub-embayment | watershed load (kg/day) | direct atmospheric deposition (kg/day) | benthic flux net (kg/day) |
|---|---|---|---|
| Meetinghouse Pond | 6.197 | 0.584 | 14.365 |
| The River – upper | 2.773 | 0.288 | 6.263 |
| The River – lower | 3.879 | 2.241 | 10.480 |
| Lonnies Pond | 2.441 | 0.225 | 1.591 |
| Areys Pond | 1.304 | 0.181 | 5.996 |
| Namequoit River | 2.737 | 0.523 | 14.570 |
| Paw Wah Pond | 1.860 | 0.082 | 3.630 |
| Pochet Neck | 8.422 | 1.767 | -0.791 |
| Little Pleasant Bay | 7.496 | 24.023 | 37.226 |
| Quanset Pond | 1.781 | 0.170 | 5.988 |
| Tar Kiln Stream | 6.123 | 0.066 | - |
| Round Cove | 4.225 | 0.170 | 8.416 |
| The Horseshoe | 0.638 | 0.063 | - |
| Muddy Creek - upper | 9.981 | 0.162 | 4.560 |
| Muddy Creek - lower | 8.477 | 0.205 | -1.226 |
| Pleasant Bay | 23.159 | 19.153 | 149.013 |
| Pleasant Bay/Chatham Harbor Channel | - | 17.786 | -40.192 |
| Bassing Harbor - Ryder Cove | 9.819 | 1.296 | 9.356 |
| Bassing Harbor - Frost Fish Creek | 2.904 | 0.096 | -0.154 |
| Bassing Harbor - Crows Pond | 4.219 | 1.389 | 0.612 |
| Bassing Harbor | 1.668 | 1.071 | -4.976 |
| Chatham Harbor | 17.099 | 14.153 | -40.208 |
| **TOTAL - Pleasant Bay System** | **127.203** | **85.693** | **184.519** |

***Dr. Brush's Response***

While I agree that the best approaches have been used given available time and resources, I am skeptical about the assumption that TN will behave conservatively. Estuaries, and perhaps especially the shallow, fringing embayments which are the focus of this work, are well known as major processors and transformers of nutrients as they move from land to sea. Particularly, estuaries are the sites of substantial removal of dissolved inorganic nitrogen (DIN) via denitrification. Additionally, N taken up by phytoplankton and benthic primary producers within these shallow systems may also represent an important sink, at least over the growing season identified here (July-September). While TN is likely to be more conservative than DIN, I am not convinced it can be considered conservative. The text notes that conservative behavior is likely over the spatiotemporal scales considered in the report; since these are small, likely rapidly flushed systems, this may indeed be the case. However, to confirm this, estimates of embayment flushing times would be helpful. Another option would be to create mixing diagrams for TN in the embayments where stations exist along the salinity gradient to test for conservative mixing. (Note: as an aside, I am unsure what is meant by the statement that "Larger losses are likely expected within the watershed ...", or how that applies to conservative behavior within the embayments.) Overall, I do believe that this effort represents a reasonable first-order approach, but I recommend following this up with additional work that tests the validity of the conservative mixing assumption, or accounts for the potential non-conservative behavior of nitrogen.

***Dr. Janicki's Response***

This assumption imposes a limit on the interpretation of the final results as it does not incorporate biological activity or sediment interactions. Comparison of predicted TN concentrations to observed TN concentration data would provide insight on the validity of the assumption of conservatism. Despite the potential shortcomings associated with this assumption, this exercise has value in the decision making process.

***Dr. Justic's Response***

This issue was briefly discussed on page E-1 of the Subtask E report but clear justification for why TN could be considered conservative was not provided. The reader is referred to the NYHOPS model documentation, where I could not find any reference regarding TN. Further, this topic is not discussed in the LIS Literature Review Memo. Thus, based on the documentation available, I cannot conclude that the assumption concerning conservative behavior of TN in LIS is justified.

I am not entirely familiar with LIS literature, but in the systems I have studied (e.g., deltaic Gulf of Mexico (GOM) estuaries) TN does not behave conservatively and its concentrations can vary over an order of magnitude due to varying sources (e.g., riverine, atmospheric, sediment resuspension, marsh erosion) and sinks (e.g., denitrification, burial) affecting both the inorganic and organic pools of nitrogen. Importantly, plots of TN versus salinity for GOM estuaries do not support the assumption that TN behaves conservatively.

TN is a critical component for estimating LIS nitrogen reductions and further analysis of the conservative/non-conservative nature of TN in LIS is recommended. Insights from other systems do not always reflect reality, but if they remain unaddressed, can lead to later challenges.

## 4.2. Review Topic 2: Empirical Modeling and Nitrogen Endpoints (Subtask F/G Draft Memorandum)

1. Comment on the overall organization, clarity, and general effectiveness of the memorandum. Is it clear what was done, why it was done, and what was learned? If not, state deficiencies and provide recommendations or suggestions on how the deficiencies might be resolved or improved (e.g., re-organization of the memorandum).

***Dr. Bierman's Response***

The overall organization, clarity, and general effectiveness of the memorandum could all be substantially improved. My general recommendations are listed below and more specific comments and suggestions are provided in my responses to other questions.

a. The memorandum confounds the definitions of important terms. Consistent with the conceptual model in Figure F-4 and USEPA (2010) guidance on stressor-response relationships, TN is the primary causal variable, chlorophyll *a*, $K_d$, and DO are the primary response variables, and eelgrass and aquatic life are the assessment endpoints. Operationally, DO was used as a surrogate for aquatic life and this makes sense. However, although the memorandum frequently refers to them as such, chlorophyll *a* and $K_d$ are not endpoints. Consistent with the conceptual model in Figure F-4, the purpose of chlorophyll *a*, $K_d$, and DO is to link TN concentrations to the assessment endpoints (eelgrass and aquatic life) via the relationships depicted in Figure F-5. Finally, TN concentrations should be characterized as threshold concentrations (e.g., Howes et al., 2003) or target concentrations, not as endpoints.

b. There is inconsistency between this memorandum (Subtasks F/G) and the Literature Review Memorandum with respect to the definitions of important terms. The latter document correctly characterizes assessment endpoints and nitrogen thresholds in a way that is consistent with the relevant USEPA technical guidance documents. The Subtasks F/G Memorandum should be revised so that it is consistent with the characterizations and terminology in the Literature Review Memorandum and USEPA technical guidance.

c. None of the 12 equations in the memorandum are numbered. All of them should be numbered for easier reference.

d. Final statistical models are presented for $K_d$ vs chlorophyll *a* (Page F-14), DO vs chlorophyll *a* (Pages F-17 and F-21), and chlorophyll *a* vs TN (Pages F-18 and F-22). All of the covariates investigated for each model should be listed, not just the covariates in the final models.

e. None of the actual values for the coefficients in any of the above final models are presented. All of these values should be presented so that the relative magnitudes of the individual terms in each of the models can be assessed.

f. Plots for "observed" vs "fitted" values are presented on Pages F-14, F-18, F-20, F-22, and F-23, but none of the axes are labeled with the parameters that are plotted. These parameters can be inferred from context, but all of these axes should be labeled for complete clarity.

g. A plot of observed data for $K_d$ vs chlorophyll *a* for embayments, along with results for the 10$^{th}$ quantile model, is shown on Page F-16 but no plots of observed data for DO vs chlorophyll *a* or chlorophyll *a* vs TN for embayments or open waters are shown. These plots of final models vs data should be presented. Statistical analyses alone are not a substitute for visual inspection of the actual observed data.

***Dr. Brush's Response***

Overall, the memorandum is very well organized and effective at presenting what was done, why it was done, and what was learned. The overview of hierarchical and multiple regression modeling was particularly excellent and very informative, as were the justifications for using each line of evidence, and general explanations of how each was developed. While the memo is generally clear, I identified some sections of text that were difficult to follow and would benefit from clarification, and also some issues regarding use of terminology that could be clarified. I detail those in my responses to the topic specific questions below.

***Dr. Janicki's Response***

The report is well written with concise explanation of the purpose, need, objectives, and approach. If there is an issue, I think it is in the lack of discussion as to specifically how these endpoints will be used. Having that sense of context may raise questions that are not discernable if the review is simply focused on the "nuts and bolts" of the approach and implementation.

***Dr. Justic's Response***

The memorandum is well organized and well written.

2. Are the TN endpoints and targets laid out in an understandable way in the *Subtask G. Nitrogen Endpoints* section of the memorandum? Are the graphs showing the hierarchical model easily understandable?

***Dr. Bierman's Response***

The TN targets for protection of aquatic life based on the Literature Review Analysis (LRA) and the Distribution-based Approach (DbA) lines of evidence are understandable because they are taken directly from Tables F-1 and F-10, respectively, and the same values are applied to each of the individual embayments.

The TN targets based on the Stressor-Response Modeling (SRM) are difficult to understand. It is not clear how the chlorophyll *a* vs TN relationships for the individual embayments are related to the final chlorophyll *a* vs TN model on Page F-18. It is not clear that the chlorophyll *a* "endpoint" value of 10 ug/L is actually not an "endpoint" but corresponds to the $K_d$ "endpoint" of 0.70 (Vaudrey, 2008) in Table F-6 which, in turn, was derived from the $10^{th}$ quantile regression relationship in Figure F-7. It is not clear that the chlorophyll *a* "endpoint" value of 5.5 ug/L was not derived using a $K_d$ "endpoint" but was taken directly from Vaudrey (2008). Finally, some of the plots for the chlorophyll *a* vs TN hierarchical models in each embayment have no observed data, some of them show no apparent relationship (or only a weak relationship) between chlorophyll *a* and TN, and many of the data lie outside the 90% confidence limits. It is difficult to understand how these SRM results are lines of evidence that can support the listed TN target concentrations.

***Dr. Brush's Response***

First, the summary of how TN endpoints were computed on p. G-1 is excellent. I also think the presentation of endpoints for each embayment or region of LIS in this subtask is excellent. The tables and graphs are easy to understand, and the supporting text and maps are similarly good. I have only minor, editorial suggestions:

- In the first column of the table, would a better entry for the STM approach be "Eelgrass protection"? That is what the approach was designed to do. Similarly, eelgrass was the target for the literature review in the embayments, although not for open water. Perhaps this gets too complicated and the first column should just be removed. Targets for protection could be summarized in a footnote instead.

- The third column heading should read "Endpoint Chlorophyll *a* Value (ug/L)" for clarity.

- Suggest changing "values or concentrations" to "concentrations" in line 4 of the "TN Endpoints Discussion" sections.

- Tables G-10 and G-12 have an extra footnote referencing a population model. Why was a different model used relative to the other tables (especially given all the data present in these two systems)?

***Dr. Janicki's Response***

The presentation of the TN endpoints and targets was adequate and should be understandable to most readers. The hierarchical modeling graphics also should be understandable to most readers.

***Dr. Justic's Response***

The TN endpoints and targets are clearly explained and the graphs are easily understandable.

3. Comment specifically on the methods used to recommend TN endpoints. Are the methods used to identify recommended TN endpoints and ranges scientifically valid and laid out in a clear way? Are the TN endpoint values reasonable for protection of the region? Are the assumptions clearly presented? What are the minimum data requirements for applying the methods to establish TN endpoints applicable to individual embayment whether for purposes of protecting Long Island Sound or the embayment itself? What considerations should be given to application of the methods to non-homogenous embayments to ensure that the TN endpoints are protective of all portions of the embayment?

***Dr. Bierman's Response***

The LRA method is scientifically valid and laid out in a clear way. It is always a good first step because it allows identification of TN concentrations and ranges corresponding to various assessment endpoints (e.g., eelgrass and aquatic life) in other similar waterbodies. It also allows identification of relevant response variables and confounding factors that should be considered in attempting to link TN concentrations to these assessment endpoints. Although the LRA method can provide a useful screening-level analysis, it should not be assumed that specific TN concentrations and ranges from other waterbodies can be directly translated to LIS because these concentrations are strongly site-specific.

The memorandum states on Pages F-2 and F-3 that a decision was made to focus primarily on TN values from the most proximate study areas (Massachusetts) and not to incorporate values from farther north (Great Bay, NH) or south (Chesapeake Bay) because those systems were considered substantially different. This approach assumed that the Massachusetts estuaries literature-based targets were appropriate for LIS, given the similarities in geography, climate, and species composition (e.g., Zostera marina) consistent with similar physical and chemical habitat requirements in both embayment as well as shallow and deeper open water habitats between the two regions. Consequently, many of my comments on the memorandum draw upon approaches, analyses, and findings from the Massachusetts Estuaries Program (MEP).

The SRM methods themselves are scientifically valid, but not laid out in a clear way in the memorandum. USEPA (2010) recommends summarizing and visualizing datasets before conducting SRM statistical analyses, but this was not done in the memorandum. In addition, the applications of the SRM methods to LIS contain conceptual flaws and questionable assumptions, and their results do not provide scientifically valid support for the TN endpoints.

The DbA is a broad, generic approach that can be useful at regional scales and is laid out in a clear way in the memorandum. Selection of TN concentration targets by using the 25$^{th}$ percentile of all TN samples in LIS embayments and open waters (Table F-10) is consistent with USEPA protocol; however, because the DbA in the memorandum did not explicitly use any site-specific data for eelgrass distributions, the primary response variables (chlorophyll *a*, $K_d$, DO) or eelgrass physical habitat requirements (sediment grain size and total organic carbon), there is no assurance that these 25$^{th}$ percentile TN targets will protect the LIS assessment endpoints (eelgrass, aquatic life).

The values from the LRA appear reasonable, but are not based on site-specific data from the LIS embayments. The values from the DbA appear reasonable, but they are based only on site-specific TN concentrations and not on any other parameters directly related to eelgrass or aquatic life. The values from the SRM are conceptually flawed and scientifically invalid (see my responses to Questions 10a – 10f for details and specific examples.

With regard to minimum data requirements, the memorandum states on Page F-1 that seagrasses (eelgrass) and other aquatic life were selected for developing nitrogen endpoints. It states that these assessment endpoints are principally reflected by water column chlorophyll *a* (through its effect on light for seagrass growth) and DO (through its effect on benthic fauna and fishes). These statements are accurate but do not reflect all of the site-specific parameters that should be considered for applying the methods to establish TN endpoints for purposes of protecting Long Island Sound or the embayments themselves. For example, as stated on Page 200 in Howes et al. (2006):

> "Determination of site-specific nitrogen thresholds for an embayment requires the integration of key habitat parameters (infauna and eelgrass), sediment characteristics data and nutrient related water quality information (particularly dissolved oxygen and chlorophyll a)."

Koch (2001) acknowledges that light and parameters that modify light (epiphytes, total suspended solids, chlorophyll *a*, nutrients) are the first factors to consider when determining habitat suitability for seagrass, but points out that these factors alone do not explain why seagrass does not occur in areas where light levels are adequate. He goes on to emphasize the importance of also considering physical-chemical factors such as current velocity, waves, tides, salinity, sediment grain size distribution (GSD), sediment total organic carbon (TOC), and sediment sulfide concentration.

In the memorandum, the TN endpoint values from the LRA are based on those developed for other, proximate systems and not on site-specific data from LIS. The values from the DbA are based only on site-specific TN concentrations and not on any of the other above parameters. The independent variables in the final SRMs include chlorophyll *a*, TN, pH, salinity, and temperature, but none of the other above parameters. It is not known whether any of these other parameters were considered in the SRMs because the memorandum lists only the independent variables in the final models, not all of those that were actually investigated.

To ensure that the TN endpoints are protective of all portions of the embayment when applying the methods to non-homogenous embayments, it would be appropriate to consider the sentinel station approach used in the MEP. As stated on Page 204 in Howes et al. (2006):

> *"The approach for determining nitrogen loading rates, which will maintain acceptable habitat quality throughout an embayment system, is to first identify a sentinel location within the embayment and second to determine the nitrogen concentration within the water column which will restore that location to the desired habitat quality (threshold nitrogen level). The sentinel location is selected such that the restoration of that one site will necessarily bring the other regions of the system to acceptable habitat quality levels."*

See my specific responses to Questions 8, 10 and 11, for related discussion on this topic, including on the manner in which the assumptions are presented in the memorandum.

***Dr. Brush's Response***

First, I strongly support the use of chlorophyll *a*, light attenuation, and DO as assessment endpoints; these are the exact endpoints used by the long-standing USEPA Chesapeake Bay Program (CBP) and were developed after extensive deliberation over many years of work. If USEPA wishes to further pursue benthic fauna, they could look into the CBP DO criteria which specifically addressed estuarine fauna by thoroughly evaluating the literature for faunal-DO relationships.

The use of a multiple lines of evidence approach to establish TN endpoints, with uncertainty ranges in the case of two methods, is in line with best practice and existing approaches, and in my view excellent. The three approaches are scientifically valid and clearly presented. The methods for each approach were also generally well explained, with some caveats provided in the relevant sections below. Some of these caveats relate to issues with textual clarity and terminology; these do not take away from the validity of the analyses and can be addressed with some relatively simple clarifications in the memo. Caveats in the Stressor-Response Modeling section raise more important methodological issues which I believe should be addressed prior to final acceptance of those TN endpoints. That said, I found the conclusions reached after each analysis to be well supported by the data and analyses.

One minor point is that the text about DO endpoints on pp. F-11 and F-12 was somewhat confusing. Endpoints from three states were reviewed, but a final DO endpoint was not selected.

Not being from the LIS region and not being intimately familiar with TN endpoints in other systems, or typical values of TN across systems, it is difficult for me to comment on whether the TN endpoints will be protective of the region. That said, I agree with the approaches used, and once the methodological issues are addressed, I believe the resulting endpoints are well supported by the data. With the caveats that I identify in my responses to the review questions about certain areas that could be clarified, the assumptions of the methods are clearly presented and discussed in the text.

There is no easy answer to respond to the question about minimum data requirements for applying methods to establish TN endpoints to individual embayments. Certainly the more data available in a given system through both time and space, the better, and ideally one would want semimonthly to monthly data at multiple stations in each embayment over several years, or at least across years with varying discharge and meteorology. In practice, however, this is going to be difficult to achieve given the practicality of sampling and the limited resources available for monitoring. Given that, I think the use of multiple lines of evidence, and the approach to pool all available data across all embayments, and use a hierarchical modeling approach that uses the global relationship to “nudge” the results in embayments with limited data, is an ideal solution that makes the most of the available data. And I think the overall amount of data used in the analyses here is impressive. Of course, for those embayments with limited or no observations, the established TN endpoints will need to be used with appropriate caution. These embayments could be prioritized for future monitoring.

Regarding the application of the methods to non-homogenous embayments, while spatial gradients in TN will occur in all embayments, I do not believe it is necessary to consider this issue in the current analysis. First, as noted above I think the approaches used are an excellent way to use all the data. Second, these embayments are small and likely well mixed, and the analysis from Subtask E indicated substantial dilution by LIS water, so I expect spatial gradients to be small. Third, given the likely high rates of mixing within embayments, I do not think it would be appropriate to relate TN and chlorophyll *a* measured at a specific station to metrics such as eelgrass or DO at that same station; an embayment-wide value is a much better approach in my view. That said, one could take advantage of those embayments with multiple stations to analyze for the presence and magnitude of spatial gradients to better inform this question.

***Dr. Janicki’s Response***

The methods used to identify the recommended TN endpoints are valid. More specific comments regarding the methods used are provided in my response to Question 10 below. The TN endpoint values are reasonable for protection of the region but it should also be noted that attaining these endpoints can only be achieved by management of TN loading. All of the significant assumptions are not clearly presented. It is important to identify the ramifications of not achieving those assumptions. Determination of the minimum data requirements for applying the methods to establish TN endpoints applicable to an individual embayment whether for purposes of protecting Long Island Sound or any embayment cannot be achieved without further analysis of the available data. Consideration of the seasonality should be included. The endpoints for non-homogenous embayments may best be expressed as a range given the spatial variability in the ambient water quality conditions.

***Dr. Justic's Response***

The multiple lines of evidence approach (i.e., scientific literature analysis, stressor-response analysis and distribution-based approaches) is well explained. However, there are several important issues related to stressor-response analysis that are discussed in my response to Question 10.

Regarding embayment non-homogeneity, implementing a coupled hydrodynamic-biogeochemical model to selected embayments could be helpful in explaining the spatial patterns in TN, $K_d$, and chlorophyll *a*, and assessing the control of sediment TN pool on water column processes.

4. Is it reasonable to group the western and eastern narrows together for modeling and endpoint development purposes?

***Dr. Bierman's Response***

Reasons for grouping different water bodies should not depend solely on geography, but also on their designated uses, assessment endpoints, extent of impairment, and data availability/representativeness. The memorandum grouped the western and eastern narrows together for modeling and endpoint development, but did not explain the rationale for doing so. Tables F-8 and F-9 show substantially more paired data for the western narrows. This could have been a practical reason for combining these areas, but this decision could be better informed by at least a visual inspection of the western vs eastern water quality data (e.g., using box plots). For the eelgrass assessment endpoint, the habitat suitability maps in Vaudrey et al. (2013), especially Figure 11 (Exclusive Band) and Figure 22 (Sum of Ranked Parameters within the Exclusive Band) both provide additional information that could be used to inform the decision on combining the western and eastern narrows. For the DO endpoint, the decision to combine these areas could be informed by their designated uses (e.g., Class SA, SB and SC), the DO criteria corresponding to these uses, and the existence and/or degree of their impairment.

***Dr. Brush's Response***

Based on my knowledge of LIS, I think this approach is entirely reasonable, as these two regions encompass the western, most impacted region of the system. It is also an ideal solution given the limited data available in the eastern segment.

This was however one portion of the report that I found a bit confusing. Since the watersheds for Western and Eastern LIS are highlighted on Fig. F-1, and embayments in these regions appear to have stations (Fig. F-20), it was unclear to me for quite a while that this effort involved developing regression models for the open waters of LIS in this region rather than embayments in the two watersheds. The text could be clarified to reflect that early on in the memo. Another point of confusion was that p. F-20 says that limited data from Eastern LIS were excluded, but the following text and Tables F-8 and F-9 suggest that these data were included.

Finally, I offer one minor observation on the regressions that were attempted in this region. If the bottom DO-chlorophyll *a* regressions had been successful, it would have been unlikely to then find significant relationships between bottom chlorophyll *a* and TN, as chlorophyll *a* in the bottom has primarily sunk from surface waters where it was fueled by surface nitrogen (i.e., these linkages are separated in space and time).

***Dr. Janicki's Response***

Presentation of the ambient data from the two areas as well as more discussion as to the similarities or lack thereof in the physical nature of those areas would help in justifying this decision.

***Dr. Justic's Response***

In the materials provided I could not find a justification for why the western and eastern narrows were grouped together. From the All Waters Map, it appears that the western and eastern narrows include several disparate parts of LIS. The western and eastern narrows have different residence times (Subtask A Report; Tables A-26, A-27) and very different nitrogen yields (Subtask A Report; Table A-2).

5. Is it reasonable to use eelgrass protection as an endpoint in both embayments and shallow open water (i.e., in the Western and Eastern Narrows)? Is the rationale for using eelgrass protection as an endpoint, both in embayments and shallow open water, well-articulated?

***Dr. Bierman's Response***

The memorandum relies upon the Long Island Sound Eelgrass Habitat Suitability Index (EHSI) model and embayment bathymetry data developed by Vaudrey et al. (2013). It is reasonable to use eelgrass protection as an endpoint in the embayments, consistent with the ranking results of the five selected parameters in the EHSI model that were weighted and depicted in Figure 22 in the Vaudrey report. These results were implicitly taken into account in the memorandum because it used a habitat suitability target of greater than 50 to estimate maximum colonization depths of suitable eelgrass habitat in each embayment.

It is not reasonable to use eelgrass protection as an endpoint in shallow open water, specifically, the Western and Eastern Narrows. Figure 11 in the Vaudrey report shows that a combination of water depth, mean tidal amplitude, and % light reaching the bottom excludes the occurrence of eelgrass in shallow open waters in these areas, even if all other parameters are optimal. Furthermore, Figure 22 in the Vaudrey report shows that only very small nearshore areas in the Western and Eastern Narrows have habitat suitability scores greater than 50. Consequently, eelgrass protection would be a reasonable endpoint in only these small areas.

***Dr. Brush's Response***

I strongly agree with using eelgrass as an endpoint in the embayments, as it is critical habitat that provides numerous ecosystem services, and we know it currently grows there (I believe we also know that it has declined from previously higher levels). Eelgrass is also a key endpoint (or an indirect endpoint via $k_d$) in other systems, including the Chesapeake. I am less certain about using eelgrass in the open water of the Narrows, but that is only because I am unfamiliar with the distribution of eelgrass in the Sound. If eelgrass grows in the Narrows, or historically grew there, then I agree with its use.

The report does not spend much time discussing the rationale for using eelgrass as an endpoint beyond the first introductory paragraph, but I'm also not sure that more text is necessary. Based on the introduction, it appears that the choice of eelgrass was made by USEPA so I do not see why Tetra Tech would need to justify it here. I do think the report does a nice job of explaining the connections between TN, chlorophyll *a*, $k_d$, eelgrass, and DO (e.g., Fig. F-4), and how protecting eelgrass will be protective of other aquatic life uses. The Literature Review document provided with the supplemental materials for this review provides extensive justification of eelgrass as an indicator, along with several other variables.

***Dr. Janicki's Response***

The use of seagrasses of many types as an endpoint for restoration of estuarine waters is well documented and very appropriate here.

***Dr. Justic's Response***

Eelgrass is an important ecological resource and the rationale for using eelgrass protection as a management endpoint is well formulated.

6. Does the model and do the data used depict a reasonable snapshot of current condition in the Sound? Could such a model be adapted to consider future conditions (i.e., higher temperatures and sea level rise)?

***Dr. Bierman's Response***

The data used for the empirical modeling approaches (LRA, SRM, and DbA) depict a reasonable snapshot of current conditions in LIS. However, these models were applied to only a small subset of the minimum data requirements for establishing TN targets applicable to individual embayments. See my response above to Question 3.

The SRM models include temperature as an independent variable and, in theory, could be adapted to consider future higher temperatures. However, these models do not compute temperature but require temperature as an input, so future higher temperatures would need to be provided from some other source such as global/regional climate change models.

None of the three empirical modeling approaches explicitly include sea levels. Different models would be required to consider the impacts of future sea level rise.

***Dr. Brush's Response***

Since the data cover 588 stations over 17 years, primarily from the period 2006–2015, I believe the model and data provide an excellent snapshot of current conditions. Since temperature is a term in many of the models, and sea level rise would be inherently included in calculations related to $K_d$ and % $i_0$, I also believe that these models could be used to explore possible future scenarios. That said, the models are empirical so caution must be exercised not to extrapolate them too far outside the bounds of the data used to develop them.

***Dr. Janicki's Response***

Not sure what specific model is being referred to so it's difficult to draw any conclusions regarding ability to address the potential effects of climate change.

***Dr. Justic's Response***

The NYHOPS model is well suited to simulate present day hydrodynamics and residence times in LIS. The model should also perform well in simulating the impacts of future higher temperatures and sea level rise on hydrodynamics and salinity distribution in LIS. However, as indicated in my responses to the questions for Review Topic 1, the present study is heavily biased towards current summertime conditions (July-September period). Higher temperatures will likely increase the duration of the growth season, which, along with stronger stratification, could exacerbate eutrophication and increase the temporal/spatial extent of hypoxia. To consider the range of future conditions, the temporal domain of the model would have to be extended to other seasons.

It should be pointed out that hydrodynamic model simulations alone are generally inadequate for water quality forecasting. Implementing high-resolution coupled hydrodynamic-biogeochemical model (e.g., a biogeochemical model forced by NYHOPS outputs) to selected LIS embayments would be very helpful in dissecting the controls of various physical and biological factors on algal growth and hypoxia. Such a model would be very valuable for developing ecologically meaningful TN management endpoints and addressing the risks associated with future climate change.

7. Is the rationale for use of in-water TN concentration (as opposed to other nitrogen endpoints such as watershed TN loading) in the stressor-response modeling well explained and documented? Are there additional considerations that should be taken into account when relating nitrogen endpoints to response variables such as chlorophyll a and DO?

***Dr. Bierman's Response***

No, the rationale for use of TN concentrations vs TN loadings in the SRMs is not well explained or documented; however, in-water TN concentrations and TN mass loadings from the watershed are different physical quantities and neither of them are endpoints. As explained above in my response to Question 1, TN concentration is the primary causal variable, chlorophyll *a*, $K_d$, and DO are the primary response variables, and eelgrass and aquatic life are the assessment endpoints. If appropriate analyses are conducted with all of the relevant site-specific data, then TN concentration targets can be developed that will protect the assessment endpoints. In turn, an appropriate site-specific, load-response model can then be used to determine TN loads from the watershed that can meet the in-water TN concentration targets. This is the approach currently being used with the linked watershed-embayment model in the 89 MEP embayments (Howes et al., 2006).

***Dr. Brush's Response***

I did not find any discussion of the use of concentrations vs. loads, so the rationale is not documented. I did find two places that mention that endpoints were based on loads and/or concentrations (p. F-1, first bullet; p F-24, first paragraph); however the memo only used concentration data so this should be corrected. I do support the use of TN concentrations as these data are commonly measured and available, are related at least to some degree to metrics such as chlorophyll *a*, DO, and eelgrass, and have been used in multiple states and estuaries to establish criteria. That said, a great body of literature exists relating estuarine chlorophyll *a* (as well as other parameters) to nitrogen loading rather than concentration (e.g., Nixon, 1992 and many others). While loads are more difficult to estimate than concentrations, they do exist for these embayments and LIS as a whole. Loading rate is what drives eutrophication and water quality response rather than concentration, so I suggest future efforts should test for relationships against loading in addition to concentration. This is particularly important in these shallow embayments, as in-water nitrogen concentrations (at least dissolved inorganic forms) in productive, shallow water estuaries can be poor indicators of loading due to rapid biological uptake and denitrification in these systems (Nixon et al., 2001).

A related comment is that I recognize that TN has frequently been used by various states and programs for setting water quality criteria. However, I have always been a little uncomfortable with this, as TN integrates across all forms of N, including dissolved inorganic N (DIN), dissolved organic N (DON), and particulate N (PN). As such it includes N in the form of autotrophic nutrients, recycled organics, and bound in living and detrital biomass. While I agree with the intent of managing for all forms of N in a system, I have always felt that aggregating it into a single pool complicates these types of analyses. Additionally, in the current analysis, the relationship between TN and chlorophyll *a* is a bit circular, in that a significant portion of TN is likely bound up in phytoplankton, which is represented as chlorophyll *a* in this analysis. I would advocate for using DIN as a target, as that is the original form in which watershed N tends to enter an estuary, and is the most bioavailable form. Therefore DIN is what will drive the eutrophication and water quality response. Using DIN also removes the circularity between TN and chlorophyll *a*. I suggest that future analyses should analyze for relationships of chlorophyll *a* against DIN (concentrations and loads) as well as TN.

As described in my response below to Question 12, flushing time can be a useful metric when testing nutrient-response relationships. Specifically, normalizing concentrations and loads to flushing time can account for hydrodynamic differences among embayments (e.g., Nixon et al., 2001). There are any number of other parameters that could also be considered, but perhaps the next most important is mean depth of the embayment. While not an estuarine example, Vollenweider's original stressor-response models between phosphorus load and chlorophyll *a* in lakes adjusted the loads to account for both flushing time and depth (see Nixon et al., 2001).

***Dr. Janicki's Response***

My experience is that a stressor-response model that entails TN loading (watershed + atmospheric deposition) to predict responses in chlorophyll *a* to changes in loadings has proven to be particularly useful. Restoration necessarily involves some degree of loading reductions and a model that includes loading provides insight into the "how much" but also the likely loading sources that are most responsible for any existing water quality degradation.

Was there any consideration of the lag effects in the stressor-response modeling? Also, was the inclusion of residence times in the stressor-response modeling?

With regard to TN concentrations as a desirable endpoint, this seems to make most sense when defining the means by which future compliance with the endpoints will be assessed.

***Dr. Justic's Response***

Water column TN is an important endpoint for managing coastal ecosystems and its use is fully justified. However, given the shallow depths of most LIS embayments (average depth = 0.1 – 14.1 m; Subtask A report), it is very likely that nutrients and carbon stored in sediments exert considerable control on water column processes, including the dynamics of water column TN, chlorophyll *a*, turbidity, and DO. Sediments and benthic communities have been referred to as "eutrophication's memory mode" and it would be useful to include some sediment-based proxies of nutrient enrichment, such as organic carbon and nitrogen content. Sediment organics are mentioned in the LIS Literature Review Memo (e.g., page 40), but they have not been adopted as a requirement.

I am not entirely familiar with LIS monitoring programs and cannot comment on whether the available data on sediment organics across multiple embayments are sufficiently dense to be effectively used in this study. However, not adequately addressing the important roles that sediments play in these shallow systems can lead to unexpected system responses and subsequent management challenges.

8. Comment specifically on the approach used for the Literature Review Analysis (LRA) Line of Evidence Method. Is this approach consistent with professional and relevant existing and/or emerging scientific practice? Is the outcome reasonable? Are the literature values selected reflective of protective values for the geographic area? Is the rationale for inclusion or exclusion of values from certain geographic areas justified and valid (i.e., Great Bay, Chesapeake Bay, etc.)? Would application of values from excluded geographic areas (i.e., Great Bay, Chesapeake Bay, etc.) be scientifically appropriate? Is the use of the MassBays reports for the literature review justified given the similar geographic location and hydrological features to Long Island Sound? Is the exclusion of Chesapeake Bay literature justified based on geographic location and hydrological features compared to Long Island Sound? Is the rationale for these decision apparent in the memorandum?

***Dr. Bierman's Response***

As stated above in my response to Question 3, LRA is a scientifically valid method and a good first step, but it should not be assumed that TN concentrations and ranges from other systems can be directly translated to LIS because these concentrations are highly site-specific. The LRA method in the memorandum focused on TN concentration targets developed as part of the MEP. Although the MEP involves development of TN thresholds in 89 embayments, it used an approach that was highly site-specific and data intensive for each of these embayments.

As stated on Pages 2 and 3 of Howes et al. (2003):

> *"An essential component of the DEP/SMAST Massachusetts Estuaries Project (MEP) is the development of site-specific critical thresholds for the coastal embayments within the study region. While the qualitative nature of these thresholds will be common to almost all embayment systems, the quantitative thresholds will vary between and within embayments. Given that general thresholds (one size fits all) for embayments would have to be tailored to protect the most sensitive systems, this approach was rejected as it tends to "over manage" the less sensitive systems. The result of "over management" is the addition of significant additional and unnecessary costs to municipalities and the Commonwealth relative to the implementation of management alternatives. In contrast, site-specific thresholds are developed on the basis of specific basin configuration, source*

> *water quality and watershed spatial features for each embayment. By being tailored to each estuary's specific characteristics, the results are more accurate and require a smaller "safety factor" in the critical nitrogen targets used for developing nitrogen management alternatives. The site-specific approach has been recommended by the USEPA in developing Nutrient Criteria for estuaries (USEPA, 2001). The MEP has already determined that total nitrogen thresholds based upon the same habitat quality can vary more than 50%, due to their specific oceanographic setting. This wide range greatly increases the need for site specific quantitative thresholds, and reinforces the cost savings projections of this approach."*

As stated on Page 16 of Howes et al. (2003):

> *"The major difficulty with determining a system's assimilative capacity is four-fold as follows:*
>
> - *Each embayment has its own capacity based upon its depth, flushing rate, surface vs groundwater inflows, and sub-ecosystems (eelgrass, salt marshes, etc.)*
> - *Coastal embayments within the temperature zone have a high degree of temporal and spatial variation, so that a large amount of data collection is required*
> - *Relatively small increases in water column nitrogen can result in significant ecological changes*
> - *Evaluations are presently through inter-ecosystem comparisons."*

In summary, the LRA line of evidence in the memorandum provides informative TN concentrations and ranges, but they should not be directly translated to the LIS without consideration of site-specific conditions in the individual embayments.

The literature values for TN concentrations in Table F-1 are all based on Massachusetts estuaries. There is evidence that they are protective for these estuaries, but it cannot be assumed that they are also equally protective for the LIS embayments.

One way to assess the protectiveness of these TN values for LIS embayments would be to compare them with existing TN values in LIS embayments for which eelgrass distribution data are available. Aerial surveys of eelgrass distributions were conducted in 2002, 2006, 2009, and 2012 (Vaudrey et al., 2013). Figure 23 in the Vaudrey report contains the locations of 21 subbasins for which these surveys were conducted. At least five of these areas overlap with the embayments in the Subtask F/G memorandum; however, none of these data were used in the memorandum.

The justification/validity of the rationale for inclusion/exclusion of values from certain geographic areas is arguable. For the purpose of a comprehensive LRA, it would have been appropriate to include

Great Bay and Chesapeake Bay; however, values from these other systems still could not have been directly translated to LIS embayments without consideration of site-specific conditions.

Use of the MassBays report for the LRA is justified. The rationale for this decision was apparent in the memorandum.

The justification for exclusion of Chesapeake Bay literature is arguable. Again, for the purpose of a comprehensive LRA, it would have been appropriate to include Chesapeake Bay; however, values from Chesapeake Bay still could not have been directly translated to LIS embayments without consideration of site-specific conditions. The rationale for this decision was apparent in the memorandum.

***Dr. Brush's Response***

The LRA is consistent with common practice and I agree with the approach and find the outcomes reasonable. As above, it is difficult for me to assess if the values are protective since I do not work in the region and do not have a good sense of typical TN concentrations across systems, but the strength of the approach is that the values are based on available data and best practice, so I am inclined to accept them as reasonable. My only suggestion is that given the abundance of data in LIS, it would be worth reviewing the available data on TN concentrations and presence/absence of eelgrass currently or historically in LIS, for comparison to the Massachusetts values.

While the restriction of the literature analysis to Massachusetts estuaries does seem a little limited, I do agree that systems too far outside the LIS region should be excluded. I thought the justifications for excluding Great Bay and Chesapeake Bay were adequate and apparent in the memo. While Great Bay may share some similarities with LIS, it is a hydrodynamically very different system, and the size, southerly location, and high turbidity of Chesapeake Bay make it in my view incomparable to LIS. I therefore agree these systems should not be included. The best comparisons will be to systems with similar latitude, underlying watershed geology and impacts (e.g., septic), and geomorphology, and I believe the LIS embayments are very similar to the small Massachusetts embayments. I therefore think that use of the Massachusetts data is justified and appropriate, and this is adequately justified in the memo. It would be nice if values could be included from Rhode Island and New Jersey, but they may not be available.

***Dr. Janicki's Response***

The overall question is whether the use of data from other estuarine systems to establish TN endpoints is valid. The literature values for the geographic areas evaluated are reflective of the conditions within each geographic area. The rationale for inclusion or exclusion of certain geographic areas incorporates unnecessary bias. Given the uniqueness of each of the estuarine systems considered, use of the LRA is not recommended.

***Dr. Justic's Response***

The literature review for the line of evidence endpoints is rigorous and comprehensive. Justification for inclusion/exclusion of certain geographical areas appears sound. However, as discussed in my response to Question 7, the approach is entirely based on water column metrics (e.g., TN, chlorophyll *a*), which could be challenging given the shallow depths of LIS embayments. Using additional sediment-based metrics (e.g., sediment organics) could strengthen the analysis.

9. In your opinion, is it scientifically valid to eliminate TN values from the LRA Line of Evidence Method that are in excess of values known to cause severe degradation and to cap recommended TN endpoint values at levels known to be protective? In your opinion, is the chosen cut-off value of 0.8 mg/L TN and above an appropriate cap value for this purpose? Note: using a degradation cut-off threshold of 0.8 mg/L TN and above resulted in a maximum literature value of 0.6 mg/L TN (i.e., the next highest value below 0.8 mg/L TN).

***Dr. Bierman's Response***

There is insufficient evidence in the memorandum to form a scientifically defensible opinion about eliminating or capping TN target concentrations. Furthermore, the limited evidence presented in the LRA was from systems other than LIS and is less relevant than comprehensive site-specific data from LIS embayments themselves.

***Dr. Brush's Response***

While one should be cautious about removing data from any analysis, I thought the approach used in the LRA, and exclusion of selected values, was appropriate and well justified. I believe the chosen cut-off and resulting threshold values are appropriate. While one always wishes for more data, and better resolved data, we can only use the information that is available.

***Dr. Janicki's Response***

Consideration of the exclusion of extreme values should be based on the relative frequency of these values. Systems can be resilient to relatively infrequent extreme values. As such, it appears that choosing a cut-off value of 0.8 mg/L is valid.

***Dr. Justic's Response***

The use of 0.6 mg/L TN as the maximum endpoint value for open water segments is well justified by the LIS Literature Review Memo and Subtask F/G Memorandum (Table F-1, Page F-3).

10. Comment on the Stressor-Response Modeling (SRM) Line of Evidence Method. Comment specifically on the method used to construct the hierarchical models, their execution, and outputs.

***Dr. Bierman's Response***

My response to Question 10 is included in my responses to Questions 10a – 10f.

***Dr. Brush's Response***

Overall, I strongly support the SRM approach and feel the various findings were justified by the analyses. However, I have some important methodological concerns and points of clarification that I believe should be addressed before accepting the derived endpoints as final. These are detailed below. Given the issues raised in Questions 10 and 12 below, and the scatter of the regression plots in Subtask G, I recommend re-evaluation of this method and its results, and exploration of some additional analyses, despite the validity and rigor of the approach.

***Dr. Justic's Response***

The hierarchical modeling approach is well justified. However, the assumed relationship among key variables (Figure F-4) is rather simplistic and does not take into account sediment organics (see response to Question 7) or the fact that water column TN also includes nitrogen stored in algal cells whose biomass is expressed as chlorophyll *a*. Further, the stressor-response relationship for bottom DO as a function of chlorophyll *a* assumed strongly stratified water column and is generally not applicable to shallow LIS embayments.

The hierarchical regression model of $K_d$ as a function of chlorophyll *a* (Figure F-6) appears to underestimate values above 1.5 $m^{-1}$. Also, the data show (Figure F-7) that high $K_d$ values (> 1 $m^{-1}$) are often associated with very low chlorophyll *a* values (0.2 – 5 $\mu g\ l^{-1}$), suggesting that other light-attenuating substances could be important. Using additional chromophoric dissolved organic matter (CDOM) and TSS data (if available) could be helpful in better informing the model.

The hierarchical regression model of chlorophyll *a* as a function of embayment TN (Figure F-9) underestimates chlorophyll *a* values above 40 μg/L. Further, it is important to note that the available embayment field data consistently point to a very week relationship between TN and chlorophyll *a* (e.g., Figures G-2, G-4, G-10, G18, G-20, G-22, G-24, G-26). Finally, the modeled TN endpoint values are consistently larger compared to the literature review endpoints and distribution based endpoints. The above issues merit further investigation.

While the recommendations below may be beyond the scope of this review, I see two potential ways how the issues raised above could be addressed:

1) Additional stressor variables (e.g., sediment organics, CDOM, TSS) could be included in a hierarchical model to see if the model predictions could be improved; and

2) The coupled hydrodynamic-biogeochemical model could be implemented to a subset of LIS embayments to examine if the numerical model results support or refute the assumptions/results of the hierarchical regression model. In the absence of further regression/modeling analysis, my recommendation would be to assume that a chlorophyll *a* endpoint could not be derived based on water column TN and use only literature analysis and distribution-based approaches, as it was done for the LIS open waters.

a. *Regarding the SRM Line of Evidence Method*. Are the selected target light attenuation values reasonable and consistent with accepted ecological science for the Long Island Sound and Southern New England regions? Do tannin-colored waters (e.g., Pawcatuck River) impact the light extinction coefficients?

***Dr. Bierman's Response***

The selected target light attenuation values, as described on Page F-8 of the memorandum, appear reasonable and consistent with accepted ecological science for LIS and southern New England.

Tannin-colored waters do impact light extinction coefficients because colored/dissolved organic matter, along with total suspended solids, generally make substantial contributions to total underwater light attenuation.

A related topic is use of these target light attenuation values to estimate maximum and average colonization depths (Tables F-2 and F-3). As described on Pages F-8 and F-9, these depths were derived using the seagrass habitat suitability map coverages and embayment bathymetry from Vaudrey et al. (2013), along with a habitat suitability target of 50. The derivation of these depths is convoluted and difficult to follow. In addition, it is impossible to visualize the locations and sizes of the potential habitat areas that are being described. It would be more informative and clear if this section of the memorandum was linked more closely to the corresponding material in Vaudrey et al. (2013), especially Figure 22 which depicts an LIS-wide map of habitat suitability scores for the Eelgrass Habitat Suitability Index (EHSI) Model. An important point that would be visualized is that only small, scattered embayment areas are potentially suitable habitat for eelgrass.

***Dr. Brush's Response***

I agree with the selection of minimum light requirements based on the Latimer et al. (2014) work, particularly the approach of selecting a mean and range, and note that these values are in line with those developed in the Chesapeake Bay (Dennison et al., 1993; Kemp et al., 2004). I agree that the Ochieng et al. (2010) seedling results should not be used in setting the minimum requirements, as the higher values reported in that study appear to be based not on minimal requirements for survival but on more stringent requirements for long-term growth (so comparing the values would be apples to oranges). It was also unclear if the results really differed from an average requirement of 22%, as the report only says that seedlings did better between 11% and 34% $I_0$. If USEPA wishes to further evaluate seedling responses, there are other papers in the literature, such as Bintz et al. (2001) from work in nearby Rhode Island.

I found the text and terminology on p. F-9 somewhat confusing; some clarification would be helpful. For example, the terms used made it unclear to me if the depths in Tables F-2 and F-3 are those with existing eelgrass, with habitat scores ≥ 50, or of mean depth throughout each embayment. Consistent terminology for these depths should be used throughout (e.g., Table F-3 appears to show average colonization depth, but the last line on p. F-9 refers to them as average embayment depth). Another issue with terminology is that the definition of mean lower low water in Tables F-2 and F-3 is incorrect. From the National Oceanic and Atmospheric Administration (NOAA) website, mean lower low water (MLLW) is "the average of the lower low water height of each tidal day observed over the National Tidal Datum Epoch" (tidesandcurrents.noaa.gov/datum_options.html). NOAA averages the lowest water level each day of a 19-year tidal epoch; the value is not related to spring tides.

I found the cutoff of habitat scores ≥ 50 to be arbitrary and not justified in the text (p. F-9). Additionally, since the Vaudrey et al. (2013) habitat scores used here already included light as the primary variable.

I agree with the overall method to convert secchi depths to $K_d$, but I suggest reconsidering the conversion factor (1.45). As the memo notes, this value varies substantially. I would not feel comfortable using a value from Chesapeake Bay which has a much more turbid, sediment-laden water column than LIS. It would be preferable to use local LIS data to develop a site-specific conversion, or to look to similar, nearby (e.g. RI, MA) estuaries for a conversion factor. (Note: There appears to be an error in text on p. F- 10 which states, "... clear and turbid seawater, ranged from 1.44 to 1.90." However the next line says that the Chesapeake value of 1.45 is "consistent with turbid seawater.")

My main concern, however, is that following all of this background work on establishing acceptable values of % $i_0$ and $K_d$, the quantile regression section on p. F-15 introduces final $K_d$ target values of 0.5 and 0.7. I found this very confusing as the report had previously developed target $K_d$ values for each embayment across a range of % $i_0$ requirements which accounted for depth (Tables F-2 and F-3). These seem to be distilled here to two, sound-wide values. I was not able to follow why this change was made, and why these two values were not used all along.

To answer the last part of this question, tannins and more generally CDOM are well known to greatly impact $K_d$ in estuaries with substantial concentrations, and I would expect this to be an issue in the Pawcatuck River. Salinity has been used as a proxy for CDOM in multiple linear regressions of $K_d$, and its inclusion in the hierarchical model as a covariate should account for this, especially given the high number of observations used from the Pawcatuck. That said, the final model used appears to be the quantile approach, and I am not able to evaluate those results without more information, particularly if salinity was included.

***Dr. Janicki's Response***

The $K_d$ targets seem to be well justified with supporting information documented in the literature that is pertinent to the area of study. Tannin- colored waters definitely impact light attenuation. This is further described in my response to Question 10b below.

***Dr. Justic's Response***

The selected target light attenuation values appear reasonable. However, as stated in my response to Question 10a, CDOM is an important component of vertical light attenuation in estuarine and coastal systems (e.g., Abdelrhman, 2017) and needs to be taken into account.

b. *Regarding the SRM Line of Evidence Method*. Comment on the quantile regression model used for chlorophyll *a* versus the light attenuation coefficient, $K_d$. Is the use of this technique sound and is it an adequate model for the goal of setting chlorophyll *a* endpoints? Are the selected chlorophyll *a* endpoints scientifically valid for the LIS?

***Dr. Bierman's Response***

See my response to Question 1 for related discussion on this topic.

Chlorophyll *a* is a primary response variable, not an "endpoint." The purpose of the quantile regression model in the memorandum is not to set chlorophyll *a* "endpoints" but to link values of $K_d$ (dependent variable) and chlorophyll *a* (independent variable) as part of the basic conceptual model depicted in Figures F-4 and F-5. A fundamental flaw in this conceptual model is that $K_d$ is assumed to depend only on chlorophyll *a* concentrations. This is not correct and is in contravention to observed data in LIS as well as in other estuaries and bays from Chesapeake Bay to Maine.

The water-column light attenuation coefficient ($K_d$) in estuarine systems is dominated by contributions from chlorophyll *a*, total suspended solids and CDOM (Batiuk et al., 2000; Cerco et al., 2010; Vaudrey et al., 2013). Using observed data for the Great Bay Estuary, Morrison et al. (2008) developed a multiple regression model and showed that the following are the component contributions to $K_d$: water (32%), turbidity (29%), CDOM (27%) and chlorophyll *a* (12%). Benson et al. (2013), cited in Table F-1 of the memorandum, asserted that the influence of nitrogen concentration on $K_d$ followed these linkages: N => chlorophyll-a => POC => $K_d$.

On Page F-15 of the memorandum it is acknowledged that suspended sediment and dissolved organic matter could have contributed to light attenuation within the LIS embayments, but it was then stated that these parameters were not included in the model for Kd because data were not available. This is not correct. It is documented in Subtask D, Summary of Existing Water Quality Data, that LIS data exist for total organic carbon (TOC), dissolved organic carbon (DOC), particulate carbon (PC), and total suspended solids (TSS). These are all of the data required to develop a site-specific multiple regression model for Kd in LIS, similar to the model developed by Morrison et al. (2008).

With respect to the quantile regression model, on Page F-15 the memorandum states that this approach is advocated for use in ecological models where a response is affected by multiple factors. It goes on to point out that the relationship between $K_d$ and chlorophyll *a* for the LIS embayments is less influenced by dissolved organic matter and suspended sediment interference at lower quantiles (Figure F-7). Following this logic, the memorandum uses the $10^{th}$ quantile regression model to associate chlorophyll *a* values with $K_d$ "endpoints" of 0.5 and 0.7 (Table F-8). In turn, it then uses the chlorophyll *a* "endpoint" of 10 ug/L (corresponding to $K_d$ = 0.7) for 12 of the 15 individual embayments in Subtask G. A literature value of chlorophyll *a* = 5.5 ug/L was used in the Nissequogue River and Mt. Sinai Harbor embayments, and no chlorophyll *a* "endpoint" was used in the Eastern and Western Narrows (combined). These chlorophyll *a* "endpoints" were then used in the embayment-specific models for chlorophyll *a* vs TN to develop the TN concentration "endpoints."

The approach in the memorandum for relating $K_d$ and chlorophyll *a* is conceptually flawed and the consequences propagate through derivation of the TN "endpoints" for all 12 of the above embayments for which it was used. It is correct that quantile regression can be appropriate for ecological stressor-response models for the purpose of deriving a numeric criterion for the independent variable. However, the objective of the $K_d$ vs chlorophyll *a* analysis in the memorandum was to accurately estimate $K_d$ (the dependent variable) for specified values of chlorophyll *a* (the independent variable), not to develop numeric nutrient criteria for chlorophyll *a*.

The consequences of this conceptual flaw can be seen by turning the logic around and visually inspecting the observed data for $K_d$ vs chlorophyll *a* in the plot on Page F-16. The derivation of the TN "endpoints" for the 12 above embayments assumes that a $K_d$ value of 0.7 corresponds to a chlorophyll *a* concentration of 10 ug/L (Table F-8). However, it can be seen from the $10^{th}$ quantile regression plot on Page F-16 that most of the observed $K_d$ values corresponding to a chlorophyll *a* concentration of 10 ug/L are greater than 0.7. Consequently, actual light attenuation in the water column is much greater than that predicted by the $10^{th}$ quantile regression fit. The underlying reason is that this model considers only the chlorophyll *a* contribution to $K_d$ and ignores the substantial contributions of suspended solids and dissolved organic matter.

In summary, none of the chlorophyll *a* "endpoints" for the above 12 embayments that were selected using the SRM are scientifically valid, nor are the corresponding TN "endpoints" that relied upon these chlorophyll *a* "endpoints."

***Dr. Brush's Response***

I am not familiar with the quantile regression approach, so it is difficult for me to evaluate this section. The text does not provide a general overview of the approach as it does for hierarchical models, which would be helpful. I found various parts of this section (p. F-15) confusing. My specific comments are as follows:

- I did not understand the first sentence (line 9). Why were the individual embayment plots not useful? This appears to set up the rationale for using quantile regression instead, but it was not clear to me why based on the preceding paragraph.

- The 5%, 10%, and 20% quantiles were examined, but without seeing all the results it is not possible to fully evaluate use of the 10% quantiles.

- The issue of terminology regarding colonization vs. average embayment depth occurs again in the second paragraph. The legend for Table F-4 was also confusing, i.e. "... embayment model, by embayment."

- See my response to Question 10a about the apparent change in target $K_d$ values in this section. Given this issue, I am unable to assess the validity of the selected chlorophyll *a* endpoints.

***Dr. Janicki's Response***

The use of quantile regression is a valid method of describing relationships that may occur at some other portion of the response distribution other than the mean as described in the text. However, the choice of the 10$^{th}$ percentile value is curious. Given that as $K_d$ increases, light availability decreases, modeling the 10$^{th}$ percentile suggests that the identified chlorophyll *a* targets would be best expressed as maximum acceptable values since at these values, 90 percent of the $K_d$ distribution is expected to be above the value predicted by the quantile model. This means that other covariates (e.g. suspended solids as described in the text) that were not modeled also contribute to light attenuation. An alternative approach if available to would be to develop estimates of the relative contribution of color, chlorophyll a and turbidity for an area where those data were available and use the relative contributions to estimate what the total $K_d$ would be on average for a given level of chlorophyll a but this may have been outside the scope of the work effort.

***Dr. Justic's Response***

The 10$^{th}$ quantile regression for $K_d$ as a function of chlorophyll *a* is well justified and the resulting chlorophyll *a* endpoint values seem scientifically valid based on the LIS Literature Review Memo.

c. *Regarding the SRM Line of Evidence Method*. Is the use of a hierarchical model appropriate for this kind of analysis? Is adequate justification provided in the memorandum for the use of this methodology? Are the statistical methods used in the hierarchical models clearly explained and technically valid? Is the goodness of fit of each modeled relationship adequately presented and interpreted? Should acceptable significance values or quality standards be made explicit? Are the nitrogen concentration endpoints developed in this model ecologically reasonable? Would they be considered protective of eelgrass in the region? Is it appropriate to show the modeled TN concentrations for two chlorophyll *a* levels (when applicable) in a single embayment?

***Dr. Bierman's Response***

See my response to Question 1 for related discussion on this topic.

Conceptually, a hierarchical model, as well as other statistical models in USEPA (2010), could be appropriate for the kinds of analyses in the memorandum. However, the methods used to construct and execute the models in the memorandum, and the outputs of these models, have numerous flaws. These are discussed above for the $K_d$ vs chlorophyll *a* relationship on Page F-14 and below for all of the other hierarchical models.

<u>DO vs Chlorophyll *a* for Embayments</u>

For the final DO vs chlorophyll *a* relationship on Page F-17, it is not clear what samples were used (e.g., grab samples, bottom water samples, profile samples). The final model for DO explained more than half (pseudo $r^2 = 0.61$) of the variability in observed DO; however, it was not possible to fully evaluate the model itself because no plots were shown for the final model with observed data for DO and chlorophyll *a*. The final model predicted increasing DO with increasing chlorophyll *a*, and relatively high DO even at extremely low chlorophyll *a*, both of which are counterintuitive. The memorandum concluded that a chlorophyll *a* "endpoint" was not able to be derived for the DO vs chlorophyll *a* relationship for the embayments.

It is not surprising that a meaningful statistical relationship could not be developed for DO as a function of chlorophyll *a*. Dissolved oxygen in aquatic systems is controlled by a complex set of physical, chemical, and biological processes that are not amenable to characterization by statistical stressor-response relationships. In fact, even the USEPA Technical Guidance Document for Stressor-Response Relationships (USEPA, 2010) does not contain a single example for DO as a dependent response variable in any of its statistical models.

<u>Chlorophyll *a* vs TN for Embayments</u>

The final model for chlorophyll *a* vs TN on Page F-18 explained less than half (pseudo $r^2 = 0.47$) of the variability in the observed chlorophyll *a* data. Again, it was not possible to fully evaluate the model itself because no plots were shown for the final model with observed data for chlorophyll *a* and TN; these data were shown for only the embayment specific plots in Subtask G.

The final chlorophyll *a* vs TN model was applied to 14 embayments (including the Connecticut River). Four of these embayments had no data. Visually, there was no apparent relationship (or only a weak relationship) between chlorophyll *a* and TN in most of the embayments with data. Many of these data were outside the 90% confidence limits of the model.

The final model for chlorophyll *a* vs TN was a key component in the selection of TN "endpoints" in Subtask G because embayment specific plots were constructed and solved for the TN concentrations corresponding to various chlorophyll a "endpoints." Using $K_d$ = 0.70 (Vaudrey, 2008) and chlorophyll *a* = 10 ($10^{th}$ quantile model) the chlorophyll *a* vs TN model predicted that eelgrass would not be protected in any of the 14 embayments, based on the range of TN values from the LRA. Using chlorophyll *a* = 5.5 ug/L (Vaudrey, 2008), the chlorophyll *a* vs TN model predicted that eelgrass would be protected in only two of the 14 embayments, based on the range of TN values from the LRA, and it predicted a TN value less than background in one embayment.

Not only are these TN "endpoints" not protective in any of the 14 embayments, it is not clear that any of them represent the full areal extents of the embayments shown in the maps in Subtask G. As I noted in my response to Question 5, the SRM relies upon the EHSI model and embayment bathymetry data developed by Vaudrey et al. (2013). Specifically, the estimated maximum colonization depths of suitable eelgrass habitat in each embayment were developed using an EHSI habitat suitability target of greater than 50. Consequently, any TN "endpoints" developed using the SRM represent only embayment areas with habitat suitability scores greater than this value. According to Figure 22 in the Vaudrey report, only very small nearshore areas in the LIS have habitat suitability scores greater than 50. To clarify this point, each of the embayment maps in Subtask G should demarcate the areas that have EHSI habitat suitability scores greater than 50 because the TN "endpoints" developed using the SRM apply only to these areas.

In summary, in combination with the conceptual flaws and questionable assumptions discussed above, the TN concentration "endpoints" developed using the chlorophyll *a* vs TN models are not scientifically valid.

On Page G-3 of the memorandum it is stated that:

> *"The embayment stressor-response models often produced TN values that were too low (below most regional background levels and thus not realistic to achieve) or too high (not protective of eelgrass). Instances where this occurred are noted in the embayment endpoint table. USEPA plans to revisit the assumptions made during the stressor-response analysis in the next phase of this work."*

It appears that even USEPA has called into question the technical validity of the statistical methods used in the hierarchical models.

DO vs Chlorophyll *a* for Open Waters

The final model for DO vs chlorophyll *a* on Page F-21 explained more than half (pseudo $r^2 = 0.70$) of the variability in observed DO. Again, it was not possible to fully evaluate the model itself because no plots were shown for the final model with observed data for DO and chlorophyll *a*. Again, as with the above DO vs chlorophyll *a* model for embayments, the final model predicted increasing DO with increasing chlorophyll *a*, and relatively high DO even at extremely low chlorophyll *a*, both of which are counterintuitive.

The memorandum stated that lack of paired bottom DO samples with chlorophyll *a* data was a limitation. Specifically, there was plenty of bottom DO data, but few chlorophyll *a* data. For the open waters in LIS this should not be surprising because significant concentrations of chlorophyll *a* usually occur in surface waters and are not co-located with the low DO, hypoxic conditions that occur in bottom waters.

Again (see above) it is not surprising that a meaningful statistical relationship could not be developed for DO as a function of chlorophyll *a*. DO in aquatic systems is controlled by a complex set of physical, chemical and biological processes that are not amenable to characterization by statistical stressor-response relationships. The memorandum concluded that a chlorophyll *a* “endpoint” was not able to be derived for the DO vs chlorophyll *a* relationship.

Chlorophyll *a* vs TN for Open Waters

The final model for chlorophyll *a* vs TN on Page F-22 explained less than half (pseudo $r^2 = 0.32$) of the variability in the observed chlorophyll *a* data. Again, it was not possible to fully evaluate the model itself because no plots were shown for the final model with observed data for chlorophyll *a* and TN. The final model predicted that chlorophyll *a* levels decrease as TN levels increase, a result that does not make sense. The memorandum concluded that a TN “endpoint” was not able to be derived for the chlorophyll *a* vs TN relationship for open waters.

***Dr. Brush’s Response***

The overview of hierarchical and multiple regression modeling was excellent and very informative. While I do not use hierarchical modeling and only have the information from the memo to rely on, I think this was an excellent way to integrate data across all systems, and leverage the global model in relatively data-poor embayments. This was well justified in the memo. One minor question I had, given the focus on independence of samples in this analysis, was if the other key assumptions were tested, namely normality and homogeneity of variance?

To evaluate the regressions, the memo includes observed vs. predicted plots and pseudo $r^2$ values. However, p-values of the overall regression and the regression statistics for each fitted parameter (i.e., fitted values, uncertainty, and p-values), are not provided. These would be important for fully evaluating the regression output. It would be somewhat helpful if acceptable significance values were chosen, although mainly that is up the reader to interpret.

As above, I am unable to evaluate if the resulting TN endpoints are reasonable and protective of eelgrass based on my own knowledge, but I find the modeling appropriate and with the caveats above I have no reason not to accept the results. I think it is fine to show two modeled TN values based on different chlorophyll *a* targets for individual embayments (Subtask G). As noted above, however, I found the related part of the Subtask F memo confusing and by the time I got to Subtask G I could not remember where the two different chlorophyll *a* values came from, or why some embayments had one value while others had two. This should be clarified in the Subtask F memo, and on the first page of the Subtask G memo. The text regarding these two chlorophyll *a* values in the "TN Endpoints Discussion" sections was helpful.

***Dr. Janicki's Response***

The use of hierarchical models is appropriate for this type of analysis and the authors justify the analytical approach for application of their hierarchical models. However, there are details of the modeling effort that should be further explained and there are no citations given anywhere in the description for their hierarchical modeling approach. This lack of detail makes it difficult to know if the models were specified correctly. Additional information is needed on the following: estimation method; model selection method and criteria used to develop the final models; covariance structure for the random effects; fit statistics; fit statistics; tables of parameter estimates, and diagnostic plots. Each of these is further described below. These comments are not to say that the models were mis-specified, only that there was not enough information presented to fully understand the model specification.

<u>Estimation Method</u>

Was maximum likelihood (ML) or restricted maximum likelihood (REML) used as the estimation method, or do they switch back and forth between ML and REML? There are important differences between these estimation methods that affect both the parameter estimates and their statistical significance. Typically, one would a) develop a full model of the fixed effects, b) model the random effects using REML, c) generate statistical tests of significance for the fixed effects using ML, and then d) report the final estimates using REML (Zuur et al., 2009).

<u>Model Selection</u>

How was it decided which fixed and random effects to retain in the model? As described above, this is typically an iterative process and Akaike Information Criteria and the likelihood ration test are typically used to evaluate both the benefits of including fixed effects terms in the model and well as the inclusion of the random effects. Again, none of this is reported.

Covariance Structure

The type of covariance structures defined for each random effect is not described. It is assumed that the "Variance Component" structure was assigned by default to estimate the group variance component of the random effect; however, it is stated several times that "random effects for station ID were included to account for data dependency". No information was given on how this was incorporated into the model structure. The error term in the provided model equations ($e_{ij}$) is not a proper specification of the inclusion of a random effect component for the station ID term as described. As described, it seems that term would be included as a nested random effect {station ID(group)} and a specific covariance structure would be specified such as compound symmetry or autoregressive covariance structure. However, either way this would result in a highly parameterized model if there are a lot of stations. Without any details of the model output it is difficult to tell. In addition, the covariance parameter estimates for the random effects should be reported. One can calculate the intra-class correlation based on these estimates to assess descriptively if the within class correlation is high and the variance component makes a valuable contribution to the modeling effort.

Model Specification

For the generalized linear models, the link function could use more explanation in general as the authors switch between distributions (gamma with natural log link for light attenuation versus chlorophyll *a*; Gaussian with identity link for DO versus Chlorophyll; gamma with natural log link for chlorophyll *a* versus Nitrogen). In particular, when modeling the open waters of LIS they state (page F-21) that a gamma with an identity link was used to model DO versus chlorophyll *a* and chlorophyll *a* versus Nitrogen. I believe these latter descriptions may be a typo as the identity link is not commonly used with the gamma distribution. Technically, the link function should be defined within the model equations provided; they are not.

There was no supporting evidence given for the choice of including the random slopes model and there should be some theoretical plausibility for inclusion of this model. Is there a plausible biological explanation for allowing slopes to vary by embayment? Perhaps this is related to residence times but it should be stated to provide support for the choice. For the generalized linear models, the random slopes term assumes that the variance component is not only a function of within group covariance but also depends on the level of the independent term (e.g., nutrient concentrations). Again, this may be a perfectly valid assumption but should be stated.

Diagnostic Information

Along with Information Criteria, the final models output should include a parameter estimates table and diagnostic plots including not only the fitted versus observed plots provided but also quantile-quantile plots and plots of the deviance residuals at minimum. These would support the choice of link function used for the final models.

<u>Model Predictions</u>

It should be stated somewhere whether the model predictions are “conditional” (i.e., based on inclusion of the random effects) or “marginal” (“population averaged” with random effects set to zero). It is assumed based on the description of the shrinkage estimates that the estimates are conditional but it should be specified.

***Dr. Justic’s Response***

Please see my response to Question 10 for related discussion on this topic. While a hierarchical modeling approach is well justified and suitable for the kind of analysis performed in this study, there are several important issues that need to be addressed. The available embayment field data consistently point to a very week relationship between TN and chlorophyll *a* and the TN endpoint values obtained using this method are consistently larger compared to the literature review endpoints and distribution based endpoints. These issues need to be addressed before informed recommendations can be made.

d. *Regarding the SRM Line of Evidence Method*. Is it reasonable to include the lower Connecticut River with the 23 priority embayments for modeling purposes? Is this inclusion ecologically and hydrologically sound? Is it reasonable to model a TN endpoint for the Connecticut River based on a hierarchical model built on water quality observations from the 23 priority embayments?

***Dr. Bierman’s Response***

See my responses to Questions 3 and 4 for related discussion on this topic. Decisions to group/not group different water bodies should be informed by comparisons of their site-specific data for the parameters in my response to Question 3d and by the habitat suitability maps in Vaudrey et al. (2013) cited in my response to Question 4.

***Dr. Brush’s Response***

As noted in my responses to Review Topic 1 (Subtask E Memorandum), I am skeptical of the approach and results for the area of influence estimation for the Connecticut River. Putting that aside, my sense is that the mouth of the Connecticut River is quite different from the other embayments, and I note that the mouths of the Housatonic and Thames Rivers were not included despite also having areas of influence estimated in Subtask E. I think the differences between these types of systems and the other embayments suggests that the Connecticut River should not be included in the present analysis.

***Dr. Janicki’s Response***

A discussion of how the hydrologic characteristics of the Connecticut River differ from those in the 23 priority embayments could provide justification for the analytical approach used.

***Dr. Justic's Response***

The summary information for the lower Connecticut River have not been presented in the Subtask A report, Subtask E report, or elsewhere in the documentation provided, and so it was impossible to assess whether combining this system with the 23 priority embayments was ecologically and/or hydrologically sound. I am not familiar with the lower Connecticut River, but it appears that it could morphologically be classified as an embayment. Apparently, the lack of paired data did not allow for this system to be modeled separately.

e. *Regarding the SRM Line of Evidence Method*. The outputs of the hierarchical model were often above 0.5 mg/L or below 0.2 mg/L. Is it regionally, ecologically, and scientifically credible to assume TN values above 0.49 mg/L are not protective of eelgrass and concentrations below 0.2 mg/L are below the background concentration for the region? Is it appropriate to give the unaltered output of the model a caveat explaining this purportedly realistic/protective range? Is it regionally, ecologically, and scientifically valid to assume TN values above 0.49 mg/L are not protective of eelgrass and concentrations below 0.2 mg/L are below the background concentration for the region?

***Dr. Bierman's Response***

See my responses to Questions 3, 10a, 10b, and 10c for related discussion on this topic. The applications of the SRM hierarchical models in the memorandum contain conceptual flaws and questionable assumptions, and their outputs do not provide scientifically valid support for any decisions on TN values to protect eelgrass in LIS.

***Dr. Brush's Response***

This is difficult for me to assess as I do not have a broad sense of typical TN values across a gradient of impact and 'ecosystem health', or specifically in LIS. The value of 0.5 mg/L (or 0.49) was based on a reasonable literature review, although admittedly limited to Massachusetts estuaries. But since the LRA approach was valid, I have no reason to doubt this estimate of an upper TN limit. The fact that so many estimated TN values from the SRM approach fell above this limit is likely due to the complexity of these systems and use of a single stressor and a single response metric, i.e. TN and eelgrass (see my response to Question 12 below). This may also be a function of the somewhat limited choice of $K_d$ values that were used in the final analysis, as opposed to the more detailed, embayment-specific values first developed in Subtask F. Since % $i_0$ is highly sensitive to depth, incorporating this depth-sensitivity into these calculations may also be useful.

I only saw a single value that fell below the proposed lower, background limit of 0.2 mg/L in Subtask G. Again I have no way of knowing if this is a reasonable background value, but the memo references Howes et al. (2006) and NHDES (2009) so this could be further explored. The DbA approach resulted in a 1$^{st}$ quartile TN concentration across the embayments of 0.27 mg/L, which is not far above 0.2, which makes me think the latter may in fact be a reasonable background estimate. But I would want to analyze TN concentrations against loading rates and flushing time to more fully explore this question.

I think the caveats currently in the Subtask G memo (i.e., Table footnotes and discussion sections) are appropriate.

***Dr. Janicki's Response***

Justification of the assumption that TN concentrations > 0.49 mg/L are not supportive of eelgrasses is lacking. Are there any ambient data from systems where eelgrasses are relatively healthy or from some historical period when eelgrasses were also relatively healthy? Presentation of the unaltered model output is warranted as assessment of the validity of the model can only be achieved with that output available for review. Post-hoc constraints of the model results can then be discussed and justified.

***Dr. Justic's Response***

Please see my response to Question 10 for related discussion on this topic. The hierarchical modeling approach has several important issues that need to be addressed. In particular, the TN endpoint values obtained using this method are consistently larger compared to the literature review endpoints and distribution based endpoints, and this problem merits further investigation before informed recommendations can be made.

f. *Regarding the SRM Line of Evidence Method*. Is the use of chlorophyll *a* corrected rather than chlorophyll *a* measurement adequately explained and justified? Are the methods used to collect chlorophyll *a* data appropriately assessed and interpreted as similarly indicative of phytoplankton biomass (e.g., considering whether measurements represent similar corrections for dead biomass that does not contribute to life processes for production or respiration) when using chlorophyll *a* for stressor-response relationships? How should dead biomass be treated?

***Dr. Bierman's Response***

The only explanation/justification given in the memorandum (Page F-13) was that there were more data available for chlorophyll *a* corrected than chlorophyll *a*. My opinion is that it is more appropriate to use chlorophyll *a* corrected because it is a better indicator of live phytoplankton biomass. Within the context of the three empirical modeling approaches in the memorandum, my opinion is that no explicit treatment of dead phytoplankton biomass is necessary.

***Dr. Brush's Response***

The use of corrected chlorophyll *a* (i.e., acidified to correct for phaeophytin) is entirely appropriate and consistent with common practice. Use of uncorrected chlorophyll *a* should be avoided; while dead biomass may contribute some to light attenuation, most of the impact will come from active chlorophyll *a*. A stronger case could be made for using uncorrected (i.e., including dead) chlorophyll *a* in regressions with DO, as dead biomass contributes to respiration. Still, common practice in the field is to use corrected values so in my view this is the proper approach.

The issue of corrected vs. uncorrected chlorophyll *a* is not really addressed in the memo; rather the memo notes that corrected values were used. I think this could be clarified slightly, particularly to explain what is meant by corrected and uncorrected chlorophyll *a*, but I do not feel that additional information is necessary.

***Dr. Janicki's Response***

The choice of whether corrected or uncorrected chlorophyll *a* data often depends upon the relative abundance of the two data types. While dead biomass may not be reflective of production, the effects of dead biomass on DO due to decomposition can be important. It seems that inspection of the goodness of fit associated with both variables would be informative.

***Dr. Justic's Response***

In the materials provided I could not find a satisfactory explanation for how the correction was made and why the chlorophyll *a* corrected value was chosen. It is unclear if the "correction" refers to pheophytin-corrected chlorophyll *a* concentrations (acidification method) or to correction for pigment loss in frozen samples (Graff and Rynearson, 2011). This needs to be clarified.

If the method itself is the issue, in spite of potential problems (e.g., Stich and Brinker, 2005), pheophytin-corrected chlorophyll *a* has largely remained a method of choice in oceanography.

11. Comment on the approach used for the Distribution-based Approach (DbA) Line of Evidence Method. Is this approach scientifically valid? Is the outcome reasonable? Is the rationale behind this approach clear? Are the TN values reflective of protective values for the Long Island Sound's geographic area?

***Dr. Bierman's Response***

In following guidance in USEPA (2001), the memorandum used 25$^{th}$ percentile values of all samples for LIS embayment waters and open waters to develop distribution-based TN endpoints. It rejected use of the 75$^{th}$ percentile values as indefensible because existing nutrient impacts on LIS made it difficult if not impossible to accurately identify or represent near-pristine conditions. As supporting evidence that the 25$^{th}$ percentile TN concentration of 0.27 mg/L (Table F-10) corresponded to desired conditions in LIS embayments, it cited the median TN concentrations in Niantic Bay (0.21 mg/L) and Mystic River (0.53 mg/L), both of which had exhibited eelgrass increases from 2002 to 2012. The memorandum stated that the concentration in Niantic Bay (0.21 mg/L) was consistent with the 25$^{th}$ percentile concentration (0.27 mg/L), but did not explain the inconsistency between eelgrass increases in Mystic River at a concentration of 0.53 mg/L, which approximated the 75$^{th}$ percentile TN concentration of 0.56 mg/L (Table F-10). Also, there was no discussion of how the value of 0.27 mg/L for all embayments relates to value of 0.40 mg/L from the LRA method on Page F-3.

This approach is scientifically valid in that it followed the guidance in USEPA (2001); however, it was limited in that it used data from only LIS. It did not explicitly include eelgrass or data from other relevant systems in the New England and mid-Atlantic regions. This analysis could be strengthened by conducting comprehensive and systematic reviews of site-specific data for these other systems, and placing emphasis on spatial classification and segmentation of each system into zones with similar flushing times, bathymetry, and sediment physical-chemical characteristics as the LIS embayments.

The outcome is reasonable in that it followed USEPA (2001) guidance and used site-specific TN concentrations; however, the outcome is of limited value because the method did not use any data for the assessment endpoints (eelgrass, aquatic life) from either LIS or from other regional systems.

The rationale behind this approach is not presented in the memorandum itself, but it is in the USEPA (2001) technical guidance.

There is no evidence in the memorandum to support an opinion on whether the TN values are protective. One way to assess the protectiveness of these TN values for LIS embayments would be to compare them with existing TN values in LIS embayments for which eelgrass distribution data are available. Aerial surveys of eelgrass distributions were conducted in 2002, 2006, 2009 and 2012 (Vaudrey et al., 2013). Figure 23 in the Vaudrey report contains the locations of 21 subbasins for which these surveys were conducted. At least five of these areas overlap with the embayments in the Subtask F/G memorandum.

***Dr. Brush's Response***

I found the DbA methods entirely valid and the outcome reasonable. The approach, the rationale for using it, and the rationale for using the first quartile were all well justified. As previously stated, it is difficult for me to evaluate the final TN endpoint except that I find it supported by the data, and the correspondence to data from Niantic Bay provides a nice confirmation. A few minor notes are as follows:

- p. F-25 refers to depth criteria from the stressor-response analysis, and that the growing season was used for consistency with the other lines of evidence. I don't recall depth criteria or the growing season being discussed in the other sections of the report.

- Regarding growing season, I was surprised that the April-September period was used here given the focus on July-September in Subtask E.

- The caption for Table F-10 uses confusing terminology.

***Dr. Janicki's Response***

The use of what is essentially a reference system approach has been shown to be problematic when establishing numeric nutrient criteria in Florida estuaries. Granted, the spatial variability of the LIS embayments is less than seen in Florida. There are many examples of the use of a DbA approach previously by USEPA. The choice of the 25$^{th}$ percentile is clearly based on professional judgment. The validity of this choice in some ways is dependent upon how the endpoints will be used. If they are to be used in a compliance assessment of future conditions, then expressing the endpoints as a range might be considered. The allowable frequency of non-compliance might also consider the uncertainty in the choice of the percentile that represents the endpoint.

Was any consideration given to applying a reference period approach? If ambient water quality data from a historical period when eelgrasses were relatively abundant were available, they could be used to define a distribution from a period of more desirable conditions.

***Dr. Justic's Response***

The DbA approach is sound and scientifically defensible. The 25$^{th}$ percentile TN values (Table F-10) for embayment waters (0.27 mg/L) and open waters (0.24/l) compare favorably with median water column TN concentrations in embayments that have historically exhibited increases in seagrass coverage (LIS Literature Review Memo).

12. Many estuaries and embayments on the central and eastern regions of Long Island Sound currently have TN and chlorophyll *a* concentrations that are near the levels recommended (chlorophyll *a* of 3-10 mg $L^{-1}$ and TN of 0.3 to 0.5 mg $L^{-1}$) by the Literature Review Analysis (LRA), Stressor-Response Modeling (SRM), and Distribution-based Approach (DbA) approach used in the analysis (examples include G1 Pawcatuck River, CT and RI, G2 Stonington Harbor, CT, G5 Mystic Harbor, CT, G6 Niantic Bay, CT, G9 Northport Centerport Harbor, NY, G10 Port Jefferson Harbor, NY, G11 Nissequogue River, NY, G12 Stony Brook Harbor, NY and G13 Mt. Sinai Harbor, NY). Despite TN and chlorophyll *a* near the target threshold values, ecosystem function and aquatic life support are still impaired in many of these systems as evidenced by reduced DO, macroalgal blooms, harmful algae blooms (e.g., annual HAB shellfish closures in Northport Harbor system), reduced benthic infauna abundance and diversity, and declining eelgrass abundance. In light of these facts, are the recommended chlorophyll *a* and TN targets justified as being protective of aquatic life? Is it adequately documented that water column TN and chlorophyll *a* targets are protective of aquatic life in embayments dominated by macroalgae?

***Dr. Bierman's Response***

The purpose of the memorandum is to develop TN concentration targets, not chlorophyll *a* targets. If ecosystem function and aquatic life support are still impaired in many of the systems with TN concentrations near target threshold values, it calls into question two underlying assumptions: (1) that TN concentration is the sole causal factor; and, (2) that TN concentration targets can be developed without conducting data-intensive, site-specific investigations in each embayment. Neither of these assumptions is valid.

See my response to Question 3 for the minimum data requirements for establishing TN concentration targets for protecting LIS embayments. These data requirements include all of the confounding factors that should be assessed in addition to in-water TN concentrations. See my responses to Question 3 (sentinel station approach) and Question 8 (data-intensive, site-specific studies) regarding actual experience in the MEP for developing TN target concentrations.

There is no documentation in the memorandum pertaining to TN and chlorophyll *a* targets in embayments dominated by macroalgae. With the exception of two TN concentrations for SE Massachusetts Embayments in Table F-1, the memorandum is silent on macroalgae.

Vaudrey et al. (2013) address macroalgae in LIS with this statement on Page 14:

> *"The inclusion of a macroalgae term (coverage of detrimental green macroalgae) was investigated in the EHSI Sub-Model, as data were collected as part of this project. It was determined that even when the macroalgae is assigned 20% of the model score weighting, the inclusion does not have an appreciable effect on the model skill (Section 7.7.2, page 141). While the inclusion of macroalgae seems theoretically sound, it appears to be an over-parameterization of the model. For this reason, inclusion of macroalgae in the model is not recommended."*

***Dr. Brush's Response***

These issues do make one question the validity of the TN endpoints established here. I think there are multiple issues to consider. First, the LRA and SRM approaches developed here focused on eelgrass in developing endpoints (the additional focus of SRM on DO was unsuccessful). Eelgrass is one of many potential indicators of a healthy ecosystem, and it should not be expected that one metric reflects an integrated picture of 'ecosystem health'. Another issue is that the methods focus on a single predictor, TN. As noted above, a more holistic approach would include DIN, and loading rates in addition to concentrations. An even better analysis might normalize loading rates to flushing time to generate expected, steady-state concentrations in the absence of biological processing, and possibly to depth or a number of other system-level characteristics.

A third issue is that of macroalgae. The competition between phytoplankton (i.e., chlorophyll *a* here) and macroalgae in shallow systems has been a long-standing topic in coastal marine ecology. While there are numerous factors that determine which will dominate an ecosystem, the conceptual model of Valiela et al. (1997) developed largely in nearby Waquoit Bay, MA indicates that dominance is a function of both nitrogen loading and flushing rate, so that a given system at a given N load could be dominated by phytoplankton or macroalge (or eelgrass) depending on flushing rate. There are numerous other factors involved too. And as noted above with reference to Nixon et al. (2001), nutrient concentrations in shallow systems can be extremely low despite high loading rates due to active plant uptake and denitrification, so that a system dominated by macroalgae would have almost no available TN in the water and low chlorophyll *a*, but still show signs of impact via macroalgal accumulation.

Beyond macroalgae, HABs develop for a number of reasons, only one being nutrient inputs. DO levels may be subject to legacy accumulation of organic matter in sediments, such that there may be a lag between reduced nutrient concentrations and improved DO. Some shallow regions of estuaries may also go hypoxic naturally, at least over diel cycles. Similarly benthic fauna may exhibit lag times in recovery, which are further complicated by the random lottery of larval supply.

Given these issues, we know that eutrophication response in coastal systems is complicated and a function of more than just nutrient concentration. Cloern (2001) presented an excellent summary of this. Despite some caveats regarding methodology, I find the TN endpoints developed in the current effort to be rooted in valid, scientifically-defensible approaches. The quality and thoroughness of the present work used the available data to the maximum extent possible given available resources. Nevertheless, the observations above suggest that the TN endpoints established here may not be indicative of a 'healthy ecosystem'. So while the current effort provides important first-order estimates of TN endpoints, it appears that additional work is needed to refine them to account for conditions within LIS, and the varied responses across its embayments.

***Dr. Janicki's Response***

Given what seems to be less than desirable estuarine health, the similarity in the endpoints to current water quality conditions leads to questions about whether the proposed endpoints are protective. If the estuarine characteristics are sensitive to small differences between the current water quality and the proposed endpoints, then the assumption that the endpoints will be protective may be justifiable. Whether the proposed endpoints will be protective of aquatic life in embayments dominated by macroalgae remains in question.

***Dr. Justic's Response***

As discussed in my response to Question 7, water column TN and chlorophyll *a* values are unlikely to fully explain the extent of eutrophication in shallow/low residence time LIS embayments. The reason is that nutrients and carbon stored in sediments likely fuel macroalgal blooms and can exert considerable control on water column processes, including the dynamics of hypoxia and occurrence of harmful algal blooms. Further, small-scale variability in estuarine hydrodynamics, stratification, turbidity, and residence times can create favorable conditions for phytoplankton blooms/hypoxia development at specific locations within an embayment. This variability cannot be adequately captured if the approach is based solely on system-wide July-September average conditions. Employing high-resolution coupled hydrodynamic-biogeochemical models for selected embayments would be helpful in dissecting the controls of various physical and biological factors on algal growth and hypoxia and could assist in developing ecologically meaningful management endpoints.

## 5. References


Abdelrhman, M.A. 2017. Quantifying Contributions to Light Attenuation in Estuaries and Coastal Embayments: Application to Narragansett Bay, Rhode Island Estuaries and Coasts 40: 994. https://doi.org/10.1007/s12237-016-0206-x

Batiuk, R., P. Bergstrom, M. Kemp, and M. Teichberg. 2000. Chesapeake Bay Submerged Aquatic Vegetation Water Quality and Habitat-Based Requirements and Restoration Targets: A Second Technical Synthesis. Printed by the United States Environmental Protection Agency for the Chesapeake Bay Program, Annapolis, MD.

Benson, J.L., D. Schlezinger, and B.L. Howes. 2013. Relationship between nitrogen concentration, light, and Zostera marina habitat quality and survival in southeastern Massachusetts estuaries. Journal of Environmental Management 131:129–137.

Bintz J.C., Nixon S.W. 2001. Responses of eelgrass *Zostera marina* seedlings to reduced light. Marine Ecology Progress Series 223:133-141.

Cerco, C.F., S-C Kim, and M.R. Noel. 2010. The 2010 Chesapeake Bay Eutrophication Model. A Report to the U.S. Environmental Protection Agency Chesapeake Bay Program and to the U.S. Army Engineer Baltimore District. U.S. Army Engineer Research and Development Center, Vicksburg, MS. December 2010.

Cloern J.E. 2001. Our evolving conceptual model of the coastal eutrophication problem. Marine Ecology Progress Series 210:223-253.

Dennison W.C., Orth R.J., Moore K.A., Stevenson J.C., Carter V., Kollar S., Bergstrom P., Batiuk R. 1993. Assessing water quality with submersed aquatic vegetation. Habitat requirements as barometers of Chesapeake Bay health. Bioscience 43:86-94.

Graff, J.R., Rynearson, T.A. 2011. Extraction method influences the recovery of phytoplankton pigments from natural assemblages. Limnol. Oceanogr.: Methods 9, 2011, 129–139.

Howes, B.L., R. Samimy, and B. Dudley. 2003. Site-Specific Nitrogen Thresholds for Southeastern Massachusetts Embayments: Critical Indicators Interim Report. Prepared by Massachusetts Estuaries Project for the Massachusetts Department of Environmental Protection.

Howes, B.L., R. Samimy, D. Schlezinger, S. Kelley, J. Ramsey, and E. Eichner. 2006. Linked Watershed-Embayment Model to Determine Critical Nitrogen Loading Thresholds for the Pleasant Bay System, Massachusetts. Prepared by Massachusetts Estuaries Project. Final report – May 2006.

Kemp W.M., Batiuk R., Bartleson R., Bergstrom P., and 12 others. 2004. Habitat requirements for submerged aquatic vegetation in Chesapeake Bay: water quality, light regime, and physical-chemical factors. Estuaries 27:263-377.

Koch, E. 2001. Beyond light: Physical, geological and geochemical parameters as possible submersed aquatic vegetation habitat requirements. Estuaries. 24(1):1-17.

Latimer, J.S., M.A. Tedesco, R.L. Swanson, C. Yarish, P.E. Stacey, and C. Garza, ed. 2014. *Long Island Sound: Prospects for the Urban Sea*. Springer Series on Environmental Management, Springer-Verlag, New York.

Long Island Sound Study. 2018. Hypoxia. Retrieved from http://longislandsoundstudy.net/about/our-mission/management-plan/hypoxia/

Morrison, J.R., T.K. Gregory, S. Pe'eri, W. McDowell and P. Trowbridge. 2008. Using Moored Arrays and Hyperspectral Aerial Imagery to Develop Nutrient Criteria for New Hampshire's Estuaries. Final Report to The New Hampshire Estuaries Project.

NHDES. 2009. *Numeric Nutrient Criteria for the Great Bay Estuary*. New Hampshire Department of Environmental Services, Concord, New Hampshire. Accessed March 2018. https://www.des.nh.gov/organization/divisions/water/wmb/wqs/documents/20090610_estuary_criteria.pdf.

Nixon S, Buckley B.., Granger S, Bintz J. 2001. Responses of very shallow marine ecosystems to nutrient enrichment. Human and Ecological Risk Assessment 7(5):1457-1481.

Nixon S.W. 1992. Quantifying the relationship between nitrogen input and the productivity of marine ecosystems. Pp. 57-83 in: Proceedings of Advanced Marine Technology Conference Vol. 5, AMTEC, Takahashi M, Nakata K, Parsons TR (eds.), Tokyo, Japan.

Ochieng, C.A., F.T. Short, and D.I. Walker. 2010. Photosynthetic and morphological responses of eelgrass (*Zostera marina L.*) to a gradient of light conditions. *Journal of Experimental Marine Biology and Ecology* 382(2):117–124.

Stich, H, Brinker, A. 2005. Less is better: Uncorrected versus pheopigment-corrected photometric chlorophyll-a estimation. Archiv für Hydrobiologie. 162. 111-120. 10.1127/0003-9136/2005/0162-0111.

USEPA. 2018a. Establishing Nitrogen Endpoints for Three Long Island Sound Watershed Groupings: Embayments, Large Riverine Systems, and Western Long Island Sound Open Water; Subtask E. Summary of Hydrodynamic Analysis. U.S. Environmental Protection Agency Region 1 and Long Island Sound Office. Prepared by Tetra Tech March 27, 2018.

USEPA. 2018. Establishing Nitrogen Endpoints for Three Long Island Sound Watershed Groupings: Embayments, Large Riverine Systems, and Western Long Island Sound Open Water; Subtask F/G. Draft Summary of Empirical Modeling and Nitrogen Endpoints. U.S. Environmental Protection Agency Region 1 and Long Island Sound Office. Prepared by Tetra Tech. April 13, 2018.

USEPA. 2010. Using Stressor-response Relationships to Derive Numeric Nutrient Criteria. EPA 820-S-10-001. U.S. Environmental Protection Agency, Washington DC.

USEPA. 2001. Nutrient Criteria Technical Guidance Manual: Estuarine and Coastal Marine Waters. EPA-822-B-01-003. U.S. Environmental Protection Agency, Office of Water, Washington, DC.

Valiela I., McClelland J., Hauxwell J., Behr P.J., Hersh D., Foreman, K. 1997. Macroalgal blooms in shallow estuaries: Controls and ecophysiological and ecosystem consequences. Limnology and Oceanography 42(5):1105-1118.

Vaudrey, J.M.P., J. Eddings, C. Pickerell, L. Brousseau., and C. Yarish. 2013. Development and Application of a GIS-based Long Island Sound Eelgrass Habitat Suitability Index Model. Final report submitted to the New England Interstate Water Pollution Control Commission and the Long Island Sound Study. 171 p. + appendices.

Vaudrey, J.M.P. 2008. Establishing Restoration Objectives for Eelgrass in Long Island Sound. Part I: Review of the Seagrass Literature Relevant to Long Island Sound. Final Grant Report to the Connecticut Department of Environmental Protection, Bureau of Water Protection and Land Reuse and the U.S. Environmental Protection Agency.

Zuur A.F., Ieno E.N., Walker N.J., Saveliev A.A., Smith G.M.. 2009. Mixed Effects Models and Extensions in Ecology with R. New York: Springer.