Preliminary Comments from Members of the Lake Erie Phosphorus Objectives Review Panel
December 10, 2014

(Comments Received as of December 5, 2014)

Table of Contents

Comments from Dr. Alber

Background: The objective of the EPA Science Advisory Board Phase I consultation is to obtain early advice on the modeling approach being applied to inform the updated phosphorus targets for Lake Erie. The four consultation questions are focused on: 1) eutrophication response indicators; 2) Lake Erie models selected; 3) Ensemble modeling approaches; and 4) Phosphorus loads and concentrations. My comments are primarily in response to question 1, with some input on question 4.

Question 1: Eutrophication response indicators: The draft document identifies four eutrophication response indicators for Lake Erie, along with information on the metrics proposed to model and track them. Although each of these indicators is reasonable, I would have liked to see more detail regarding how the metrics will be measured and what spatial and temporal scale will be used. The points I would like to see addressed are detailed below:

1. Basin-specific, summer (June-August) average chlorophyll a concentrations are proposed as a measure of overall phytoplankton biomass. This relates to Lake Ecosystem Objective # 6 (maintain mesotrophic conditions in the open waters of the western and central basins of Lake Erie, and oligotrophic conditions in the eastern basin).
    a. Spatial distribution - there was no information provided regarding where phytoplankton would be sampled: is there a set number of established stations? Are stations located in the offshore areas only or do they include the nearshore and the coastal margin? How are stations distributed amongst the basins? What is the process used to get an average for each basin?
    b. Sampling – how often are stations sampled? Are they obtained from surface water only or are they depth-averaged over the photic zone? Are all samples discrete samples or are sondes or satellites used?
    c. Summer – I am concerned that both inter-annual variation and long-term changes in temperature and precipitation may affect the amount of phytoplankton measured in the June-August window. If there is an early spring, for example, sampling starting in June may miss peak biomass. Perhaps a longer interval should be considered (e.g. May – September), and/or a more flexible time period (i.e. the maximum average concentration over 3 consecutive months of each year). It is also surprising to see Jun-Aug identified here, as other Lake Erie studies use March-June as the spring period and define July-Sept. as summer.

2. Cyanobacterial blooms (including Microcystis sp.) is proposed as a measure of harmful algal blooms in the Western Basin. This relates to Lake Ecosystem Objective #4 (maintain cyanobacteria biomass at levels that do not produce concentrations of toxins that pose a threat to human or ecosystem health).
    a. Spatial distribution – same considerations as above regarding the number and distribution of stations. I recognize these are all within the Western Basin, but in this case to they include nearshore and the coastal margin? Here (and elsewhere) a map would be useful.
    b. Sampling – same considerations as above with respect to sampling frequency. Additionally, how will cyanobacterial cells be identified? Is dry mass estimated from cell counts (and biovolumes?) or are they physically separated from other plankton onto filters? How will dry mass be calculated in the models?

   c. Summer – Same considerations as above with regard to definition of a bloom period.

3. Hypoxia in the hypolimnion of the Central Basin is proposed as a measure of the extent to which the ecosystem is impacted by low oxygen conditions. This relates to Lake Ecosystem Objective #1 (minimize the extent of hypoxic zones in the waters of the Great Lakes).
   a. Spatial distribution – same considerations as above regarding station location and whether these are established stations. (Again, maps would be useful.)
   b. Sampling – are there sondes at these stations? How is stratification characterized? What is the cut-off for hypoxia?
   c. Summer – Same considerations as above

4. Cladophora in the nearshore areas of the Eastern Basin as a measure of the extent of nuisance algae. This relates to Lake Ecosystem Objective #2 (maintain the levels of algal biomass below the level constituting a nuisance condition).
   a. Spatial distribution – Are there established stations? How will observations of Cladophora (P content or biomass) be scaled up? Will algal extent be mapped?
   b. Sampling – how often are stations sampled? What time period is being considered for this metric?
   c. P vs. biomass - I am not sure why this metric is focused on stored P as opposed to biomass. I recognize that Cladophora can have luxury uptake of P and that DRP can be related to growth potential, but the Lake Ecosystem Objective is not how much P they have or their growth potential, but rather the accumulation of nuisance algae. Cladophora dry weight seems like it would be a more straightforward metric.

Additional considerations regarding indicators:

1. Lake Ecosystem Objective #3 (maintain algal species consistent with healthy aquatic ecosystems in nearshore waters) is not directly addressed by these proposed indicators, as there is no mention of sampling and identifying algae in nearshore waters. It is also not clear if “algal species” refers to phytoplankton, macroalgae, or both.
2. Although it appears to be written in a parallel format to #5, Lake Ecosystem Objective #6 does not specify “algal species”. Assuming this means phytoplankton, there is no mention of samples collected to meet this objective.
3. If there are historic data on BOD (biochemical oxygen demand) available, it would be useful to evaluate BOD, whether it has changed over time, and whether it is related to nutrient input or any of the eutrophication metrics. BOD provides information on how quickly the microbial populations that are present can break down the substrates in the sample (including decaying algal blooms), and hence may be a way to assess the availability of organic matter and the potential for hypoxia. In a BOD assay, water is incubated at a standard temperature and length of time and the resulting decrease in oxygen is recorded. Although 5-day BOD incubations (BOD_5) are traditionally performed, comparing BOD_5 with a 20-day analysis (BOD_{20}) is a measure of the relative lability of the substrates present in the sample. In a sense, this mimics differing residence times: during shorter incubation times there may be time for only the most labile substances to be broken down and consumed, whereas with longer incubations the

more refractory substances may be used as well. Mallin et al. (2006) found that in North Carolina rivers, lakes, and streams, BOD_5 was often more strongly correlated with chlorophyll *a*, while BOD_{20} was often more strongly correlated with turbidity, total suspended solids, total phosphorus, and total nitrogen.

Question 4: Phosphorus loads and concentrations. The charge to the SAB asks about the challenge of establishing loading as compared to concentration targets. Establishing loading targets makes sense from a management perspective. The difficulty is that the organisms respond to concentrations, not loads, and concentration in a given area depends on both physical factors (e.g. residence time, stratification) and internal processing. The same loading rate could therefore cause problems under one set of conditions and not others. There is also difficulty in the time scales – with hypoxia setting up in summer in response to springs inputs, for example, whereas other responses might be more immediate. In order to explore this topic further the scenarios chosen for model runs could be expanded to address variation in both the timing and location of phosphorus inputs, thereby generating a suite of responses that will allow one to determine how the season and the location of inputs affect phosphorus concentrations, and , consequently, potential eutrophication impacts. (The current proposed suite of proposed model runs only varies phosphorus species and total load.)

I also have suggestions/comments regarding phosphorus species:

1. Dissolved reactive phosphorus (DRP) – Nutrient loads are almost always driven by discharge as opposed to concentration, and the observed increase in DRP loads in recent years is very likely a result of an increase in discharge. In particular, the high DRP load in spring of 2011 occurred at a time of very high discharge and this observation exerts undue influence in regressions when it is the final point in the series. It would be useful to evaluate trends in discharge to see if that is the main reason for the observed phenomenon. It would also be useful re-examine the discharge/DRP load relationships with additional years to see if these trends have continued. In addition, it would be instructive to examine the relationships between discharge and both TP and DRP concentration to evaluate the extent to which concentrations vary with discharge (in other words, is the increase in loads really just due to an increase in discharge or is the concentration also changing?) Finally, plotting the concentration of DRP over time and its concentration as a percentage of TP would allow one to determine whether there has been a shift in constituents.

2. The primary focus of this effort is on the stimulation of primary producers by inorganic P. However, heterotrophs (and some phytoplankton) take up DOP. It would be interesting to evaluate whether DOP concentrations, or their percentage of TP, have changed over time, and whether there is any relationship between DOP loads and the eutrophication response indicators. Understanding this might be helpful with regard to determining whether these alternative pathways are worth evaluating further.

References:

Mallin, M.A., Johnson, V.L., Ensign, S.H., MacPherson, T.A., 2006. Factors contributing to hypoxia in rivers, lakes, and streams. Limnology and Oceanography 51, 690-701.

Comments from Dr. Ammerman

Eutrophication Response Indicators

(1) Overall phytoplankton biomass as represented by chlorophyll a -

- Basin-specific, summer (June-August) average chlorophyll concentration

(2) Cyanobacteria blooms (including *Microcystis* sp.) in the Western Basin –

- Maximum basin-wide cyanobacteria biomass (mass dry weight)
- Summer total basin-wide cyanobacteria biomass (mass dry weight integrated over summer bloom period)

(3) Hypoxia in hypolimnion of the Central Basin –

- Number of hypoxic days
- Average areal extent during summer
- Average hypolimnion DO concentration during stratified lake conditions

(4) *Cladophora* in the nearshore areas of the Eastern Basin –

- Stored P Content

Comments: The challenge with response indicators for models is keeping them simple and easy to measure, and at the same time also of maximum utility for modeling and prediction. The current response indicators are relatively simple and easy to measure, but may no longer be of sufficient utility to address the eutrophication problems of Lake Erie. Indicators 1 and 2 should be measured with remote sensing satellites whenever possible, as is currently being done by the Lake Erie HAB Bulletin, and new remote sensing tools to distinguish *Microcystis* and other cyanobacteria species should be developed. Different cyanobacteria respond differently to nutrient concentrations and availability, some can fix nitrogen, for instance, so information on specific types is important. This is particularly important in western Lake Erie where cyanobacterial blooms are the major problem.

The hypoxia indicators appear appropriate, though better ones may eventually be developed. The areal extent of hypoxia is often used as the major indicator in estuarine and marine areas (Long Island Sound, Louisiana Coast), as well as Lake Erie. This indicator is most useful where repeated hypoxia surveys can be made throughout the summer, such as in Long Island Sound. In locations like the Louisiana Coast, however, only one comprehensive hypoxia survey is made every summer. The total volume of hypoxia has often been considered for measurement, as it would provide a better measure of total oxygen consumption compared to the parameters above, but it is usually dismissed as too challenging in terms of resources or technical capacities.

It is clear that more work is needed to develop indicators that can predict *Cladophora* and the phosphate that it stores, as described in the Draft Technical Approach. Again, remote sensing tools might be useful to at least estimate *Cladophora* biomass. Such estimates would still require appropriate assumptions to convert remote-sensed chlorophyll to biomass and phosphorus content.

While phosphorus is clearly the dominant limiting nutrient in Lake Erie, I do think it has been overemphasized, and the current increase in *Microcystis* blooms and other challenges indicates that the ecosystem is more complex than previously believed. Total phosphorus levels have not increased, yet *Microcystis* blooms have. Part of this is due to the invasion of zebra and quagga mussels and other invasive species. However, not all cyanobacteria are alike, as mentioned previously, *Microcystis* does not fix nitrogen but appears to be well-equipped to scavenge low concentrations of dissolved inorganic and organic phosphate. Recent molecular and other data suggest that nitrogen should also be considered in discussions of nutrients, at least more than has been to the present. In addition, efforts to provide time-series monitoring of nutrient concentrations (both nitrogen and phosphate) with automated instrumentation should be considered.

Comments from Dr. Bartell

Question 2: *Please comment on each of the models chosen to evaluate the eutrophication response in Lake Erie? Are the models appropriate for representing the eutrophication response indicators? Do the models reflect the best available scientific knowledge?*

Detailed comments will be developed to focus initial discussion of the individual models for the upcoming December 10 meeting. The subsequent and necessary evaluation of each model will require referring to original literature descriptions of each model, which will not be completed prior to the meeting.

In more general terms, as summarized in Table 1 of the Technical Approach document, only one of the proposed models, the ELCOM-CAEDYM 3D, addresses all of the identified eutrophication responses for Lake Erie. Several of the models appear useful for describing at least two response indicators. Two of the models are relevant only to the Cladophera response. The WLEEM 3D model appears comprehensive and more broadly applicable, but focuses only on the western basin of the lake. Taken as a collection, the models outlined in the Technical Approach appear to present the scientific capability to examine the Lake Erie eutrophication responses in relation to the stated modeling and management objectives.

Evaluating the state of the science represented by the models will require detailed examination of the governing equations, process formulations, parameterization, and key underlying simplifications and assumptions. The nature and source of input data (e.g., initial conditions, environmental factors, P loads, hydrodynamics, etc.) will also have to be examined in detail to characterize the strengths and limitations of the proposed models. Previous model performance (i.e., accuracy, precision) will also importantly contribute to the evaluation of the state of the science represented by the proposed models. Presumably these activities will be completed during the course of the model review.

In evaluating the models, it will be important to distinguish between fundamentally different modeling approaches and different coding of essentially the same governing equations (e.g., continuity equations, bioenergetics-based growth) and process formulations (e.g., Monod nutrient limitation, Holling type II trophic functions). To what extent are the models really different? Answering this question will also be important in evaluating the intended ensemble modeling approach: for example, the ensemble might be justifiably narrowed to a subset of the ten proposed models without sacrificing technical capabilities in addressing the overall modeling objectives. The ensemble approach is fundamentally appealing. However, substantial effort might be invested in understanding the similarities/differences among the models, rather than focusing finite resources on defining justifiable and acceptable P loadings to Lake Erie. “Right-sizing” the ensemble might help reduce this potential drawback of the ensemble approach.

Comments from Dr. Chen

4. How can we ensure that the phosphorus concentrations and loading targets are internally consistent with respect to the ecological response indicators of concern?

The use of target values of phosphorus load and phosphorus concentration to address the eutrophication issues in Lake Erie has been the most implementable approach for this complex and multi-factor problem. However, the use of this approach with the target values set in the 1980's appears to be no longer sufficient to limit algal blooms and hypoxia in the lake. It is thought that this is potentially because the spatial resolution of earlier models is no longer adequate and/or the ecosystem has changed in structure and function since the earlier modeling was conducted. The poor efficacy of current target values suggests that other factors have complicated the relationship between phosphorus loads/concentrations and algal blooms and hypoxia/anoxia. Many of these other factors are discussed in the reports produced since 2010.

The efficacy of new target values that would be based on the new ensemble of models is dependent on whether the models capture the critical variables that drive the eutrophication response indicators (ERI) (questions 2 and 3 of the SAB charge). The suite of models appear to include most of these factors that influence or relate to non-toxic algal blooms and Harmful Algal Blooms (HABs) in the western basin, hypoxia/anoxia in the Central basin, and *Cladophora* blooms in the Eastern basin. However, as shown in Table 1 of the document, "Annex 4 – Draft Technical Approach (final Nov 12)", each of the models addresses some ERIs but not all. Moreover, the ERIs are specific to basin or offshore/inshore dynamics suggesting that the modeled responses will also be spatially distinct from one another. To ensure that P concentrations and loading targets are internally consistent with respect to the ERIs, the differences in the model outcomes will need to be resolved and the spatially distinct processes identified in the models will need to be linked.

The expectation that new loading and concentrations targets will be "internally consistent" with the intended outcomes of the ERIs assumes that external P loading alone is driving the ecosystem responses associated with the ERIs, and does not necessarily account for other forcing factors, changing environmental conditions, and internal sources of P to the L. Erie basins. Among the underappreciated factors that could confound this internal consistency are: a) the relationship of total P concentration to concentrations of bioavailable P (dissolved reactive P (DRP); b) groundwater inputs of P and internal recycled P; b) unexpected effects of increased temperatures and episodic storm events in algal blooms; c) temporal and spatial linkages of P loading, light, temperature, and algal blooms; c) the role of increased resting stages of HABs in sediments influencing the increased frequency and intensity of HABs; d) role of nitrogen (N) and N:P ratios in controlling algal blooms. Most of these factors are mentioned in documents produced since 2010 but not necessarily included in the ensemble of models. Each of the four ERIs is discussed in relation to these factors.

1) Overall phytoplankton biomass (basin specific summer average):
Primary and secondary productivity will determine standing phytoplankton biomass and existing data for the western basin suggests that primary production is related to the bioavailable from of P, DRP which is difficult to measure accurately, rather than the more easily quantified total P. The

relationship between total P and DRP concentrations is variable and each varies considerably within and between watersheds of western L. Erie. Correctly estimating the bioavailable fraction of DRP is critical to the effectiveness of P targets and concentrations. In fact, setting both total P and DRP loading targets and concentrations would seem to be most effective as suggested in the IJC LEEP report of Feb. 2014.

Increased temperatures over time due to climate change will also likely favor increase primary production, earlier stratification, and more intense hypoxia in the central basin. Increases in episodic storm events will also cause greater delivery of P from the watershed into the rivers both from agricultural land and urban stormwater runoff, both non-point sources that are difficult to quantify. It isn't entirely clear whether these long term changes are incorporated into the models.

2) Cyanobacteria blooms in western basin (max. basin-wide biomass, summer total basin-wide max):
Cyanobacteria blooms, particularly *Microcystis aeroginosa*, in the western basin already appear to be responding to the increases in DRP which are not associated with increases of total P. Thus, the importance of quantifying DRP and controlling its inputs to the western basin is particularly critical to controlling Harmful Algal Blooms (HABs). In addition, the relationship of N to P concentrations determines the switching of *Microcystis* blooms to *Anabaena*, a N fixing species. It appears that N limitation is related to the decline of *Microcystis* which does not fix N. In some studies, toxic *Microcystis* has been shown to be limited by both N and P. Monitoring N and N:P ratios would be useful in understanding the nature of the HABs.

In other systems containing HABs, the increasing frequency and intensity of blooms is enhanced by increasing densities of resting stages of cyanobacteria in sediments which fuel even more intense blooms. In the case of Microcystis in L. Erie, the population in the Maumee River appears to be different from the Western basin population (Kutovaya et al. 2012). Therefore, the blooms in the western basin are not being washed in from the Maumee River. The positive feed back loop of the increase of the "egg bank" in either the river or the lake does not appear to be accounted for in any of the current models.

3) Hypoxia in hypolimnion in central basin (# of hypoxic days, area extent, average hyplimnion concentration during stratification):
Internal loading of P is highest in the central basin due to the long retention time of this basin relative to the western basin (635 days v. 51 days). The differences in flushing rates of the two basins along with seasonal changes in water influx from tributaries and water temperature create complex space and time relationships that determine the presence and severity HABs. P loading from tributaries occurs in spring while the later warming of the Western basin water stimulates algal resting spores to germinate and cell division to accelerate bloom creation. This results in the production of organic carbon that fuels the hypolimnetic hypoxia in summer the Central basin. The linkages between space and time will determine whether optimal conditions exist for bloom formation and the resulting organic carbon load to sediments in the Central basin. These links between space and time need to be captured in the modeling in order for bloom formation and hypoxia events to be optimally predicted. Only those models that capture multiple basins over monthly or daily time steps will be able to address these space/time links.

Lastly, the hypoxia ERI is not only an important indicator of poor ecosystem condition but also has the most important impacts on fish habitat as well as increasing the potential for recycled inputs of P from the sediments.

4) *Cladophora* in nearshore areas of the Eastern basin (stored P content):
The ability of the model outputs to be consistent with actual *Cladophora* proliferation in the Eastern Basin will be dependent on both the water clarity created by the Dreissenids and the P bioavailability coming from external and internal sources upstream in the Western and Central basins. This dependence on boundary conditions of the Eastern Basin makes the outcome of this ERI difficult to predict. Only one of the models (Lam's 9 Box model) appears to incorporate interbasin transport between the Central and Eastern basins. The Auer model relies on boundary conditions between the Central and Eastern basins from the Leon model. Since there was not a description of the Higgins model in the Model Overview, it is difficult to know how the upstream affects will be addressed. The spatial connection of this ERI to the processes in other basins needs to be rigorously made in the modeling.

Comments from Dr. Connolly

Please comment on each of the models chosen to evaluate the eutrophication response in Lake Erie? Are the models appropriate for representing the eutrophication response indicators? Do the models reflect the best available scientific knowledge?

The modeling approach is conceptually appealing. Applying several models to inform the choice of loading targets and allocations will allow the team to understand an array of relevant metrics affected by changes in phosphorus loading and the uncertainties and biases inherent in the models.

The challenge will be to choose a set of models that meet the objective of being good predictors of the improvements in water quality metrics as phosphorus loads are reduced. The stated approach appropriately specifies that the models take into account ".. the bioavailability of various forms of phosphorous, related productivity, seasonality, fisheries productivity requirements, climate change, invasive species and other factors, such as downstream impacts, as necessary." It seems appropriate that the candidate models be evaluated with regard to these factors. The stated evaluation criteria do not explicitly include such evaluation. Several of the factors seem to be critical to modeling water quality responses to load reduction. For example, the ability of a model to account for the impacts of changes in phosphorus form that may be associated with load reductions will be important if the options for load reduction produce differing mixes of phosphorus forms.

Another point to consider with regard to model evaluation criteria is model performance at the low end of the range of phosphorus loadings associated with calibration and validation. Some of the models chosen for evaluation may exhibit biases in predicting water quality response to lower phosphorus loading. Such bias may not be evident in the evaluation criteria as currently structured. It is conceivable that a model may exhibit a correlation between measured and modeled water quality metrics that exceeds 50% despite have a consistent bias at the low end of phosphorus loading. Since the objective is to use the models to predict improvements in water quality metrics in response to load reductions, consistent bias at the lower loadings would seem to be unacceptable.

Whether the candidate models reflect the best available scientific knowledge seems like a question inconsistent with the idea of using models that range from fairly simple constructs that make numerous simplifying assumptions to spatial explicit models that include arrays of state variable and processes. I am more comfortable addressing two variants of this question. Does each model have sufficient predictive ability to provide reasonably accurate predictions of responses to load reduction? Does each model conform to the principles of Occam's Razor? Is the complexity of the model warranted given our understanding of the included processes, their impact on predictions and the availability of lab and field data to adequately constrain the defining parameters? In addressing these questions, I suggest focusing on the intention to predict future conditions and the ability to define the forcing functions that affect the processes.

My initial review of the candidate models leads me to believe that not all meet the requirements that I have posed here. I look forward to more specific discussions on the individual models in our December 10 meeting.

Comments from Dr. DiGuilio

Full disclosure – I am an environmental toxicologist, not a modeler nor an expert on eutrophication.

Overall, I found the draft document to be well-written and that it made a strong case for the approach taken, appeared to make appropriate use of existing data, and provided a good background for the various models that are being evaluated. The supporting documents also provided useful background information concerning the history of the issue and its management, and explanations for the increase in problems since the 1990's. (I just wish that somewhere there had been included a good map of Lake Erie, indicating the key rivers feeding the lake, municipalities etc.)

One potential gap appeared to me to be a better, quantitative explanation of the Lake Ecosystem Objectives (page 2). I think linking phosphorous loading and concentrations to "Eutrophication Response Indicators" is an excellent approach. And the indicators selected appear very well suited to the critical issues posed by eutrophication in Lake Erie (i.e., hypoxia, algal biomass, algal species composition, cyanobacteria biomass, and trophic status). However, the targets for these indicators were left quite vague. For example – "maintain cyanobacteria biomass at levels that do not produce concentrations of toxins that pose a threat to human or ecosystem health in the Waters of the Great Lakes." Statements of this nature were used for most of the indicators (except trophic status).

I was expecting that further down in the document, some actual measures of what are considered acceptable levels of cyanobacteria etc. would be provided, but they were not. In terms of extending the modeling efforts into management decisions and monitoring programs, this would seem important.

Comments from Dr. Diaz

The weight of evidence as to what key factor is driving eutrophication in the Great Lakes and degrading Lake Ecosystem Objectives points to excess phosphorus. A management program directed at a combination of controlling loads and concentrations of phosphorus will provide the best strategy for meeting the Lake Ecosystem Objectives.

Weight of evidence is a reasonable approach to setting numerical nitrogen concentration criteria. This approach when combined with best professional judgment can be a powerful tool for drawing conclusions in many areas of water quality management.
How was Spring mean phosphorus concentration selected for substance objectives?

What are the minimum data requirements for an adequate characterization of a Spring mean?

The models seem to range widely in complexity. How does this affect minimum dataset required and strength of predictions?

Are the six models that predict hypoxia able to predict number of hypoxic days, area of summer hypoxia, and oxygen concentration?

In the 9-box eutrophication model, why was the settling rate changed?

Successful ensemble model analysis will depend on uniform presentation of output. What steps will be taken to insure a common format for curves and data tables?

It seems each model works at a different temporal and spatial scale, and each model predicts a different subset of response indicators. How will these differences be dealt with?

What metrics will be used to determine model performance and accuracy?

Comments from Mr. Endicott

1. Please comment on whether the eutrophication response indicators proposed sufficiently address and provide the scientific foundation for the Lake Ecosystem Objectives for Lake Erie. During your evaluation of the eutrophication response indicators, identify other metrics appropriate for measuring eutrophication response in Lake Erie and other Great Lakes that should be considered, and whether there is a method (model) available to measure this response.

I am struck by the absence of phosphorus concentration as an indicator, since it was the mass balance for phosphorus and the load-concentration relationship that was so central to the prior effort to establish loading targets to control eutrophication in Lake Erie and the other Great Lakes. Also, for at least the first generation of models, P was often the most accurately predicted state variable. I suppose the issue is whether phosphorus concentration is an ecological response indicator.

The 4 response indicators (basin-specific chlorophyll a biomass, Western Basin cyanobacteria biomass, Central Basin hypoxia and Eastern Basin Cladophora) appear to address the Lake Ecosystem Objectives for Lake Erie. It may be easier to relate some of the response indicators to ecosystem objectives than others. For example, high cyanobacteria biomass does not always result in high toxin concentrations. However, limiting biomass should prevent cyanotoxin occurrences.

The phytoplankton biomass response indicator uses summer average chlorophyll concentration as the metric. Is this concentration measured in the photic zone, and integrated water column, or a surface sample?

The Cladophora response indicator (stored P) makes sense in the context of Canale and Auer's Cladophora model, but wouldn't biomass or areal extent be a more meaningful indicator for this nuisance? I am concerned that it may be difficult for a broad audience to understand the significance of this indicator.

2. Please comment on each of the models chosen to evaluate the eutrophication response in Lake Erie? Are the models appropriate for representing the eutrophication response indicators? Do the models reflect the best available scientific knowledge?

The Draft Technical approach has selected an appropriate set of models for Lake Erie P load-response modeling. There are similarities and differences between the 10 models in terms of complexity, spatial/temporal resolution, processes and state variables. Diversity between the models is desirable and useful. The cyanobacteria models (Obenour and Stumpf) seem fairly similar, as do the Cladophora models (Auer and Higgins). Both DePinto and Leon et al. models are both highly spatially resolved and complex in terms of processes and numerous state variables. Chapra and Lam 9-box models are comparatively simple. The Zhang and Rucinski models seem to fall somewhere in-between.

Table 1 in the Draft Technical approach indicates which response indicator can be predicted by each model. Assuming that each model has been calibrated and confirmed for the corresponding response indicators, then it should be appropriate for representing those response indicators in this exercise.

In terms of reflecting best available scientific knowledge, certainly each model has its strengths and weaknesses. I still need to review Kim et al. (in press) “A commentary on the modelling of the causal linkages among nutrient loading, harmful algal blooms, and hypoxia patterns in Lake Erie” to better understand the structure of each model. The Draft Technical approach identifies the importance of model validation as an objective test of acceptability and utility.

I could not find the reference (Lam et al. 2002) discussing application of the 9-box model to pre- and post- zebra mussel infestation.

3. Please comment on the appropriateness of the ensemble modeling approach to examine the suite of eutrophication response indicators. Are the models included in the ensemble, when used either singly or combined, sufficient to provide a scientifically grounded basis for the required update of phosphorus load targets for Lake Erie?

The ensemble modeling approach is based on using predictions from each model to derive load-response curves, using a standardized set of P loadings and hydrologic/meteorologic conditions from the 2008 reference year. Obviously, there will be differences between the load-response curves produced from different models, and it is not entirely clear how these differences will be interpreted and communicated. Some other issues to consider are whether the load-response curves are based on steady-state predictions (if that matters), the use of inter-annual variability to explore confidence bounds for the load-response curves, and consideration of spatial/temporal allocation of P loads.

4. An anticipated outcome of the modeling exercise is to better understand and quantify what types of conditions would be expected in the lake based upon different levels of phosphorus loading, and to use that information to inform selection of phosphorus loading targets needed to meet the nutrient Lake Ecosystem Objectives. The phosphorus loading targets could be converted to concentration targets, particularly for river mouths/nearshore zones. Please comment on efficacy and value of establishing target values for both phosphorus loads and concentrations in order to meet to the Lake Ecosystem Objectives. How can we ensure the phosphorus concentration and loading targets are internally consistent with respect to the eutrophication response indicators of concern?

I’m not sure I completely understand this question. If the loading targets are intended to attain desirable levels of response indicators in lake basins (open water), then why would it be reasonable to expect that converting these loadings to concentration targets would be protective for localized impacts in river mouths and nearshore zones? Targets expressed in terms of TP are probably the easiest to confirm by monitoring, but you still need to monitor the response indicators, for reasons that include generating data to improve the models.

Comments from Dr. Heath

1. Eutrophication Response Indicators: The ERI proposed are reasonable choices for assessing attainment of the Lake Ecosystem Objectives. Total phytoplankton biomass, cyanobacterial biomass, central basin hypoxia, and *Cladophora* biomass not only are at the heart of the concern but also have a sufficient data history to allow comparison with the past. I am concerned that none of these variables provides an insight to the biochemical status of the cyanobacterial and algal assemblages. The models are predicated on the assumption that these communities are strongly P-limited and that external loading is effectively the sole source of available phosphorus. With increasing frequency since 2002, reports have been given that suggest this is not the case, especially late in the growing season. The recent reports by Chaffin regarding the nutrient status of *Microcystis* as being P-limited in the spring but N-and-P co-limited or N-limited during the July – September interval, is cause for concern regarding this assumption. Also, I am concerned that microcystin concentrations are not included in the ERI, as they are increasingly a major source of concern for managers.

I am concerned with some of the metrics chosen.

A. Chlorophyll *a* has a long history of collection and is a good choice in that regard. But it should be recognized that it is an unreliable surrogate of algal biomass. The chlorophyll content of many phytoplankton species is a variable function of nutrient and light conditions. A better measure would be a direct measure of total algal biomass, especially to be compared with cyanobacterial biomass.

B. Cyanobacterial biomass (dry weight) is an excellent indicator, but it can be difficult to measure in the presence of mixed phytoplankton communities. Also, given the patchiness that generally occurs in cyanobacterial blooms, I question the reliability of "basin-wide measures" unless they can be done by remote sensing techniques. As mentioned above, I'm surprised that <u>microcystin concentration</u> is not included as an ERI. I can appreciate that it is expensive and less than straightforward to measure and that besides cyanobacterial biomass, those factors salient to its appearance are unclear, making it problematic to model predictively, but it is at the heart of the concern for those who depend on Lake Erie as a drinking water source.

C. Hypolimnetic Hypoxia in the Central Basin. At what DO is "hypoxia" defined? It could be 4 ppm because that is the generally accepted limit of fish tolerance. Or it could be <0.4 ppm at the sediment-water interface, because of the Fe and Mn biogeochemistry involved at that point. Both of these measures (DO ≤ 0.4 ppm; DO ≤ 4 ppm) could and should be monitored and modeled. Determining the areal and temporal extent of these measures is good. The usefulness of "average hypolimnetic DO during stratification" is unclear to me. The only measure that I'd suggest adding would be the depth at which DO ≤ 4 ppm, because that would estimate of the extent to which deep, cold water fish habitat is constrained.

D. I am uncertain whether this means total P in *Cladophora*, or P-storage compound stored in the alga that is readily hydrolyzed and released. Does this variable have a sufficient history of reliable collection? Is it currently being monitored regularly and reliably sufficient to be used as an ERI? A more biochemically meaningful measure would be P-debt of *Claldophora*, a measure of how much more P the alga could assimilate from the environment. However, I am unaware that this measure is being taken or could be included in any of the models.

2. Comments on each of the models proposed:

i. Chapra TP Model: The strength of this model is that it relates well to the earlier work done in the 1970s. It is based on the implicit assumption that internal loading will be negligible in determining average annual TP concentration in the various segments of the lakes. I agree that expanding the model to include predictions of Chl *a* and central basin hypoxia would be beneficial and should be encouraged. My concern with phenomenological models such as this is that they work as long as the system remains unchanged, but when the ecosystem changes in a major way, especially changes that alter internal loading (e.g. zebra mussels, internal recycling, global climate change) they may no longer be reliable. The intent to assess model uncertainty and a sensitivity analysis is an especially good idea here. An evident weakness of this model is that it is unable to be extended to the nearshore.

ii. WLEEM: I like fine-grained, process-based models because they are good for tasks such as this and can be readily revised by recent scientific findings. From the descriptions given, it seems that this model is built on a wealth of hydrodynamic, chemical and biological science. The strengths are that the A2EM sub-model can be used to predict a variety of events in the food web and thereby be validated in a number of ways. I like the inclusion of zooplankton grazing on edible algae, as it provides a realistic process-based view of how standing algal biomass occurs. I especially like the inclusion of benthic events, including dreissenid bioenergetics and nutrient cycling. It also has been extended to *Cladophora* growth. The description indicates that it has been sufficiently calibrated and validated with appropriate field data sets. My main misgiving with such complex models is that they can propagate errors and be unstable. An uncertainty assessment and sensitivity analysis should address these issues adequately. The proposed procedure for validating and applying the model through sensitivity analysis is very well outlined for this model.

iii. ELCOM-CAEDYM: The model is not described very well in the materials given. As with WLEEM, it is a process-based complex model. It will be interesting to see how the ERI response curves of this model and WLEEM compare. I assume that both these fine-grained models can be aggregated appropriately to allow comparison with more aggregated models.

iv. EcoLE: As with those above, this is process-based complex model that includes hydrodynamics. I like the explicit inclusion of suspended sediment release/uptake of P and N-dynamics, including nitrification and denitrification. This should provide another useful prediction of events in the open waters of the lake. It is unclear how useful this model would be in nearshore communities and especially in addressing *Cladophora*. I like that this model has been used successfully to predict western basin hypolimnetic oxygen depletion rates. My main concern with the model is that it apparently has only PO_4 (approximated by measures of SRP) as a variable, without considering other forms of P that have been shown to be available.

v. UM/GLERL: As I understand it, this model will give a yearly prediction of cyanobacterial bloom size. It does this with monthly Maumee River discharges. I like the use of Bayesian inference to determine the model parameters. This model is most likely to provide a comparison with the Chapra model, but I question its usefulness in nearshore situations. On p. 23 they note that they will recalibrate the model to "bioavailable" phosphorus loads. This sounds good, but they fail to identify how "bioavailability" will be determined or modeled. There is a significant if aging literature on "bioavailable P" that indicates that this variable is a function largely of how long the assay runs; that is, it is an indistinct variable. Also, model predictions are for an overall bloom size in MT. The problem is that management concerns are more fine-grained (e.g. size of bloom at water intakes).

vi. NOAA HAB Forecasting Model: This model provides a prediction of cyanobacterial bloom severity. It is unclear from the description given whether "severity" is predicted as the greatest acute size (amount of standing crop at bloom peak) or integrated bloom size (integrated from 1 July through 15 Sept, for example). Both measures can be useful, although I suspect that lake ecosystem managers are more concerned about the size of the bloom at its peak, as this would likely contain the greatest amounts of microcystin. I agree with the final sentence of the description: the model should include July, as that is when the blooms begin to form.

vii. Environment Canada 9-Box Eutrophication Model: I like the tie with the past in the use of this model, and I like the recalibration achieved post-dreissenid invasion. I like the inclusion of meteorological data into the model and that it implicitly includes internal loading. The representation of P-dynamics is thin. Generally, TP = particulate P + soluble P (P which passes through a 0.45 µm or 0.1 µm pore-size filter); and the soluble P is then articulated into SRP and SUP (soluble unreactive P, which is often designated "organic-P"). SRP approximates PO_4, which is readily assimilated by bacteria and phytoplankton; SUP can be biologically available, as well, under conditions that release PO_4 from it. It is much more useful to designate particulate P as a state variable because much that is phytoplankton biomass or detritus. To lump particulate P and SUP can be very confusing because SUP is a variable portion of this lump (i.e. SUP can at times be much greater than particulate P, and not so at other times).

viii. One-Dimension Central Basin Hypoxia Model: This model seems to be calibrated to achieve reliable predictions of SOD given external loading as the sole forcing. I am surprised not to see an internal loading function, given that this model expressly models those conditions under which internal loading would be expected.

ix. Great Lakes *Cladophora* Model: This model seems to be appropriately calibrated to yield the information required. It seems that this model essentially extends the ELCOM-CAEDYM model to prediction of *Cladophora* biomass, etc. It would be useful if this model contains physiological or biochemical variables that could be used to validate the outcomes (alà Healey and Hendzel). It would be especially good if the model predicted either the amount of Storage-P or it's complement, P-debt.

3. Ensemble Modeling Approach The ensemble modeling approach makes a great deal of sense. To my eyes, there is yet much to be learned about the Lake Erie ecosystem processes. I believe strongly that an ecosystem approach is needed because it is in the context of an ecosystem that these changes are occurring. Specifically, it is unclear whether this is "the same old problem" revisited on us, or a new problem with some familiar symptoms, yet fundamentally different. The former suggests that these recent problems arise largely from external loading alterations (e.g. altered agriculture practices) and so can be managed by managing external loading. The latter view questions whether that will succeed, and suggests that internal processes may have become more important than they were in the past (internal loading, increased regeneration rates, accelerated warming due to GCC).

If in a debate, I could argue either side, but I'm inclined to favor the latter view, because there is too much evidence to suggest that the "everything is P-limited all the time" view is incorrect. I believe that the models chosen are based on the best science available and that is currently broadly accepted. There is little in these models that I would consider "offbeat" or controversial. As a biogeochemist, the models that I prefer are those that are best articulated biogeochemically (iv, ii, and iii, in that order). I think that successful fine-grained models will in the end be more useful than highly aggregated models because many management concerns are placed-based (e.g. drinking water intakes). I like the inclusion of meteorologically driven models (vii and viii) because in many ways this is a storm-driven ecosystem. I like that the models chosen represent a wide variety of model complexity. I don't see any of these models as being the *sine qua non* in addressing the current issues or predicting the ERI.

The ensemble I believe represents a reasonable synthesis of the best available science regarding the function of the Lake Erie Ecosystem. These models depend on a body of knowledge built up over the past 30 years, but it is a growing body of knowledge. The ensemble approach will provide response curves akin to those generated by the various GCC models use to predict future climates. I expect that as various models provide a given ERI response curves, the curves will differ. The collective judgment will be the response envelope of the set of response curves. I believe that managers will be able to look at that response envelope and determine "best case – worst case" scenarios. I see this as far preferable to selecting a single model to make these predictions and response curves.

4. Efficacy and Value of Establishing Target Loads These models singly and collectively will provide P-loading ERI response curves from which targets can be set. They can't do otherwise; that is what the models are hardwired to do. I believe that the Model Evaluation and Application plan is sound. I question whether the "correlation suggesting that 50% or greater of the variability in the relationship between TP or DRP and the ERIs" will satisfy management needs. Whether those target loads achieve their intended purpose, remains to be seen, of course. Achieving the intended LEOs will depend on (1) the success of the science behind the models (correctness of the assumptions, collective wisdom in constructing the models, etc.) and (2) the political and social will to constrain external P-loading to the identified target loads. I believe there are reasons to question both of these possibilities.

First, the target loads will be successful if phytoplankton communities and *Cladophora* stands are strongly P-limited in their growth and physiological responses and if external loading is the major source of available P (vs. internal loading and recycling). Because of the success in reclaiming the Lake Erie ecosystem in the 1980s by controlling P-loading, and because of the wealth of data from bioassays and biochemical indicators that phytoplankton communities were strongly P-limited at that time, attention was appropriately focused on understanding P-dynamics in Lake Erie phytoplankton communities. Relatively little attention has been given to the interplay among P and other elements (especially N and C) and processes in food web architecture (e.g. grazing food chain vs. microbial food web). What is known from studies since the 1980s is that P-limitation of phytoplankton provides a graded response in physiological indicators, rather than an all-or-nothing response. Since at least 2000, the physiological indicators of phytoplankton community P-limitation have weakened (P-debt, Phosphorus Deficiency Index, Alkaline Phosphatase Specific Activity, SRP *in situ* concentration, etc.). There are increasing reports that while western basin phytoplankton communities are P-limited in the spring (May – June), they become N-and-P co-limited or N-limited in July and August. These reports are based largely on Schelske-type nutrient amendment bioassays. For these reasons I emphasize the need for adaptive management strategies in revising these models in response to new scientific findings. Furthermore, there are reports that appearance of microcystin in a cyanobacterial bloom is most strongly correlated with N-limitation, rather than P-limitation.

Ensuring that the phosphorus concentration and loading targets are internally consistent with respect to the ERIs of concern will require that the ERIs be monitored and continue to respond in the lake as predicted by the models. The stated plan in the Draft Technical Approach seems adequate for this task. Additionally, the physiological indicators of P-limitation should be monitored if and when external loads are decreased. The anticipated results would be an increase in P-limitation physiological responses as external P-loading decreases.

Secondly, I believe it is wise to consider that the external P-loads will be decreased to target loads not because the science mandates it but because social, economic and political will permits it. What if the LEOs can be achieved only by a 75 percent reduction in current P-loading? What social and political changes will need to be made to achieve that and at what cost? Discussions leading to GLWQA in 1972 and 1978 were permeated more by socio-politico-economic considerations than by science, as I recall. It may well be that a new kind of ecosystem modeling effort will be required to achieve progress. It may be necessary to develop models that include social, political and economic behavioral parameters, as well as biogeochemical parameters. Such models would identify the lake ecosystem outcomes in the context of political decision making necessary to make it happen. Such decisions would be an optimum between the benefits accrued from attaining LEO, the costs of achieving socially acceptable changes in behavior (much as relinquishing the use of P-detergents was in the 1970s), and the economic and political feasibility of making those changes.

Comments from Dr. La Point

Comment on whether the eutrophication response indicators proposed sufficiently address LEOs for Lake Erie.

(1) Overall phytoplankton biomass as represented by chlorophyll a -
Basin-specific, summer (June-August) average chlorophyll concentration; a traditional indicator of lake trophic status (i.e., oligotrophic, mesotrophic, eutrophic).

(2) Cyanobacteria blooms (including Microcystis sp.) in the Western Basin –
Maximum basin-wide cyanobacteria biomass (mass dry weight)
Summer total basin-wide cyanobacteria biomass (mass dry weight integrated over summer bloom period)

(3) Hypoxia in hypolimnion of the Central Basin –
Number of hypoxic days
Average areal extent during summer
Average hypolimnion DO concentration during stratified lake conditions

These response indicators are appropriate and should cover the ORP loadings, as well as key algal growth indicators. In the list of models, the ones (in my opinion) to be given the highest priority are those that integrate loadings, total concentrations, and benthic P turnover as a function of algal-Dreissena interactions. The models that incorporate loadings, turnover, and consequences (WLEEM, ELCOM-CAEDYM 3D, and EcoLE) should be given high priority. Also, as the ultimate issues are loadings and benthic turnover, the models should link to empirical data developed by related studies (e.g., Ohio Lake Erie Phosphorus Task Force II Final Report, Nov 2013).

One aspect of the overall proposed modeling exercise seems to call for a “model Tsar”. The inputs for the various models call for similar empirical data: concentrations, loadings, Cladophora growth, etc. Hence, as various inputs are developed and added, it seems a strong effort should be made to compare model outputs and use bootstrapping techniques to find out where the model variances are: what processes or conditions have the largest influence on model predictions? Which models have the largest variance and what inputs can be gleaned from other models to enhance the focus? Finally, there is always the need for frequent ground-truthing, to compare model predictions to reality (diatom vs non-diatom fluxes). One key approach may be to incorporate satellite surveillance approaches (MERIS, Stumpf, et alia 2012)

In the reports from USEPA, Canada and Ohio EPA, the flux of benthic P arises often. As the overall P concentrations are not increasing, the focus must come to understanding the algal-Dreissena-DRP connection. One piece of evidence for this is the fact that, although average P concentrations have not increased, the consequences have. Hence, maybe annual average loadings should not be as highly emphasized (e.g., the Chapra model)? Wetzel 2001 (and references therein) note the strong influence of bacterial decay in benthic cycling of orthophosphorus. This key dynamic should be included in at least one of the selected models and incorporated overall.

One recommendation I have for the question of "response indicator sufficiency" is that the algal-bacterial-benthic P dynamics need to be taken into account. Measures of biomass (phytoplankton, Microcystis) are necessary, but not sufficient to understand the problems. One needs to incorporate measures of P uptake and deposition – even if empirically derived from production-respiration chambers established under laboratory conditions or, preferably, *in-situ* from within the western and central basins of Lake Erie. Such data would enhance the utility of the WLEEM model.

An approach may be to enhance the present efforts by incorporating a GIS-based approach, such as the USGS SPARROW modeling approach, to determine sources of P for Lake Erie basins. Although the reports from EPA and Ohio EPA have determined the loadings for tributary rivers, Detroit, Maumee,

ISSUES:
Storm flows; upstream controls on P contributions; urban contributions (lawn care); educating the public! Tertiary treatment for Detroit WWTP?

2. Comment on each of the models chosen to evaluate eutrophication response in Lake Erie.

As I mentioned above in Question 1, the models are good ones, as far as their design purposes go. I, personally, regard models of particular use that are of use in all three Lake Erie basins and that incorporate hydrodynamic inputs as loadings, as well as benthic fluxes among algae, water, and mussels. WLEEM and EcoLE stand out as two good models to use, specifically for daily loadings and nearshore wave-perturbation dynamics. From the USEPA Draft Technical approach, WLEEM…" includes such aspects as interactions between solids transport and lower food web dynamics, explicit modeling of growth and associated grazing/nutrient cycling processes associated with zooplankton functional groups, incorporation of a benthic algae growth (e.g., Cladophora sub-model), incorporation of a dreissenid bioenergetics/phytoplankton filtering/nutrient cycling sub-model for multiple age classes of two mussel species, and kinetic adsorption/desorption of orthophosphate to particulate inorganic solids from tributary loads and bottom sediment resuspension". In my opinion, this is a critical need to determine the consequences of benthic P fluxes. The EcoLE model also attempts to derive benthic P fluxes.

The ELCOM-CAEDYM model is also useful for planktonic production/uptake of P in the three basins. Complex models should require simulations to address variance components. Knowing these can lead to focused research on the parameter inputs to enhance model predictions.

The Obenour model (UM/GLERL Probabilistic Cyanobacteria Model) is particularly useful in linking Maumee River loadings to the western basin cyanobacterial blooms. It would appear to me that this model could also be applied to the major loadings stemming from WWSTPs on the Detroit River?

The NOAA (Stumpf) model has the value of being matched to satellite data. In any of the models proposed, satellite data should be used to ground-truth model predictions. The model predictions here can support/enhance those of the Obenour model.

The Environment Canada 9-Box Eutrophication Model is of high use, in my opinion. The usefulness stems from incorporating fluxes between epi- and hypolimnion and the daily steps in benthic biological activity. The short time-steps are important in understanding the relatively fast cycling of P by bacteria, algae and mussels.

The Central Basin Hypoxia model is of valuable use for planktonic responses. I would prefer to see it incorporate benthic fluxes, as these seem to be important for regenerating P in the lake.

In any case, models selected for use should be able to support and/or enhance predictions developed by the other models. This would be a great benefit of the "ensemble modeling approach." (see my reponse to Question 3, below).

In addition, I would suggest looking at the approach detailed in the SPARROW model by USGS (Ator, *et al. 2011)*. The use of basin-level data to discern loadings for N and P would be of use in Lake Erie, particularly in regards to P loadings from the Maumee and Detroit Rivers.

3. Comment on the appropriateness of ensemble modeling approach to examine the suite of eutrophication response indicators for Lake Erie.

I have addressed much of this question in the prior answers (particularly No. 2). However, in short, I think that the ensemble approach is a very good one and I highly recommend it. My only caveat is one that would recommend incorporating more empirical data to enhance model predictions (or "ground-truth" them). Measures of algal or bacterial biomass, alone, are not sufficient. Measures of biological activity are needed. For example, measures of benthic production/respiration would provide a solid index of the biotic interactions (e.g., algal-cyanobacterial-mussel) responsible for P re-introduction into the aquatic environment. From the readings, this is a key modeling need, as annual averages of concentration and loadings are not sufficiently predictive.

4. Comment on the efficacy and value of establishing target values for P loads and concentrations to meet Lake Ecosystem Objectives.

It is absolutely imperative, in my opinion, to come to understand the loadings from tributary sources. However, it is as important to understand the role the biota play in benthic P flux. As Wetzel 2001 (and recent papers summarized in the LaMP report) suggest, the amount of P available in a lake may be much more influenced by bacteria in the sediments than by tributary stream loadings. In Lake Erie, much of the P inputs are either agricultural (Maumee) or urban (Detroit). The P loadings are critical and it is recognized that the concentrations in tributary streams must be lowered (LaMP report). However, as the nuisance algae are continuing to increase, despite interim goals having been met, this calls for better information and knowledge of what is happening in the benthos. How do bacterial and algal production, Dreissena production and uptake, and subsequent bacterial decay all influence aqueous P concentrations? These remain key research needs – and are reflected in the LaMP report.

References:

Ator, S.W., Brakebill, J.W., and Blomquist, J.D., 2011, Sources, fate, and transport of nitrogen and phosphorus in the Chesapeake Bay watershed: An empirical model: U.S. Geological Survey Scientific Investigations Report 2011–5167, 27 p.

Ohio Department of Agriculture, DNR, EPA, Lake Erie Commission. 2013. Final Report, Ohio Lake Erie Phosphorus Task Force II. November 2013. 96pp.

Wetzel, R.G. 2001. Limnology, 3rd ed. Academic Press, NY. 1006 pp.

Comments from Dr. McLauglin

The following are my preliminary thoughts and comments on the Draft Technical Approach For Lake Erie Load-Response Modeling dated November 4, 2014 (referred to here as “the Approach”). Because the task of providing numeric phosphorus loading and/or concentration targets that are meaningful with respect to achieving Lake Erie ecosystem objectives is a very complicated one, an adaptive approach to research, monitoring, modeling, and management of important nutrient loads and responses is an appropriate way to proceed. Using an ensemble modeling approach also could be very valuable. However, which models will be used in a truly “ensemble” approach (as discussed, for example, in the Science and Technical Advisory Committee workshop report for the use of multiple models in managing Chesapeake Bay, February 25-26, 2013 Annapolis, Maryland), and how that will be accomplished are still unclear.

In my view, it is important to the success of the overall Lake Erie nutrient management effort that monitoring and modeling are closely integrated so that it is possible to make the most out of an adaptive management approach. It’s hard to tell from the Approach how well the modeling effort and proposed Eutrophication Response Indicator (ERI) metrics have thus far considered the extent to which future field activities can support the corroboration of model predictions. If it hasn’t been done already, it may be very instructive to explore in some detail the type and amount of field sampling that would be necessary to validate model predictions with varying degrees of confidence. This would also provide additional opportunities for the broader stakeholder community to provide input on the choice of ERIs and metrics, ensuring that there is a process in place for the most relevant metrics to be incorporated into monitoring and modeling efforts.

The draft approach should more clearly link, perhaps in a table, the individual eutrophication response indicators (ERIs) and their metrics to each Lake Erie Lake Ecosystem Objective (LEOs). As I tried to create my own table during this review, I was left thinking that the Approach may not fully address some LEOs. For example, how will the proposed models quantitatively link cyanobacteria biomass with the concentration of algal toxins, the latter being important to meeting the LEO “maintain cyanobacteria biomass at levels that do not produce concentrations of toxins that pose a threat to human or ecosystem health”? Also, LEO #3 “maintain algal species consistent with healthy aquatic ecosystems in the nearshore” suggests that some measure other than chlorophyll a concentration, cyanobacteria biomass, and/or Cladophora (stored P content) may be needed, such as an algal community metric/index.

I also believe that it is important for EPA, working with other affected parties, to give substantially more attention to establishing agreed-upon quantitative goals for the predictive performance of each model, especially in terminology that is transparent and meaningful to a broad audience of stakeholders and decision-makers (see, for example, related discussions in the environmental modeling guidance document, EPA 2009, referred to in the Approach). The Approach already reflects this need to some degree; however, I think much more can be done that would add considerable value to the Approach and its products.

I offer three ideas in this regard. The first is to provide more detail on how various statistical measures such as correlation coefficients, coefficients of determination, mean square error, and/or others may be used to assess model performance. This would include, where appropriate, some numeric criteria for each that can guide judgments as to the acceptability of model performance (e.g., acceptable, acceptable with caveats, not acceptable). The Approach begins this process. For example, on page 19 (referring to the page number shown in the Adobe software menu bar), the Approach states that one measure used to corroborate model predictions with field data will be a correlation of 50%, and in some cases a correlation of less than 50% will be accepted. These statements are a step in the right direction, but more detail is appropriate. For example, what type of correlation analysis will be conducted? A description of the basis for selecting 50% would also be helpful and improve the transparency of the overall modeling effort. In addition, it may be important to consider that many types of relationships can yield a specified overall correlation, even though different localized regions of the relationship may have very different degrees of correlation and levels of predictive accuracy. This becomes especially important when using load-response relationships to derive loading targets by choosing specific regions of the relationship that correspond to desired response conditions. Therefore, model performance assessment should include more than basic correlation analysis.

The second idea involves broadening the metrics used to assess predictive performance to include metrics that more directly communicate the implications of various degrees of correlation to non-modelers and non-statisticians. For example, a relevant question when identifying a phosphorus loading target based on model predictions is: "how precise is the estimate of the load that is expected to achieve the response objective?" This might be expressed as confidence limits on the target load that would reflect the degree of scatter (i.e., uncertainty) in the load-response relationship, with more scatter resulting in a wider confidence interval. The degree of scatter might also be expressed as the likelihood of "false positive" and "false negative" decision error rates as described in the EPA 2009 modeling guidance, and EPA's 2006 guidance titled "Guidance on Systematic Planning Using the Data Quality Objectives Process". For example, a quantitative answer could be provided to the question "Based on a given load-response relationship, if phosphorus load target 'X' is selected with the expectation that a desired response condition will be achieved, what is the estimated likelihood that the desired condition will not, in fact, be achieved? And, what is the confidence interval for that estimate?" Addressing such questions can help "translate" correlation measures into practical information about model uncertainty that is more broadly understood and useful to many.

The third idea follows from the second: provide preliminary, if not final, numeric ERI metric thresholds to aid model evaluation. Clearly, the development of such thresholds is anticipated though they may not yet be available. While not presented in the Approach, their importance is discussed in one of the supplemental documents provided by EPA, i.e., "Approach for Determination of Phosphorus Objectives and Targets Loads in Lake Erie". On page 43, for example, the document states:
"…there still will remain a management decision to be made on: what is the target objective for each of the eutrophication metrics so that the load-response curves can be used to identify the revised target loads to be published in the Agreement?"

Ultimately, the practical value of developing predictive phosphorus load-response relationships is in being able to identify phosphorus loads that are likely to yield desired response conditions. The absence of such response metric thresholds from the Approach presents, in my view, a significant challenge to implementing the Approach. As an alternative to using final ERI metric thresholds, EPA might consider identifying a set of preliminary thresholds used to support model evaluation efforts as described in "idea #2" above. Several "what-if" thresholds could be located at different points on each load-response relationship to estimate, for example, confidence intervals around corresponding "what-if" target phosphorus loads at each point. Not only would this provide a more complete assessment of model performance, it could also provide useful benchmarks for comparing the uncertainties associated with each model and for assessing model improvements that may be made as part of the adaptive management approach anticipated by the Annex 4 workgroup.

Using preliminary thresholds also could provide important context for correlation metrics that may, on their own, suggest that model uncertainty is potentially too high for the models to be useful. Clearly, a correlation coefficient of 0.5 or less (the criterion mentioned in the Approach) may indicate the presence of a considerable amount of variation explained by factors other than phosphorus load, and that the model may therefore be of little value. However, it may also be the case that confidence in the phosphorus load prediction associated with a given ERI metric threshold is high enough to be useful to managers relative to existing loads. Establishing even preliminary ERI metric thresholds can be an important part of assessing how useful the load-response model predictions may be for management purposes under different conditions or scenarios.

Comments from Dr. Reckhow

The comments below focus on Question 3 (ensemble modeling approach), but some of them also address other related issues.

1. **Ensemble Modeling Approach –** Applying multiple models to inform decision making is a good idea, particularly if the models characterize different meaningful endpoints and have different structural components. Many of the proposed models are over parameterized, which prevents a thorough uncertainty analysis; thus, it will not be possible to have an uncertainty-weighted composite prediction. As a consequence, the predictions for each model in the ensemble must be assessed based on the merits of the model.

2. **Independence of Models in the Ensemble** – How different are the models? For example, perhaps the 3D models are so similar that it should be expected that predictions from these models will not differ by much. If so, then this similarity needs to be taken into account when the ensemble of models is applied.

3. **Adaptive Management Approach** – Strictly speaking, adaptive management is a learning process that is expected to lead to improved information/models through observation of the effectiveness of initial management actions. Under this definition, some of the implemented management actions should be selected based on opportunities for learning. For phosphorus management in Lake Erie, we would like to learn the effectiveness of initially-implemented phosphorus control strategies as quickly as possible, so that changes can be made for subsequent management actions. Two challenges for effective adaptive management:
    a. The lag time between implementation and meaningful effect (e.g., HAB) can be quite long, particularly for agricultural BMPs.
    b. Detection of the signal (water quality improvement in Lake Erie) from the noise (background variability) may take several years of sampling, even in the absence of a long lag time.

    In principle, there are several points in the causal chain between management action and meaningful endpoints where the impact can be assessed for adaptive management. For example, at one extreme, the implementation of a management action can be considered a measure of impact; at the other extreme, change in a HAB metric can be considered a measure of impact. In between, phosphorus load and concentration are intermediate measures of impact. The closer the measure is to a meaningful impact (HAB), the more meaningful the measure becomes. Correspondingly, the closer the measure is to the management action, the sooner the measure is observable. Unfortunately, statistically-observable measures of impact are closest to the phosphorus source and furthest from the meaningful endpoints (e.g., HABs), making inferences on the effectiveness of management actions hard to assess for meaningful endpoints. This suggests that monitoring to inform adaptive management may be most useful if the monitoring occurs near implemented management actions (e.g., BMPs) for phosphorus loads/concentrations, since changes in phosphorus near the source should occur sooner and be more definitive than will changes in meaningful lake endpoints.

4. **Eutrophication Response Indicators** – The Draft Technical Approach briefly mentions the substantial uncertainty in the prediction of *Cladophora* due to changes in phosphorus load. To address this, it is proposed that the "models will explore the relative impacts of loads recommended for other eutrophication response indicators on *Cladophora* growth potential. DRP will be used as an input to the response curve model, and stored P content as the response measure of *Cladophora* biomass accumulation/growth." It is important to note that the prediction error for *Cladophora* growth remains, regardless of this indirect assessment. I recommend that the *Cladophora* growth model also be applied, regardless of error accumulation, since this is the meaningful endpoint.

5. **Model Prediction Error** – I find it interesting that the important issue of prediction uncertainty for *Cladophora* growth is of concern, but there is virtually no comment about the prediction uncertainty for the 2D and 3D models. I realize that overparameterization of these models limits the use of error propagation, since the covariance matrix for the model parameters cannot be estimated from the data. It appears that "skill assessment" (statistical comparisons between predictions and observations) will be used to assess model goodness-of-fit. Reckhow et al. (1990) should be consulted for guidance on the applications of these methods and for the construction of hypothesis tests.

6. **Model Verification** – In the development and application of water quality models, it is standard practice to set aside data not used in calibration, for model verification purposes. This approach is based on the reasoning that the set-aside-data provide a test of the model under new conditions and thus reflect how the model will perform when applied for prediction. How plausible is this reasoning?

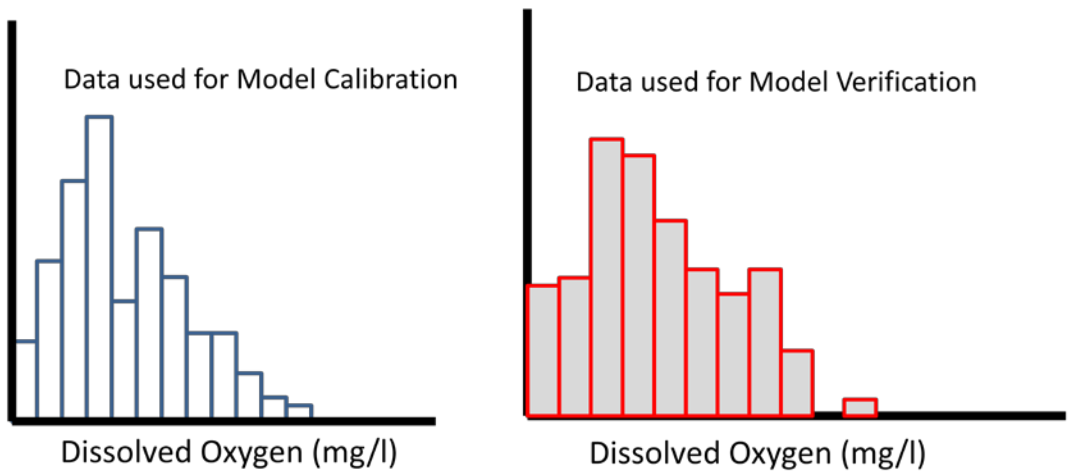
   Consider the situation where a model is calibrated with data from 2010-2011, and then data from 2012 are used for verification. What is likely to be different between these calibration and verification data sets? Will these differences be sufficient to give us confidence that the calibrated model can be relied upon for predictions when important forcings/inputs (e.g., pollutant loadings to a waterbody) change?

   In essentially all cases, the major differences between 2010-2011 and 2012 datasets are likely to be natural forcing functions such as hydrology, temperature, and solar radiation. It is extremely unlikely that the forcing functions that are the focus of the model application, such as LULC changes in a watershed or point source pollutant discharges, will change very much. To the extent that pollutant loads to a waterbody change over this time period, it will largely be due to changes in hydrology.

   So, conventional water quality model verification has become basically a charade. This situation is not the fault of modelers; rather, it is simply the consequence of limited available data. Nonetheless, water quality modelers who employ this approach to model verification need to be more candid about the limited value of conventional model verification.

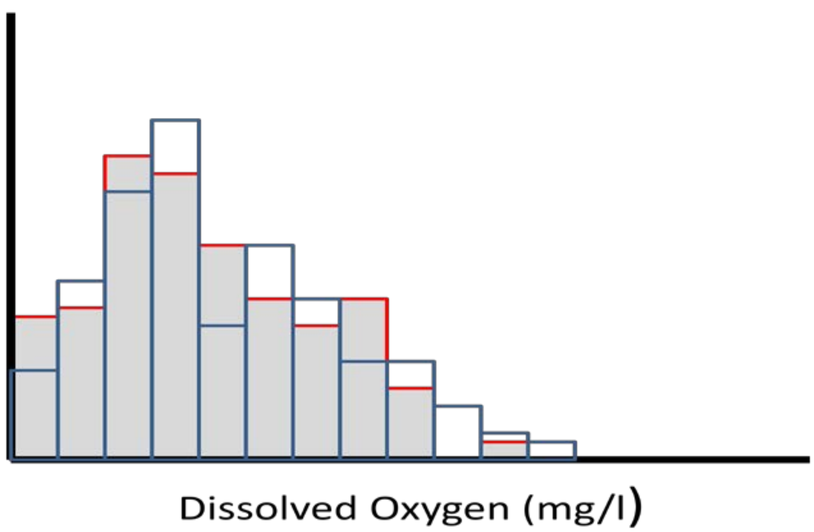
As an alternative, here is the basis for a statistical test that could provide a measure of the rigor in model verification. To begin, consider the figure below displaying histograms of dissolved oxygen data for model calibration and verification:

Evaluation of the Strength and Importance of Model Verification
Case 1

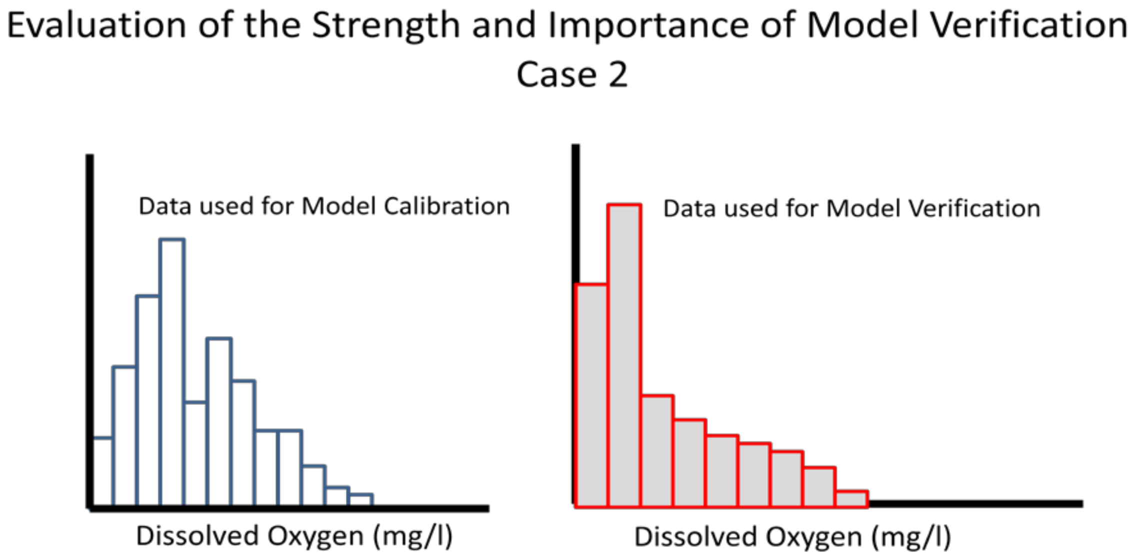


The next figure overlays the calibration and verification histograms for Case 1; notice how similar they are. The lack of difference between these two data sets indicates that “verification” lacks rigor; essentially, the model is being re-assessed with calibration-like data.

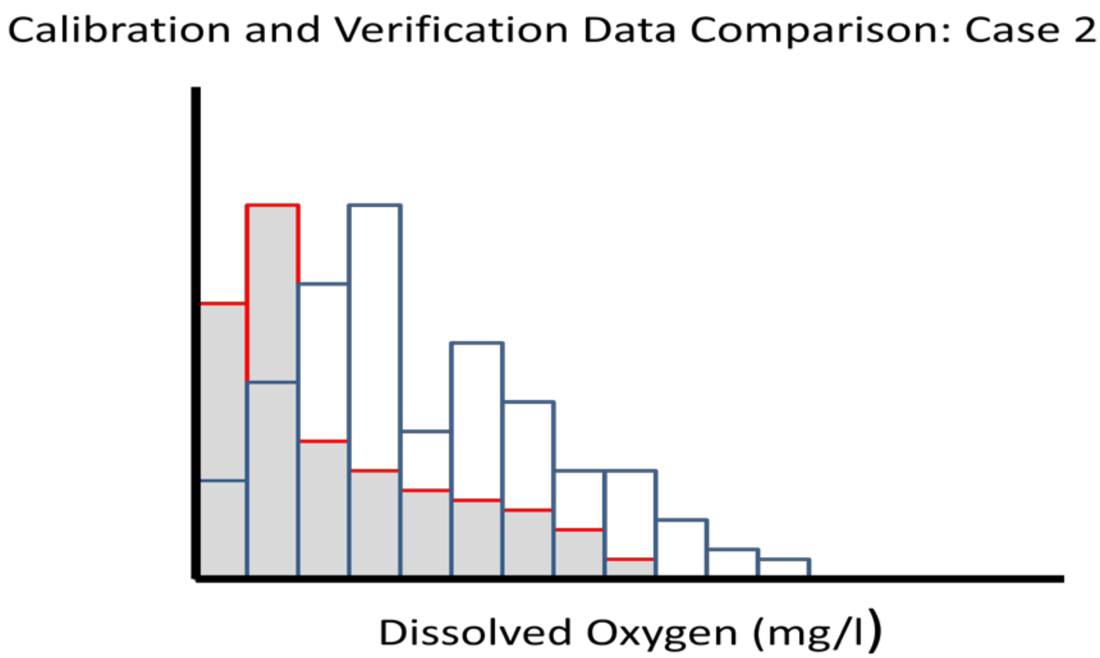
Calibration and Verification Data Comparison: Case 1



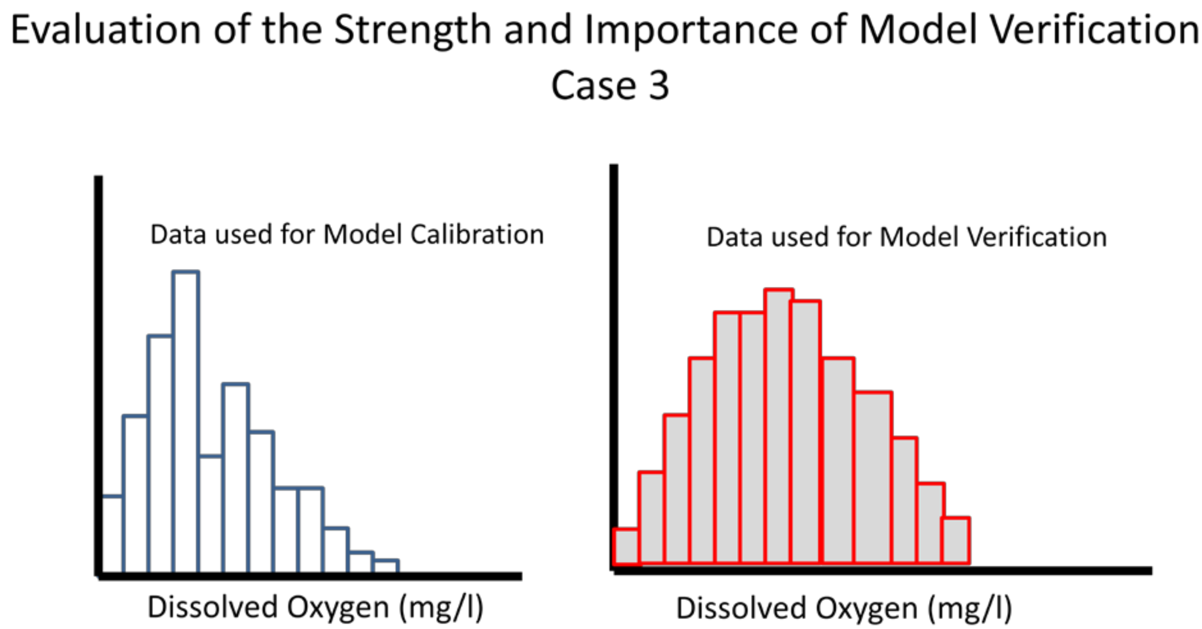
Now consider Case 2 below:

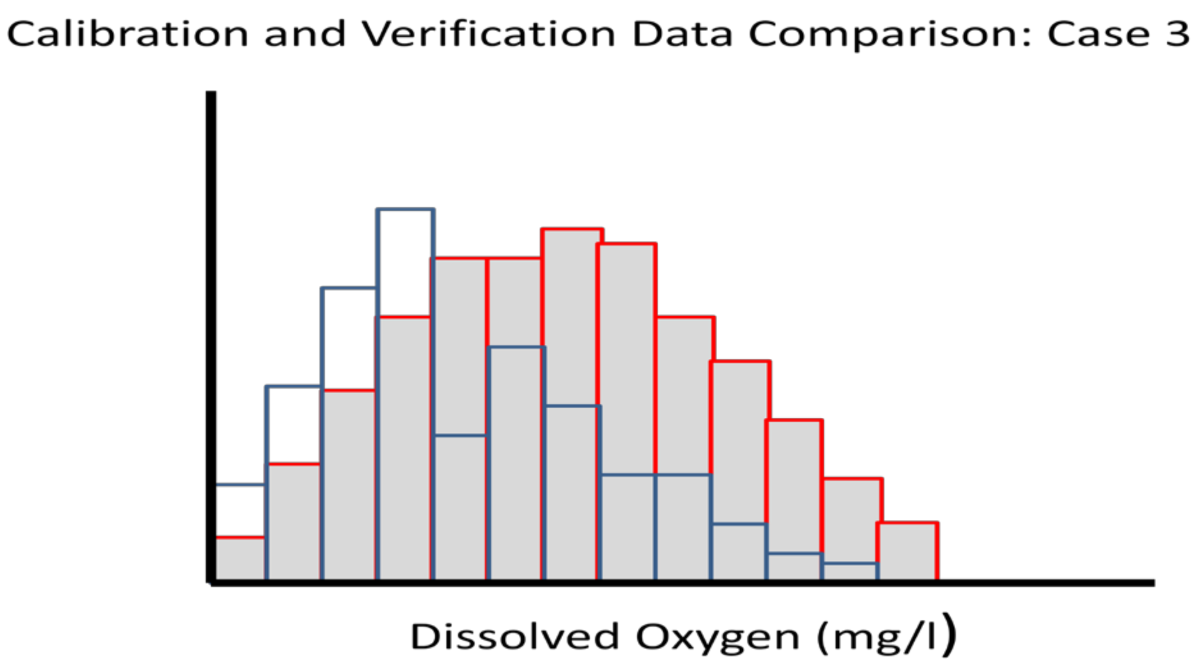


An overlay of the two histograms, shown below, indicates that the calibration and verification data sets are different, which suggests that verification is more rigorous than in case 1. However, note that the verification data in case 2 show DO to be lower than for model calibration. Since model applications are quite likely to address improved water quality and higher dissolved oxygen, the verification test may be rigorous but it does not reflect conditions expected for model use.



Now consider Case 3 below:





In case 3, the histogram of verification data is different from the histogram of calibration data, and this time the verification DO is higher than the calibration DO, which is a more likely prediction scenario.

In conclusion, to evaluate the rigor of the verification exercise, I recommend that modelers apply a Kolmogorov-Smirnov test, or a Chi-Square test, to quantitatively assess the difference between the calibration and verification data sets. If this becomes routine practice, the accumulated results will provide us with a comparative basis for having confidence that a water quality model can be used to reliably predict water quality in response to management changes.

References

Reckhow, K.H., J.T. Clements, and R.C. Dodd. 1990. Statistical Evaluation of Mechanistic Water Quality Models. *Journal of Environmental Engineering.* 116:250-268.

Comments from Dr. Rosi-Marshall

Phase 1 Consultation Questions:
1. Please comment on whether the eutrophication response indicators proposed sufficiently address and provide the scientific foundation for the Lake Ecosystem Objectives for Lake Erie. During your evaluation of the eutrophication response indicators, identify other metrics appropriate for measuring eutrophication response in Lake Erie and other Great Lakes that should be considered, and whether there is a method (model) available to measure this response.

The approach of eutrophication indicators seems to provide the foundation for meeting the Lake Ecosystem Objectives. The main indicators include: 1) overall phytoplankton biomass as indicated by chl a; 2) Cyanobacterial blooms in the Western Basin; 3) hypoxia in the hypolimnion of the Central Basin; and 4) the P content of *Cladophora* in the Eastern Basin. These eutrophication indicators appear appropriate although it would be useful to illustrate in the document how these four specific eutrophication indicators have responded over time to the changes in P concentration and loading.

I would like further clarification for the reasoning behind segregating the last three topics in space. I gather that these three areas of Lake Erie function somewhat independently but this could be described more thoroughly in the Technical Approach document; on pages 5-6 a further description of this approach are would be helpful. In addition, text to clarify why the metrics are useful in different sections of the lake would be a useful.

It would also be helpful to know how robust the data are for investigating the selected eutrophication indicators. For example, what are the frequency of the data collected or how spatially extensive are the available data? For example, when the technical report says "basin-wide" what exactly does this mean? The plan is to apply the models for 2008 because "the lake exhibited representative metrics for the eutrophication indicators". Although the data to be used as input variables is clearly indicated on pages 16-18, it is not clear from this description that there are robust data on the eutrophication response indicators for 2008. I presume that these data are available, but it would be useful to know how extensive and comprehensive these data are. This would help reviewers evaluate how effective they will be as indicators. Maybe an additional description could be included in this section or on the section describing the indicators.

It would also be helpful to provide information on data being collected via monitoring that was rejected for use as eutrophication indicators. Were these 4 indicators selected because there is sufficient data that measure these indicators? Are there data on other potential indicators that were rejected from inclusion for some specific reason? Although these indicators seem sufficient it would be helpful to know if there are other things that would be useful indicators, but for which data do not exist. This type information would be useful to evaluate the choices of data presented in the document and would provide insights into whether new variables should be included in future monitoring efforts.

2. Please comment on each of the models chosen to evaluate the eutrophication response in Lake Erie? Are the models appropriate for representing the eutrophication response indicators? Do the models reflect the best available scientific knowledge?

The models selected seem appropriate and are wide-ranging in their approaches, i.e. they range from mass balance to spatially explicit models. Using multiple modeling approaches seems like an effective approach to explore the multiple ways of examining this issue. The models are all well established and the basis for the model structure design and utility are published in easily accessible peer reviewed literature. It would be useful if the document indicated whether other models for Lake Erie exist but were not selected. Was the ensemble modeling approach simply to take all available models of Lake Erie and P or was some criteria used to evaluate whether models should be included or not? Providing this information would be helpful for the review process.

3. Please comment on the appropriateness of the ensemble modeling approach to examine the suite of eutrophication response indicators. Are the models included in the ensemble, when used either singly or combined, sufficient to provide a scientifically grounded basis for the required update of phosphorus load targets for Lake Erie?

The ensemble approach seems well reasoned. If there are multiple approaches to use, it makes good sense to try them. Using multiple models in combination seems to be a useful approach for developing updated P concentrations and loading targets.

On page 25, the Anticipated Results section states that the report will include "Examples of load-response curves and comparisons of load-response curves and, to the extent the model lead to differing conclusions, a discussion of the differences". Does this team envision a systematic approach for determining exactly how to deal with model outcomes that are very different? How will these differences be evaluated to make policy determinations? Is there an approach *a priori* that can be used to evaluate how different model outcomes can be evaluated? The different models apply to different segments of the lake and it seems that the multiple models will allow for assessing P loading in different basins and/or the near shore). So it seems the models are not all equivalent, but rather provide insights into different aspects of P loading and eutrophication. This could be more clearly articulated in the Technical Approach Document.

Aside from conducting the modeling simultaneously with the same input data, there does not seem to be a plan to actually integrate the models. The ensemble approach seems to use multiple models and see what the outcome is. The term "ensemble" implies that the models will somehow be integrated. It was not clear from the document that there is a plan to combine the models or model output in a systematic fashion. If there is a plan to do this, a description of the approach would be a useful addition to the draft plan. If this is not the approach planned, it is more appropriate to state that approach being used is a "multiple and independent modeling" endeavor using 9 different models. In addition, as stated above, the models will provide output and insights about different eutrophication indicators. It would be useful to clarify this in the text.

4. An anticipated outcome of the modeling exercise is to better understand and quantify what types of conditions would be expected in the lake based upon different levels of phosphorus loading, and to use that information to inform selection of phosphorus loading targets needed to meet the nutrient Lake Ecosystem Objectives. The phosphorus loading targets could be converted to concentration targets, particularly for river mouths/nearshore zones. Please comment on efficacy and value of establishing target values for both phosphorus loads and concentrations in order to meet to the Lake Ecosystem

Objectives. How can we ensure the phosphorus concentration and loading targets are internally consistent with respect to the eutrophication response indicators of concern?

The plan to establish revised target values for P concentrations and loading as a means to address eutrophication in Lake Erie is scientifically defensible. Phosphorus has consistently been shown to be the limiting nutrient that results in the impairment and eutrophication of Lake Erie. Although other nutrients and factors may contribute, controlling P concentrations and loading is defensible based on the best available science. The restrictions put in place in the 1970s did successfully reduce the P loads and reduced eutrophication in the 1980's.

Although the loading has been reduced, often below the target the current conditions in Lake Erie demonstrate that eutrophication continues to be a concern for the water quality. Although the restrictions in the past were effective at reducing total loads, since that time changes in the lake require revisiting of the P concentrations and target loading because this reduced loading is still leading to algal blooms. The report states that the models employed in the past did not sufficiently address the nearshore eutrophication issues nor were they designed to incorporate zebra and quagga mussel dynamics (invasion occurred after the modeling and target loads were in place) and climate change in basin may have altered P cycling dynamics, currents, ice cover, etc. The proposed approach (described on page 5) is to establish eutrophication response indicators, conduct the ensemble modeling and then apply an adaptive management approach wherein P concentrations and load targets are revisited periodically seems reasonable.

The plan to revisit the targets using an adaptive management approach is wise because conditions in the lake may change in the future. However, it is not clear from the document how often this "revisiting" will occur and what "periodically" means. It would be useful to clarify the timing of adaptive management and who is tasked with the adaptive management activities. From documentation from the LaMP, the plan is to revisit and revise on an annual basis (page 17, Lake Erie LaMP 2011), but this should be clarified in the Technical Approach section on the use of adaptive management.

In terms of ensuring that "the phosphorus concentration and loading targets are internally consistent with respect to the eutrophication response indicators of concern" future monitoring, research and adaptive management strategies for revisiting the concentration and load targets is advisable. If the P concentration limits and load targets do not achieve the goals, then revisiting and rethinking of the processes controlling eutrophication should be done. For example, if the influence of internal P cycling changes over time, either due to changing dynamics in the lake, new invaders or climate change, meeting the concentration and load targets may not achieve the desired outcomes.

References:

Lake Erie LaMP. 2011. Lake Erie Binational Nutrient Management Strategy: Protecting Lake Erie by Managing Phosphorus. Prepared by the Lake Erie LaMP Work Group Nutrient Management Task Group.

Comments from Dr. Smith

I was expecting to see more on the combination of the model predictions. It would be possible to combine the estimates of the relevant metric using a weighted combination with weights connected to model predicted ability. A difficulty is the scale of the different models although one might smooth the predictions or aggregate predictions. I see model uncertainty to be less important than prediction uncertainty, especially at loads that are critical.

To what extent are the different metrics redundant? Can one of the metrics be used to predict the others?

I like the approach of Theil (Theil, H. 1966. Applied econometric forecasting. North-Holland, Amsterdam) for model prediction/data comparisons. This is based on the regression of predicted and observed (Waller et al., Ecological Modelling 164 (2003) 49–63 and Pineiro et al., Ecological Modelling 216 (2008) 316–322 provide some citations and comments). It might be possible to consider aggregated measures to deal with space/time scale issues to compare different models. I think graphing the predicted and observed values is also valuable (not mentioned in the document but I assume it will be done). This is better than simply using a metric of agreement especially if concern is over a subregion of the data.

One can measure redundancy of Model A to Model B by using one model's output as an independent variable in the regression goodness of fit of another model.

The description of the UM/GLERL model in my opinion needs more details. The model has a linear effect for time (year) however it would seem that time would probably be better modeled as categorical unless there is evidence of a linear relationship. There is also a comment at the end of the description that indicates that cross-validation will be used although the Bayesian approach leads to a predictive distribution rather than a single value. Perhaps the intention is to use the expectation as the predicted value.

Comments from Dr. Stubblefield

The following responses to the "Draft Technical Approach For Lake Erie Phosphorus Load-Response Modeling" should be considered preliminary and will be revised following the objectives review panel meeting (10 December 2014).

Consultation Questions:
1. Please comment on whether the eutrophication response indicators proposed sufficiently address and provide the scientific foundation for the Lake Ecosystem Objectives for Lake Erie. During your evaluation of the eutrophication response indicators, identify other metrics appropriate for measuring eutrophication response in Lake Erie and other Great Lakes that should be considered, and whether there is a method (model) available to measure this response.

Eutrophication Response Indicators (ERI) identified in the Technical Approach Document include:

- *Overall phytoplankton biomass as represented by* ***chlorophyll a***
    - *Basin-specific, summer (June-August) average chlorophyll concentration*
- ***Cyanobacteria blooms*** *(including Microcystis sp.) in the Western Basin –*
    - *Maximum basin-wide cyanobacteria biomass (mass dry weight)*
    - *Summer total basin-wide cyanobacteria biomass (mass dry weight integrated over summer bloom period)*
- ***Hypoxia in hypolimnion*** *of the Central Basin –*
    - *Number of hypoxic days*
    - *Average areal extent during summer*
    - *Average hypolimnion DO concentration during stratified lake conditions*
- ***Cladophora in the nearshore*** *areas of the Eastern Basin –*
    - *Stored P Content*

These response indicators reflect the state-of-the-science with regard to assessment of lake eutrophication and should be appropriate for measuring eutrophication response in the Great Lakes. Other endpoints may be available; however, sufficient "high-quality" data may not be available to prove useful in conducting the ensemble modeling approach proposed. It is also important to note that although phosphorus (TP or DRP) are no doubt the driving parameters affecting eutrophication response, there may be other parameters (e.g., micronutrients) that play a role in the proliferation of species. Consideration of these parameters may help reduce uncertainty or explain variability in model predictions.

2. Please comment on each of the models chosen to evaluate the eutrophication response in Lake Erie? Are the models appropriate for representing the eutrophication response indicators? Do the models reflect the best available scientific knowledge?

I am not an expert in eutrophication nor am I a modeler; therefore I am not sufficiently familiar with each of these models at this point to be able to comment on the relative utility of each of the specific models. Incorporation of a table in the approach document that compares each of the models and provides technical pros and cons for the models would be extremely useful. It will be important that each of the models provide output that is comparable and in a form that will be useful to regulatory *authorities in establishing acceptable phosphorus objectives and programs designed to ensure compliance with those objectives.*

3. Please comment on the appropriateness of the ensemble modeling approach to examine the suite of eutrophication response indicators. Are the models included in the ensemble, when used either singly or combined, sufficient to provide a scientifically grounded basis for the required update of phosphorus load targets for Lake Erie?

The ensemble modeling approach seems appropriate given the scope of the models and their current status regarding calibration and validation. Given the fact that none of the models are currently available in the public domain, the ensemble approach seems to be the only practical way to evaluate the various models. It will be important to ensure that no "proprietary" issues hamper the evaluation of the models or how they handle the various functions. It is also important to realize that the outcome of this exercise may well be to derive a "hybrid" model that is built on the best aspects of one or more of the models in order to provide a usable regulatory tool.

The second part of the question posed seems to be: "are the proposed models fit for purpose?" It appears that this is the overall goal of the proposed effort, i.e., the evaluation of the models and an assessment of their utility for informing decision-makers about regulatory approaches to be used in controlling eutrophication in the Great Lakes. Clearly the models must be sufficient to relate bioavailable phosphorus exposure concentrations to the stated eutrophication response indicators, the models must be validated with real-world observations, and we must have an idea of the uncertainty associated with the models. Ultimately, we want the models to be predictive so that we can evaluate the potential utility of various eutrophication control strategies and we want to be able to evaluate the success or failure of the strategies once implemented.

4. An anticipated outcome of the modeling exercise is to better understand and quantify what types of conditions would be expected in the lake based upon different levels of phosphorus loading, and to use that information to inform selection of phosphorus loading targets needed to meet the nutrient Lake Ecosystem Objectives. The phosphorus loading targets could be converted to concentration targets, particularly for river mouths/nearshore zones. Please comment on efficacy and value of establishing target values for both phosphorus loads and concentrations in order to meet to the Lake Ecosystem Objectives. How can we ensure the phosphorus concentration and loading targets are internally consistent with respect to the eutrophication response indicators of concern?

The value of establishing target values for phosphorus loads and concentrations can only be assessed once we truly know the relationship between phosphorus concentrations and their effect on the Lake Ecosystem Objectives. Controlling phosphorus may take us a long way toward controlling eutrophication in the Great Lakes; however, it may not be the only parameter that matters and may not represent the "silver bullet" that controls eutrophication and permits us to achieve the LEOs. For example, some reports have suggested that in certain situations nitrogen or essential micronutrients may be factors controlling eutrophication and the phosphorus is of lesser importance. To the extent that TP and DRP are the main factors controlling Lake eutrophication, then establishing river mouth/nearshore loading targets for phosphorus may be useful, especially in the nearshore environment. Assessment of the extant field data may provide the means to correlate TP and/or DRP with the ERIs but it may be *necessary to employ more reductionist approaches (i.e., lab/mesocosm level techniques) to assess and confirm controlling factors.*

Comments from Dr. Valett

page #6: (2) Do you actually mean integration? That would result in MXT? You mean simply summation? right?
page #6: (3) This is akin to a disturbance regime: duration, extent, intensity - are other measures of regime to be included? frequency, resilience, resistance, others?

page # 6: (4) what about wave disturbance? Caldophora holdfast issues? growth potential? Is this a rate?
page #8, regarding the Chapra model - '…calculated as a function of time..' Do you mean residence time? Or that there is temporal resolution of the response variables?
page #8, regarding the Chapra model - I assume that there is a biological component not well articulated here.
page #10, regarding WLEEM model: There is a tense change here that is interesting. 'WLEEM has been calibrated and corroborated…. Documentation of the corroboration process will be accomplished….Model confirmation and uncertainty assessment will be…'. Has the 'calibration and corroboration' been done yet?
page #16, regarding the Great Lakes Cladophora Model – does N come into play with this taxa? What record of bloom behavior exists vs. TN?
page #16 and 17 – regarding use of 'primary' and 'secondary' data – On Page #16, under #3 'Data Sources' the document indicates that 'All models being applied for this project will use secondary data…'. On page #18, first line of full paragraph indicates that 'Primary data sources to develop, calibrate, and validate….'. This seems like a contradiction since the document earlier indicated that only secondary data are to be used. Is it correct to interpret that here the term 'primary' means 'most important' and not 'primary' in the sense that it is collected by field workers?

page #17, regarding P loads - Besides these annual TP load estimates, all the data are to be at least 9 years old...but see Chapra (2012). Why wouldn't current data be used?

page #17, regarding 'water quality data' - Are these data current? From matched time frames as the loads? Do they need to be?

page #17, regarding 'Tributary Nutrient Data' - How are these data different or the same as those identified as 'load estimates' above?

page #19, regarding 'Model Evaluation and Application' - I did not know that the program was designed to eliminate models. See objectives on page #5. See also the bottom of this page, first line in second to last paragraph.....'Each of the models selected for the analysis'. Nowhere prior to page #19 is this indicated to be the case. What is the logic or intent employed here?

page #20, regarding the final paragarph and rationale for 2008 baseline data, reason 3 - What does this late sentence mean? 'The lake exhibited representative metrics for the eutrophication response...' It had better. Is there a case where is doesn't? The metric value may be zero, but I am not sure what to make of this statement. Surely a response metric can't disappear?

page #22, final paragraph before iii. ELCOM-CAEDYM, regarding the activities intended for WLEEM - I didn't see anything that said WLEEM can do a labile C output. What's the point of this? It seems like a tag-on.

page #23, first paragraph addressing UM/GLERL model - This is the first place that 'bioavailability' is being addressed. Why here...and why not other places?

page #23, also first paragraph, regarding #2 - Does this mean more data? What is meant by 'data products' in this sense? Or is this a proposed model alteration to incorporate or more of the same.

page #24, ix. Great Lakes Cladophora Model – regarding first bulleted item - What does it mean to ‘accommodate physical features and biokinetic considerations...’ for the model? What does it currently employ for assessment of these features? This doesn’t seem trivial.

page # 25, regarding ‘Anticipated Results’ – This approach to model comparison seems unnecessarily qualitative. Relying on ‘discussions of outcomes’, ‘comparisons of the load-response curves’, and ‘discussion of differences’ as the basis for conclusions seems to fall short of a quantitative assesment of model performance. If, in fact, the modeling team is charged with ‘selecting’ models, don’t they need quantitative bases for the process? Moreover, how does this process relate to model selection. It is not clear how these initial tasks relate to outcomes that will be used to address load sources that may be targeted for alteration. Is this the ultimate goal?