Response to Independent Technical Reviewer Comments and Public Comments on *Subtasks F and G. Summary of Empirical Modeling and Thresholds (April 13, 2018 Draft)*

October 1, 2020

A. Expert Technical Review Comments on the April 13, 2018 Draft F/G Memo

This section contains the original responses written by each of the expert technical reviewers.

Question 2-1

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).

Comment 2-1 Bierman

Comment Tracking ID #30

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.

> *Response: See response to comment tracking ID #31.*

Comment Tracking ID #31

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.

Response: We made the terminology consistent in the literature review and the F/G memos. We also added a definitions box. In these memos, for example, EPA used the following terms:

- *Management goals: Reestablish and maintain water quality and habitat conditions to support diverse self-sustaining commercial, recreational, and native fish, water-dependent wildlife, and shellfish.*
- *Assessment endpoint: Estuarine eelgrass habitat abundance and distribution; Benthic and pelagic community diversity and abundance*
- *Primary response variable: chlorophyll a-corrected and DO*
- *Primary causal variable target concentration: total nitrogen*

This decision was made based on the following EPA guidance:

- *USEPA. 1998. Guidelines for Ecological Risk Assessment. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC.*
- *USEPA. 2010. Using Stressor-response Relationships to Derive Numeric Nutrient Criteria. U.S. Environmental Protection Agency, Washington DC. EPA 820-S-10-001.*

Comment Tracking ID #32

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

Response: We added equation numbers to the memo.

Comment Tracking ID #33

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.

Response: We added text to each section indicating that the predictors pH, salinity, and temperature were included to see if they significantly improved the fit of the model.

Comment Tracking ID #34

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.

Response: We added new tables that list the final model coefficients for each model.

Comment Tracking ID #35

f. Plots for "observed" vs "fitted" values are presented on Pages F-14, F-18, F-20, F-22, and F23, 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.

Response: We updated these figures to provide the requested detail.

Comment Tracking ID #36

g. A plot of observed data for K_d vs chlorophyll *a* for embayments, along with results for the 10th 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.

> *Response: We added population level plots for models that figure in the final TN values so viewers can see the entire model relationship as requested.*

Comment 2-1 Brush

Non-Substantive Comment Tracking ID #NS-1

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.

> *Response: Thank you for your comment. Responses to detailed comments provided in the topic specific questions are provided with those comments below.*

Comment 2-1 Janicki

Comment Tracking ID #37

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.

> *Response: As stated on the inside cover to the report "This Tetra Tech technical study was commissioned by the United States Environmental Protection Agency (EPA) to synthesize and analyze water quality data to assess nitrogen-related water quality conditions in Long Island Sound and its embayments, based on the best scientific information reasonably available. This study is neither a proposed Total Maximum Daily Load (TMDL), nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance, but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies."*

Comment 2-1 Justic

Non-Substantive Comment Tracking ID #NS-2

The memorandum is well organized and well written.

> *Response: Thank you for your comment.*

Question 2-2

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?

Comment 2-2 Bierman

Comment Tracking ID #38

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 10th 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.

> *Response: We clarified the use of terminology in the text to reduce confusion on the linkage between these values. We also clarified language on where chlorophyll values come from (Kd targets and literature values). We also added all the population data to each plot (as well as a population model plot), so viewers can see how each embayment dataset influences the adjustment of the hierarchical model curve for each embayment. Lastly, we clarified where the model confidence intervals come from (the entire population model). Please note that the confidence intervals are around the model slope; they are not prediction intervals so it is not uncommon for points to lie outside regression confidence intervals.*

Comment 2-2 Brush

Comment Tracking ID #39

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.

> *Response: We changed the format of columns in the table to add clarity and consistency with terminology. An example is provided below.*

Management Goal	Assessment Endpoint	Lines of Evidence	Chlorophyll *a*-Corrected Primary Response Variable (µg/L)	TN Primary Causal Variable Target Concentration (mg/L)
Seagrass Protection and Restoration	Chlorophyll *a*-corrected	Stressor–Response Model for Individual Embayments Mean (80th Percent Confidence Interval)	5.5	
			10	
		Literature Review Median (Range)		
		Distribution-Based Approach – All Embayments 25th Percentile		
Other Aquatic Life Protection and Restoration		Literature Review Median (Range)		
		Distribution-Based Approach – All Embayments 25th Percentile		

Comment Tracking ID #40

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

> *Response: We changed the column heading to "Chlorophyll a-Corrected Primary Response Variable (ug/L)".*

Comment Tracking ID #41

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

> *Response: We changed "values or concentrations" to "concentrations."*

Comment Tracking ID #42

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)?

> *Response: We used the population fit (grey line) because the blue line did not intersect the Chla values. After updating the models, this footnote was no longer needed.*

Comment 2-2 Janicki

Non-Substantive Comment Tracking ID #NS-3

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.

> *Response: Thank you for your comment.*

Comment 2-2 Justic

Non-Substantive Comment Tracking ID #NS-4

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

Response: Thank you for your comment.

Question 2-3

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?

Comment 2-3 Bierman

Comment Tracking ID #43

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.

Response: The literature review approach is one of three complementary lines of evidence (along with distributional-based and stressor response) being used to identify target TN concentrations in this analysis. All three approaches are well established and their application widely described in existing technical guidance (e.g., USEPA 1999, 2001, 2010, 2012, 2015). EPA made every effort to focus only on relevant literature from comparable ecosystems to LIS. Originally, we focused on available literature from the Chesapeake Bay to Maine. For the reasons stated by the review, namely the effect of site and region specific differences as one moves further afield, a decision was made to focus primarily on 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; the northern systems being farther from the Virginian province and the southern being a substantially different estuarine system in terms of size, geography, hydrodynamics, salinity structure, and climate.

USEPA. 1999. Protocol for Developing Nutrient TMDLs. EPA 841-B-99-007. 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, Washington DC.

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

USEPA. 2012. Authorization to Discharge under the National Pollutant Discharge Elimination System: The Town of Newmarket, New Hampshire. NPDES Permit #NH0100196. U.S.

Environmental Protection Agency. Accessed February 2017. https://www3.epa.gov/region1/npdes/permits/2012/finalnh0100196permit.pdf.

USEPA. 2015. Authorization to Discharge under the National Pollutant Discharge Elimination System: The City of Taunton, Massachusetts, Department of Public Works. NPDES Permit # MA0100897. U.S. Environmental Protection Agency. Accessed February 2017.

Non-Substantive Comment Tracking ID #NS-5

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).

Response: Thank you for your comment.

Comment Tracking ID #44

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.

Response: The requested data are thoroughly summarized in the data description in Memo D. Per this comment, we added more descriptive plots, more statistical diagnostics, and attempted to clarify the path of model construction. We added additional analysis to the Kd modeling and hierarchical modeling sections to address this reviewer's comment about "conceptual flaws and questionable assumptions", especially adding TSS and DOC (using salinity as a surrogate) into the Kd models and clarifying the hierarchical model descriptions.

Comment Tracking ID #45

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 25th 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 25th percentile TN targets will protect the LIS assessment endpoints (eelgrass, aquatic life).

Response: It is unclear what the reviewer means by site-specific data for eelgrass distributions, primary response variables, or physical habitat requirements. The EPA guidance that the reviewer cites in support of the validity of this approach does not require that information be used when using the distribution-based approach. That guidance recommended use of the 25th *percentile in the absence of least disturbed reference population conditions, which was the case in LIS. For the distribution-based approach, EPA used data from waterbodies that could support the management goals and assessment endpoints, which is sufficient for applying this approach (based on the guidance). Lastly, we used nutrient concentration from two study embayments (Niantic and Mystic) which had displayed some recovery of eelgrass as supplementary support for this analysis and added data from a third, Stonington Harbor. Please see response to comment tracking ID #72 for details.*

Comment Tracking ID #46

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.

Response: It is unclear how LRA values could be based on site-specific data since most literature review derived values are from external literature and not from the data one is using, however as noted in the text, we did search for literature from the LIS. The DbA approach followed EPA guidance and does not typically incorporate parameters related to response, such as eelgrass or aquatic life. Please see response to comment tracking ID #45 for a response for how this was addressed. Please see responses to comment tracking ID #44 and responses to questions 10a-10f for additional specific responses.

Comment Tracking ID #47

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.

Response: Please see the response to comment tracking ID #48 for details about how these factors are incorporated inherently into the analysis via the EHSI.

Comment Tracking ID #48

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."

Response: We used the Eelgrass Habitat Suitability Index (EHSI) developed by Vaudrey et al. (2013) and detailed in the Literature Review and F/G memo. We used this index to help identify light targets for eelgrass growth. These habitat suitability index values are derived from many of the factors described by the reviewer and, therefore, the stressor-response approach implicitly includes the habitat factors important to the assessment endpoint being used and for the embayment areas being modeled.

Site specific data of the kind being recommended are not universally available for the embayments in every location nor do resources allow for a discretized analysis of each embayment. Although the lines of evidence used in this project may not consider all of these site-specific details, they have been satisfactorily and defensibly applied to similar systems in this way.

Sentinel station modeling is commonly applied in water quality modeling settings (i.e., finding the most sensitive or response grid cell and layer and modeling such that reductions meet desired conditions at that critical cell). However, resources were not available to do the type of water quality modeling that relies on this approach and as was applied in the MEP (Howes et al. 2006).

Howes, B.L., S.W. Kelley, J.S. Ramsey, R. Samimy, D. Schlezinger, and E.M. Eichner. 2006. Linked Watershed-Embayment Model to Determine Critical Nitrogen Loading Thresholds for Nantucket Harbor, Town of Nantucket, Nantucket Island, Massachusetts. SMAST/DEP Massachusetts Estuaries Project. Massachusetts Department of Environmental Protection, Boston, MA. Accessed March 2018. http://www.oceanscience.net/estuaries/report/Nantucket/Nantucket_Hbr_MEP_Final.pdf.

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.

Non-Substantive Comment Tracking ID #NS-6

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.

Response: Thank you for your comment. Responses to this comment appear under questions 2-8, 2-10, and 2-11.

Comment 2-3 Brush

Comment Tracking ID #49

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.

Response: Thank you for your comment and support. As detailed in the memo, EPA did pursue the use of aquatic life assessment endpoints via DO, but given limitations in the data, was unable to produce models sufficient to derive TN concentrations from that relationship using the stressor-response approach; however, the literature-based and distribution-based approaches implicitly include protection for aquatic life including benthic fauna.

Non-Substantive Comment Tracking ID #NS-7

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.

Response: Thank you for your comment.

Comment Tracking ID #50

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.

Response: Before completing the DO-chlorophyll models, all potential DO endpoints were explored including those from the 3 states. Had the models produced statistically significant results, EPA would have modeled to the appropriate DO endpoints (e.g., for each state's waters, using the appropriate state criterion). However, this was rendered unnecessary because of the *lack of a defensible DO-chlorophyll stressor-response model for the stressor response approach line of evidence.*

Non-Substantive Comment Tracking ID #NS-8

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.

Response: Thank you for your comment.

Comment Tracking ID #51

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.

Response: Thank you for your comment and support.

Comment Tracking ID #52

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.

Response: Thank you for your comment and support for the approach. We agree with the observations about mixing across these small and, based on the salinity analysis in Memo E, highly diluted embayment segments being modeled. We used all available surface water quality which, as noted, was highly variable across embayments and largely insufficient in space or time to do the types of gradient analysis requested. While it might be possible to do such for one or *two embayments, application of the approach appropriately across all waterbodies is precluded by the limited data.*

Comment 2-3 Janicki

Comment Tracking ID #53

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.

Response: Thank you for your comment and support for the reasonableness of the values and this region. Further effort is planned to convert concentrations to loads and identify load reductions as recommended in this comment. To help clarify some of the assumptions, we added additional details on model assumptions pursuant to this and other comments by the same reviewer related to modeling details. We believe that data limitations were overcome using the approach applied for the stressor-response modeling and are not relevant to the literature review and not an issue for the distribution-based approach used. EPA believes that minimum data requirements were exceeded for the stressor-response approach, especially the global model. Please also see response to comment tracking ID #52 for additional response about non-homogeneity in embayments.

Comment 2-3 Justic

Non-Substantive Comment Tracking ID #NS-9

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.

Response: Thank you for your comment. Responses to your comments can be found under the response to comments on Question 10.

Comment Tracking ID #54

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.

Response: Site-specific biogeochemical models were not possible given resources for the project.

Question 2-4

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

Comment 2-4 Bierman

Comment Tracking ID #55

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.

> *Response: EPA guidance (USEPA 2001) recommends classification be done to reduce the variability or noise associated with natural gradients in biological, geomorphological, hydrologic, and chemical phenomena that affect nutrient concentrations or nutrient cycling and response. The EPA estuarine and coastal guidance largely focuses on geomorphic factors, important drivers of estuarine nutrient response, thus the distinction in this work between embayments and open water. EPA guidance does not recommend that designated uses nor impairment status nor data availability be used as classification considerations. In considering combining the western and eastern narrows, which are human constructs and not ecological constructs, EPA considered these underlying natural factors as well as the expected ecological composition both in terms of expected biological assemblages and ecosystem function of these two areas. EPA's decision was that, ecologically, these two segments would be expected to support similar assemblages and to exhibit generally similar nutrient inputs, uptake, processing, and response in the absence of human disturbance, and thus could be combined, ecologically, for the basis of this analysis.*
>
> *USEPA. 2001. Nutrient Criteria Technical Guidance Manual: Estuarine and Coastal Marine Waters. EPA 822-B-01-003. U.S. Environmental Protection Agency, Washington DC.*

Comment 2-4 Brush

Comment Tracking ID #56

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).

> *Response: Clarifying text was added to the Section "LIS Open Water Models." Please note that previous Table F-8 ("Paired Observations and Station Counts for the Dissolved Oxygen vs Chlorophyll Open Water Model, by Open Water Group) refers to the "Eastern Narrows", not to be confused with the "Eastern LIS". Data from the "Eastern LIS" were removed, while data from the "Eastern Narrows" (part of the Western LIS) were included in the open water models.*
>
> *We agree with your comment concerning DO. The text did not clarify that the DO-chlorophyll relationships were bottom DO versus surface chlorophyll and this distinction is indicated in the revised memo. EPA believes and agrees with the reviewer that this mismatch in space (and likely time) between surface chlorophyll and DO demand in the bottom of the water column, is an important component in the lack of a relationship.*

Comment 2-4 Janicki

Comment Tracking ID #57

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.

> *Response: We refer the reviewer to Memos A and D, which provide ambient data on these two waterbodies, and to the response to comment tracking ID #55 for more on the basis for combining these two segments.*

Comment 2-4 Justic

Comment Tracking ID #58

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).

> *Response: We refer the reader to the response to comment tracking ID #55 for more on the basis for combining these two segments. We agree with the observations regarding estimated residence time. The nitrogen yield differences are overwhelmingly driven by anthropogenic inputs and are not a reliable basis for making classification decisions.*

Question 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?

Comment 2-5 Bierman

Comment Tracking ID #59

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.

> *Response: Thank you for your comment. The open water portion of the LIS, as defined for this work, includes waters of sufficient minimum depths and clarity to currently support and to have likely historically supported eelgrass. The term "open water" may be misleading but is not restricted to only deep, wind exposed areas. It also includes shallow, leeward, coastal waters outside of embayments that are well mixed with deeper, exposed areas. As the cited Figures 11 and 22 from Vaudrey et al. (2013) indicate, there is suitable, shallow habitat in ample portions of this open water of LIS, for example around Fisher Island and northern portions of Long Island, off of the Connecticut River, etc. and there are documented beds existing in some of these areas (Tiner et al. 2013). Additionally, eelgrass is used in combination with other aquatic life, which accounts for other assessment endpoints in the open waters of the Eastern and Western Narrows.*
>
> *Tiner, R., K. McGuckin, and A. MacLachlan. 2013. 2012 Eelgrass Survey for Eastern Long Island Sound, Connecticut and New York. U.S. Fish and Wildlife Service, National Wetlands Inventory Program, Northeast Region, Hadley, MA.*
>
> *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.*

Comment 2-5 Brush

Comment Tracking ID #60

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.

Response: Thank you for your support of using eelgrass as an endpoint. We agree that eelgrass is an important management goal for LIS including in the Eastern and Western Narrows. Figure 11 in the Vaudrey et al. (2013) document indicates that the depth, tidal, and light environmental is sufficient in this area if all other parameters are optimal. Some parameters are not due to human disturbance associated with excess nutrients (e.g., low dissolved oxygen). Still, some upper embayment areas in the Narrows score in the mid-range for habitat suitability (Figure 22, Vaudrey et al. 2013). Note that in the analysis, eelgrass is used in combination with other aquatic life, which accounts for other assessment endpoints in the open waters of the Eastern and Western Narrows.

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.

Comment 2-5 Janicki

Non-Substantive Comment Tracking ID #NS-10

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

Response: Thank you for your comment.

Comment 2-5 Justic

Non-Substantive Comment Tracking ID #NS-11

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

Response: Thank you for your comment.

Question 2-6

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)?

Comment 2-6 Bierman

Comment Tracking ID #61

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.

> *Response: Thank you for your comment about how the data used for the empirical modeling approaches depict a reasonable snapshot of current conditions in LIS.*

Comment Tracking ID #62

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.

> *Response: Thank you for your comment regarding applicability of these data for future conditions.*

Comment Tracking ID #63

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.

> *Response: Thank you for your comment regarding applicability of these data for future conditions.*

Comment 2-6 Brush

Comment Tracking ID #64

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.

> *Response: Thank you for your comment about how the data used for the empirical modeling approaches depict a reasonable snapshot of current conditions in LIS.*

Non-Substantive Comment Tracking ID #NS-12

Comment 2-6 Janicki

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.

> *Response: Thank you for your comment. Climate change was not a focus of the analysis.*

Comment 2-6 Justic

Comment Tracking ID #65

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.

Response: Thank you for your comment regarding applicability of these data for future conditions. Note that in response to this and similar comments, additional time periods were considered in the revised analysis. Please refer to responses to comments on Memo E regarding time periods (e.g., Dr. Brush's comments on question 1-2).

Comment Tracking ID #66

It should be pointed out that hydrodynamic model simulations alone are generally inadequate for water quality forecasting. Implementing high-resolution coupled hydrodynamic-biogeochemical models (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.

Response: Site specific biogeochemical models were beyond the resources available for this project. Thank you for your feedback regarding applicability under future conditions.

Question 2-7

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?

Comment 2-7 Bierman

Comment Tracking ID #67

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).

Response: We added the rationale for using TN concentrations versus loads to Memo F/G. We used concentrations instead of loads because we did not have temporally resolved load measures or estimates (i.e., annual estimates) for each embayment that could be matched to response conditions, concentrations could be compared to the temporally varying response variables discussed in Memo D, concentrations are directly related to organism response, and concentrations are consistent with EPA guidance. However, we might explore load-based models

> *if appropriate data can be identified. For a response on terminology, please see the response to Dr. Bierman's comments on question 2-1 (comment tracking ID #30-36). Also see the response to Dr. Justic's comments on question 2-3 for a response on models (comment tracking ID #54).*

Comment 2-7 Brush

Comment Tracking ID #68

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).

> *Response: We removed reference to loads on pages F-1 and F-24. Regarding the use of loads instead of concentrations, please see the response to comment tracking ID #67.*

Comment Tracking ID #69

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).

Response: We used the total nitrogen fraction to be consistent with guidance. Arguments as to dissolved versus total fractions are well known, with defensible positions to be made on both sides. In deference, we chose to continue to be consistent with existing guidance and policy to use the total fraction. The total fraction incorporates all the nitrogen able to contribute to biomass responses currently and based on any future recycling. It also incorporates all the past effect and future nutrient inertia. It also avoids arguments concerning dissolved fractions as either too ephemeral to quantify accurately or reflective of excess and therefore, unpredictive fractions.

With regards to flushing or its corollary residence time, we explored incorporating residence time as a predictor, however reliable estimates of residence time for each embayment were difficult to estimate accurately. To the extent residence time is reflected in salinity and the embayment specific modeling effect, it is incorporated. With regards to depth, it was incorporated into the Kd modeling effort via seagrass colonization, and, again, to the degree embayment specific differences in depth are incorporated via the hierarchical model, is also considered there.

Comment 2-7 Janicki

Comment Tracking ID #70

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.

Response: Thank you for your comment. Please see response to comment tracking ID #67 with regards to load versus concentration.

We did not consider lag effects in the stressor-response models. The lack of temporally resolved data for each embayment precludes our ability to adequately incorporate lag effects.

Please see response to comment tracking ID #69 for a response with regard to residence time.

Comment 2-7 Justic

Comment Tracking ID #71

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.

> *Response: Thank you for your comment. We are unaware of comprehensive, LIS wide embayment sediment organic carbon and nitrogen content datasets for use in this application. The sediment organic carbon data used in the Vaudrey et al. (2013) study was dated and primarily concentrated in open water areas (see Figure 16 from that publication). The same study did not mention sediment nitrogen data. For these reasons, we did not pursue incorporating sediment data into the modeling effort.*

Question 2-8

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?

Comment 2-8 Bierman

Comment Tracking ID #72

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.

> *Response: Thank you for your comment. We agree that the MEP is site-specific, which is why we did not use specific values from MEP, but rather the range of those values as a guide. The literature review approach was not intended to be site-specific and it is part of a multiple lines of evidence approach including site-specific stressor-response modeling. Site specific hydrodynamic and mechanistic modeling was beyond the resources available for this project.*
>
> *We used data from the eelgrass surveys referenced (Tiner et al. 2013); the same work referenced by Vaudrey et al. (2013). As noted in the F/G memo, we identified two embayments (Niantic Bay and Mystic Harbor) from these surveys that are also priority embayments within which seagrass coverage increased the most between 2002 and 2012 (Tiner et al. 2013). We since added data for Stonington Harbor but excluded data from the Connecticut River Area that showed essentially no gain (Connecticut River Area). Nutrient concentration data from within these*

watersheds were compiled and reviewed as supporting information. As seen below with water quality data extracted from tables in Memo D, the median values for Mystic, Niantic, and Stonington are 0.53 mg/L, 0.26 mg/L, and 0.33 mg/L respectively. In comparison, the median value of 0.39 mg/L for Massachusetts estuaries (described in Table F-1 of the F/G Memo), is similar to the average of these three medians (0.37 mg/L).

Extracted Rows from Tables in Memo F/G.

Parameter Name in Database	Parameter Description	Data Collection Period	# of Stations	# of Samples	# of Samples by Depth			Values		
					Bottom	Middle	Surface	10th Percentile	90th Percentile	Median
Mystic										
TN_mgL	Total nitrogen [mg/L]	2004–2015	4	114	44	1	69	0.39	0.75	0.53
Niantic										
TN_mgL	Total nitrogen [mg/L]	2002–2014	18	112	4	1	107	0.17	0.38	0.26
Stonington Harbor										
TN_mgL	Total nitrogen [mg/L]	2008–2015	5	77	28	0	49	0.23	0.45	0.33

Tiner, R., K. McGuckin, and A. MacLachlan. 2013. 2012 Eelgrass Survey for Eastern Long Island Sound, Connecticut and New York. U.S. Fish and Wildlife Service, National Wetlands Inventory Program, Northeast Region, Hadley, MA.

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.

Comment Tracking ID #73

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.

Response: Thank you for your comment and support of using data from Massachusetts estuaries. A decision was made to focus primarily on 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; the northern systems being farther from the Virginian province with different climate and the southern being a substantially different estuarine system in terms of size (e.g., Chesapeake Bay is 3.5x larger), geography (e.g., Chesapeake Bay is 143 miles south, nearly twice as long, oriented north-south), hydrodynamics (differences in residence time), salinity structure (e.g., the two estuaries have similar volumes, but Chesapeake Bay has 4x larger watershed leading to a greater salinity gradient and greater influence of freshwater), and climate. The approach assumes that literature based targets from these Massachusetts estuaries were more appropriate for LIS, given the similarities in geography, climate, and species composition (e.g., Zostera marina) that require similar physical and chemical habitat requirements in both embayment as well as shallow and deeper open water habitats between the two regions.

To address the commenter's concern about the literature review not being directly transferable to LIS, the literature review was not intended to be site-specific and it is part of a multiple lines of evidence approach. For more on this, please see response to comment tracking ID #72.

Comment 2-8 Brush

Comment Tracking ID #74

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.

Response: Thank you for your comment and support of excluding the Chesapeake Bay and Great Bay and including Massachusetts estuaries. During the literature search, we did not come across similar information for Rhode Island and New Jersey. We looked for additional literature published in the last year and found nothing further from either state that merited addition.

With regards to the use of data from embayments with known eelgrass information, EPA did rely on available information on eelgrass and water quality: please see the response to comment tracking ID #72.

Comment 2-8 Janicki

Comment Tracking ID #75

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.

Response: We believe, consistent with guidance, that use of the literature review line of evidence is defensible and recommended. Such research provides a useful line of evidence that has been employed successfully in many nutrient target concentration setting applications, especially when used within a multiple line of evidence framework such as that employed here. With regards to the appropriateness of the regions selected for use, please see the response to comment tracking ID #73.

Comment 2-8 Justic

Comment Tracking ID #76

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.

Response: Thank you for your comment. The response to your comment about water column metrics can be found in the response to comment tracking ID #71.

Question 2-9

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).

Comment 2-9 Bierman

Comment Tracking ID #77

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.

Response: Please refer to the response to your comments on question 2-8 (comment tracking ID #72), which addresses LIS site-specific data and the literature review line of evidence. To address *your concern about the level of evidence to support eliminating or capping TN target concentrations, EPA included a note with each table in Section G of the F/G Memo explaining protective concentration levels ("As per literature review and noted in Table F-1, values exceeding 0.49 mg/L are not considered protective of eelgrass and above 0.60 mg/L are not protective of other aquatic life. Values below 0.20 mg/L are considered below background levels [Howes et al. 2006; NHDES 2009]).") EPA updated the Howes et al. 2006 reference and made it clearer where each value came from with a new footnote ("As per literature review and noted in Table F-1, values exceeding 0.49 mg/L are not considered protective of eelgrass [Howes et al. 2013] and above 0.60 mg/L are not protective of other aquatic life [Howes et al. 2010]. Values below 0.20 mg/L are considered below background levels [NHDES 2009]."). Additional details are provided below.*

- *The value of 0.20 mg/L can be found on page 66 of the NHDES 2009 report (74 of the PDF)*
- *The value of 0.49 mg/L can be found on page 176 of the Howes et al. 2013 report (203 of the PDF)*
- *The value of 0.60 mg/L can be found on page 137 of the Howes et al. 2010 report for the Parkers River Embayment System, Yarmouth, Massachusetts (165 of the PDF)*

Howes, B.L., E. Eichner, R. Acker, R. Samimy, J. Ramsey, and D. Schlezinger. 2013. Linked Watershed-Embayment Model to Determine Critical Nitrogen Loading Thresholds for the Westport River Embayment System, Town of Westport, Massachusetts. SMAST/DEP Massachusetts Estuaries Project. Massachusetts Department of Environmental Protection, Boston, MA. Accessed August 2019. https://www.mass.gov/files/documents/2016/08/wj/mep-westport-bb.pdf

Howes, B.L., S. Kelley, J.S. Ramsey, R. Samimy, D. Schlezinger, and E. Eichner. 2010. Linked Watershed-Embayment Model to Determine Critical Nitrogen Loading Thresholds for the Parkers River Embayment System, Yarmouth, Massachusetts. SMAST/DEP Massachusetts Estuaries Project. Massachusetts Department of Environmental Protection, Boston, MA. Accessed August 2019. http://www.yarmouth.ma.us/DocumentCenter/View/1435/2010-Final-Mass-Esturaries-Project-Report-?bidId=.

NHDES. 2009. Numeric Nutrient Criteria for the Great Bay Estuary. New Hampshire Department of Environmental Services, Concord, New Hampshire. Accessed August 2019. https://www.des.nh.gov/organization/divisions/water/wmb/wqs/documents/20090610_estuary_criteria.pdf.

EPA notes that the other three expert technical reviewers supported the sufficiency of evidence and defensibility of excluding the values indicated.

Comment 2-9 Brush

Non-Substantive Comment Tracking ID #NS-13

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 cutoff 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.

Response: Thank you for your comment and support of using the selected cutoff value.

Comment 2-9 Janicki

Non-Substantive Comment Tracking ID #NS-14

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.

Response: Thank you for your comment and support of using the selected cutoff value.

Comment 2-9 Justic

Non-Substantive Comment Tracking ID #NS-15

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).

Response: Thank you for your comment and support of using the selected cutoff value.

Question 2-10

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.

Comment 2-10 Bierman

Non-Substantive Comment Tracking ID #NS-16

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

Response: Response to the reviewer's comments on question 10 can be found in responses to questions comment tracking ID#: 79, 80, 88, 89, 92, 93, 94, 95, 106, 110, and 114.

Comment 2-10 Brush

Non-Substantive Comment Tracking ID #NS-17

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.

Response: Thank you for your comment. Please refer to the responses to comment tracking ID# 81, 82, 83, 84, 85, 86, 90, 96, 97, 98, 107, 111, and 115. In addition, we revised the stressor response approach, incorporating additional analyses and approaches, based on comments from a number of reviewers.

Comment 2-10 Justic

Comment Tracking ID #78

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 µg 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.

   Response: In response to this and other comments, we added language clarifying how the quantile model accounts for these additional light-attenuating substances. Moreover, we updated the Kd~Chl hierarchical model to explicitly incorporate TSS and DOC (via salinity) into estimates of embayment specific chlorophyll targets per reviewer request. We updated the Chl~TN hierarchical model, which has improved model fit and we show the population model which has precision comparable to many coastal Chl~TN models. Please see responses to comment tracking ID's #54 and 71 for additional relevant response details.

Question 2-10a

***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?**

Comment 2-10a Bierman

Comment Tracking ID #79

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.

> *Response: Thank you for your comment supporting the selected light attenuation targets. Please refer to the response for comment tracking ID #78 for a response to the comment regarding incorporating CDOM and TSS into the Kd modeling efforts.*

Comment Tracking ID #80

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.

> *Response: We added a figure to Memo F/G identifying the eelgrass habitat suitability index scores (<50, 50-88, and >88). This is derived from the same dataset as the Vaudrey Figure 22 map, but focused on specific EHSI cutoffs.*

Comment 2-10a Brush

Comment Tracking ID #81

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%. 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.

> *Response: Thank you for your support of the light level values selected. The Ochieng study did not just show that seedlings did better between 11% and 34%, they concluded that levels below 34% were likely adverse, as cited. But the reviewer raises an important point – the value was somewhere between 11% and 34% and the Bintz and Nixon (2001) study was a valuable recommendation. We added and discussed that citation in the section.*

Comment Tracking ID #82

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 ε 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.

> *Response: The depths associated with Tables F-2 and F-3 are for areas with habitat scores >=50. We updated the text to consistently use the term "colonization depth". We also corrected the table note definition for MLLW.*

Comment Tracking ID #83

I found the cutoff of habitat scores ε 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.

> *Response: We provided an explanation for the value of 50 in the text. Habitats with EHSI of 50 were known to support seagrasses, but values >88 were recommended for restoration sites since that was the minimum EHSI score for restoration sites where growth was seen. However, Vaudrey et al. (2013) iterate many times that "existing eelgrass beds are also found in grids with a model prediction of 50 or greater". We acknowledge that the EHSI includes light as a predictor, but the authors did not provide an option to quantify suitability in the absence of light and it was beyond resources to calculate a new EHSI without light.*

Comment Tracking ID #84

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.")

> *Response: We recalculated the Secchi-Kd product equation using LIS-specific data and will use that value moving forward. We updated the text in the memo to reflect this new analysis.*

Comment Tracking ID #85

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.

> *Response: This was explained on page F-15 as due to noise in the individual embayment chlorophyll vs Kd plots for individual embayments. At that time, a decision was made to use a*

global fit. We revisited the models, incorporated TSS and DOC (using salinity as a surrogate) effects, and now have hierarchical model specific chlorophyll targets for each embayment based on individual embayment Kd needs and characteristics. Please also see response to comment tracking ID #78.

Comment Tracking ID #86

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.

Response: Please see response to comment tracking ID #78. Pursuant to reviewer comments, the hierarchical model was added back in and incorporated TSS and DOC (via salinity) effects per recommendation. We combined the results with other analyses to derive chlorophyll targets to protect Kd.

Comment 2-10a Janicki

Non-Substantive Comment Tracking ID #NS-18

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.

Response: Thank you for your comment and support of the Kd targets. Please see response to comment tracking ID #78, 86, and 91 for information regarding revisiting the Chl~Kd model based on this and similar comments.

Comment 2-10a Justic

Comment Tracking ID #87

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.

Response: Thank you for your comment and support of the Kd targets. Please see response to comment tracking ID #78, 86, and 91 for information regarding revisiting the Chl~Kd model based on this and similar comments.

Question 2-10b

***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?**

Comment 2-10b Bierman

Comment Tracking ID #88

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 K_d 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 K_d in LIS, similar to the model developed by Morrison et al. (2008).

> *Response: We responded to Dr. Bierman's comments on Question 1 under responses for comments with Comment Tracking IDs #30 through #36. Please see responses to comment tracking ID #78, ID #85, and ID #86 indicating how we responded to this and other related comments on the chlorophyll~Kd models.*

Comment Tracking ID #89

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 10th 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 10th 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 10th 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.”

> *Response: We disagree that the use of quantile regression is conceptually flawed in this case. We do agree that we are not deriving a criterion, but we disagree that quantile regression is only appropriate for deriving criteria. Quantile regression is an appropriate method to derive a value for stressors controlling responses when the response surface varies across the distribution of stressor values (see Cade and Noon reference as well as papers by Xu et al. 2015a, 2015b); in our case, chlorophyll values associated with Kd, not, as the reviewer stated, Kd values associated with chlorophyll. Quantile regression is ideally suited to deriving values for stressors that control a response in that portion of the distribution minimally affected by other factors. In the case of Kd, TSS and CDOM do increase light attenuation, as we and the reviewer acknowledge, and this is reflected in the noise seen in the median and upper percentile responses of Kd to Chl. However, at the lower quantile, the clear factor ceiling (or basement in this case) response of Kd to Chl reflects the primary limiting effect of chlorophyll on Kd, minimally influenced by the other factors. For this reason, we believe quantile regression to be a defensible method for interpolating those Chl values associated with eelgrass Kd requirements. This having been said, in response to this and other review comments with regards to the Kd~Chl models, we have developed hierarchical models that incorporate TSS and DOC (using salinity as surrogate) to derive additional chlorophyll values’ please see response to comment tracking IDs #78, 85, and 86 for more details.*

> *Cade, B.S., and B.R. Noon. 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1:412–420.*
>
> *Xu, Y., A.W. Schroth, and D.M. Rizzo, D.M. 2015. Developing a 21st Century framework for lake-specific eutrophication assessment using quantile regression. Limnology and Oceanography: Methods 13(5):237–249.*

Xu, Y., A.W. Schroth, P.D. Isles, and D.M. Rizzo, D.M. 2015. Quantile regression improves models of lake eutrophication with implications for ecosystem-specific management. Freshwater Biology 60(9):1841–1853.

Comment 2-10b Brush

Comment Tracking ID #90

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.

Response: In response to this and other comments, we provided additional detail on the quantile regression model in the text and clarified the language. The reviewer is encouraged to read the cited literature upon which this approach is based, especially the Cade and Noon paper which is targeted towards ecologists. In response to individual questions:

- *The original paragraph explained that the resulting embayment specific plots yielded uninformative plots of Kd vs chlorophyll in many embayments due to noise in the models, making it difficult to derive chlorophyll targets for many embayments without extrapolation. As a result, a sound wide, population level model was developed. As stated above, we improved both the explanation and models in the revised memo.*
- *We improved the display and discussion of the quantile model in the next version.*
- *We edited the section to improve clarity and we had erroneously used average depth instead of average colonization depth. We corrected this.*
- *Please see response to comment tracking ID #85.*

Cade, B.S., and B.R. Noon. 2003. A gentle introduction to quantile regression for ecologists. Frontiers in Ecology and the Environment 1:412–420.

Comment 2-10b Janicki

Comment Tracking ID #91

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 10th percentile value is curious. Given that as K_d increases, light availability decreases, modeling the 10th 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.

> *Response: This interpretation is correct: the lower quantile models that portion of the relationship that appears to represent the strongest constraint on Kd by chlorophyll, unencumbered by the effects of other factors known to attenuate light (TSS and CDOM). The latter factors would, indeed, result in Kd values larger than those predicted by chlorophyll at any chlorophyll concentration. In essence, it is a way to evaluate what Kd value would exist in the absence of those other confounding effects, as the clearest approximation of constraint. Please see response to comment tracking ID #78, ID #85, and ID #86 for how we improved the chlorophyll estimates to protect Kd using TSS and DOC (using salinity as surrogate) pursuant to this and other similar comments.*

Comment 2-10b Justic

Non-Substantive Comment Tracking ID #NS-19

The 10th 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.

> *Response: Thank you for your comment.*

Question 2-10c

***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?**

Comment 2-10c Bierman

Comment Tracking ID #92

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.

DO vs Chlorophyll *a* for Embayments

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.

> *Response: Thank you for your comment and insights on the difficulty of DO modeling in mixed systems. We added a population model of DO~Chl as requested to illustrate the overall model. Please see response to comment tracking ID #78, 85, and 86 for responses to the Kd model concerns.*

Comment Tracking ID #93

Chlorophyll *a* vs TN for Embayments

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 (10th 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.

> *Response: In response to this comment, we added a population wide Chl~TN model so the model can be evaluated. We also added additional model diagnostics so the model can be further evaluated. We updated the hierarchical LIS model with new data (additional embayments) and adapted predictors so that we now have more data per embayment. Also, in revising the chlorophyll values to address issues (see comment tracking ID #78), we revised our chlorophyll targets. Lastly, we now plot the population dataset on each plot so viewers can see how the hierarchical model merely adjusts the overall model to the embayment specific data and is not actually a "poor fit" through any embayment dataset. We explained the 90% confidence interval issues in comment tracking ID #38. That was a misunderstanding of hierarchical models, perhaps. There is a clearly evident relationship between TN and Chlorophyll across embayments, as expected (See Figure F-17). We also improved the model explanation. The results of this work (based on reviewer comments) is a model that is more variable, but a better fit for resolving TN values for chlorophyll values for the chlorophyll values (3 to 6 µg/L range) to protect embayment averaged Kd values and maximum associated with the upper limit Kd=0.7 LIS goal (10 µg/L). In addition, the final Memo G values are TN ranges rather than specific values, using multiple lines of evidence, in recognition of the variability from the stressor-response line. We would argue, as other reviewers suggested, that the resulting multiple lines of evidence based ranges provide scientifically defensible protective TN values for these embayments.*
>
> *With regards to the question of EHSI values and extent of protection, we direct the reviewer to comment tracking ID #80 and ID #83 and the text of the memo, which articulates the choice of EHSI value and how that incorporates the areas for protection in each embayment.*
>
> *Finally, we removed the text pursuant to revisiting the statistical analysis and reviewer comments.*

Comment Tracking ID #94

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.

> *Response: Thank you for the comment. We agree with the difficulty in modeling DO and in linking surface chlorophyll to DO, especially benthic DO, in dynamic systems. Please also see response to comment tracking ID #92.*

Comment Tracking ID #95

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.

> *Response: Thank you for the comment. We added more plots in the main text to show the relationships, as requested.*

Comment 2-10c Brush

Comment Tracking ID #96

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?

> *Response: Thank you for your comment. Yes, the key assumptions were tested.*

Comment Tracking ID #97

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.

> *Response: We agree that p-values are largely open to interpretation. We included statistics for each parameter and added more overall model performance metrics, as requested.*

Comment Tracking ID #98

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.

> *Response: Thank you for your response. We clarified text across the document in response to this and similar comments. We added text to Memo F/G reminding readers of the source of the chlorophyll a values and updated that based on continuing analyses.*

Comment 2-10c Janicki

Comment Tracking ID #99

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.

> *Response: Thank you for the comment. We added the requested additional model information in the main text and supporting tables and figures in response to this concern.*

Comment Tracking ID #100

Estimation Method

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).

> *Response: See response to comment tracking ID #99. We added the detail in the memo that RML methods were used.*

Comment Tracking ID #101

Model Selection

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.

> *Response: We removed fixed and random effects with a p-value >0.05 during the model fitting process. After fixed effects were finalized, we evaluated random effects (random embayment intercept and random slope per embayment) using a likelihood ratio test comparing the full model to the nested model without random effects. The model predictions used to calculate R-squared and the blue trend lines in the scatterplots are conditional (based on the random effects). We added text to the memo to clarify these points.*

Comment Tracking ID #102

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.

> *Response: See response to comment tracking ID #99. In response to this comment, we added additional model specification text, tables, and diagnostics that provide the model output information requested.*

Comment Tracking ID #103

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.

> *Response: See response to comment tracking ID #99. The model links are now shown within the equations as requested. Link information has been double checked and confirmed in the text.*

Comment Tracking ID #104

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.

> *Response: See response to comment tracking ID #99. We added the requested parameter estimate tables and diagnostic plots.*

Comment Tracking ID #105

Model Predictions

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.

> *Response: See response to comment tracking ID #99. We added statements about conditional model predictions to the F/G Memo.*

Comment 2-10c Justic

Non-Substantive Comment Tracking ID #NS-20

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.

> *Response: We revisited these models, but as with any similar stressor-response models using field based data, they are likely to continue to have high variability. This variability should factor into the decision making based on tolerance for variability. The intent of this effort is to derive*

lines of evidence based on sound reasoning that link TN targets to desired conditions. Given that, we focused on ranges of TN values derived from the multiple lines of evidence rather than single specific values. We agree with the comment that hierarchical modeling is justified and suitable for this application. We improved the language related to model information and interpretation to allow decision-makers to factor this information as thoroughly as possible into decision-making.

Question 2-10d

***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?**

Comment 2-10d Bierman

Comment Tracking ID #106

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.

Response: Please see responses to comment tracking IDs #43–48 and ID #55 for responses to these referenced comments.

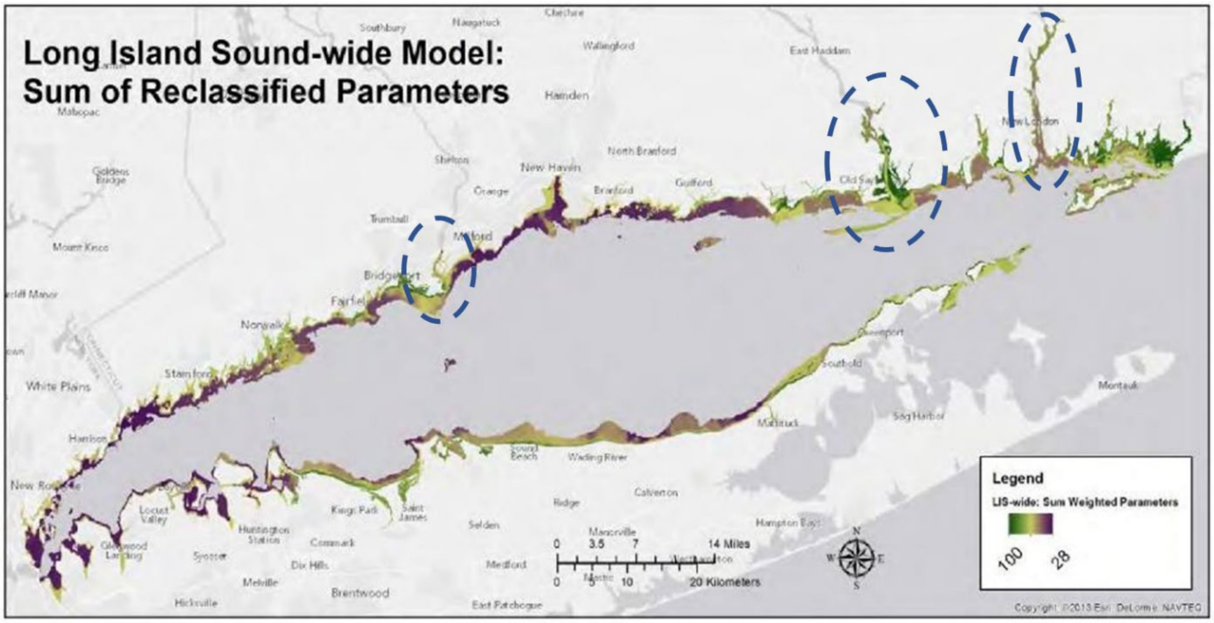
Comment 2-10d Brush

Comment Tracking ID #107

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.

Response: We agree that the thalweg of the Connecticut River (and other two main tributaries) are not like embayments in that their residence times and general physical characteristics differ substantially. However, the Connecticut River ecosystem is composed of adjacent marsh, tidal creek and sub-embayment habitats with substantially higher residence times and characteristics not unlike those of adjacent embayments and those habitats require protection (see Barrett et al. 1997. Moreover, as Figure 22 from Vaudrey et al. 2013 indicates (see below with tributaries highlighted in hatched blue ovals), all three tributary ecosystems contain suitable seagrass habitat. We added language clarifying this in the sections on the Connecticut River.

Barrett, J.M., M. Prisloe L. Gianotti, and N. Barrett. 1997. Distribution and abundance of submerged aquatic vegetation in the lower, tidal Connecticut River. Funded by the Connecticut Department of Environmental Protection Long Island Sounds Research Fund. http://www.lisrc.uconn.edu/DataCatalog/DocumentImages/pdf/Barrett_et_al_1997.pdf Accessed May 2020.



Comment 2-10d Janicki

Comment Tracking ID #108

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.

> *Response: See response to comment tracking ID #107.*

Comment 2-10d Justic

Comment Tracking ID #109

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.

> *Response: See response to comment tracking ID #107.*

Question 2-10e

***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?**

Comment 2-10e Bierman

Comment Tracking ID #110

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.

> *Response: Please see responses to comment tracking IDs #s43–48; IDs #79–80; IDs #88–89; and IDs #92–95.*

Comment 2-10e Brush

Comment Tracking ID #111

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 1st 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.

> *Response: Thank you for your comment. We updated the table footnotes and associated citations in the Subtask G memo to provide additional clarity for the referenced information. We also revisited the Kd analysis pursuant to the expert technical review and adjusted the resulting values based on this re-visitation.*

Comment 2-10e Janicki

Comment Tracking ID #112

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.

Response: We included the model output and diagnostics as requested and in response to comment tracking ID #s 99–105. Please also see comment tracking ID #s 45 and 72 indicating where EPA used TN concentration data from embayments with seagrass recovery.

Comment 2-10e Justic

Comment Tracking ID #113

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.

Response: Please see the response to comment tracking ID #78.

Question 2-10f

***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?**

Comment 2-10f Bierman

Comment Tracking ID #114

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.

Response: Thank you for your comment. We agree that corrected chlorophyll data (removal of pheophytin interference through acidification or narrow band pass filters) are the correct data to use and will continue to use those. We revisited the underlying data, reconfirmed which sources are corrected, and updated the analyses accordingly.

Comment 2-10f Brush

Comment Tracking ID #115

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.

Response: See response to comment tracking ID #114.

Comment 2-10f Janicki

Comment Tracking ID #116

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.

Response: See response to comment tracking ID #114.

Comment 2-10f Justic

Comment Tracking ID #117

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.

Response: See response to comment tracking ID #114.

Question 2-11

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?

Comment 2-11 Bierman

Comment Tracking ID #118

In following guidance in USEPA (2001), the memorandum used 25th percentile values of all samples for LIS embayment waters and open waters to develop distribution-based TN endpoints. It rejected use of the 75th 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 25th 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 25th 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 75th 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.

> *Response: Thank you for your comment and support of this line of evidence. We included information from Massachusetts estuary in the separate, literature-based line of evidence. However, we decided to forego combining surface water quality data from other estuaries in with LIS for the distribution-based line because of differences in estuarine characteristics (hydrology, climate, etc.) and because this had already been incorporated via the literature-based line. Most of the other reviewers agreed with this decision. Please see the response to comment tracking ID #45 and ID #72 for a response to the comment on including information from embayments with existing seagrass information and clarification on the results from those embayments.*

Comment 2-11 Brush

Comment Tracking ID #119

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.

Response: Thank you for your support of the distribution-based approach. Concerning pg F-25: we added text to the F/G memo describing the input dataset. We also reevaluated the seasonality in response to this comment and other comments and reconciled these differences by using a seasonality for Memo E that brackets that used in Memo F/G. Finally, we clarified the caption for Table F-10.

Comment 2-11 Janicki

Comment Tracking ID #120

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 25th 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.

Response: Thank you for your comment. We are developing these TN values for protection of assessment endpoints and not for compliance assessment, but the focus of the Memo G values is on a reported range and not specific TN values. Given the temporal extent of nutrient impacts on LIS and state of impacted seagrasses relative to available water quality data, a temporal reference period was not an option for LIS. However, please see the response to comment tracking ID #45 and ID #72 for comments on referencing information from embayments with existing seagrass information.

Comment 2-11 Justic

Non-Substantive Comment Tracking ID #NS-21

The DbA approach is sound and scientifically defensible. The 25th 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).

Response: Thank you for your comment and support of the distribution-based approach.

Question 2-12

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?

Comment 2-12 Bierman

Comment Tracking ID #121

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."*
>
> *Response: Please see responses to comment tracking ID #48. With regards to macroalgae, we added language to the F/G Memo indicating which assessment endpoints and response measures are being used for each line of evidence and why.*

Comment 2-12 Brush

Comment Tracking ID #122

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 macroalgae (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.

> *Response: Thank you for the comment and the support on the defensibility of the approach. This work is part of a larger effort by many entities to identify defensible targets and to make progress on restoring LIS. With regards to DIN and normalized loading, please see responses to comment tracking ID #67 and ID #69. With regards to endpoints like macroalgae and HABs, we added language to the F/G Memo indicating which assessment endpoints and response measures are being used for each line of evidence and why.*

Comment 2-12 Janicki

Comment Tracking ID #123

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.

Response: It is unclear that the endpoints are similar to current conditions: 21 of 22 of the embayments with water quality data have median TN concentrations above the distribution-based line values, 13 of 22 are above the literature-based line, and 6 out of 13 of first round embayments are at or above the stressor-response lines of evidence. Seeing as the endpoints are presented as a range including all three lines of evidence, it is defensible to say most study embayments are currently above at least 2 of the 3 lines of evidence-based endpoints.

Comment 2-12 Justic

Comment Tracking ID #124

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.

Response: Thank you for the response. Please see the responses to comment tracking ID #71 and ID #54.

B. Public Comments on the April 13, 2018 Draft F/G Memo

This section contains the original comments included in the public comments.

B.1 Technical Comments

Additional Analyses to Consider

Comment Tracking ID #125 (Public Comment 1); Northhampton DPW
Furthermore, this study should include analyses of the relative impact of the distance (river-miles) between the source of nitrogen and the embayment being protected. There may be deeper anoxic zones with the Connecticut River or wetlands that serve to reduce N-loads.

There may also be significant silt and nitrogen loads associated with large storm events that may be creating difficult to measure impacts on seagrass growth.

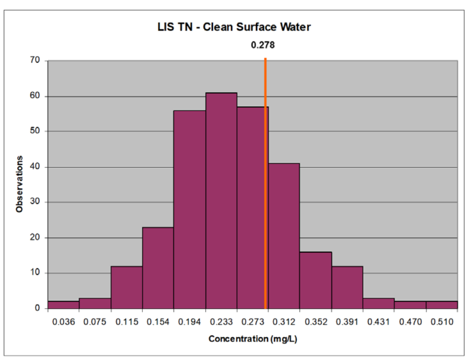
> *Response: The focus of the analysis is to derive protective average concentrations for the selected embayments. The distance to sources and large storm events, which likely contribute to generating the average concentrations, are not relevant to deriving in-site concentrations that relate to in-situ responses.*

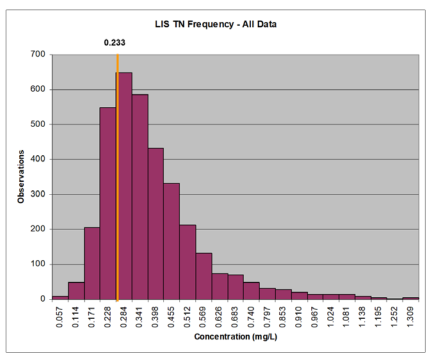
Comment Tracking ID #126 and 127 (Public Comment 2); Footprints in the Water
The analysis demonstrated the best fit to the data for median concentrations, but could not avoid the one size fits all application, which avoids Type I and Type II error. CT DEEP has struggled with this weakness of an empirical statistical approach, with all the underlying uncertainties and natural variability that may force an incorrect trajectory of change for a specific waterbody. Subtask G analysis shows the risks with the “best fit” numbers as targets whether derived from literature (i.e., about 0.4 mg TN/L) or distribution-based using LIS data (0.24 – 0.27 mg TN/L). The consistency among the literature findings presents a good argument for the literature values although the individual estuary analyses in Task G suggest the importance of site-specific evaluation and targets as well as the need for adequate, local monitoring data. The best fit results should not be broadly applied and assumed to be useful for the variety of sites and conditions found in LIS. However, CT DEEP looked at phosphorus in upland, freshwater systems and found land cover/land use (LC/LU) attributes to present a more viable and accurate way to assess current condition, stressor levels and potential for management expressed as a Best Attainable Condition (BAC). This could be applied to coastal embayments with the presumption that the external stress from land degradation is a dominant driver of stress that reflects nutrients and a full suite of external stressors as an independent “dose” variable. This was developed into a narrative nutrient criterion translator for phosphorus as described in the attached Becker and Dunbar white paper.

I recommend that relationships between desired endpoints and LC/LU be explored more fully in the strategy and considered for devising endpoints that reflect combined stress and enrichment factors that might greatly reduce uncertainty and are simple to apply on a site-specific basis with the benefit of potentially higher performance at smaller-scale applications that avoids the “one-size fits-all” dilemma of current protocols. Adequate land cover data are available, and provide a ready historical reference condition by using a high level of forestation to calculate nutrient loads as an individual pollutant stressor, or as an index of combined stress as has been effectively done with impervious cover models.

This avoids errors of independent control variables since land is external to the aquatic system, and uncertain translations of nutrients to chlorophyll-a that have occurred on the individual watershed/embayment analyses. An enrichment factor of 1 represents the reference state and in the case of phosphorus in the attached, was calibrated to a threshold of about 2 (double the natural load) as a changepoint. Anthropogenic enrichment in the range of 2-9 was related to a Tier 3 level using the biocondition gradient (BCG) tiered approach. Recovery potential could be assessed for TN as limit of technology for point source discharges and land available for recovery or rehabilitation to a forested state to assess a BAC. This approach provides a more actionable outcome, and a pragmatic approach based on normalizing to site-specific factors such as TN load, and predicting recovery potential in a tiered approach related to biocondition. It avoids natural and other uncertainties that are difficult to quantify with the Task F and G methods and sets site-specific targets that are not only relevant, but have an actionable path forward and can be assessed for attainability. It also sets the stage for an effective trading program based on Natural Capital and Ecosystem Service outcomes that is more transferrable throughout the entire watershed than the complexities on Nitrogen loading and transformations to delivery, and provides local benefits to meet local water quality goals as well.

I do note that the distribution-based approach as applied in the Task F/G report was limited to a 25% level of all data because no sites were deemed to be reference sites. Using the DEEP approach, and a watershed stress index that can be hindcast to natural condition based on land cover, truer reference and current assessment datasets can be devised. Or, as I did many years ago, the Eastern Sound data can be used to provide a clean distribution as is shown in the first graph below, where a 75% concentration is at 0.278 mg TN/L, and a total, current distribution of data throughout the sound as shown in the second graph below where a 25% concentration is at 0.233 mg TN/L. This is remarkably similar to the Task F/G report numbers of 0.27 and 0.24 mg TN/L for embayments and open waters, respectively, suggesting the distribution approach may provide a reasonable concentration target for LIS systems and subsystems. However, there are still the translations to loading and site-specific needs for embayments that would need to be overcome that are avoided with the CT DEEP approach.





Response: Thank you for your comment and support of the distribution-based approach. The focus of this work was on defining TN concentrations that are protective of uses and the stressor-response does pair the generic values of the distribution based and literature-based evidence to site specific factors. The LU/LC approach being advocated is not really appropriate for this application.

Comment Tracking ID #128 (Public Comment 3); CTDEEP

DEEP would like EPA to consider developing alternate biological indicators (i.e. other than eelgrass) for estuaries in Long Island Sound. DEEP has recently proposed the *Development of Sampling Methodology and Tools to Assess Embayment Health in Long Island Sound* that focuses on benthic macroinvertebrates. The objectives of the project would be to 1) develop a sampling methodology to collect benthic macroinvertebrates from embayments in Long Island Sound, 2) collect benthic macroinvertebrate data and water quality parameters from a minimum of 30 embayments, and 3) conduct data analysis to evaluate existing multimeric biological indices such as AZTI Marine Biotic Index (Borja et al 2000 [footnote 2: Borja, A., Franco, J. Perez, V. 2000. A marine biotic index to establish the ecological quality of soft-bottom benthos within European estuarine and coastal environments. Mar. Poll. Bull.40, 1100-1114.]) or develop a new index for states in Long Island Sound watershed to use in their assessment programs to aid in embayment health assessments to meet Clean Water Act goals.

Benthic macroinvertebrates are the preferred biological indicator for Long Island Sound estuaries because 1) they provide a direct measure of the designated use 2) integrate ambient environmental conditions and are known to respond to environmental gradients, 3) they have widespread distribution and expected to occur in aII embayments in Long Island Sound, and 4) other states in the geographic area such as New Jersey, and states in the Chesapeake Bay Program have been successful in the development of invertebrate indices to measure embayment health.

> *Response: We would be interested in exploring a benthic index as a response measure, but consistent and available macroinvertebrate data across the entire sound was unavailable for this analysis. The Chesapeake Bay Program and Gulf of Mexico both have bay-wide invertebrate indices and those have been used in similar contexts, so we imagine one could be used in LIS as well.*

Concern with TN Endpoint Value

Comment Tracking ID #129 (Public Comment 4); Northhampton DPW

This document attempts to model LIS embayment health (as seagrass growth) in response to Total Nitrogen (TN) loading. The relationships presented are tenuous. The focus is on dissolved oxygen (DO) and light penetration as measures of predicting embayment health (in form of Seagrass health).

The literature review indicates that TN levels up to 0.50 mg/L do not seem to impact DO (with reported corresponding DO >5 mg/L). Therefore, starting this analysis with the presumption that reducing TN levels of 0.40 mg/L is required appears excessively conservative and not supported by the science. This presumption may also have significant cost impacts, reducing the funds available for addressing more significant environmental issues.

Obviously, greater light penetration will result in greater seagrass growth in the form of seed germination or rhizome growth (the most probable means of generating seagrass growth in deeper waters). The report attempts to develop a tenuous relationship between TN and light penetration using Chlorophyll-a as a bridging factor. Obviously, there are multiple factors involved that may impact light penetration and chlorophyll-a concentrations: season, time of day (day/night), weather, time of year, existing plant growth, other shading factors, sediment entrainment, stratification of the water column, water temperature, salinity, pH, etc. There may also be other factors that impact seagrass growth independent of light penetration; e.g. pollutants, mechanical impacts (tidal influences and actions from storms or other sources), other flora/fauna, or disease. As such, significantly greater data collection and analyses are required to establish a relationship of TN in the embayment to seagrass growth.

Response: Thank you for your support of the light endpoints. With regards to oxygen, please note that the literature review also indicates that moderate impairment also beings at TN of 0.50 mg/L and more protection for oxygen is assured as TN declines. Moreover, oxygen is but one endpoint we considered. We had to use existing data for the model and revisited the hierarchical model in a subsequent analysis to improve the model with additional predictors. Please note that chlorophyll a is well established as limiting Kd, as noted in the literature review and has been targeted for use in restoring light in several cited examples. As excess algal growth (measured as chlorophyll a) is a factor known to limit light, and therefore seagrass growth, it will have to be mitigated, potentially along with other factors, to restore seagrasses. We recognize progress on other limiting factors will likely also be necessary and encourage broad participation in reducing those stressors as well. Please see response to comment tracking ID #48 concerning other factors accounted for in suitability index utilization.

Comment Tracking ID #130 (Public Comment 5); Springfield Water and Sewer Commission

TN Target: The proposed TN target of 0.40 mg/Lis less than half of the USGS estimate of national background nutrient concentrations for streams (1.0 mg/L). [footnote 4: https://pubs.usgs.gov/circ/circ1225/pdf/nutrients.pdf] Such a target is unrealistic and unnecessary as it relates to human contributions of nitrogen through point and non-point sources.

Response: We believe this is an incorrect interpretation of the cited USGS report. This same USGS report says (p. 34) "Background nutrient concentrations can vary considerably from region to region, or even within watersheds, because of differences in hydrology and in naturally occurring nutrient levels in soils, rocks, and the atmosphere. The data analyzed for this report are insufficient to define background nutrient concentrations on a regional basis. Thus, all available data from undeveloped areas were combined to derive national background concentrations. The national background concentrations are higher than most concentrations measured in relatively undeveloped areas across the Nation and may not be applicable for use in regional or local analyses." The same report, on page 53, has a callout box discussing continuing nutrient problems in the lower Long Island Sound watershed, including increasing nitrate trends. Other USGS NAWQA studies estimate background concentration as substantially lower, (e.g., 0.58 mg/L in Dubrovsky et al. 2010, The quality of our Nation's waters—Nutrients in the Nation's streams and groundwater, 1992–2004: U.S. Geological Survey Circular 1350, 174 p. http://water.usgs.gov/nawqa/nutrients/pubs/circ1350). Lastly, the NAWQA studies including the cited report targeted streams and not estuarine systems.

Comment Tracking ID #131 (Public Comment 6); CTDEEP

Table F-10: How does 0.27 mg/L of Total Nitrogen compare to concentrations in the study embayments and presence/absence of eelgrass (where applicable)? How does 0.24 mg/L of Total Nitrogen compare to concentrations of TN offshore of the embayments in LIS?

Response: Please refer to the CDF of TN value in embayments and open water for a comparison of TN concentrations for offshore areas. Please also see response to comment tracking IDs #45 and #72 with regards to concentrations paired with seagrass data.

Connecticut River

Comment Tracking ID #132 (Public Comment 7); Springfield Water and Sewer Commission
Independent evaluation for the Connecticut River: The Report states:

"Thirteen embayments and one riverine system (Connecticut River) were identified by EPA as watersheds to focus on. Data for the Connecticut River area of influence were modeled along with the embayment data, rather than modeled separately, due to the sparsity of paired data for the Connecticut River and the fact that the Connecticut River estuary is essentially an embayment" (F-7).

While it might be reasonable to treat the Connecticut River like a single embayment, there is no reason to expect the metrics for this system to behave like all the other individual embayments. The Connecticut River, the largest fresh body discharging to the LIS, should be evaluated independently.

> *Response: Please see response to comment tracking ID #107.*

Comment Tracking ID #133 (Public Comment 8); CTDEEP
A major comment is that the Connecticut River is treated as an embayment in EPA's nitrogen reduction strategy and only the lower portion of the river is considered relative to endpoints. We reiterate that open water and large rivers are distinct and unique water bodies, separate from embayments and as such, the assessment and measurement endpoints for these areas should not be assumed to be equal to those developed for embayments. An independent and equal evaluation of open waters and large rivers is important to establish appropriate water quality and management targets for these areas, providing support for necessary regulatory actions affecting all sources within the larger Long Island Sound watershed (including upstream states) which contribute nutrients to open waters and large rivers.

> *Response: See response to comment tracking ID #107.*

Current Conditions

Comment Tracking ID #134 (Public Comment 9); Springfield Water and Sewer Commission
Further, according to the summer 2017 sampling data, the Connecticut River is currently delivering TN in concentrations similar to national background levels.

> *Response: See response to comment tracking ID #130.*

Comment Tracking ID #135 (Public Comment 10); Springfield Water and Sewer Commission
Connecticut River at/near natural background levels: The water quality results from the Connecticut River embayment sampling from 2017 are remarkable. Nitrate is well under natural background levels for streams (max observed= 0.36 mg/L). Ammonia is non-detect. TN is therefore also less than typical natural background levels (max= 0.61 mg/L). Total phosphorus and Ortho Phosphorus are also near detection levels and very low (max observed = 0.056 and 0.037 mg/L, respectively). There is little quality gradient from sampling location CTR01 to sampling location CTR07. Whatever nutrient loads are delivered to the Connecticut River, they are removed to background levels by the time the river reaches the estuary.

It is unclear what additional nitrogen reductions are expected to be achieved, above those already required by the LIS TMDL.

> *Response: The purpose of this effort to identify concentrations that are protective of seagrass and other aquatic life, not to determine reductions. Regarding your comment about background concentrations, please see the response to comment tracking ID #130.*

Data Concerns

Comment Tracking ID #136 (Public Comment 11); NACWA

Concerns with EPA's Reliance on Old Data and Flawed Analyses: NACWA remains concerned about the reliance on old data and information to develop the nitrogen endpoints in the Memorandum. In addition, while EPA presents the results of multiple analyses, the approaches EPA's contractor employs have all faced criticism in the past:

Use of literature values is inadequate because it fails to take into account the many other variables that differ from one waterbody to the next. Without factoring in these other variables, literature values alone cannot demonstrate an actual cause and effect link between nitrogen loads and impacts on the endpoints for the waterbody in question.

> *Response: At the time the analysis was conducted, we used all available data that met quality assurance protocols, specifically those data with an approved Quality Assurance Project Plan (QAPP). There was no specific evidence provided as to what constitutes old data; data for this study was collected between 2000 and 2016. Additionally, we are using a multiple lines of evidence approach that is approved by EPA, the Science Advisory Board, and the independent expert technical reviewers who reviewed the draft analysis. The literature review approach was not intended to be site-specific and it is part of a multiple lines of evidence approach including site-specific stressor-response modeling and distribution-based approach. Moreover, the logic behind the statement that other studies do not take into account the variables that vary from one waterbody to the next is unclear. Many of the studies referenced did indeed account for the effects of variables that influence nutrient responses in estuaries. Lastly, cause and effect impacts of nutrients in estuaries, including LIS, have been clearly demonstrated experimentally and are not the focus of this effort.*

Comment Tracking ID #137 (Public Comment 12); Springfield Water and Sewer Commission

TN endpoints are below national background concentrations: The Report utilizes a "multiple lines of evidence" approach to propose TN endpoints for embayments to protect eelgrass and other aquatic life. Chlorophyll-a, clarity, and dissolved oxygen were used as surrogates for eelgrass and aquatic life.

The report provides inadequate and inconsistent embayment specific data regarding each of the three lines of evidence to support the TN endpoint results, especially with regard to the Connecticut River.

The scientific literature analysis provided, catalogues assessment endpoints used for Massachusetts embayments and estuaries. One suggests that "severe ecological degradation" begins at 0.80 mg/L TN. Another says that 0.91 mg/L TN provides "benthic habitat protection." These are values used by others for specific purposes, with no assessment of cause and effect. All of them arrive at TN concentrations that are less than the national background nutrient concentration (i.e., unaffected by human activities) of 1.0 mg/L TN for streams and rivers[2]. [footnote 2: https://pubs.usgs.gov/circ/circ1225/pdf/nutrients.pdf]

Response: See response to comment tracking ID #130.

Comment Tracking ID #138 (Public Comment 13); Springfield Water and Sewer Commission
Of major concern, is that data from the Connecticut River was not included in the analysis, and limited data from summer 2017 does not align with model expectations. Given the questionable results shown in Subtask F, the application of stressor-response modeling provided in Appendix G needs to be critically examined to determine its worth. Given the importance of such effort and critical impacts to communities tributary to the Connecticut River, the largest fresh waterbody discharging to LIS, it is prudent to gather adequate water quality data from the Connecticut River Embayment before applying stressor-response analysis.

Response: See response to comment tracking ID #107.

Comment Tracking ID #139 (Public Comment 14); Springfield Water and Sewer Commission
Lack of assessment: The Report did not provide eelgrass measurements and assessments performed during the embayment monitoring, thereby making health comparisons challenging.

Response: The report provided citations to the Tiner et al. 2013 and Vaudrey et al. 2013 reports, which were the available information on eelgrass measurements and assessments and those have been incorporated into the distribution-based analysis.

Comment Tracking ID #140 (Public Comment 15); Springfield Water and Sewer Commission
Lack of Connecticut River Data: On Page G-30 of the Report, EPA states with respect to the Connecticut River: *"No paired data was available for the embayment within the growing season (April-September). Therefore, the global fit using data from all 1,335 embayment observations LIS-wide was used for stressor-response analysis."*

The lack of data for the Connecticut River brings into question the use of this methodology and its applicability to Connecticut River embayments.

Response: See response to comment tracking ID #107.

Comment Tracking ID #141 (Public Comment 16); NHDES and VTDEC
On December 15, 2017, NHDES, VTDEC and MADEP submitted comments (see email below) on the November 17, 2017 draft of Tasks F and G. One of our concerns is that the proposed TN threshold for the Connecticut River embayment did not account for any paired data collected in the embayment but that we understood that EPA was embarking on a two-year study (beginning in late-summer 2017) to collect more data in the Connecticut River estuary.

On April 13, 2018, EPA issued a response to our comments [Appendix F1: Response to Technical Comments on Subtasks F and G. Summary of Empirical Modeling and Thresholds (November 15, 2017)]. On page F1-65 of the response to comment document, EPA stated that the “…data gathered during the 2017 survey was used to refine the Subtasks F and G memo” and that the “…data from the 2018 survey may be used at a later date.” However, on page G-30 in section G.15 Connecticut River, CT of the April 13, 2018 draft of Tasks F and G, it states that “No paired data was available for the embayment within the growing seasons (April – September) and Figure G-29 on page G-32 shows this graphically (see below). An explanation as to how the 2017 data was used and when the 2018 data will be included in the analysis to set end points would be appreciated.

> *Response: Please note that the analysis did incorporate the 2017 and 2018 Connecticut River data as well as updated data for the Housatonic. The statement "no paired data was available for the embayment within the growing seasons (April – September)…" refers to the fact that the stressor response modeling only used data where there were paired observations of interest (e.g., chlorophyll a and Kd or chlorophyll a and dissolved oxygen). At the time, existing method information suggested the Connecticut River chlorophyll a data was not consistent with the other program methods, so it could not be paired to nutrient measures for the stressor response modeling. We have updated the models with the additional paired Connecticut River data. For additional information, please also see the response to comment tracking ID #107.*

Comment Tracking ID #142 (Public Comment 17); CTDEEP

Please make reference to the time frame that data used to establish nitrogen endpoints is from.

> *Response: We added text identifying the years associated with each dataset.*

Comment Tracking ID #143 (Public Comment 18); CTDEEP

Page F-20, second paragraph: "Paired data for Eastern LIS was limited (n=31; 2.5% of the sample) and was not included in this analysis." Does Tetra Tech mean the eastern narrows or the eastern region of LIS which is typically referred to as Eastern LIS? Was this data excluded because it was not labeled "corrected" chlorophyll? Depending on the peer review outcome, this section may need to be revisited.

> *Response: In the F/G memo, "Eastern LIS" refers to the portion of LIS that is east of the "East Narrows." The limited amount of Eastern LIS data was based on paired data using corrected chlorophyll. We have resolved issues involving pheophytin free chlorophyll a data and have updated the data and models accordingly.*

Comment Tracking ID #144 (Public Comment 19); CTDEEP

Page F-22, first paragraph: "The lack of paired bottom DO samples with chlorophyll data was a limitation. There was plenty of bottom DO data, as evidenced by the hypoxia maps drawn for LIS, but this analysis was unable to find adequate paired bottom DO with chlorophyll samples to build this relationship." Again, depending on the peer review outcome, this section may need to be revisited.

> *Response: We used all available data. Please note that paired data was not as abundant as the reviewer suggests. Although there were DO and chlorophyll a data, such data had to be available at the same date and location to be paired. Additionally, expert technical reviewers agreed that pairing surface chlorophyll a and bottom DO was not reasonable.*

Comment Tracking ID #145 (Public Comment 20); Battelle

Section F would benefit from additional clarity on empirical modeling inputs, i.e. raw data selection criteria such as seasonality and depths and time of day, outlier filtering etc. Consider adding more content akin to the six sentences on page F20.

> *Response: We added text describing the input dataset criteria (depth and growing season) to each stressor response model. We identified any extreme values (outliers) that were removed*

and that is described in Memo D. We also added language better explaining the models, as well as model performance.

Data Request

Comment Tracking ID #146 (Public Comment 21); CTDEEP
In addition to the comments included in Appendix 1, CT DEEP would like to request a copy of the embayment dataset assembled in Subtask D. As you know, we have initiated special studies for our priority embayments and access to this data would expedite the initial information gathering task. Information regarding CT DEEP' s embayment approach is included as Appendix 2.

Response: EPA will make this data available on the LISS website.

Defensibility of Relationship between TN and Other Variables

Comment Tracking ID #147 (Public Comment 22); Springfield Water and Sewer Commission
Comments on Lack of Regulatory Basis to Use Eelgrass to Determine TN Endpoints:
The Report provides a discussion related to determining total nitrogen endpoints for LIS estuaries and embayments. Certain TN endpoints are associated with a healthy population of eelgrass, and a healthy population of eelgrass is equated to protecting designated uses in the watershed. In conjunction with the Report, EPA provides on their website, information stating that although it is premature to determine if the endpoints developed in the Report are appropriate for use in setting effluent limitations at treatment plants, treatment facilities should be prepared for effluent limitations.

The Report does not provide a meaningful basis to support the use of eelgrass as a determining factor for meeting designated uses. That being said, and more importantly, the Report has not shown a meaningful relationship between TN endpoints that support healthy eelgrass and TN restrictions on point source discharges. Another concern is that the TN endpoints developed in the Report are less than TN background levels (1.0 mg/L[2]) [footnote 2: https://pubs.usgs.gov/circ/circ1225/pdf/nutrients.pdf] for U.S. streams and rivers, as estimated by the USGS.

Response: Please refer to the Literature Review Memo, which is available on EPA's Nitrogen Strategy website and has been updated based on public and expert technical review comments of the Draft F/G Memo. In the Literature Review Memo, EPA details why eelgrass is a defensible endpoint. Additionally, the expert technical reviewers echo the defensibility of eelgrass in their comments. Concerning your comment about TN background levels, please see the response to comment tracking ID #130.

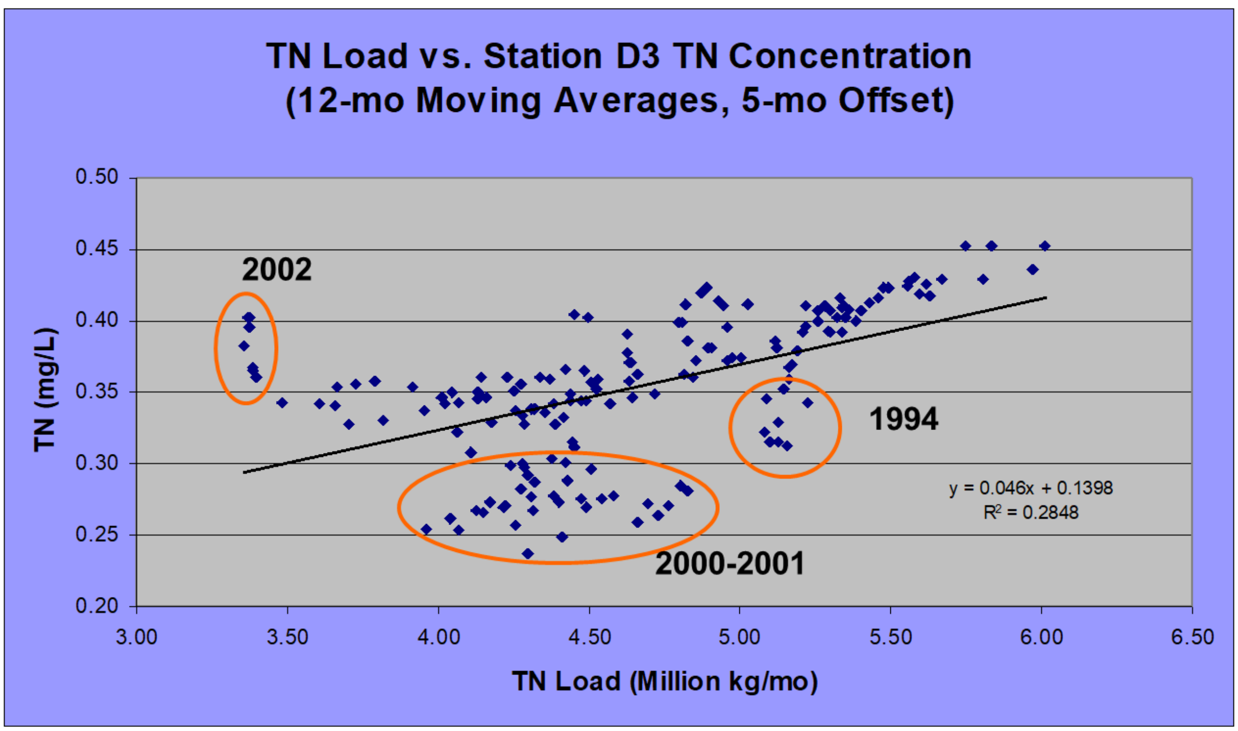
Comment Tracking ID #148 (Public Comment 23); Springfield Water and Sewer Commission
Specific Comments
Regulatory basis to apply TN endpoints developed for eelgrass, to point source discharges:
Subtask F/G memo states: "EPA selected seagrasses and other aquatic life for developing nitrogen endpoints." However, there is no technical or scientific basis presented, that shows why seagrasses are appropriate to develop TN endpoints, nor is there any discussion provided that shows a relationship between limiting TN at point sources and the health of seagrasses (specifically eelgrass). This linkage is critical in order to develop a sound regulatory basis for the imposition of TN limits at points sources, developed from the TN endpoints in the Report.

Response: See response to comment tracking ID #147 concerning your comment about the technical or scientific basis for using seagrasses. Please also see the response to comment tracking ID # 135 concerning your comment related to TN reductions, which were not the focus of this effort.

Comment Tracking ID #149 (Public Comment 24); Footprints in the Water

The presumption that TN concentration, a state variable and the strategic endpoint for this work, is representative of TN loads, the actual forcing factor, has not been tested with the data. This calls into question the utility of using TN as the independent stressor, as noted above. An earlier descriptive analysis of the relationship between TN loading and concentration revealed a complicated relationship that only seemed quantitatively related at Station D3 and only if the data were manipulated to reflect a 12-month moving average and TN concentrations were offset from loading by a few months. In the chart below, a 5-month offset provided the best correlation. Specific years seem to fall off the trend, e.g., 2002 was a very dry year, but 2000-2001 and 1994 are less easily explained and may be forced by other factors or sampling or analytical error. I can't find my plot of TN load and concentration vs. Chlorophyll-a, but the Task F/G analysis provides more detailed relationships that are probably superior representations of the relationships, but with the same sources of error likely to come into play as described above.

Given these concerns, I expect the necessary translation between TN concentration and TN loading will have to be something other than a dilutional relationship. Management planning cannot be done without that translation, which is very similar to the numeric translation of narrative nutrient criteria into numerical criteria. Unfortunately, it adds another layer of uncertainty to the process, and is likely to



underestimate the TN load because TN concentration is "processed" nitrogen, subject to attenuation and biological processing, including denitrification, that lower concentrations at varying rates depending on local conditions and the suite of forcing factors that are not considered in a single-pollutant empirical statistical analysis. However, with a well-calibrated and verified mechanistic model, the relationship between true forcing factors and the TN outcome indicator can be strengthened to be sure management is aimed in the right direction, and the chances of Type I and Type II error are reduced, Type II error probably representing the greater risk. A risk-analysis should be part of the strategy.

> *Response: Please refer to comment tracking ID #135 for a response related to the sequencing of this effort, which focuses on identifying protective concentrations first. This effort was not concerned with translating these to loads or reductions. Please also refer to comment tracking ID #67 on loads.*

Distribution-based Approach

Comment Tracking ID #150 (Public Comment 25); NACWA

Concerns with EPA's Reliance on Old Data and Flawed Analyses: NACWA remains concerned about the reliance on old data and information to develop the nitrogen endpoints in the Memorandum. In addition, while EPA presents the results of multiple analyses, the approaches EPA's contractor employs have all faced criticism in the past:

The distribution-based approach (aka, the reference-based approach) has been roundly criticized since EPA first attempted to use it when establishing national criteria recommendations. Few states have opted to use this approach given its limitations in the nutrient context. The usefulness of distributions of data – in this case very old data – is extremely limited when trying to describe the complex interactions involved with nutrients.

> *Response: Please see response to comment tracking ID #136 regarding the age of the data. EPA is using a multiple lines of evidence approach, which is recommended by and has been approved by the Science Advisory Board, and the expert technical reviewers of the drafts of Memos E and F/G. The distribution-based approach is one line of evidence and it is not intended to be site-specific. Additionally, in EPA's experience, nearly all states use the distribution-based approach as a line of evidence.*

Comment Tracking ID #151 (Public Comment 26); Springfield Water and Sewer Commission

The distribution-based approach is questionable as well, since cause and effect is not considered in the evaluation. Since over 90% of embayment samples exhibit TN concentrations less than 1.0 mg/L TN, this indicates that the LIS TMDL has been highly effective in lowering TN contributions from inlet watersheds.

> *Response: Please refer to comment tracking ID #150 for a response concerning the distribution-based approach. Also refer to comment tracking ID #130 for a response concerning background concentrations.*

Comment Tracking ID #152 (Public Comment 27); Battelle

Seems incomplete to derive 30% of the N endpoint with a distribution-based method that is innately quantitative yet does not include an uncertainty estimate. “No uncertainty estimates around these values were calculated.” pageG1

Response: No single value was selected for the analysis, so this is not 30% of value. Moreover, uncertainty is implicit in the selection of the percentile.

Editorial Comment

Comment Tracking ID #153 (Public Comment 28); CTDEEP
Page F-7: The term paired data is mentioned in the second paragraph. It should be defined for readers that are not familiar with the Subtask D memorandum.

Response: We added clarifying text.

Comment Tracking ID #154 (Public Comment 29); CTDEEP
Page F-13: Tetra Tech used "corrected" chlorophyll because the number of observations of "corrected" chlorophyll is larger than "uncorrected" chlorophyll. The CT DEEP LIS monitoring dataset contains paired chlorophyll data and we understand that this data was excluded from the analysis because it was not specified as "corrected" chlorophyll. We also understand that this issue will be considered by the peer review panel. Depending on the peer review outcome, references to limited data relative to "corrected" chlorophyll throughout the report may need to be revisited.

Response: Please refer to the response for tracking comment ID #114.

Comment Tracking ID #155 (Public Comment 30); CTDEEP
Page F-25, second paragraph: " ... use of the 75th percentile of reference waters seemed indefensible given that such waters would be difficult if not impossible to accurately identify or represent." Suggestion to reword this sentence to reflect that the 25th percentile was selected based on ...

Response: We changed the sentence regarding "75th percentile values as indefensible" to why the 25th percentile was selected (instead of why it was not).

Comment Tracking ID #156 (Public Comment 31); CTDEEP
Figure F-12: Include a table to identify the embayments, the number of observations and the number of sample locations that were included in the distribution-based analysis (ex. Table F-7).

Response: We provided reference to these data as requested.

Comment Tracking ID #157 (Public Comment 32); CTDEEP
Subtask G. If the gray line "LIS Population Fit" is synonymous with the "Global Fit" and "embayment observations LIS-wide", please choose one common term to identify this model throughout Subtask G.

Response: We changed the term "global" to "population" for consistency.

Comment Tracking ID #158 (Public Comment 33); CTDEEP
Subtask G. Figure G-28, Connecticut River: Please outline the Connecticut River land drainage area applicable to this methodology.

Response: The inset map in this figure outlines the drainage area.

(Comment Tracking ID #159 (Public Comment 34); New York State Department of Environmental Conservation)

The disclaimer in the beginning of the Technical Memo clearly stated that the purpose of the document is a pure scientific research paper. It provides up to three lines of evidence (scientific approaches) that states could use as references to derive numeric water quality criteria for nitrogen in marine systems. However, the Response to Comments document uses language in multiple places that imply the nitrogen endpoints in the Technical Memo can be used to achieve state's water quality standards. These statements in the Response to Comments appear to contradict the intent of the Technical Memo and is akin to interpreting state's narrative water quality standards. DEC strongly recommends the language be changed to clarify that "these nitrogen endpoints are lines of evidence that could be consulted with when developing water quality goals to restore or protect the designated use of the waterbody." The following are locations where DEC staff has found the questionable language.

1. Page F1-2, the third paragraph;
2. Page F1-34, the first paragraph of the Response to Comment 2;
3. Page F1-35, second paragraph of the Response to Comment 3;
4. Page F1-38, first paragraph last sentence; and
5. Page F 1-61, first paragraph.

Response: Thank you for your comment. The Response to Comments Document referenced in this question responds to comments made on earlier draft memos for Tasks E and F/G. Updates have been made to the current version of the memos to address this comment. EPA has retained the disclaimer that indicates the limits and appropriate use of the document.

General Comment

Non-Substantive Comment Tracking ID #NS-22 (Public Comment 35); Footprints in the Water

I have a few other comments, mostly of a more minor nature and points of clarification, but have run out for time for now. I also did not have time to edit this, so I apologize for grammatical errors and typos, but please contact me if you have questions or need clarifications. As a poor substitute for other, more general comments, I have attached a comment letter from April 2010 submitted to EPA in response to the proposed Florida Rule on nutrient criteria that provides an overview of concerns that may be associated with setting numeric endpoints for nitrogen in Long Island Sound.

I thank you for all your great work, and hope that we can discuss some of these more direct, and simpler options for setting and attaining not just TN goals, but ecosystem health goals for Long Island Sound and its watershed. Now that I am back in CT and engaged in the LISS STAC, I hope to provide more support for these efforts, and attend appropriate meetings to join in the discussion.

Response: Thank you for your comment.

Non-Substantive Comment Tracking ID #NS-23 (Public Comment 36); CTDEEP

Connecticut Department of Energy and Environmental Protection (CT DEEP) is pleased to have this opportunity to submit comments on the above referenced technical draft deliverable prepared by Tetra Tech on behalf of the Environmental Protection Agency (EPA) and dated April 13, 2018.

In January 2018, CT DEEP submitted comments through the Technical Advisory Group which included recommendations to strengthen the technical approach presented in Subtask F. We appreciate EPA' s consideration of our comments and note the minor modifications made to Subtask F as a result. Attached as Appendix 1 is a list of comments and questions specific to the public draft Subtasks F & G document.

Subtask F explores a methodology for developing nitrogen endpoints for embayments, Long Island Sound open water, as well as the Connecticut River (a large riverine system). While we have some comments and concerns, the methods and nitrogen endpoints explored by EPA and Tetra Tech provide useful information for CT DEEP to consider as part of our Second Generation Nitrogen Strategy.

Response: Thank you for your comment.

Information on State Efforts

Comment Tracking ID #160 (Public Comment 37); New York State Department of Environmental Conservation

DEC would like to reiterate the importance of aligning the total nitrogen endpoint research with the locally led LINAP Subwatersheds Wastewater Plan (SWP) process. As you may know, the locally led Suffolk County SWP is making meaningful progress and is anticipating the completion of the plan in 2019 that will include nitrogen reduction goals for all the embayment's around Suffolk County. The process that the SWP has used for waterbody specific nitrogen reduction goals is in some ways similar to some of the methodologies used by the EPA. Overall the SWP process for establishing ecological endpoints and using them to develop watershed specific nitrogen reduction goals:

- Focuses on a number of ecological endpoints (Chi-a, water clarity, HABs and dissolved oxygen) that all contribute to improved water quality;
- Uses local reference waterbodies to establish nitrogen loading rate for achieving all the ecological endpoints;
- Uses all local water quality data;
- Recognizes that eelgrass is not the only desired ecological endpoint;
- Establishes mass loading rate driven load reduction goals in lieu of in-water total nitrogen concentration targets; and
- Takes into consideration the critical role individual waterbody hydrodynamics plays in a waterbodies response to a nitrogen load;

Response: Thank you for your comment. EPA regularly communicates with NYSDEC and Suffolk County on development of the SWPs and welcomes more detailed site-specific assessments. An outcome of that communication was agreement to compare the calculated concentrations under the proposed load reductions for Suffolk County with the multiple lines of evidence used in this study. Tetra Tech found the literature review endpoints found to be protective of seagrasses (0.40 mg/L), as well as to the value from EPA's 25th percentile approach (0.27 mg/L), to be comparable to the Suffolk County results. They are also similar to several stressor-response based endpoints from the stressor-response models using the population models reported in the F/G Memo, which ranged from 0.4 to 0.57 mg/L and used primarily surface water TN and chlorophyll a concentration, solving for different chlorophyll targets based on meeting seagrass light needs. As a result, EPA believes that the two independent methods corroborate each other's *results. EPA encourages more site-specific assessments and the application of those results to develop watershed specific nitrogen reduction goals.*

Limitations of the Analysis

Non-Substantive Comment Tracking ID #NS-24 (Public Comment 38); Footprints in the Water
The report provides a competent statistical analysis of empirical data to establish nitrogen endpoints for embayments, large riverine systems and Western Long Island Sound open water using three lines of evidence: literature review, stressor-response analysis and a distribution-based approach. Limitations of the analysis are largely a result of the spatial and temporal extent and density of available data and the almost exclusive use of state variables to presume forcing factors by correlation since mechanistic modeling was beyond the scope of the work. However, the data were compared to state variable model output using the SWEM model and embayment mixing models and watershed loading models to try to piece together and verify relationships between state variables and verify that state variables used as indicators of stress, especially concentrations of total nitrogen (TN) were representative of actual stress from external loading of nitrogen.

Response: Thank you for your comment and support of the analysis.

Stressor-response Approach

Comment Tracking ID #161 (Public Comment 39); NACWA
Concerns with EPA's Reliance on Old Data and Flawed Analyses: NACWA remains concerned about the reliance on old data and information to develop the nitrogen endpoints in the Memorandum. In addition, while EPA presents the results of multiple analyses, the approaches EPA's contractor employs have all faced criticism in the past:

- The empirical stressor-response analysis was criticized by EPA's own Science Advisory Board in 2010, when the Board noted that "considerable unexplained variation can be encountered when attempting to use the empirical stressor-response approach to develop nutrient criteria" and that "statistical associations may not be biologically relevant and do not prove cause and effect" (*Report to EPA Administrator Lisa Jackson, SAB Review of Empirical Approaches for Nutrient Criteria Derivation, April 27, 2010*).

  Response: See response to Tracking ID #136 with regard to the age of the data, lines of evidence, and cause and effect. We are using a multiple lines of evidence approach, which is approved by EPA, the Science Advisory Board, and the expert technical reviewers of the drafts of Memos E and F/G. The stressor-response approach is one line of evidence being used to derive the range of values for each waterbody.

Comment Tracking ID #162 (Public Comment 40); Springfield Water and Sewer Commission
The majority of the Report discusses how endpoints to protect eelgrass were developed. To summarize, EPA modeled three different relationships (called stressor-response relationships), and used the results of those relationships, to develop TN endpoints.

The three modeled stressor-response relationships are as follows:

a) Kd vs chlorophyll was modeled to determine what level of chlorophyll is associated with what levels of Kd that protect eelgrass (Kd is the light attenuation coefficient, a function of water clarity);
   See comment 9 below, the results of the analysis are of limited value.
b) DO vs chlorophyll was modeled to determine what level of chlorophyll is associated with levels of DO that protect aquatic life;
   See Comment 10 below, usable results from this analysis are questionable.
c) Chlorophyll vs TN was modeled to determine what level of nitrogen is associated with levels of chlorophyll that protect both eelgrass and other aquatic life.
   See comment 11 below, usable results from this analysis are questionable.

   Response: Thank you for your comment. Please see responses to your referenced comments where they were responded to directly in the relevant comment tracking ID#.

Comment Tracking ID #163 (Public Comment 41); Springfield Water and Sewer Commission

Limited value of first stressor-response relationship: Figure F-6 of the Report shows the observed vs fitted relationship between K_d (light attenuation coefficient) and chlorophyll, the first of three stressor-response relationships. The scatter is not random; the fitted values over predict at low observed values and under-predict at high observed values.

Page F-15 of the Report states: "*However, K_d vs chlorophyll embayment fitted plots predicted moderate k_d levels even at extremely low chlorophyll levels. One potential explanation ... suspended solids and dissolved organic matter ... were contributing to light attenuation and increasing the error in the chlorophyll vs. K_d relationship. Suspended solids or dissolved organic matter data were not available and could not be modeled."*

For these reasons, this stressor-response relationship is of limited utility, and should not be relied upon to drive point source upgrades that may be unnecessary and achieve no significant environment benefit.

Failure to establish endpoint in second stressor-response relationship: The second stressor-response relationship evaluated was DO vs chlorophyll-a. On page F-18 and F-22 of the Report, EPA states: "*The model fit reasonably well, as seen in the above plot. However, the coefficient for chlorophyll was positive, suggesting that DO levels increase as chlorophyll increases. Also, the fitted model predicted relatively high values of DO even at extremely low chlorophyll levels ... Grab samples of DO are, therefore, of little utility in gaging the complete manifestation of metabolic effects on DO ... There were sparse data available for paired samples taken at the bottom of the water column across LIS (40 observations) ...* ***For these reasons, a chlorophyll endpoint was not able to be derived for the DO vs. chlorophyll relationship***."

It is not uncommon for productivity to improve average DO conditions. DO minima from nighttime or diurnal monitoring would be needed to assess the impact of chlorophyll-a concentrations on DO. This represents a significant data gap that essentially precludes this type of analysis from being performed.

Failure to establish nitrogen endpoint in third stressor-response relationship: The third stressor-response relationship evaluated was Chlorophyll vs TN. On page F-23 of the Report, EPA states: *"The results of this model were subpar ... Additional variables were explored to see if the model fit could be improved. Temperature was found to be insignificant and was removed from the model. In addition, the coefficient for nitrogen was negative, suggesting that chlorophyll levels decrease as nitrogen increases.*

Therefore, a nitrogen endpoint was not able to be derived for the chlorophyll vs nitrogen relationship for open waters."

This analysis (Figure F-9) includes the same issue as the Kd v chlorophyll-a: that is scatter is not random; the fitted values over predict at low observed values and under-predict at high observed values. This relationship is of limited utility.

> *Response: Please see the response to comment tracking ID #78. We did not use the models from any of the comments above in the original analysis.*

Comment Tracking ID #164 (Public Comment 42); Springfield Water and Sewer Commission

Failure of stressor models to provide feasible TN values: On page G-33 of the Report, EPA states with respect to the Connecticut River: "*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. EPA plans to revisit the assumptions made during the stressor-response analysis in the next phase of the work.*"

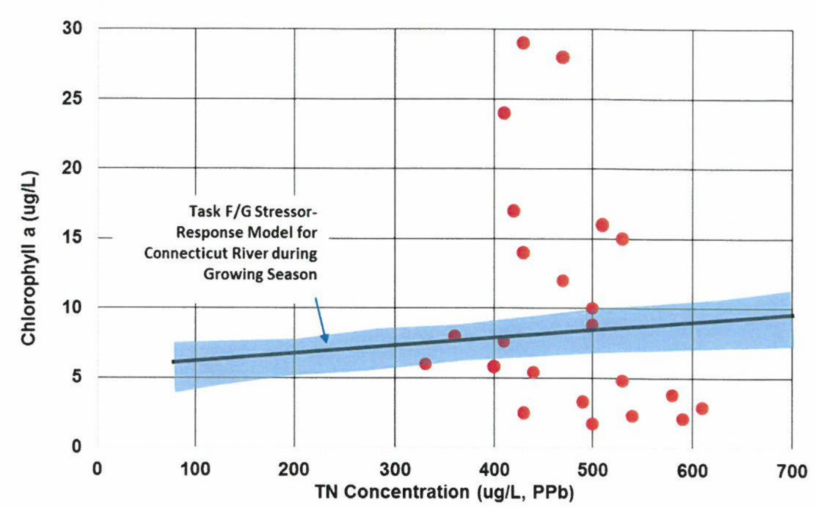
Given these findings, this regulatory effort should be re-evaluated. It is very likely this indicates that chlorophyll-a in the Connecticut River embayment is not sensitive to TN loads.

Level of uncertainty incompatible with usable results: On page F-20 of the Report, EPA states, regarding the fitted plots and 90-percent confidence intervals of the chlorophyll versus nitrogen relationship presented in Subtask G: *"The uncertainty in the predicted values stems from the uncertainty in the estimated model parameters."*

Parameter uncertainty in this case is dwarfed by model uncertainty. The stressor-response models developed are so uncertain that any conclusions derived from them is questionable.

Chlorophyll-a vs. nitrogen model meaningless: The validity of open water stressor-response relationships are of equally significant concern. DO vs chlorophyll-a shows a positive relationship because diurnal minima were not available. EPA states that bottom DO paired with chlorophyll-a would strengthen the analysis, but this is technically unsound due to bottom hypoxia. The chlorophyll-a vs nitrogen model (Figure F-11) showed an extremely poor relationship, and it's use remains questionable.

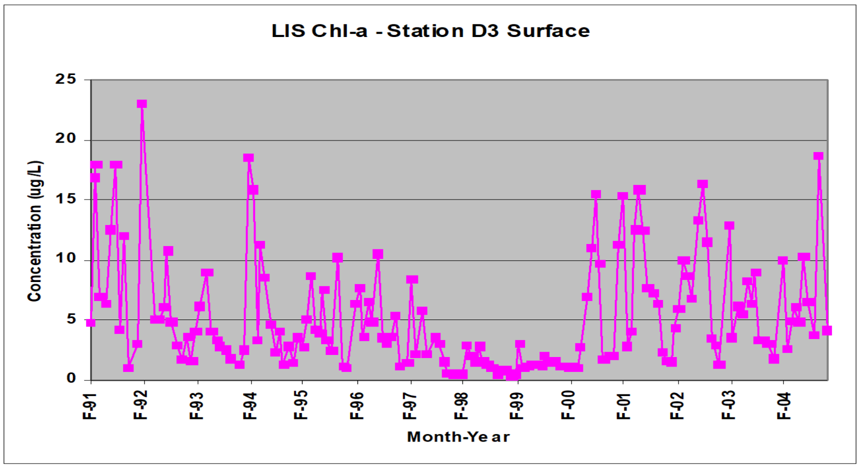
Recent data shows lack of relationship between TN and chlorophyll-a: Sampling performed in summer 2017 provides 23 data pairs in the Connecticut River estuary. Without the LIS-wide model parameters developed by Tetra Tech, we cannot predict Chlorophyll-a based on these measurements. However, we can plot the pairs of TN and chlorophyll-a over the "Chlorophyll vs. Total Nitrogen Relationship for the Connecticut River Area of Influence" (Figure G-29). Doing so shows a lack of relationship between total nitrogen and chlorophyll-a in the Connecticut River.



Response: Please see response to comment tracking ID #136 for comments on cause and effect, which is not the focus of this effort. EPA revisited these models and updated the memos to reflect that. Note that the TN range for the CT River alone is limited, which is why EPA used a hierarchical modeling approach, so the data presented are out of context from what a larger gradient, that included naturally lower TN concentrations, would show in terms of algal response. Lastly, please note that another, arguably defensible reason for the lack of a TN response may be that the CT River has been saturated with excess TN.

Comment Tracking ID #165 (Public Comment 43); Footprints in the Water

The methods and lines of evidence used standard analytical protocols and some innovative applications short of mechanistic modeling to strengthen the utility of the nitrogen endpoints. However, the use of state variables with an uncertain relationship to stress, and the reliance primarily on TN as an indicator of stress limited the ability to categorize and attribute uncertainty to specific sources of error such as sampling error, natural variability and non-linear relationships between “stressor” and “response” state variables. This is typical of these types of analyses where uncertainty forms a cone with lower uncertainty for physical relationships, e.g., salinity due to mixing, and increases for non-conservative pollutants like nitrogen, and then greatly widens when the uncertainty of response variables like chlorophyll-a as an indicator for primary productivity are layered on. There is also error created by the lack of synoptic data of equivalent stressor-response density, or temporal and spatial scale representation that match adequately to be considered a stressor-response correlative relationship. Further error enters from assuming independence of stressor (dose) – response in the analysis, which is rarely true for nitrogen, which quickly moves from compartment to compartment and is gassed off if conditions are right. The rates and relationships can be highly variable on a daily basis, and have seasonal peaks and valleys as well that may be confounded by other limitations such as light, temperature, other nutrients (e.g., silica, which is often ignored and can be important for diatom limitation), and substrate. Sampling nitrogen effectively across those pools is difficult with water column constituents more easily assessed in the TN pool, and benthic or macroalgae less often assessed in the TN pool. Finally, representing algal production with biomass estimates based on chlorophyll-a or even carbon may be grazed at varying rates or senesce rapidly producing unreliable estimates of production. Even use of an index period to qualify data can be problematic as Long Island Sound goes through cycles or sequences of primary producers with diatom-dominated blooms in the pre-Spring months of February and March followed by soft algae such as dinoflagellates and benthic algae and macrophytes in mid-to-late summer and fall. The graph below shows monthly chlorophyll-a data for Station D3, for example, from 1991-2004, which exhibited a steady decline in phytoplanktonic chlorophyll-a from1991 until 2000 and then rebounded that could not be explained by laboratory or sampling error or changes in loading or synoptic nitrogen data. Best guess is that it represented an actual condition of standing crop of phytoplankton that may have been processed by zooplankton grazing trends or cycles, perhaps related to warming temperatures, but could not be tested.



Response: Thank you for your comment.

Comment Tracking ID #166 (Public Comment 44); CTDEEP

Page F-13, Fourth paragraph: "One assumption of many statistical methods is that the data are independent." Is that an assumption of the statistical response relationships that Tetra Tech employed here?

Response: Data were assumed independent unless they belonged to a defined group. For example, data collected from the same station ID were not assumed to be independent from other data collected from the same station ID. A random intercept for station ID was included in the modeling process to account for any data dependencies within stations.

Comment Tracking ID #167 (Public Comment 45); CTDEEP

Subtask G**:** Some embayments include a stressor-response model endpoint of 5.5 ug/L in addition to 10 ug/L. Please provide a brief explanation as to when the 5.5 ug/L endpoint will be used and why the embayment adjusted fit line does not always intercept the 5.5 ug/L endpoint.

Response: These are currently modeled as equally valid potential endpoints bracketing a range of tolerable conditions based on different sources. Neither one is particularly recommended for one case or another at this time. EPA revisited these analyses pursuant to review comments and *has updated the memo; as a result these embayment specific values have changed and most intercept these values.*

Comment Tracking ID #168 (Public Comment 46); CTDEEP

Subtask G**:** Why is the Total Nitrogen endpoint from the global fit stressor-response analysis of all LIS data different for Norwalk Harbor than the other embayments where only the global fit was used (Farm River, Southport Harbor, Connecticut River)?

Response: The embayment (blue) and population (grey) trends both depended on the values of the covariates (pH and temperature). The median values were calculated using the same dataset filtered for depth and growing season. However, non-paired data were used in order to take advantage of additional univariate data. If no data for a covariate were available for a given embayment, then LIS-wide values were used. Some embayments may have lacked TN and Chlorophyll data, but had pH or temperature data, which would have modified the model slightly. However, these models have all been updated with new data and revised models. Only a very few embayments used the population only model.

The population fit for Norwalk Harbor used available temperature data to calculate the median. The population fits for Farm River, Southport Harbor, and Connecticut River used LIS-wide medians.

We added text to better describe the calculation of the trend lines.

Comment Tracking ID #169 (Public Comment 47); CTDEEP

Subtask G**:** In the Subtask G stressor-response figures included for each embayment and Western LIS, should the global fit analysis (gray line) always be the same? Please explain why it is or is not always the same.

Response: See response to comment tracking ID #46. The population (grey) trend depend on the median values of the covariates (pH and temperature) for the embayment. The differences in pH and temperature across embayments accounts for the change in the population trend.

Comment Tracking ID #170 (Public Comment 48); Battelle

Pages G3,5,7,11,13&15etc – "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)." Initially, this sentence appeared to be an unacceptable punt which indicated inability to explain model sensitivity or representativeness. Recurrent use throughout leaves one to beg the question of the consultant's use of the method at all. Meanwhile, the equal weight appears to continue to be given to the stressor-response method line of evidence (page G1)?

Response: No single value was derived for any waterbody, so weighting is not an issue. The values are simply presented as part of the range of values. We revisited the stressor-response model to attempt to improve variability and their ability to provide interpolated values for the stressor-response line of evidence, but any model is limited by the variability inherent in the data and the ability of the models.

Comment Tracking ID #171 (Public Comment 49); Battelle

Why is achievability cited in the data discussion? The extent to which an endpoint threshold is realistic to achieve has no bearing on its derivation. It would be a helpful finding in a UAA, but it's doubtful that is the intent here.

> *Response: The language in the F/G Memo has been updated to address this comment. In the previous draft, the F/G Memo was clarifying that due to various factors, including model variability, some of the values from the S-R analysis were outside the experience of the model, lower than what are observed background levels, or exceeded protective limits.*

Comment Tracking ID #172 (Public Comment 50); Battelle

Norwalk's, Farm River's, Southport's, and other endpoints should be labeled as generic LI Sound north coastal endpoints because there is no "embayment-specific" evidence in use.

> *Response: In the updated draft document, chlorophyll* a *targets have been specified based on averages of Kd targets that include individual embayment specific light level requirements using bathymetry and hierarchical models for chlorophyll* a *targets. Many of these values are the same, but did incorporate embayment specific depth information. We do note where the population level model was used in embayments with no site specific paired data, and have made clarifications in the tables and text where this occurs.*

Comment Tracking ID #173 (Public Comment 51); Battelle

The lack of fit in Figures G2,4,10,12,18,20,24,26 warrants discussion.

> *Response: These figures have all been updated with new data and revised hierarchical models. Model fit for any one waterbody cannot be interpreted using only data from that embayment, per discussion of the method in the text. We have now plotted the population dataset so viewers can see how the model fits the population and is adjusted for embayment specific data.*

B.2 Non-Technical Comments

Applicability of Analysis for Decision Making

Comment Tracking ID #174 (Public Comment 52); NACWA

In addition, NACWA continues to believe that the reports related to the Strategy – including the recent Memorandum on Subtask F&G – are not yet technically adequate or defensible and should not serve as the sole or major basis for requesting additional nitrogen reductions or Clean Water Act permit revisions. EPA argues that these documents will not have direct regulatory impacts, but permitting decisions are already being influenced by the incomplete work on the Strategy.

> *Response: EPA enlisted the help of four experts to conduct an independent expert technical review of the analyses presented in the memos for Subtasks E and F/G. The technical reviewers concluded that the methodology used was technically sound and offered some suggestions to strengthen the analysis. Based on the suggestions, EPA has updated the analysis and the memos to address comments received.*

Comment Tracking ID #175 (Public Comment 53); NACWA

Due to the flaws discussed above and in our previous letter, LIS analyses, alone or in tandem, or any preliminary information from this exercise, should not be used as the basis for any current action, including setting or informing permit limits.

> *Response: The results of this work do not set permit limits. They are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations of nitrogen that are protective of seagrass and aquatic life.*

Comment Tracking ID #176 (Public Comment 54); Springfield Water and Sewer Commission

The Springfield Water and Sewer Commission and our team of consulting scientists, engineers, and legal professionals have reviewed the above-referenced report, together with relevant technical documents and rules, regulations and policies of the United States Environmental Protection Agency (EPA) and associated state agencies.

We have found this report to be incomplete, both with respect to the lack of regulatory basis for the premise of the report itself, as well as the lack of an acceptable technical basis for the conclusions derived therein.

> *Response: The technical study is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies. EPA enlisted the help of four experts to review the analysis. These experts offered feedback to strengthen the analysis, which EPA incorporated into an updated analysis and corresponding memos. The result of this work is not related to regulation. Rather, the results are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations that are protective of seagrass and aquatic life.*

Comment Tracking ID #177 (Public Comment 55); Springfield Water and Sewer Commission

EPA provides conflicting information on uses of the study:

The Report provides the following guidance for future use of the TN endpoints developed:

> *The TN endpoints established are "not a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance, but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies." (page F-i)*

However, EPA's website containing Subtask F/G implementation is as follows:

> *EPA is now developing Total Nitrogen endpoints (Subtask F & G memo) for each waterbody grouping that are protective of designated uses. The endpoints are intended as a source of relevant information to be used by federal and state water quality managers and stakeholders in developing nitrogen reduction strategies, including nitrogen reduction targets and allocations.*
> *Although Water Quality Based Effluent Limits (WQBELs) for National Pollutant Elimination Discharge Elimination System (NPDES) permits will not be products of the work, facilities should start planning now for future water quality based nitrogen limits, especially if they are planning upgrades to existing treatment plants*

EPA supports use of a single total nitrogen load a/location for each state. NPDES permits issued in each individual state will have to demonstrate consistency with achieving the load allocation for that state to meet water quality standards in a particular sub-watershed.

As technical products under the Nitrogen Strategy are completed, EPA will review permits in a particular sub-watershed to develop a permit strategy for EPA-issued permits, and will work with States, municipalities and regional bodies to develop and implement strategies to attain nitrogen reduction targets.

Subsequent regulatory actions such as NPDES permit actions, if any, based on the modeling efforts or data from this study would require a formal comment and public notice period.
EPA and state permitting programs will use a combination of permit renewals and/or revisions to existing NPDES permits as appropriate to incorporate revised nitrogen limits where necessary.
[footnote 3: https://www.epa.gov/npdes-permits/frequently-asked-questions-long-island-sound-watershed-permitting]

EPA needs to provide clarity regarding the intended uses of the TN endpoints developed in this study. The intention to use the Report to support the imposition of water quality based effluent limitations on point source discharges is in contradiction to the preamble of the Subtask F/G Report, and in contradiction to CFR 122.44, the federal regulations that establish a basis for the implementation of water quality based effluent limitations in NPDES permits.

Response: The statement at the beginning of this report is still accurate and clarifies the intended use of this study: This study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.

The intent of the analysis is not to set effluent limitations on specific point source discharges nor set wasteload allocations among groups of point source discharges. Rather, the results are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations that are protective of seagrass and aquatic life. Water quality managers can then consider what mix of nitrogen source reduction actions, if any, are needed for a particular watershed.

EPA is working to update the Frequently asked Questions on the NPDES website to make it consistent with this Response to Comments document.

Comment Tracking ID #178 (Public Comment 56); Springfield Water and Sewer Commission

Based on the results provided in Subtask F and G, we do not believe this new research provides an adequate technical basis on which to establish new TMDL allocations for TN, especially in the Connecticut River.

Response: This study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study *might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.*

The results of this analysis do not set new TMDL allocations. Rather, the results are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations that are protective of seagrass and aquatic life. The information based on three lines of evidence (literature values, stressor-response modeling, and a distribution-based approach). EPA enlisted the help of four experts to conduct an independent expert technical review of the analyses presented in the memos for Subtasks E and F/G. The technical reviewers concluded that the methodology used was technically sound and offered some suggestions to strengthen the analysis. Based on the suggestions, EPA has updated the analysis and the memos to address comments received.

Comment Tracking ID #179 (Public Comment 57); Springfield Water and Sewer Commission

In summary, EPA needs to establish a more sound regulatory framework through which the results of this Report could be utilized in a meaningful way, absent the development of a new TMDL for the protection of aquatic life, using eelgrass as the indicator. EPA also needs to provide clarity and transparency to the public and regulated community regarding future uses of the Report. We remain concerned that NPDES program directors may utilize this report as a basis to impose water quality based effluent limitation for TN, above those required by the LIS TMDL. This is of particular concern since the water quality endpoints chosen are 0.40 mg/L TN, which are far more stringent than natural background levels or feasibly achievable through wastewater treatment technologies.

The three (3) stressor-response models chosen by EPA to evaluate appropriate endpoints were all determined to have either failed to demonstrate a usable relationship, or to have provided meaningful data. We would encourage EPA to review the basis of its effort to develop TN endpoints through eelgrass evaluation. Since ambient level of TN are below background levels, it is unclear why EPA is focusing on further nitrogen reductions to improve aquatic health in the Sound.

Response: This study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.

The intent of the analysis is not to set water quality-based effluent limitations on specific point source discharges nor set wasteload allocations among groups of point source discharges. Rather, the results are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations that are protective of seagrass and aquatic life. Water quality managers can then consider what mix of nitrogen source reduction actions, if any, are needed for a particular watershed.

In the revised Memo for F/G, EPA provides the results for all three lines of evidence for each embayment. These values provide a range of target concentrations for TN likely to be protective of seagrass and aquatic life. As a result, no single value is reported from this report.

Comment Tracking ID #180 (Public Comment 58); NHDES and VTDEC

As mentioned in our December 15, 2017 email, we appreciate EPA's efforts to collect additional data and request that EPA not proceed with establishing final reductions, allocations and/or NPDES permit limits for facilities in the Connecticut River basin until this new data has been collected and state agencies have had an opportunity to review any updates to the TN threshold as well as any proposed reductions, allocations and/or NPDES permit limits based on the thresholds.

Response: This study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.

EPA has incorporated additional data from the Connecticut River and Housatonic River into the analysis. EPA has also incorporated feedback from state agencies, the public, and expert technical reviewers into the analysis to strengthen the technical analysis and address any technical deficiencies. EPA is currently looking at additional sources of data to include in any future updates.

Comment Tracking ID #181 (Public Comment 59); CTDEEP

Appendix 2—Description of CT Approach to Addressing Water Quality and Nutrients in LIS Embayments
In order to improve water quality restoration and protection outcomes, CTDEEP undertook a public process to identify water quality concerns and specific water bodies for development of TMDLs or other Action Plans to address water quality restoration and protection. That process, called Integrated Water Resource Management, was an outgrowth the 303d Vision process developed collaboratively between the States and EPA. As part of that process, CT identified that it would develop plans to restore degraded waters and protect higher quality waters, focusing on impacts from nutrients and from stormwater, as well as on designated uses including aquatic life uses support, shellfishing and recreation. As part of this effort, CT coastal embayments were also identified as priority areas. A subset of waterbodies were identified for development of TMDLs/ Action Plans under this process.

Since these priorities were established in 2016, CT has been working to identify a strategy to develop TMDLs/Action Plans to address these priority concerns and areas. The approach currently under consideration is based on developing a plan consistent with Connecticut Water Quality Standards, focusing on designated uses for the embayments, primarily recreation, shellfishing and aquatic life use support. We are planning to have a phased approach for evaluating and managing nutrient loads to each embayment.

Phase 1

Phase l would focus on using existing water quality criteria or interpreting existing narrative standards to restore and protect water quality in each embayment. CT has numeric and narrative water quality standards which would form the basis for the TMDLs/Action Plans. Specifically, CT has numeric water quality criteria for dissolved oxygen and certain toxic parameters and narrative standards for nutrients, toxicity in surface waters and sediments and for biological condition. Water quality goals for Phase 1 will focus on dissolved oxygen, water clarity and providing habitat conditions support of designated uses. We may consider evaluating the natural trophic tendency for each embayment when setting appropriate water quality targets. In waters shared with other states, the more restrictive of the criteria between the two state would be used.

We are evaluating the use of paired water quality models to evaluate pollutant loadings to embayments and link these to water quality conditions, dissolved oxygen and water clarity, in each embayment. We have been working to identify which models would work best for this project. For the watershed model, we are considering the use of the HSPF model. If that is not possible, we would likely use the AVGWLF model. We prefer HSPF because it has previously been calibrated for both CT and RI within other projects. It is a robust model which would allow for inclusion of state-specific data and provide a detailed analysis of point and nonpoint sources contributing pollutant loads to the embayment. Both the HSPF and AVGWLF models have been previously used within the context of the LIS program. We have scheduled a 2-day training to become more familiar with the HSPF model and the BASINS environment.

For estuarine in-water models, we are expecting to use the model recommended by the Niantic River Project, provided it is sufficient to allow for modeling of dissolved oxygen and water clarity within each embayment. If that model is not sufficient, we may evaluate the AQUATOX model or other models as a replacement. By running the paired models, we should be able to identify embayment-specific nutrient levels or other constituents or conditions associated with attaining water quality goals for each embayment.

We are using the Pawatuck River estuary complex for development of our approach to embayments and have partnered with Rhode Island to work collaboratively to develop an approach suitable for our shared resource. CT would then extend that approach to other CT embayments.

Phase 2

Phase 2 focuses on extending beyond the use of the existing narrative water quality standard for biological integrity and developing a refined approach to establishing biological criteria for estuarine embayments. After such criteria are developed, the TMDLs/Action Plans previously developed under Phase 1 would be re-evaluated to insure consistency with the new biological criteria for embayments.

Much of the focus of Subtasks F & G are based on endpoints related to aquatic life protection that are based on literature values and stressor response models relying heavily on eelgrass research. The designated use in Connecticut Water Quality Standards Regulations for Class SA and SB waters that include estuaries of Long Island Sound that most closely match "aquatic life protection" as expressed in Subtasks F & G is "*habitat for marine fish, other aquatic life and wildlife*". DEEP does not have a subcomponent or tiered designated use for eelgrass because data are not available to support this designation in Long Island Sound embayments at this time. Initial seagrass distribution modeling projects (e.g. Vaudrey et al 2013 [footnote 1: Vaudrey, J.M.P., J. Eddings, C. Pickerell, L. Brousseau, 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.]) are a step in the right direction, but do not provide enough justification to develop endpoints and thresholds for use in Long Island Sound.

> *Response: EPA has provided an argument based on well-established and peer reviewed research linking the importance of eelgrass habitat to aquatic life. This is iterated in the documents, which were then, themselves, peer reviewed by an expert panel of estuarine ecologists who acknowledged support for and the defensibility of eelgrass protection to aquatic life. As the document established, the eelgrass assessment endpoint was used for protection and restoration*

of eelgrass per se as well as its importance for the protection and restoration of aquatic life in LIS.

The final results are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations of nitrogen that are protective of seagrass and aquatic life. The recommendations will be based on three lines of evidence (literature values, stressor-response modeling, and a distribution-based approach).

Change in Direction

Comment Tracking ID #182 (Public Comment 60); Springfield Water and Sewer Commission

The purpose of Subtasks F & G was to develop TN endpoints. Whereas previously EPA has focused on developing endpoints that are protective of the water quality standards for nitrogen and DO, this Report develops TN endpoints that are protective of aquatic uses, using eelgrass as the indicator. This is a major shift in direction that should be examined more critically, involving the public and the regulated communities.

Response: This study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.

EPA is involving the public and regulated communities. EPA has included a Technical Stakeholder Group in the process from the beginning of the analysis. This group has had a chance to comment on previous analyses and memos and EPA has incorporated feedback. Additionally, EPA has conducted public webinars about the project. For the March 2018 documents, EPA initiated both a public review process and an independent expert technical review panel. The public process allowed for stakeholder engagement.

Concerns with Comment Process

Comment Tracking ID #183 (Public Comment 61); NACWA

The National Association of Clean Water Agencies (NACWA) represents over 300 public clean water utilities across the country, including more than 40 utilities in Regions 1 and 2. NACWA's members treat and reclaim the majority of the wastewater generated each day nationwide, providing an essential public service that protects human health and the environment. NACWA previously wrote Regional Administrators Dunn and Lopez to express concern over the lack of stakeholder engagement on the LIS Nitrogen Reduction Strategy (Strategy) efforts, the overall flaws in the process EPA was using – outside of the total maximum daily load (TMDL) context – and the technical deficiencies with some of the earlier reports. That letter is attached for your reference.

We appreciate the opportunity EPA has provided to comment on the technical report entitled, "Establishing Nitrogen Endpoints for Three Long Island Sound Watershed Groupings" (Subtask F&G Memorandum or Memorandum). The fact that EPA is now seeking public comment on this document signals an important improvement over how the Agency had been proceeding with work on the Strategy in the past. Concerns remain, however, over the process EPA is using to conduct this work. For instance, while EPA is now seeking comments from a broader array of stakeholders, it states on its website for the Strategy that there are no plans to respond to the comments that are submitted. Seeking input without thoughtfully responding to that input is not meaningful stakeholder engagement.

Response: This study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.

Regarding the public comment process. EPA has included a Technical Stakeholder Group in the process from the beginning of the analysis. This group has had a chance to comment on previous analyses and memos and EPA has incorporated feedback. Additionally, EPA has conducted public webinars about the project. For the March 2018 documents, EPA initiated both a public review process and an independent expert technical review panel. The public process allowed for stakeholder engagement.

Regarding the commenters concerns over technical deficiencies in earlier reports, EPA conducted an independent expert technical review of the analyses conducted under Subtask E and F/G to strengthen the analyses, address any technical deficiencies, and attend to any technical concerns raised by the technical review experts. Please refer to comment 18 for more information on how the technical review was conducted. Since receiving comments, EPA has updated the analysis and revised the reports to address all comments received (both public and technical review). Additionally, EPA is publishing response to comments documents for Memo E, Memo F/G, and Policy Comments, all of which individually respond to all comments received.

Comment Tracking ID #184 (Public Comment 62); Springfield Water and Sewer Commission

Failure to provide Appendices F1 to F4: Appendices F1-F4 to the Report (identified below) are not publicly available. While Appendix F3 and F4 were provided over private email communication, EPA advised that Appendix F1 and F2 (comments on the Report from the Technical Stakeholder Group) would not be publicly available until after close of the public comment period.

Appendix F1: Response to Comments on Task F/G Technical Comments
Appendix F2: Compilation of Comments
Appendix F3: Endpoint Values Found in Massachusetts Estuary Project Reports
Appendix F4: Paired Data for Stressor-Response Modeling

Response: EPA will ensure that all related appendices for Memo F/G are posted to the LISS website.

Expert Technical Review

Comment Tracking ID #185 (Public Comment 63); NACWA

NACWA also has questions about EPA's planned peer review process, how that will work, and how the peer reviewers will be selected. Submitting these documents for peer review does not obviate the need for a formal review and comment process, with meaningful engagement from the Agency, including responding to stakeholder comments.

> *Response: EPA solicited comments from the public and also requested an independent expert technical review of the work done under Subtasks E and F/G. The independent contractor was HydroAnalysis, Inc. As summarized in their technical report to EPA:*
>
> *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 Memo E and Memo F/G and given 20 specific questions to respond to. 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, nor did they communicate with EPA or with Tetra Tech during the review process or during the development of the summary report.*
>
> *EPA 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. EPA 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.*

General Concern about Modeling

Comment Tracking ID #186 (Public Comment 64); NACWA

To avoid arbitrary decision-making when using any model, an agency must be able to draw a rational connection between the factual inputs, modeling assumptions, modeling results and conclusions drawn from these results. Sierra Club v. Costle, 657 F.2d 298, 332-33 (D.C.Cir.1981). A reviewing court also will reverse an agency action that relies on a model, "if the model is so oversimplified that the agency's conclusions from it are unreasonable." Appalachian Power Co. v. EPA, 249 F.3d 1032, 1052 (D.C. Cir. 2001) (citations omitted). When a model is challenged, EPA must provide a full analytic defense. Eagle-Picher Indus., Inc. v. U.S. EPA, 759 F.2d 905, 921 (D.C.Cir.1985). EPA must be able to explain the assumptions and methodology used in preparing the model. Small Refiner Lead Phase-Down Task Force v. EPA, 705 F.2d 506, 535 (D.C. Cir. 1983). Further, proceeding without a fully developed model of the Sound is contrary to EPA's own recommended water quality criteria for nutrients which state: "wherever possible, develop nutrient criteria that fully reflect localized conditions and protect specific designated uses." 66 FR 1671, 1673 (Jan. 9, 2001).

> *Response: EPA has incorporated feedback from the expert technical reviewers and public commenters in memos for Subtasks E and F/G. In these memos are included assumptions and*

methodology used. EPA has provided additional clarifying detail in the memos to support those interested in the details of the analysis.

Impact on States

Comment Tracking ID #187 (Public Comment 65); Northhampton DPW

As stated above, any "design" endpoints will have significant economic impacts. We should not rush to regulate without significant time to collect relevant data, develop appropriate science and allow greater public input to any proposed regulation.

> *Response: This study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.*
>
> *EPA does not plan to propose any regulation as a result of this analysis. Rather, the results are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations that are protective of seagrass and aquatic life. Water quality managers can then consider what mix of nitrogen source reduction actions, if any, are needed for a particular watershed. Additionally, EPA has taken the time to collect all available relevant data and develop appropriate science (confirmed by the independent expert technical review) and has allowed for public input through comments collected in drafts of the analyses.*

Comment Tracking ID #188 (Public Comment 66); NHDES and VTDEC

Further, we continue to remain skeptical about the significance that most wastewater treatment facilities in the upper basin states have on degradation of LIS and feel that should significant loading restrictions be placed upon them in the future, it will have little benefit on water quality in LIS but have significant impacts to the small communities in our states.

> *Response: This study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.*
>
> *This results of this work are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations of nitrogen that are protective of seagrass and aquatic life. Water quality managers can then consider what mix of nitrogen source reduction actions, if any, are needed for a particular watershed.*

Link between TMDL and Strategy

Comment Tracking ID #189 (Public Comment 67); NACWA

EPA's Avoidance of the TMDL Program is Inappropriate

EPA acknowledges on its website that the Strategy is intended to address remaining impairments, but will proceed outside of the TMDL process:

> *EPA is implementing a strategy to aggressively continue progress on nitrogen reductions, in parallel with the States' continued implementation of the 2000 Total Maximum Daily Load (TMDL), and achieve water quality standards throughout Long Island Sound and its embayments and near shore coastal waters. The strategy recognizes that more work must be done to reduce nitrogen levels, further improve dissolved oxygen (DO) conditions, and address other nutrient-related impacts in Long Island Sound. The nitrogen reduction strategy complements the 2000 TMDL in important ways. Foremost, while the 2000 TMDL is premised on achieving water quality standards for DO in the open waters of LIS, the EPA strategy expands the focus to include other nutrient-related adverse impacts to water quality, such as loss of eelgrass, that affect many of LIS's embayments and near shore coastal waters.*

While EPA is careful to refer to "impacts" in its discussion of the Strategy, avoiding the use of the word "impairment", the underlying implication that serves as the foundation for all of this work is EPA's view that certain waters need additional work to meet water quality standards. In other words, they are "impaired." The Clean Water Act provides a clear process to follow when addressing impaired waters. EPA used the TMDL program when it first established the dissolved oxygen TMDL in 2000, but it has now chosen to avoid a new TMDL process in favor of a non-transparent 'strategy' that it argues allows for more "adaptive management."

We are very concerned about how the work underlying the Strategy may be used – especially to the extent that it may be used to support new effluent limits imposed on regulated parties, despite the fact that those parties are currently covered by wasteload allocations in the TMDL. In its August 2018 response to NACWA's April 2018 letter, EPA says this about the studies it is conducting to support the Strategy: "The studies, once finalized, may not apply to a particular situation or circumstance, but are intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies." That does not allay our concern.

If "nitrogen reduction strategies" referred to by EPA include an assumption that certain parties need to reduce their discharges, those strategies could lead to an attempt to impose new, more stringent permit limits. That would be illegal. EPA's permit rules are directly linked to and required to implement the TMDL program. Indeed, EPA guidance directs permit writers to use WLAs to derive water quality based effluent limitations ("WQBELs") for permits. NPDES Permit Writers' Manual at 107. The regulations specifically provide that when developing WQBELs, EPA must ensure that "effluent limits developed to protect a narrative water quality criterion, a numeric water quality criterion, or both, are consistent with the assumptions and requirements of any available WLA for the discharge prepared by the State and approved by EPA pursuant to 40 CFR 130.7." 40 CFR 122.44(d)(1)(vii)(B).

Any new limits based on a "nitrogen reduction strategy" that uses the new LIS studies would be blatantly inconsistent with the TMDL's "assumptions and requirements," and would therefore not comply with the Agency's permitting regulations. Although federal rules also refer to a "reasonable potential" analysis, the Agency has stated in guidance that a separate reasonable potential analysis is unnecessary where the state has already approved a TMDL and WLAs, because the water quality determination is inherent in the TMDL and WLA decision. *Questions and Answers on the Great Lakes Water Quality Guidance,* Set 2 (March 20, 1996) at Q&A No. 21. Discharges that were covered in the LIS TMDL should be based on the WLA's in that TMDL, until EPA and/or the States develop a new or revised

TMDL. The studies that EPA is currently performing, even if they result in the creation of a "nitrogen reduction strategy," cannot form the basis for new permit limits unless they are used in developing that new or revised TMDL, with all of the process and public comment that is required.

In its recent letter to NACWA, EPA notes that the work it has underway is not developing any requirements or other provisions that are legally binding. This position ignores the fact that any additional water quality improvements expected to result from development of this Strategy are unlikely to be realized if it is not used by states and regional offices to develop permit limits and other requirements that are legally binding. EPA should instead use the established process for reviewing and revising TMDLs that require improvement.

> *Response: As stated by the comment, this study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.*
>
> *The intent of the analysis is not to set effluent limitations on point source discharges. Rather, the results of this work are intended to are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations of nitrogen that are protective of seagrass and aquatic life. Water quality managers can then consider what mix of nitrogen source reduction actions, if any, are needed for a particular watershed.*

Comment Tracking ID #190 (Public Comment 68); Springfield Water and Sewer Commission
EPA is actively implementing a strategy to reduce nitrogen concentrations in the Long Island Sound (LIS), in conjunction with an EPA approved 2000 TMDL. The TMDL was developed in order to achieve water quality standards for dissolved oxygen (DO) in open waters of the LIS. Prior to the 2000 TMDL, hypoxia (low levels of DO) were present in the LIS. In order to achieve compliance with DO, the TMDL calls for total nitrogen (TN) reductions in point source, non-point source and air deposition, such that the resulting levels of TN in the LIS, are compatible with the water quality criteria for DO.

Significant progress toward attaining the DO water quality standard have already been obtained. As per the *LIS Year in Review* (2017), the average peak area of waters with "unhealthy" DO is less than half of the pre-TMDL levels. The area of water with less than 3 mg/L of DO in 2015 and 2017 were the second and third smallest recorded in the past 31 years of monitoring. In addition, there have been no open waters below 1 mg/L DO in seven of the eight past years. As a result of nitrogen reduction efforts, there are 45 million fewer pounds of nitrogen discharged annually to the Sound from human sources (a 59% reduction). [footnote 1: Newsletter of the Long Island Sound Study, Spring 2018. http://longislandsoundstudy.net/wp-content/uploads/2018/05/2017YearinReview_03-singles-second-printing-14-aug-18.pdf.]

While substantial progress has been made toward achieving the water quality standard for DO through nitrogen reductions as required by the TMDL, EPA has expanded its focus beyond DO, and include an evaluation of overall aquatic health for the LIS. While this scope is beyond the current LIS TMDL, EPA is using the TMDL to support this expanded goal. In order to evaluate the overall health of the LIS, EPA has determined that the presence of healthy eelgrass communities are appropriate indicators of overall aquatic health.

Subtask Memo F/G, which is the subject of our review, deals exclusively with assessing eelgrass, and the relationship of eelgrass to an overall healthy ecosystem in the LIS. As mentioned above, it is important to note that eelgrass is not discussed in the LIS TMDL, whether as a stand-alone indicator of water health, or a variable that could address hypoxia (the purpose of the TMDL). In effect, EPA has changed the endpoint of the TMDL from addressing low DO, to now addressing the presence of eelgrass.

The above notwithstanding, and even accepting EPA's premise that healthy eelgrass populations are representative of healthy aquatic ecosystems - the Subtask F/G Report (Report) does not demonstrate a relationship between TN endpoints derived for eelgrass health in embayments and nitrate restrictions on point source discharges.

> *Response: This study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.*
>
> *The intent of the analysis is not to set effluent limitations on point source discharges. Rather, the results of this work are intended are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations of nitrogen that are protective of seagrass and aquatic life. Water quality managers can then consider what mix of nitrogen source reduction actions, if any, are needed for a particular watershed.*
>
> *Regarding the commenter's concerns over technical deficiencies in earlier reports, EPA conducted an independent expert technical review of the analyses conducted under Subtask E and F/G to strengthen the analyses, address any technical deficiencies, and attend to any technical concerns raised by the technical review experts. Please refer to comment 18 for more information on how the technical review was conducted. Since receiving comments, EPA has updated the analysis and revised the reports to address all comments received (both public and technical review). Additionally, EPA is publishing response to comments documents for Memo E, Memo F/G, and Policy Comments, all of which individually respond to all comments received.*

Comment Tracking ID #191 (Public Comment 69); Springfield Water and Sewer Commission

The determination of appropriate TN endpoints to support a healthy eelgrass population in embayments, is an entirely distinct and separate focus from the LIS TMDL to address low DO in open waters. The Report fails to establish a regulatory basis that relates the LIS TMDL water quality goal of addressing hypoxia, to EPA's current strategy in Subtask F/G to determine chlorophyll-a, clarity, and dissolved oxygen (DO) endpoints that support the growth of eelgrass. The Report also fails to establish the regulatory basis for the use of eelgrass as a measure of whether water quality and designated uses are met. Finally, it appears that EPA is using independent TMDL efforts with different end points (DO vs. eelgrass via chlorophyll-a, clarity, and DO) to protect different critical locations (LIS vs. tributary embayments). The tasks completed under the F/G Report should not be considered LIS TMDL implementation, but rather a new or modified TMDL development.

In order to establish the regulatory basis that provides for establishing TN endpoints that support a healthy eelgrass population in tributary embayments, EPA must address the following:

a) Identify the water quality aspect of the tributary embayment that is impaired in accordance with Section 303(d) of the CWA;
b) Identify the particular pollutant of concern that directly prevents the attainment of the water quality criteria identified above;
c) Develop a TMDL to allocate loads to point and non-point sources, such that the water quality standard identified above can be attained.

Response: This study is neither a proposed TMDL, nor proposed water quality criteria, nor recommended criteria. The study is not a regulation, is not guidance, and cannot impose legally binding requirements on EPA, States, Tribes, or the regulated community. The technical study might not apply to a particular situation or circumstance but is intended as a source of relevant information to be used by water quality managers, at their discretion, in developing nitrogen reduction strategies.

The results of this work are intended to are intended to provide data, analysis and other relevant information for helping watershed managers set target concentrations of nitrogen that are protective of seagrass and aquatic life. Water quality managers can then consider what mix of nitrogen source reduction actions, if any, are needed for a particular watershed.

Support for EPA Approach

Comment Tracking ID #192 (Public Comment 70); The Nature Conservancy

Thank you for the opportunity to review and comment on "Establishing Nitrogen Endpoints for Three Long Island Sound Watershed Groupings (Subtasks F&G Memorandum)". On behalf of The Nature Conservancy's bi-state Long Island Sound Program, we applaud the efforts of the U.S. Environmental Protection Agency (EPA) to advance the Long Island Sound Nitrogen Strategy to further reduce nitrogen pollution throughout the Sound, in parallel with the States' continued implementation of the 2000 Total Maximum Daily Load. Now, we urge you to leverage the results of your leadership and the progress you have made toward mitigating the impacts of excess nitrogen in Long Island Sound embayments and nearshore coastal waters by implementing Subtasks H&I - translating waterbody-specific endpoints into percent reductions and summarizing example nitrogen load distributions attaining endpoints by source category.

The Conservancy strongly supports EPA's goal of achieving water quality standards that protect and restore the ecological conditions required to maintain existing populations of eelgrass and support recovery of historic eelgrass habitat locations. We also acknowledge the complexity involved in establishing nitrogen endpoints across Long Island Sound's large and varied geography. However, we cannot afford to slow the progress that has been made.

As you know, results from the U.S. Fish and Wildlife Service's 2017 aerial survey of eelgrass extent revealed a 12 percent loss of the remaining seagrass in the Sound, and a staggering loss of 43 percent of eelgrass in Little Narragansett Bay over the preceding five years. *Cladophora* blooms continue to plague southeastern Connecticut embayments, while a rust tide and fish kills were reported off the coasts of Darien and Westport in September. There is an urgent need to accelerate nitrogen reduction in our nearshore waters to keep pace with increasing temperatures and escalating eutrophication impacts to the Sound's habitats and coastal communities.

While we appreciate the specific technical concerns and desire for additional detailed analysis, monitoring and model development among water managers, we support EPA's approach to assessing potential endpoints based on existing Sound-wide data. Subtasks F&G provide an integrated, analysis useful for advancing adaptive management actions as well as identifying additional research and monitoring needs. There is a critical, ongoing need for coordination and alignment between EPA's efforts and more detailed, geographically specific analyses underway in Connecticut and New York, however we encourage EPA to carry out Subtasks H&I as soon as possible to help identify nitrogen thresholds and "no regrets" actions. While achieving these reductions will be challenging, delaying action to wait for improved data unnecessarily risks further impacts, loss of the Sound's vulnerable eelgrass ecosystem and threatens the economy and wellbeing of our coastal communities.

> *Response: Since EPA received your comments, we have been revising the analysis to incorporate feedback from an independent expert technical review panel that strengthen the work done under Subtasks E and F/G. This work is near completion and EPA will next complete drafts of Subtasks H and I.*

Attachments Provided with Comments (No Response Needed)

Attachment to NACWA Comments: Letter from NACWA to EPA Regional Administrators Dunn and Lopez, dated April 2, 2018

Attachment #1 to Footprints in the Water Comments: Connecticut Methodology for Freshwater Nutrient Management Technical Support Document

Attachment #2 to Footprints in the Water Comment: RE: DOCKET ID NO. EPA-HQ-OW-2009-0596, WATER QUALITY STANDARDS FOR THE STATE OF FLORIDA'S LAKES AND FLOWING WATERS