Individual Comments from Members of the Lake Erie Phosphorus Objectives Review Panel
January 29, 2015

Table of Contents

Comments from Dr. Alber 2
Comments from Dr. Ammerman 5
Comments from Dr. Bartell 7
Comments from Dr. Carrick 9
Comments from Dr. Chen 13
Comments from Dr. Connolly 16
Comments from Dr. DiGuilio 18
Comments from Dr. Diaz 19
Comments from Mr. Endicott 20
Comments from Mr. Fitzpatrick 31
Comments from Dr. Heath 33
Comments from Dr. Johnson 39
Comments from Dr. Klump 42
Comments from Dr. La Point 46
Comments from Dr. McLauglin 50
Comments from Dr. Reckhow 54
Comments from Dr. Reddy 60
Comments from Dr. Rosi-Marshall 66
Comments from Dr. Smith 69
Comments from Dr. Stubblefield 71
Comments from Dr. Valett 74

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 do 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. For instance, production of the microcystin toxin increases with increased nitrogen concentrations. In addition, efforts to provide time-series monitoring of nutrient concentrations (both nitrogen and phosphate) with automated instrumentation should be considered. The Administration has recently released a Nutrient Sensor Challenge through the Alliance for Coastal Technologies to stimulate the development of affordable, accurate, and reliable nutrient sensors.

In the long run, automated time-series measurements for a variety of parameters should be increasingly emphasized throughout the critical areas of Lake Erie, deploying both in-water sensors and remote sensing. These include measurements of physical parameters (including oxygen), nutrients, chlorophyll, pigments and other parameters indicative of specific populations, as well as eventually process measurements such as phytoplankton productivity.

Comments from Dr. Bartell

Question 1: Eutrophication Response Indicators

The proposed response indicators appear useful in addressing eutrophic conditions in Lake Erie and the Great Lakes in general. However, some modifications of the proposed indicators might be considered. For example, chlorophyll-a:biomass ratios are not constant in phytoplankton, but vary in relation to the population growth dynamics of individual algal and cyanophyte species included in this aggregate measure. Interpretation of measured cyanophyte biomass (dry weight) in relation to corresponding total phytoplankton chlorophyll-a values might be made difficult because of the physiological dynamics of chlorophyll-a. Additionally, changes in the overall regional climate relevant to Lake Erie might require some modification of the proposed averaging period (i.e., June – August) if the increased growing season presents "bloom" conditions earlier (April-May) or later (September-October) in the year.

The P content of *Cladophora* appears to be of lesser utility as a proposed indicator of eutrophication in Lake Erie and the Great Lakes in general. Spatial-temporal variations in this metric might reduce its overall usefulness in assessing relationships between P loadings (or concentrations) and trophic status of nearshore conditions in Lake Erie. Only two of the proposed models in the ensemble appear relevant for this indicator. It might prove more useful to focus on overall productivity and biomass of *Cladophora* instead of (or in addition to) internal P content of this plant. Alternatively, this indicator might be simply replaced with increased monitoring of nearshore P concentrations (i.e., total P, orthophosphate).

The Lake Erie ecosystem has changed substantially (e.g., introduced dreissenid mussels, altered zooplankton and fish community structures, etc.) since the earlier work (1970s) on P loadings and subsequent eutrophication responses in this complex system. Current internal processing of phosphorous and other nutrients (e.g., N) might be sufficiently different from the historical P cycling conditions and associated data used to develop, implement, and evaluate models included in the proposed ensemble.

Question 2: Selected Models

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 *Cladophora* 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).

Question 3: Ensemble Modeling Approach

The ensemble modeling approach is fundamentally appealing. However, to what extent are the models really different? Answering this question will be important in evaluating the intended ensemble approach: For example, given apparent similarities among several of the process models, the ensemble might be justifiably reduced to a subset of the ten proposed models without sacrificing technical capabilities in addressing the overall modeling objectives. If all ten models are used, substantial effort might be invested in understanding the similarities/differences among the model results, rather than focusing finite resources on defining justifiable and acceptable P loadings (concentrations) for Lake Erie. "Right-sizing" the ensemble might help reduce this potential drawback of the ensemble approach.

Question 4: Target Loadings and Concentrations

Target values for P loadings (and concentrations) developed from the proposed ensemble modeling approach can provide useful information for managing phosphorous in Lake Erie watersheds with an ultimate goal of reducing eutrophication. However, the efficacy of developing and using such model-based targets will ultimately depend on the accuracy and reliability of the models. The relationships between P loadings (and other nutrient loadings) and associated ecological responses within Lake Erie will likely remain dynamic, depending on the varying influence of other factors (e.g., climate change, invasive species, chemical contaminants) that potentially affect productivity within this complex system. Therefore, the models and their projected target loadings might require periodic evaluations to assess the continued relevance of the targets in meeting management objectives. Use of an adaptive management framework in developing, applying, and re-evaluating the efficacy of modeled target loadings is highly recommended.

Comments from Dr. Carrick

1. Justification and Need for this Work

Clearly there is a need to reevaluation the role phosphorus (P) has on the ecology of Lake Erie. This factor should be included in any management objectives for Lake Erie. Given the recent changes in the lake and significant bloom events of 2008, 2011, and 2014, this study is needed in the short term, and should be considered urgent. The factors that support this idea are as follows:

1. There have been changes in regional climate in the Great Lakes, which have been well documented (see McCormick and Fahnenstiel 1999).
2. Recent food web changes have occurred:
   Proliferation of invasive Dreissenid mussels
   Proliferation of invasive zooplankton- *Bythotrephes*
   Declines in native species
   Shifts in phytoplankton species composition
   Occurrence of *Cladophora* in nearshore regions
3. P loadings to the lake have changed over the past 20 years
   Ratio of dissolved P to TP has gone up
4. Hypoxia (occurrence, extent) in the central basin has not shown a significant decline/reduction as expected (see Scavia et al. 2014).
5. Several large harmful algal bloom events in the western basin have occurred over the past 10 years (2008, 2011, 2014, see Millie et al. 2014).

2. Evaluation of Approach to this Work

2a. Modeling Effort

Strength

- Based upon the number and magnitude of the changes to the Lake Erie ecosystems, it was logical and sensible to revisit the models used to evaluate the response to P loading in the lake.

- The application of this series of models is a good idea. The notion of using multiple lines of evidence is a sound and robust approach. Moreover, many of the models have already been used in the Great Lakes, calibrated against the conditions in these lakes, and thus have either been built for this purpose or adapted to this end.

Chapra- Mass balance model,
WLEEM model- water quality: Chl, Cyanobacteria
ELCOM- water quality Chl, Cyanobacteria, hypolimnetic oxygen concentration
EcoLE- hypoxia, chlorophyll, Cyano chl
UM-GLERL- Spatial-temporal model for Cyanobacteria coverage
NOAA forecasting- Spatial-temporal model for Cyanobacteria coverage
9-Box Eutrophication model- TP, DO, Chl
1-D Central Basin- hypoxia
Great Lakes Cladophora model- Cladophora coverage

Limitations

- The biology of the lake has changed significantly, so that applying known models to this new situation may be misleading. Even relatively simple relationships between phytoplankton chlorophyll and P-loading appeared to have changed, such that the mechanisms for this remain unknown.

- The assumptions of each model should be spelled out more clearly and directly so that the suitability of each application can be better considered. These models each work under differing time and spatial scales, and with this comes assumptions about how realistic it is to use them under certain circumstances, and then compare to output to one another. The central basin hypoxia model is geared to seasonal-scaled phenomenon, while the ELCOM and EcoLE models are water quality models that can general daily output. The UM-GLERL model predicts spatial coverage of cyanobacterial blooms, which is a very different quantity.

- Most of these models do not truly weight the relative importance of variables in regulating the metrics of interest.

- Algal blooms data are difficult to predict because they do not adhere to conventional statistical and mathematic treatments (zero-inflated data).

2b. Choice of Metrics

Most of this metrics have been used before in relation to evaluation lake productivity and cultural eutrophication (see Wetzel 2001). That said, there is very little innovation here or room to consider anything new that may have been developed over the past 20 years. The key metrics being used are as follows:

Chlorophyll is an average June-August concentration
Cyanobacteria blooms (Cyanobacteria numbers, biomass)
Hypoxia- hypolimnetic oxygen concentrations in the central basin
Cladophora in the nearshore Eastern Basin

Strengths

- Many of the metrics being used here are relatively straightforward to measure and have been measured over a long time period in the Great Lakes. This is very importance, because information generated here can be compared against previous historic conditions in the lake.

- Given the routine nature of these metrics, numerous models have been developed to make these predictions; this allows several, difference models to be used and their output compared against one another. Thus, the preponderance of evidence approach can be used to achieve or achieve consensus.

Limitations

- The error related to each metric produced by the array of models being used here may be different and should be considered when interpreting results.

- These metrics may be sensitive enough in order to detect meaningful change. The use of hypolimnetic oxygen does not seem to be responsive to P loading, perhaps because it is regulated by so many other physical-chemical factors (e.g., temperature, water movement).

3. Questions and Need for Clarification

- •The justification for the selection of this subset of model should be strengthened. Some of the models appear to have been chosen because the researchers were comfortable with them (have used them previously).
- I expected to see a set of models used that have been developed to predict (evaluate) the occurrence of individual populations (niche models, multivariate ordinations). This makes sense, given that specific groups (or species) of algae are being modeled.
- How were these target concentrations chosen (why spring TP)?
- Are the Lake Erie target numbers too high for TP- these values seem indicative of relatively eutrophic conditions (see Wetzel 2001)?
- Did the Depinto et al 2006 study consider changes in the plankton that have occurred in contemporary Lake Erie?
- Given that some base conditions in Lake Erie have changed, such as temperature (McCormick and Fahnenstiel 1999) and base chemistry (see Barberio and Tuchman 2004), can we assume these metrics will behave similarly as modeled previously?
- Are the key parameters used in these models still relevant to current day conditions in the lake? For example, did the uptake kinetics for Lake Erie phytoplankton (current) change from the 1980's or are the same as those measured algal cultures that were used to establish some of these input parameters?
- How is internal loading addressed and can phytoplankton (chlorophyll) be internally loaded (see Carrick et al. 2005)?
- Why are nutrients other than P being considered? There is evidence that both Si and N can limit phytoplankton growth (see Moon and Carrick 2007).
- How will the winter dynamics in Lake Erie be incorporated into the modeling effort? Some work showed that winter diatom blooms has contributed significantly to summer hypoxia (see Lashaway and Carrick 2010; Wilhelm et al. 2013).

4. Literature Cited

Barberio, R.P., and M.L. Tuchman. 2004. Long-term dreissenid impacts on water clarity in Lake Erie. ***Journal of Great Lakes Research*** 30: 557-565.

Carrick, H.J, J.B. Moon, and G.F. Gaylord. 2005. Phytoplankton dynamics and hypoxia in Lake Erie: Evidence for benthic-pelagic coupling in the central basin. ***Journal of Great Lakes Research***.31: 111-124.

Lashaway, A.R., and H.J. Carrick. 2010. Effects of light, temperature, and habitat quality on meroplanktonic diatom rejuvenation in Lake Erie: Implications for seasonal hypoxia. ***Journal of Plankton Research***, 32: 479-490.

McCormick, M.J. and G.L. Fahnenstiel. 1999. Recent climatic trends in nearshore water temperatures in the St. Lawrence Great Lakes. ***Limnology and Oceanography*** 44: 530-540.

Millie, D.F., G.R. Weckman, G.L. Fahnenstiel, H. Carrick, E. Ardjmand, W.A. Young II, M. Sayers, and R. Shuchman. ***2014.*** Using artificial intelligence for CyanoHAB niche modeling: Discovery and visualization of *Microcystis*-environmental associations within western Lake Erie. ***Canadian Journal of Fisheries and Aquatic Sciences***, 71: 1642-1654.

Moon, J.B., and H.J. Carrick. 2007. Seasonal succession of phytoplankton nutrient limitation in the central basin of Lake Erie. ***Aquatic Microbial Ecology***. 48: 61-71.

Rucinski, D.K., Beletsky, D., Depinto, J.V., Schwab, D.J., Scavia, D. 2010. A simple 1-dimensional, climate based dissolved oxygen model for the Central Basin of Lake Erie. ***Journal of Great Lakes Research*** *3*6: 465-476.

Scavia, D., Allan, J.D., Arend, K.K., Bartell, S., Beletsky, D., Bosch, N.S., Brandt, S. B., Briland, R.D., Daloğlu, I, DePinto, J.V., Dolan, D.M., Evans, M.A., Farmer, T.M., Goto, D. Han, H., Höök, H.O., Knight, R., Ludsin, S.A., Mason, D., Michalak, A.M., Richards, R.P., Roberts, J.J., Rucinski, D.K., Rutherford, E., Schwab, D.J., Sesterhenn, T., Zhang, H., Zhou, Y. 2014. Assessing and addressing the re-eutrophication of Lake Erie: Central Basin Hypoxia. ***Journal of Great Lakes Research*** 42: 226-246.

Wilhelm, S. W., G.R. LeCleir, G.S. Bullerjahn, R.M. McKay, M.A. Saxton, M.R. Twiss, and R.A. Bourbonniere. 2013. Seasonal changes in microbial community structure and activity imply winter production is linked to summer hypoxia in a large lake. ***FEMS Microbiology and Ecology*** 1-11.

Comments from Dr. Chen

Question 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 loading to P concentration and ultimately 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):
While P loadings are the regulated variable, the resulting P concentrations are what phytoplankton respond to, and thus the consistency between loadings and concentration is critical. 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 (Sonzogni et al. 1982; Baker et al. 2014), 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 (Chaffin et al 2014). 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 feedback 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 hypolimnion 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. Moreover, the oxygen depletion in central Lake Erie appears correlated with DRP load once again raising the importance of the bioavailable form (Ruscinski et al. 2010). The linkages between space and time and bioavailable P 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.

Baker, D.B., R. Confesor, D.E. Ewing, L.T. Johnson, J.W. Kramer, B.J. Merryfield. 2014. Phosphorus loading to Lake Erie from the Maumee, Sandusky and Cuyahoga rivers: The importance of bioavailability.

Chaffin, J.D., T.B. Bridgeman, D.L. Bade, C.N. Mobilian. 2014. Summer phytoplankton nutrient limitation in Maumee Bay of Lake Erie during high-flow and low-flow years. J. Great Lakes.

Kutovaya, OA, RML McKay, B.F.N. Beall, S.W. Wilhelm, D.D. Kane, J.D. Chaffin, T.B. Bridgeman, G.S. Bullerjahn. 2012. Evidence against fluvial seeding of recurrent toxic blooms of Microcystis spp. In Lake Erie's western basin. Harmful Algal 15: 71-77.

Rucinski, D. K., D. Beletsky, J.V. DePinto, D. J. Schwab, D. Scavia. 2010. A simple 1-dimensional, climate based dissolved oxygen model for the central basin of Lake Erie. J. Great Lakes Res. 36: 465–476.

Sonzogni, W.C., S.C. Chapra, D.E. Armstrong, T.J. Logan. 1982. Bioavailability of phosphorus inputs to lakes. J. Env. Quality 11: 555-563.

Comments from Dr. Connolly

Comments Pertaining to SAB Charge 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?

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. That said, the chosen models include models that may have limited predictive ability because they are too simple or too complex. The lake has undergone changes in the last two decades caused by the proliferation of Dreissenid mussels, increasing temperature, changes in phosphorus load speciation, and changes in algal species. The simple models have limited ability to account for these factors. For example, the impact of Dreissenid mussels is indirectly accounted for by increasing the settling speed of algae. In doing this, the settling speed becomes an unconstrained calibration variable. Does the calibration value represent a constant that will be unaffected by the reductions in phosphorus loading? I suspect we do not know. The most complicated considered model, ELCOM-CAEDYM, has so many state variables that cannot be constrained by data that there is a risk of the modeling being a curve-fitting exercise that can produce an impressive calibration, but may have little true predictive ability.

The EPA guidance on environmental modeling[1] takes a reasoned position on model complexity noting: "Models tend to uncertainty as they become increasingly simple or increasingly complex." The document includes an informative figure that I have copied and include below. It illustrates the important point that model uncertainty is great for simple models because of framework uncertainty and for complex models because of data uncertainty. I suggest that this concept be incorporated in the evaluation of the chosen models.

The approach taken to modeling includes specifying 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." Additionally, the models should be able to account for the species shifts that have occurred and may continue as phosphorus loads are reduced. I suggest that the evaluation of the model predictions include assessments against the above factors, taking in to account my comments on model complexity. Such assessments may help determine the reliability of the chosen models as predictors of the improvements in water quality metrics as phosphorus loads are reduced.

[1] USEPA 2009. Guidance on the Development, Evaluation, and Application of Environmental Models. EPA/100/K-09/003, March 2009.

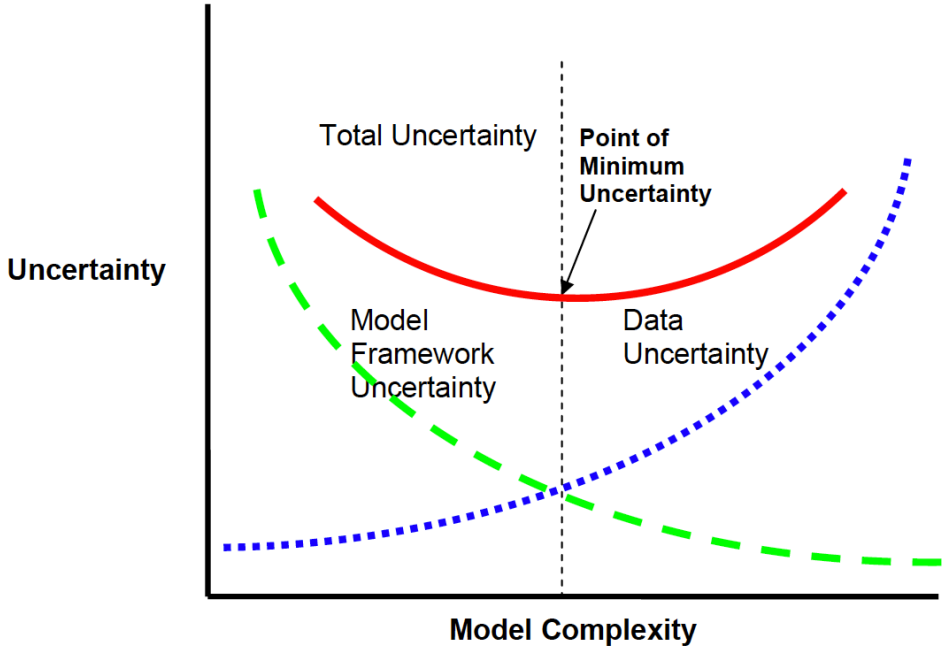


Figure 3. Relationship between model framework uncertainty and data uncertainty, and their combined effect on total model uncertainty. (Adapted from Hanna 1988).

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 cover a range of complexity. I think it may be helpful to address two variants of this question that come from my comments on model complexity. 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.

Comments from Dr. DiGuilio

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.

Similarly, approaches for actually integrating the models selected into an ensemble model were very vague. Several models were very distinctive in terms of inputs required, outputs, uncertainty analysis, incorporation of emerging issues such as climate change etc. I was left with the impression that merging these models into one efficient and accurate ensemble model will be very challenging.

Another concern that emerged in the course of the December 10 meeting in Chicago is the potential importance of nitrogen in some of the adverse phenomena observed in Lake Erie. It seems the agency needs to seriously consider this nutrient before focusing solely on phosphorous.

Comments from Dr. Diaz

The weight of evidence, in both peer-review and gray literature, as to what key factors are driving eutrophication in the Great Lakes and degrading Lake Ecosystem Objectives point to a complex interaction of physical processes and biological processes driven by 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 concentration criteria for either phosphorus and/or nitrogen, or both. This approach when combined with best professional judgment can be a powerful tool for drawing conclusions in many areas of water quality management.

With Spring mean phosphorus concentration selected for substance objectives, it is important to set criteria as to what would be minimum data requirements for adequate characterization of a Spring mean. As the models seem to range widely in complexity, how are the minimum dataset requirements for each model determined?

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

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

Comments from Mr. Endicott

I would first like to say that I appreciate the opportunity to serve on this panel. As a kid who grew up swimming, fishing and sailing in western Lake Erie, I have a deep love for that body of water. I would also like to compliment the other panelists and invited experts who made our meeting in December so informative.

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.**

The Lake Ecosystem Objectives related to nutrients are:

1. minimize the extent of hypoxic zones in the Waters of the Great Lakes associated with excessive phosphorus loading, with particular emphasis on Lake Erie;

2. maintain the levels of algal biomass below the level constituting a nuisance condition;

3. maintain algal species consistent with healthy aquatic ecosystems in the nearshore waters of the Great Lakes;

4. 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;

5. maintain an oligotrophic state, relative algal biomass, and algal species consistent with healthy aquatic ecosystems, in the open waters of Lakes Superior, Michigan, Huron and Ontario; and

6. maintain mesotrophic conditions in the open waters of the western and central basins of Lake Erie, and oligotrophic conditions in the eastern basin of Lake Erie.

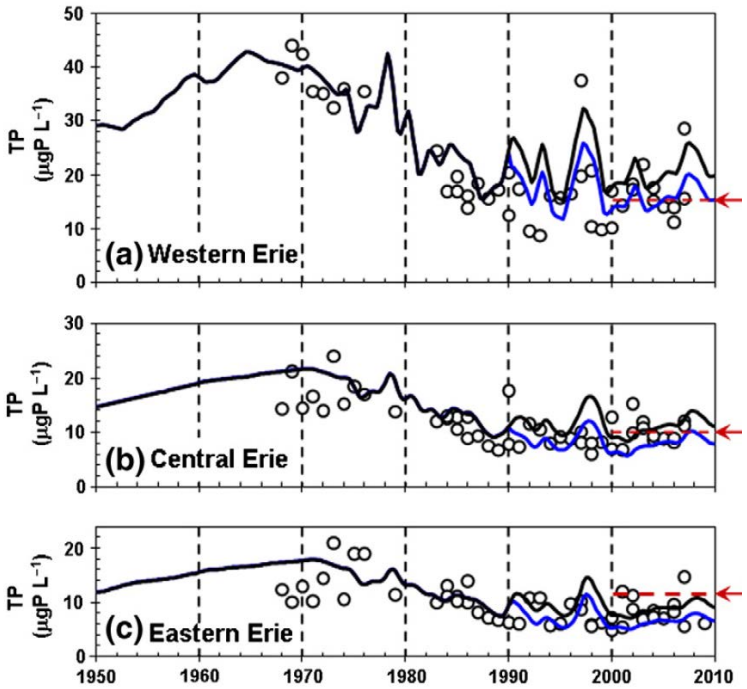
The 4 response indicators (basin-specific chlorophyll a biomass, Western Basin cyanobacteria biomass, Central Basin hypoxia and Eastern Basin Cladophora) appear to address most of the Lake Ecosystem Objectives for Lake Erie. It is easier to relate some of the response indicators to ecosystem objectives than others. For example, the Central Basin hypoxia response indicator (although it fails to define/quantify hypoxia) directly addresses Objective #1. The cyanobacteria biomass response indicator partially addresses Objective #4, although it falls short of defining cyanobacteria biomass levels below which cyanotoxins are not produced. The basin-specific chlorophyll a biomass indicator addresses Objective #6 and partially addresses Objectives 2 and 3. The Eastern Basin Cladophora indicator partially addresses Objective 2. I think EPA should be able to develop preliminary targets or thresholds for the response indicators at this time, and should do so.

Integration of modeling and monitoring activities, as others have mentioned, should be pursued in order

to maximize the value of this work. Useful monitoring that has not been discussed in this context includes (1) data from water quality sampled at drinking water intakes in Ontario (Winter et al. 2014) and (2) the 12-hour monitoring of algal toxin by the City of Toledo[2].

Details are lacking in regard to how the response indicators will be monitored, including sampling methods, spatial and temporal resolution, and quantification methods. In particular, the spatial resolution of response indicators may not be adequately defined. Of the 4 response indicators, only eastern basin Cladophora is not basin-wide; the adequacy of data to determine "basin-wide" is not addressed (e.g., is it an average? does it include or exclude nearshore areas?). The Cladophora indicator is apparently applied to the nearshore region, but even that is vague.

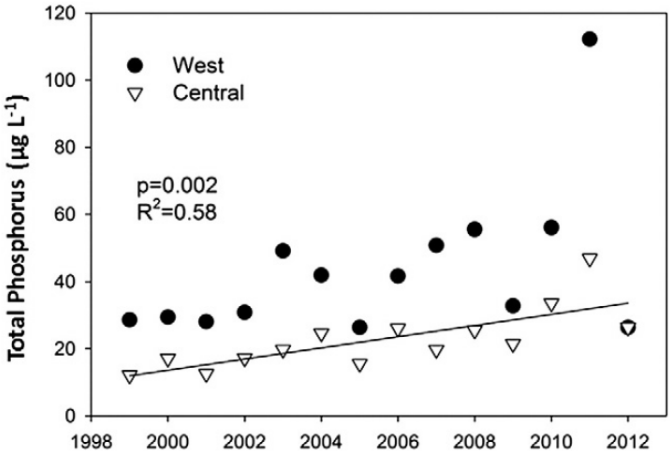
Why these details matter can be illustrated by the "basin-wide" trends in total phosphorus (TP), rendered by 2 different efforts. Chapra and Dolan (2012) used data from Environment Canada and GLNPO to compute open-lake annual median values for TP concentrations. They did not pool the EC and GLNPO data, but computed separate annual medians for years in which both entities were sampling a particular basin. They also excluded near-shore samples, assuming that open-lake (offshore) waters best indicate long-term trends. Not surprisingly, their data indicate that TP levels for each of the Lake Erie basins has dropped significantly over the decades following the GLWQA, although target concentration levels are still exceeded for some years. Their plot of annual open-lake medians (below) shows little trend since 1990.



Source: Chapra and Dolan (2012)

[2] http://www.toledo.oh.gov/services/public-utilities/water-treatment/algal-toxin-tap-level-reports/

Scavia et al. (2014) also compute open-lake annual values for TP concentrations. However, they calculated mean concentrations from a dataset cited from the Ecofore-Lake Erie Forage Task Group Report. It is unclear whether these data included concentrations sampled in nearshore waters. Their graph of the trend in annual mean TP (reproduced below) shows a statistically-significant increase in central basin since 2000, and what appears to be a general concentration increase over this time period in the western basin as well.
There must be differences in the dataset underlying these statistics for the resulting trends in TP to appear so different from those of Chapra and Dolan (illustrated above).



Source: Scavia et al. (2014)

The technical approach should clarify sampling representativeness of “basin-wide” indicators (e.g., do these include or exclude nearshore areas, areas in vicinity of tributaries, etc?). Is the sampling design random, fixed stations, etc? As shown above for TP, different trends and interpretations can result from these choices. Resampling (bootstrap and jackknife) can be used to determine the adequacy of data intended to calculate average waterbody concentrations (e.g., Lesht 1988). I also point this out because others have stated (with regards to water quality trends in Lake Erie), “Total phosphorus levels have not increased, yet Microcystis blooms have”. This generalization is problematic on a number of levels, including the veracity of stating there to be no increase in total phosphorus concentrations. Moreover, the science appears compelling in that Microcystis blooms are a response to increasing loads and concentrations of SRP. If SRP is an increasing fraction of TP, as monitoring data from Ohio tributaries suggest, then the observations made above may well be internally consistent.

In terms of Objective 4, it would seem that the most significant indicator protective of human health would be microcystin concentrations at the location of drinking water intakes. As many as 4 of the models may be capable of simulating Microcystis density at those locations, which could possibly be translated into probable microcystin concentrations, as others have suggested. Should microcystin concentration be an additional response indicator? Note that I would recommend retaining a basin-wide cyanobacteria indicator to address accumulation in biota and dog (wildlife?) poisoning.

Translating Microcystis density into probable microcystin concentrations is possible, and is being pursued in other eutrophic ecosystems. In the James River, a chlorophyll-a study (Bukaveckas 2013) is

developing site-specific information on the occurrence of HABs, coupled with information documenting their impacts on living resources and human health, to provide a basis for deriving HAB-based chlorophyll a criteria for freshwater and saltwater regions of the James. Quantitative relationships are being developed between nutrient loading, chlorophyll a, and the abundance of harmful algal species as well as an assessment of their toxicity. Microcystis cell densities and microcystin concentrations showed significant positive relationships with chlorophyll a on a Bay-wide scale and for individual tributaries. Research undertaken in 2012 documented the presence of microcystin in water, sediment and biota in the tidal freshwater segment of the James River.

The phytoplankton biomass response indicator uses summer average chlorophyll a concentration as the metric. Opinion of the panel appears divided in terms of preference for chlorophyll a vs. a direct measure of biovolume. Chlorophyll a has the advantage of being readily measurable in laboratory, field (data sondes w/ fluorescence detectors) and remotely via satellite.

Regarding central basin hypoxia, external phosphorus loading may not be the major driver of hypolimnetic oxygen depletion in the central basin, at least in the short term. However, it is at least hypothetically controllable, in comparison to climactic and meteorological forcing and sediment depletion. It should be possible to expand the modeling scenarios to consider other years in order to explore how this forcing can perturb relationships between phosphorus loading and response indicators. Inter-basin (western to central) transport of phytoplankton and detrital carbon are important (and possibly under-explored) factors in determining the extent of central basin hypoxia.

The Cladophora response indicator (stored phosphorus) makes sense as it is a state variable within the Cladophora growth 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. Also, is there a particular management concern related to storage of phosphorus in Cladophopra? I don’t think so, and it isn’t articulated in the Draft technical approach.

Here again, there is an opportunity to improve the product by integrating monitoring and modeling. I think a plan to systematically monitor Cladophora should be developed and implemented. Could this be done by incorporating periphyton sampling at reference locations and remote sensing?

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

The Draft Technical approach has selected an appropriate set of models for Lake Erie phosphorus load-response modeling. There are similarities and differences between the 9 models in terms of complexity, spatial/temporal resolution, processes and state variables. In terms of reflecting best available scientific knowledge, certainly each model has its strengths and weaknesses. In my opinion, diversity between the models is desirable and useful. The empirical cyanobacteria models (Obenour and Stumpf) seem to be fairly similar and are based on comparable observations. Among the mechanistic models, both the DePinto and Leon models are highly spatially resolved and complex in terms of processes, parameters and state variables. The Chapra and Lam models are comparatively simple, and both have been used to simulate eutrophication in Lake Erie back to pre-1970 conditions. The Zhang and Rucinski models fall somewhere in-between in terms of complexity and apply novel spatial resolutions. Auer’s model is unique in its singular focus on Cladophora growth.

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. The Draft Technical approach identifies the importance of model validation as an objective test of acceptability and utility.

Results of assessing the models in greater detail are presented in the attached tabulation (also available on the Lake Erie Ensemble Modeling Approach Consultation web site). I must give credit to the recent JGLR publication of Kim, Zhang, Watson and Arhonditsis (2014) for the general layout and content of the table, which was then expanded to include all of the models under consideration here as well as additional attributes (nearshore resolution, meteorologic forcing, state variable details, cyanobacteria, Dressenids, SOD, sediment P release, and features under development). I did not include simulation of ice cover as an attribute in the table, an oversight since ice cover is discussed fairly often as a factor related to phytoplankton dynamics in Lake Erie. Ice cover is included in only the ELCOM-CAEDYM model.

Based on this tabulation, the following comments are offered about each of the models:

Chapra and Dolan (TP mass balance)
The TP mass balance model dates back to the 1978 Great Lakes Water Quality Agreement. As the authors state: “our model is primarily designed to predict the annual average concentrations of (total phosphorus) in the offshore waters of the Great Lakes as a function of external loadings”. I suppose this model is rudimentary by the standards of the evaluation used at the December meeting; others might prefer “parsimonious”. It lacks many features of the other models: TP is the only nutrient, no vertical or nearshore resolution, no explicit phytoplankton state variable, no DO, no sediment P release. On the other hand, this model has a long track record, has been applied to all of the Great Lakes, and incorporates the impact of Dressenids by increasing the TP settling rate. Chloride was used to calibrate interbasin transport, a historical approach in Lake Erie dating back as least as far as Boyce and Hamblin (1975). Its simplicity means it can run long simulations a large number of times, which has advantages in terms of sensitivity and uncertainty analyses.
Planned extension of model predictions to chlorophyll a may be difficult, due to the variability of CHL:TP ratios in Lake Erie over the past decades.

Yerubandi (Lam 9-box)
The 9-box model is the other veteran of the 1978 Great Lakes Water Quality Agreement and, like the TP mass balance model, it has a long track record and extensive calibration and validation. The model simulates temperature, dissolved oxygen, SRP and organic phosphorus. There is no explicit phytoplankton state and the spatial scale is again basin-wide although epilimnion, metalimnion and hypolimnion are resolved in each. The 9-box model does respond to meteorological forcing functions via calibration of a 1-D vertical thermal balance and wind-driven interbasin transport. Dreissenids are represented by increasing the phosphorus settling velocity. Sediment oxygen demand and regeneration of phosphorus from the sediment are modeled as constant proportions of TP loading to the lake. It may be difficult to confirm whether these functions have predictive value.
In general, 9-box model has strengths and weaknesses similar to the TP mass balance model.

Rucinski (Central basin hypoxia)
The central basin hypoxia model is a 1-D vertical thermal and mass balance model driven by the Princeton Ocean Model. Water quality constituents include dissolved oxygen, available (DRP) and unavailable (organic) phosphorus, one phytoplankton functional group, and a zooplankton functional group. Spatial resolution is central basin-wide, which captures climatic effects on vertical stratification but not the perturbations resulting from storm-induced mixing. This model does not include cyanobacteria, Dreissenids, or sediment P release. SOD is modeled as a Michaelis-Menten-like function of phosphorus loading. Phosphorus loadings to the western basin were significantly manipulated (accounting for settling loss, bioavailability, etc.) before they were used as model input. Results of calibrating this model to temporal dynamics and vertical resolution of DO are impressive through annual simulations over the period 1987-2005. Predictions of DRP and chlorophyll a were not so encouraging. Running the POM in 1-D vertical mode is computationally efficient, although the model is not run continuously over multiple years due to neglecting ice cover. Modeling results demonstrated the impacts of interannual variability in hydrometeorology vs. phosphorus loading. For this "ensemble" exercise, the central basin hypoxia model will be coupled to WLEEM predictions of phosphorus and organic matter transport from the western lake basin.

Zhang (EcoLE)
EcoLE is based on the Corps of Engineers' QUAL-W2 reservoir model which integrates hydrodynamics and water quality simulations. EcoLE resolves Lake Erie vertically and longitudinally (along the long axis of the lake). Like the proceeding models, it offers no nearshore predictive capability. It is driven by meteorological forcing, and has relatively complex representations of nutrient, phytoplankton (including cyanobacteria) and zooplankton state variables. The model was applied to a fairly narrow time window (1997-1999). Dreissenids (both zebra and quagga mussels) are represented by grazing and nutrient excretion forcing functions; the abundance and size distribution of each species were treated as spatially variable but temporally constant. SOD is modeled as a function of DO and temperature, and sediment P release under anoxic conditions is included.

EcoLE was applied by Zhang et al. (2011) to examine the basin-scale impacts of Dreissenids on phytoplankton in Lake Erie. As noted in the table, the 2-D spatial discretization created difficulties in representing the nearshore vs. offshore distribution of Dreissenids, as well as the impact of nutrient loading on nearshore water quality. However, EcoLE is computationally efficient in comparison to fully 3-dimensional models.

Diagnosis of EcoLE predictions showed Dreissenids to have weak direct grazing impacts on algal biomass. Excretion of bioavailable nutrients (PO_4 and NH_3) had a much greater impact: 23-56% of the SRP in the western basin water column. Model results were inconclusive regarding the composite effect of Dreissenids on cyanobacteria.

DePinto (WLEEM)
WLEEM is a highly-resolved 3-D model of western Lake Erie that links simulations of hydrodynamics, thermal budget, surface waves, sediment transport and ecosystem (nutrients, eutrophication and sediment diagenesis) dynamics. Of the models included in this ensemble, WLEEM appears to be the most capable in terms of addressing all of the response indicators (chlorophyll, cyanobacteria, DO and benthic algae), with the limitation that it is applied to only the western basin. This model includes all of the tabulated features and attributes, and has been calibrated and validated to data over the period 2005-

2013. WLEEM incorporates a full (and calibrated) simulation of sediment phosphorus recycling, something I could not confirm for any of the other models. Chloride was used again as a tracer, although for this model it is used to validate hydrodynamic transport. Based on presentations I have seen (e.g., Verhamme et al. 2014) WLEEM predictions of TP, SRP, and chlorophyll a agree well with western basin monitoring data. Notably, the model simulates the CYN blooms observed in 2011 and 2013.

This model should be extended to lake-wide application, although it is unclear how feasible this would be computationally. WLEEM cyanobacteria predictions should also be carefully compared to the Obenour and Stumpf models, as these are truly independent modeling approaches.

Leon (ELCOM-CAEDYM)

ELCOM-CAEDYM is the other "heavy iron" 3-D computational model that links simulations of hydrodynamics and eutrophication. Unlike WLEEM, however, this model is applied at the scale of the whole lake. Thus, ELCOM-CAEDYM is unique in resolving water quality in the nearshore waters of each lake basin. The model is driven by meteorological forcing, has complex representations of nutrient and phytoplankton (including cyanobacteria) state variables, and the thermal balance simulates ice cover. SOD is modeled as a function of DO and temperature, and sediment P release is included. Dreissenids and zooplankton are apparently being added to the model, although details were not provided. Benthic algae is not modeled.

Similar to any complex ecosystem model, ELCOM-CAEDYM is over parameterized and the calibration approach has been criticized (Kim et al. 2014), especially the parameterization of growth parameters for the phytoplankton functional groups. The simulation of 190 days in 2002 seems quite limited, perhaps related to the considerable computational expense (4 ½ days on a PC). Loadings from 11 major tributaries were included in the model, accounting for 87% of the estimated total phosphorus loading to the lake. Because the model simulation did not begin until April, spring maximum phosphorus loadings from the smaller tributaries were accounted for as initial conditions. Over prediction bias in summer phosphorus concentrations in the central and eastern basins was attributed to uncertainty regarding the bioavailability of phosphorus in these tributary loadings.

Spatial (top 5m) predictions of chlorophyll a and total phosphorus in August were compared to data, and appeared to capture major trends in the observations. Temporal predictions of chlorophyll a, multiple nutrient states and light extinction were also compared to data collected at a station in the eastern basin, again with mostly favorable results. DO predictions were not presented. Other results presented in Leon et al. (2011) include dynamics of phytoplankton succession in western and eastern basins, lakewide distributions of vertically-averaged phosphorus (TP and SRP) and chlorophyll a, and nearshore vs. offshore dynamics in these water quality variables. Although these latter modeling results are unconfirmed by data, they suggest that ELCOM-CAEDYM simulations are consistent ("in the correct range") with general trends and spatial patterns of water quality in the lake.
Although cyanobacteria are simulated, prior applications of the model have not predicted significant bloom concentrations. As with WLEEM, comparison of ELCOM-CAEDYM 's cyanobacteria predictions to Obenour and Stumpf models should be informative.

Auer, Higgins, Canale (Great Lakes Cladophora /Cladophora growth model)
Like the name says, this is a mechanistic process model that specifically addresses the growth of Cladophora. The model also predicts biomass density, internal P content and biomass sloughing. There are several versions of this model which, except for the original application by Canale and Auer (1982), have used external concentrations of SRP as a forcing function rather than a state variable. Application of the model by Tomlinson et al. (2010) in Lake Michigan indicated that Cladophora biomass production increased by 67% due to greater light penetration resulting from Dreissenid filtering. Tomlinson's results also showed that SRP concentrations would need to be less than 1 μg/L in order to manage Cladophora growth to acceptable levels. Auer at al. (2010) applied the model in Lake Erie and predicted that Cladophora growth potential increased 81% due to increased water transparency together with higher SRP concentrations. These results demonstrate the utility of the Cladophora growth model in the context of Lake Erie ecosystem management and the specific objectives of Annex 4 of the GLWQA.

For this application, the Cladophora growth model will be linked to nearshore eastern basin nutrient boundary conditions predicted by ELCOM-CAEDYM. The model also uses temperature as a boundary condition and, in some applications, TSS to estimate the light (PAR) extinction coefficient. Wind, wave action and/or depth are also used to predict shear stresses that are, in conjunction with temperature, responsible for biomass sloughing. Details regarding linkage of the Cladophora growth model to ELCOM-CAEDYM were not provided, so significant issues such as the spatial and temporal resolution of this linkage could not by reviewed.

The last 2 models are empirical regressions that relate Maumee River loadings directly to CYN bloom in the western basin. These models are appealing in their conceptual simplicity and direct use of observations for tributary loading and cyanobacteria response. Their weakness is the potential unreliability of forecasts if the controlling processes are non-stationary (in the case of Stumpf model) or depart from trends defined in prior years (Obenour). In both cases, it is also unclear how these results will be scaled to account for other external loads of phosphorus to Lake Erie.

Stumpf (Cyanobacteria Bloom Intensity)
The Stumpf model is a regression of summed areal Cyanobacterial Index (CI, a remote sensing estimate of surface Microcystin concentration) against total spring discharge or P loading from the Maumee River. The model was developed using western basin data from 2002 to 2011. Spring discharge and TP loading from the Maumee were both found to be suitable dependent variables. As noted by Stumpf et al. (2012), the time lag between spring loading and peak biomass is sufficient to allow recycling of total phosphorus and basin-wide dispersion to support growth under optimal temperature and light conditions. The relatively small annual sample size (n = 10) used in the regression may be a weakness in terms of forecast performance and broader application. According to the Draft technical approach, this model will be updated and used to directly calculate TP and DRP loading targets.

Obenour (Bayesian cyanobacteria)
The Obenour model forecasts the size of annual cyanobacteria blooms (dry weight in MT) from a monthly-weighted average of spring total phosphorus load from the Maumee River and temporal change to bloom susceptibility. Cyanobacteria bloom size is derived from remote sensing and plankton tow data. The model includes a trend indicating that the lake has become more susceptible to large cyanobacteria blooms over the period 2002-2013. Development of this regression model in a Bayesian

framework produces considerable insight regarding uncertainty in the relationship between total phosphorus load and cyanobacteria bloom size.

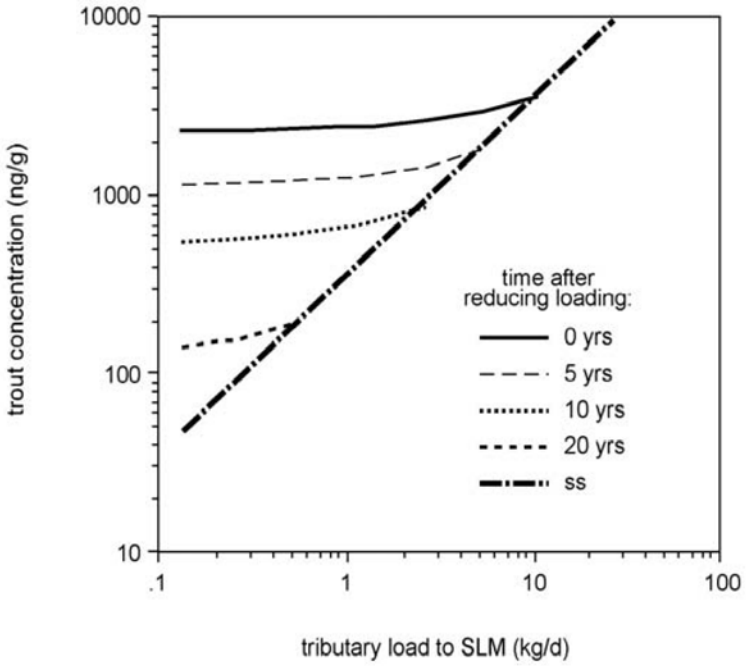
According to the Draft technical approach, the Obenour model is being recalibrated to include bioavailable (DRP) phosphorus loads from the Maumee River as an independent variable. The revised model will be used to calculate load-response (TP and DRP vs. peak bloom size) curves.

3. Please comment on the appropriateness of the ensemble modeling approach to examine the suite of eutrophication response indicators.

The ensemble modeling approach is based on using predictions from multiple models to derive load-response curves, using a standardized set of P loadings and hydrologic/meteorologic conditions from the 2008 reference year. As demonstrated above and in the Draft Technical approach (Table 1), no singe model offers the capability of predicting all 4 response indicators. Thus, it is necessary to use multiple models to address all response indicators. Moreover, predicting response indicator 4 (Cladophora in the nearshore of the eastern lake basin) will require combining the Cladophora growth model with ELCOM-CAEDYM nutrient boundary conditions. Similarly, predicting response indicator 3 (Central Basin hypoxia) with the Central basin hypoxia model will require input of western basin transport fluxes from WLEEM. Beyond that, there is redundancy in the model predictions for the other 3 response indicators: 6 models predicting phytoplankton biomass (in 1 or more lake basin); 5 predicting western basin cyanobacteria, and 5 predicting central basin hypoxia.

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. Weighting averaging according to uncertainty and the likelihood of not achieving a desired response condition (for a given P loading target) seem to be practical options. Some other issues that are not adequately addressed include 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. I recommend that loading scenarios be expanded to consider the timing and location of phosphorus loading.

A number of the models include release of phosphorus from the sediment (“internal loading”). This will decouple annual phosphorus loading from in-lake phosphorus concentrations and response variables. A similar phenomenon occurs in models of persistent toxic chemicals in the Great Lakes under conditions of significantly reduced external loadings. As shown in the load-response graphic (below) for the predicted response of lake trout polychlorinated biphenyl (PCB) concentration at various times after reducing PCB load, the load-response curve for times shorter than steady-state (5-20 years for PCBs) depart from linearity due to decoupling of water column and sediment dynamics. I expect the load-response curves in models including sediment phosphorus release will be similar.



Source: Endicott et al. 1992

4. 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.

Loading targets for both TP and SRP are preferable. In-lake phosphorus concentrations reflect both external loading and recycling from sediment. As a practical matter, how could a target value for SRP concentrations be confirmed, given the difficulty in measuring concentrations below 1 µg/L? Furthermore, how would establishing target concentrations for TP be any different from including TP as an ecological response indicator?

References

Auer, M.T., Tomlinson, L.M., Higgins, S.N., Malkin, S.Y., Howell, E.T. and H.A. Bootsma. 2010. Great Lakes *Cladophora* in the 21st Century: Same organism – different ecosystem, Journal of Great Lakes Research, Journal Great Lakes Research, 36(2):248-255.

Boyce, F.M. and P.F. Hamblin. 1975. A simple diffusion model of the mean field distribution of soluble materials in the Great Lakes. Limnol. Oceanogr. 20(4):511-517.

Bukaveckas, P.A. 2013. Executive Summary; First Year Findings from the James River Chlorophyll-a Study. Prepared for the Virginia DEQ Science Advisory Panel by Dr. Paul A. Bukaveckas, Virginia Commonwealth University. July 15, 2013.

Chapra, S.C., Dolan, D.M. 2012. Great Lakes total phosphorus revisited: 2. Mass balance modeling. J. Great Lakes Res. 38(4): 741–754.

Verhamme, E.M., J.V. DePinto, T.M. Redder and J. Grush**.** 2014. Western Lake Erie Ecosystem Model (WLEEM): Insights from 2011, 2012, and 2013. Presented at the 57th annual conference on

Great Lakes Research. Hamilton, Ontario, Canada. June 2014. (www.limno.com/limnotech/pdfs/2014_IAGLR_Verhamme.pdf)

Endicott, D.D., Richardson, W.L. and D.J. Kandt. 1992. MICHTOX: A mass balance and bioaccumulation model for toxic chemicals in Lake Michigan. US Environmental Protection Agency, Office of Research and Development, ERL-Duluth, Large Lakes research Station, Grosse Ile, Michigan. 183 pp.

Kim, D.K., Zhang, W., Watson, S. and G.B. Arhonditsis. 2014. A commentary on the modelling of the causal linkages among nutrient loading, harmful algal blooms, and hypoxia patterns in Lake Erie. Journal of Great Lakes Research, Volume 40, Supplement 3, 2014, Pages 117-129.

León, L.F., Smith, R.E.H., Hipsey, M.R., Bocaniov, S.A., Higgins, S.N., Hecky, R.E., Antenucci, J.P. and S.J. Guildford. 2011. Application of a 3D hydrodynamic–biological model for seasonal and spatial dynamics of water quality and phytoplankton in Lake Erie. J. Great Lakes Res. 37, 41-53.

Lesht, B.M. 1988. Nonparametric evaluation of the size of limnological sampling networks: Application to the design of a survey of Green Bay. Journal of Great Lakes Research 14(3):325-337.

Scavia, D., Allan, J.D., Arend, K.K., Bartell, S., Beletsky, D., Bosch, N.S., Brandt, S. B., Briland, R.D., Daloğlu, I, DePinto, J.V., Dolan, D.M., Evans, M.A., Farmer, T.M., Goto, D., Han, H., Höök, H.O., Knight, R., Ludsin, S.A., Mason, D., Michalak, A.M., Richards, R.P., Roberts, J.J., Rucinski, D.K., Rutherford, E., Schwab, D.J., Sesterhenn, T., Zhang, H., Zhou, Y. 2014. Assessing and addressing the re-eutrophication of Lake Erie: Central Basin Hypoxia. *Journal of Great Lakes Research*. Vol. 42, Issue 2: 226-246. doi: 10.1016/j.jglr.2014.02.004.

Tomlinson, L.M., Auer, M.T., Bootsma, H.A. and E.M. Owens. 2010. The Great Lakes Cladophora Model: Development, testing, and application to Lake Michigan. *Journal of Great Lakes Research.* Vol. 36. pp. 287-297.

Winter, J.G., Palmer, M.E., Howell, E.T. and J.D. Young. 2014. Long term changes in nutrients, chloride, and phytoplankton density in the nearshore waters of Lake Erie. Journal of Great Lakes Research, In Press, Corrected Proof, Available online 23 December 2014.

Comments from Mr. Fitzpatrick

Eutrophication Response Indicators

(1) Overall phytoplankton biomass as represented by chlorophyll-a provides a reasonable eutrophication response indicator, but could be supplemented by algal species identification and cell counts. These data could provide information concerning changes in community composition and whether the phytoplankton community is shifting to a balanced or desired community structure. It appears that this sampling may already be taking place in order to identify cyanobacteria biomass blooms referenced in (2) below. It appears from the modeling descriptions that the WLEEM, CADDYM and EcoLE models can distinguish between cyanobacteria biomass and non-cyanobacteria biomass.

(2) Cyanobacteria blooms measured as maximum basin-wide biomass and summer total basin-wide biomass (both as mass dry weight) are good response indicators. Consideration should be given to sampling of microcystin in an effort to see if microcystin production can be linked to total cyanobacteria biomass or perhaps to nutrient ratios or other environmental drivers or data. At this point it does not appear that any of the available models are capable of computing or predicting microcystin production. Perhaps, however, consideration can be given to a mixed deterministic/ statistical approach, wherein the available deterministic or mechanistic models predict cyanobacteria biomass and nutrient ratios and other environmental drivers or data and then a statistical model is used to relate these variables to the likelihood of microcystin production and concentration.

(3) The use of the number of hypoxic days, average areal extent during the summer and average hypolimnetic DO during stratified conditions provide good response indicators. However, EPA may also wish to consider adding DO-area-days and DO-volume-days, binning the DO in 1 mg/L units (ex., 0-1 mg/L, 1-2 mg/L, 2-3 mg/L, etc.). While the data are not likely to be available with which to make these estimates, the WLEEM and CAEDYM models can provide these types of outputs and may be useful in assessing load reduction response and developing knee-of-the-curve response to load nutrient load reductions. DO-area-days and DO-volume-days were used to perform scenario screening in the USEPA Long Island Sound Study. DO-area-days can be used to assess impacts to the benthic community, while DO-area-days can be used to assess impacts on pelagic communities.

(4) Lake resource managers may wish to consider implementation of *Cladophora* sampling programs that consider measuring *Cladophora* abundance using transects to estimate biomass (dry weight and P-content) and percent cover. This information could be used to help establish numeric thresholds for assessing eutrophic condition (Sutula, 2014; Scanlan, 2007).

Models and Ensemble Modeling

A complete review of the models is not possible at this point without further details of the model frameworks, calibration/validation and skill assessment. From Table 1 presented in the Draft Technical Approach it does not appear as if one model can address all of the response indicators and regions of interest, hence an ensemble modeling approach may make the most sense to assist the binational workgroup in establishing Lake Erie phosphorus objectives.

Such an approach has been suggested for Chesapeake Bay by the Chesapeake Bay Science and Technical Advisory Committee (http://www.chesapeake.org/pubs/324_Weller2014.pdf). A determination as to whether the models that are included in the ensemble provide a scientifically grounded basis for the required update of phosphorus load targets for the Lake. Information provided at the December 10 meeting, as well as a review of the individual models either from peer-reviewed papers and information provided by the Annex 4 workgroup will help inform that determination.

Phosphorus Loads and Concentrations

While establishing loading targets is easier to do from a management and modeling perspective, the question is being asked is, "Does it make sense to establish concentrations targets, particularly for river mouths and nearshore zones?" I believe that establishing nutrient concentration targets for the mouths of rivers may not make sense as it is the loading that is more important. A concentration of 10 ug P/L at a river flow of 50,000 cfs is going to have more impact on the Lake than a concentration of 10 ug P/L at a river flow of 500 cfs. In addition, other factors such as year-to-year variation in residence time and vertical stratification may have as great an impact on the response indicators as does an individual concentration target. While the plan is to evaluate the indicator response to loads for one year (2008) it would be useful to provide the review panel with information with respect as to whether 2008 can be considered a dry, wet, or average year and how does the stratification of the lake in 2008 compare to other recent years.

Scanlan, C.M, J. Foden, E. Wells, and M.A. Best, 2014. The monitoring of opportunistic macroalgal blooms for the water framework directive. Marine Pollution Bulletin 55:162-171.

Sutula, M., L. Green, G. Cicchetti, N. Detenbeck, and P. Fong, 2014. Estuaries and Coasts 37: 1532-1548. Thresholds of adverse effects of macroalgal abundance and sediment organic matter on benthic habitat quality in estuarine intertidal flats.

Comments from Dr. Heath

The following responses are based on my preliminary responses along with knowledge gained during the face-to-face meeting in Chicago, IL at GLNPO.

1. Eutrophication Response Indicators: The ERI proposed are reasonable choices for assessing attainment of the Lake Ecosystem Objectives. Total phytoplankton biomass, cyanobacterial biomass, and central basin hypoxia, 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 this is not the case with *Cladophora*, as I discuss below. I am concerned that none of these variables provides an insight to the biochemical status of the cyanobacterial and algal assemblages, which is much closer to the heart of the matter. The chosen 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. 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 prediction of 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 if is 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 $\leq$ 0.4 ppm; DO $\leq$ 4 ppm) could and should be monitored and modeled. The meeting failed to clarify this matter. Determining the areal and temporal extent of these measures is good because those measure the extent of damage to benthic communities, especially to insect larvae. 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 $\leq$ 4 ppm, because that would estimate of the extent to which deep, cold water fish habitat is constrained.

D. I am still 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.

E. I believe that the production of microcystin should be considered as an important ERI. This has become an important issue following the drinking water problem in Toledo in summer 2014. This will need additional research because microcystin production has been noted to be a function of N-limitation, as well as many other variables (light, Fe, micronutrients, T, pH, and alkalinity).

2. Comments on each of the models proposed:

General Comments: As a group, these models are all in one way or another, stimulus-response models, where the stimulus is external P-loading and the responses are the ERI. The possible interplay between and among other elements such as N and C is largely ignored.

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 and are possibly misleading. 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 can 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 dynamic models is that they can propagate errors and be unstable. Because of the mathematical structure of models such as this, it is unclear that an uncertainty analysis can be accomplished. I am also concerned that the validation and calibration years may not have differed sufficiently to warrant this as a valid "validation." Because Lake Erie is a storm-driven ecosystem, a better approach would be to calibrate using a dry year, and validate using a wet year.

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 and at times to make a significant contribution to P-availability.

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. This is important because it is likely that the dreissenids have changed the functioning of the ecosystem in ways that influence internal loading and recycling - and eventually the ERI. 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 sometimes, 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 to be significant.

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 its complement, P-debt.

3. Ensemble Modeling Approach The ensemble modeling approach makes a great deal of sense, at face value. "Ensemble" in the sense of chamber music, implies that the models are working together to achieve a result that none of them alone could. I don't see any of these models as being the *sine qua non* in addressing the current issues or predicting the ERI. I like that the models chosen represent a wide variety of model complexity. I see this as far preferable to selecting a single model to make these predictions and responses.

On the other hand, at the GLNPO meeting it was very unclear how these models will work "ensemble," and the terse responses from the staff indicated that this is a "work in progress" nowhere near the end. The models chosen differ widely in their mathematical architecture, and so in the ways they can best be used. Instead of trying to find a way to use all the models in the same way, it will be better to use each model to its best advantage. For prediction of ERI on a basin-wide basis models must be chosen that can be analyzed for "uncertainty." The large dynamic models on the other hand may be best used for heuristic purposes. The fine-grained models may best be used over short terms to predict concentrations of P and microcystin at drinking water intakes. I see this as the most important task facing the GLNPO staff: how will these models be used "ensemble?"

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 much 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 strongly 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 like the inclusion of meteorologically driven models (vii and viii) because in many ways the Lake Erie Basin is a storm-driven ecosystem. I believe the set of models chosen represents a reasonable incorporation 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 and likely the ecosystem has changed and continues to change in ways not fully understood.

4. Efficacy and Value of Establishing Target Loads These models singly and collectively will provide P-loading vs. 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 current 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 econometric 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. Johnson

My overall impression is that the plan lacks sufficient detail to assess whether the actual implementation will be successful or not. The indicators lack specificity with respect to details such as sampling methods, sample frequency, sample density, location (within the water column and within the particular basin), or the data source if part of an ongoing monitoring program. Indicators serve multiple purposes; one of those is communicating information to the public, the other is communicating progress towards GLWQA targets. The agencies (EPA and EC) need to assess whether the proposed indicators serve the needed functions of communicating progress to both the Parties and the public. Some interpretation may be required to translate the proposed indicators to values that the public recognizes and appreciates.

It would be helpful to create a matrix of inputs needed for each individual model, populated with information such as: parameter, units, sampling season, sample frequency, horizontal and vertical sampling frame. Also useful would be information about which indicator the metrics will inform, which biological assemblage(s) are targeted, and which Annex objective is informed. This exercise will identify gaps and overlap in efforts, and will help to ensure consistency in model inputs and outputs for reporting purposes.

For example, Table 5-1 in "an Approach for Determination of Phosphorus Objectives and Target Loads for Lake Erie" could be modified to include further details about each of the metrics. This information would be guided by the information on data sources (Section 3 of the report).

I have an overriding concern that the boundary conditions of these models do not sufficiently address potential changes due to climate change. This Great Lakes region is undergoing profound changes in precipitation patterns, and winter-time conditions appear to be especially impacted. What are the impacts of less ice cover (duration and spatial extent)? Are the loading models constructed to account for the combined effects of changing climate and changing land use conditions? Of particular concern to me is that it appears (from my perhaps somewhat superficial knowledge of the models) that the watershed-based loading of both dissolved and particle-bound nutrients is treated as a point source at the mouths of rivers. It is not apparent to me whether and how the watershed-scale processes that influence loading are determined and whether or not factors such as climate change, land use change, management / stewardship practices are considered.

Charge Question 1:
The indicators are relevant for addressing the Annexes, however, it is not clear that the data are available in the form and frequency needed to be input to each of the models. (see comments above on the development of a data input matrix.) I did not thoroughly read the documentation for each model, but it is incumbent upon the modeling team to identify the specific inputs needed for each model so that load-response relationships can be measured to calculate the appropriate indicators.

Specific comments about indicators:

(1) Overall phytoplankton biomass as represented by chlorophyll a -
Basin-specific, summer (June-August) average chlorophyll concentration.

No specific information was offered in terms of 1) how these will be measured 2) the spatial scale, 3) temporal frequency and time frame of the measurements. These decisions are required before modeling the load-response relationships.

Annual Chlorophyll a measures are a "blunt instrument" that provide little diagnostic power for causes of eutrophication and creation of dead zones. One of the features of the Great Lakes system is the shifting seasonal pattern of precipitation due to climate change, with larger precipitation events resulting in large runoff events. These runoff events are responsible for a large proportion of the external phosphorus loads, and result in large blooms. While summer averages can tell us something about annual trends, they do not provide a link to potential mechanisms (e.g., urban stormwater runoff; combines sewer overflows; agricultural runoff) that could be linked to spring runoff, or individual storm events. With earlier spring met and runoff, the period of measurement should be examined to assess whether the time frame of June – August represents the reality of current and (near future) tributary flow and bloom dynamics.

While annual Chlorophyll a measures are one indicator of productivity a potentially more important issue is the changing composition of the phytoplankton community. The shifting composition of the phytoplankton community, especially relative to historic patterns, is confounded by the changing climate (i.e., shorter duration of ice cover and warmer temperatures) and changing nutrient input (see comments above, paragraph 4, regarding watershed processes) and cycling patterns.

Chlorophyll a can be measured from water samples as well as remotely using remote sensing. It is important to capture both seasonal signals as well as event-based samples. The use of remote sensing technologies can be used to calibrate load-response relationships, especially as they relate to storm events.

(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)

No specific information was offered in terms of 1) how these will be measured 2) the spatial scale, 3) temporal frequency and time frame of the measurements. These decisions are required before modeling the load-response relationships.

Using remotely sensed high-resolution, hyper (or multi-) spectral imagery, it is possible to determine the area encompassing the bloom. Such measurements could potentially become incorporated into a routine monitoring program encompassing temperature, chlorophyll, and cyanobacterial pigments. (see Kutser et al. 2006 "MERIS bands 6 and 7 allow detecting phycocyanin absorption feature near 630 nm and a small peak in reflectance spectra near 650 nm characteristic to only cyanobacteria. Thus, MERIS can be used in detecting cyanobacteria if they are present in relatively high quantities." More routine use of fluorescence sensors would also be useful.

I have not examined the cyanobacteria models in detail, and wonder if the models can be spatially referenced to predict risk of blooms in particular locations based on water quality as well as climatic conditions.

Is there a reason why the direct measure of microcystin concentrations is not considered as an indicator?

(3) hypoxia in the Central Basin:
The criteria for defining hypoxia has not been provided. A biologically relevant measure of hypoxia that targets fish is now used to define the relationship between oxygen concentrations and critical temperatures known to limit some fish species. That metric (called TDO3; Jacobson et al. 2010) defines the temperature at which the DO reaches a level of 3 mg/L. A biologically relevant measure such as this may be useful for this application.

(4) Cladophora in the nearshore areas of the Eastern Basin –
Stored P Content

There are currently remote sensing tools for mapping the extent of Cladophora in the coastal waters. This seems a more appropriate measure than stored P content, which is likely to change with the "condition" of the mat. Is there sufficient knowledge to understand the metabolic relationship between the P content of the alga and the water column C, N and P concentrations? Does the age and condition of the mat influence the amount of stored P?

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?*

Model ensembles are useful when the outputs of the models can be compared directly, so that the results can be summarized in terms of the statistical properties of the outputs (range, average, etc.). Multiple lines of evidence are useful, but only if the underlying biases and behavior of the models are understood and can be interpreted appropriately, and that each can reliably derive load-response curves that can be interpreted in the context of each indicator.

The current set of proposed models range from rather simple box models to deeply complex process models with hundreds of input parameters. Coherent results among models varying in complexity can ultimately provide managers and policy makers with greater confidence about load-response relationships that may increase the potential for implementation of difficult decisions and expensive management options. However, appropriate protocols should be implemented to verify model outputs.

What is missing from the current report is a discussion of how the ensemble results will be compiled, compared, and communicated for appropriate audiences. The modeling team may wish to use the IPCC reports of model ensemble results as a model.

Comments from Dr. Klump

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?*

In summary there are essentially 4 proposed:

1) Basin specific, summertime (June-Aug) phytoplankton biomass via chl a. This is as good a general indication of eutrophication as we have, although the time period seems too narrow. Some of the worst bloom conditions observed in the last few years have occurred later in September and even into October. Remote sensing has proven a powerful tool for observing the spatial extent of bloom conditions in the western and central basins.

2) Cyanobacteria blooms in the Western Basin. Also a key indicator, and the same comment applies here with respect to the period of observation being extended into October. With the projected lengthening of the stratified period given by some climate models, one could reasonably argue for an earlier start to the observation period as well, e.g. May – Oct. It is worth noting that not all cyanobacteria produce toxins.

3) Hypoxia in the central basin, including duration, areal extent and average hypolimnetic DO concentration. This appears reasonable and tractable with an observing system using current technology.

4) Cladophora in the nearshore of the Eastern Basin – stored P content. This strikes me as the least diagnostic of the group, and the restriction to the Eastern basin does not, without further explanation, appear to make sense. Extensive cladophora growth has occurred in Lake Michigan nearshore areas in some of the most pristine areas of the lake, a direct result of dreissenid filtration providing nutrients and increased water clarity. In Lake Michigan cladophora growth would not be an indicator of eutrophication. Other nuisance algae exist in Lake Erie in addition to cladophora, e.g. Lyngbya in the western basin may have a bigger adverse impact on nearshore zones (see e.g. Bridgeman and Penamon 2010, JGLR 36:167).

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 indicator? Do the models reflect the best available scientific knowledge?*

I would answer the last question with "yes", the ecosystem modeling effort on Lake Erie is probably as well developed as in any equivalent system, and the effort has been ongoing for a long time. I am not familiar with them all, but some of these models are fairly sophisticated and complex, with a high degree of spatial resolution. Clearly some have strengths in certain areas. For example, models for forecasting hypoxia will require full incorporation of the physics of stratification, since it is the

hydrodynamics which has the biggest influence on the duration and extent of hypoxic conditions. The impact of climate change is projected to be significant, particularly since one of the major projections is a significant lengthening of the stratified period. Obviously, a reduction in benthic respiration as a result of decreased nutrient loading, may in fact be confounded by a longer stratified period.
The question is whether the models capture the relevant dynamics? And models will only provide results based on those dynamics. This is one reason why the impacts of invasive species remain virtually impossible to predict.

Simplicity is attractive. It is not clear that a proliferation of terms estimating time-dependent rate functions gains ground for management. This harkens back to indicators. If MT of algae is the target, then Chl or biomass will suffice and models which output that metric will be informative.

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 appears attractive and reasonable. It may prove more difficult in practice since it is to a significant extent an apples and oranges problem, if the intent is to somehow integrate these models. However, the ensemble approach in its simplest form may be a benefit when it comes to examining the congruence of results. One might argue it is analogous to the ensemble of GCMs for climate that largely model the same system, but formulate the basic physics differently, although that is an oversimplification. Relying on a single model would be risky. Some are likely to deliver very similar results since the formulation of the models are empirically based and derive from similar sets of equations of state.

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 the efficacy and value of establishing target values for both phosphorus loads and concentrations in order to meet 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?*

Ultimately it is the input of “new” nutrients which drives production in this system, although there is a caveat with respect to internal cycling. While the flux is key, obviously the timing and stoichiometry of inputs play a major role. To my mind establishing a target based on loading is the most defensible, and translates most directly to the likely management practices that will be needed in order to reduce that load. Although algae respond to concentrations, the magnitude of their response will be determined by inputs. Loading, furthermore, in these watershed systems, tends to be very event driven, e.g. with perhaps ~70% of the load occurring within a handful of days during which precipitation and flow is the highest. This implies that timing is critical in determining the delivery from the watershed.
Comparison of the 2011 wet spring, super bloom with the 2012 drought conditions is very instructive,

with the 2012 bloom conditions something like 10% of the 2011 observations. These 2 years would be good test for these models to see whether they capture this difference. This also implies that the response coupling for loading is strong for the western basin in particular.

Obviously, the volume of delivery is highly variable. The volume of the receiving reservoir, however, is not. It is essentially constant. Hence concentrations are of problematic regulatory or management utility, particularly if one considers that management will ultimately have to target loads.

There is already well known evidence that the stoichiometry of forms of phosphorus has shifted to an increasing proportion of the total being delivered a soluble reactive phosphate – resulting in an increasing bioavailable P pool. In addition, it is unclear from the brief model descriptions the extent to which these models incorporate internal loading, and whether it is even known if internal loading (recycling) has increased in the face of excessive external loading. Has the P inventory within the system increased over time or have dreissenids increased the residence time of P in the western basin – both would imply that internal loading might become proportionately more important. It appears that the "seed" stock of microcystis in the sediments may have expanded, allowing the western basin bloom to "light up" quickly and over a very broad area rather than expansion outward via growth. Resuspension and P regeneration may be important processes and important to include in the modeling frameworks. Recent data from PO4 sensors, for example, implies a link between resuspension, water column SRP increases, and subsequent chl a increases.

This also leads to the issue of phosphorus vs. nitrogen which has received attention in the literature (e.g. Paerl et al. Science 10 Oct 2014, p175). Microcystis does not fix nitrogen, and denitrification is likely a major N sink. On the other hand P does not disappear and is potentially available via regeneration, both in the short term and in the longer term. Bioassays have shown that P is limiting in the spring, and while the microcystis dominated phytoplankton community can become N limited later in the summer, the implication is that there is little correspondence between N limitation and the size of the microcystis bloom. It may not be this simple, however. Again the differences between 2011 and 2012 are instructive. The argument that it may not be sufficient to regulate P alone may have merit. With the preferential loss of N vs. P via denitrification, N loads could presumably affect these bloom dynamics, including size. P regeneration may be an importance process. Based upon the discussion it appears that internal loading is largely an unknown and not adequately modeled in most of the model frameworks. P may be building up in this system and represent an internal reservoir with the potential to exacerbate future conditions.

In addition, hypoxia may be driven in significant part by the longer term loading to the system and the longer term accumulation, as well as by the arrival of recent organic matter as a consequence of annual inputs, i.e. even without significant recent inputs of labile organic matter the sediments might still consume hypolimnetic oxygen at appreciable rates for some time to come. One question, therefore, is whether any of these models take into account the accumulation of excess material over a period of years, i.e. what is the time frame of these models? This also leads to the comment that the system, especially in the case of hypoxia, will likely take several years in which to respond to reductions in loading, or to put it in this framework, can or do these models make that calculation, can they provide an estimate of the likely lag between reduction and response for hypoxia? The differences in 2011 and 2012 show a remarkable difference in loading response of algal production, and cause for some optimism that the target loadings would be consistent with the eutrophication indicators of concern.

Loading has fluctuated dramatically in the last 20 years. Loads in 2012 were ~20% of the 2011 loads and the bloom was 10-20% of the 2011 bloom. Looking at the Obenour data – and the lake is providing a strong signal - if the defined goal for a nuisance condition is < 10^4 MT of algae, then loading should approach < 300 MT/month or < 4000 MT per year. This represents a reduction in excess of 50%. That will not be easy. Variability is also fairly high, and leads to the question of whether a "bad" year (v. high loads) can be tolerated. Is meeting a target 75%, 90%, etc. of the time sufficient?

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. al.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 response 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. McLaughlin

The following are my 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. 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. There is a considerable amount of written guidance available that can support an adaptive and integrated approach for addressing the challenges of understanding and managing Lake Erie nutrients. See, for example, references 1 through 7 provided at the end of these comments.

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. Providing more detail on this aspect of the approach should be a high priority for EPA.

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, and listed as reference 7 below). 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. Several references are included below that can help inform improvements to the Approach (for example, is see references 7 through 14).

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 selected 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 (see, for example, references 7, 8, 13, and 14). 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 prediction 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. Among the statistical methods that could be used to evaluate "what-if" scenarios, EPA should consider receiver operating characteristic (ROC) analysis as a method for describing the uncertainty in load:response relationships in terms of the confidence that specific load reductions will achieve desired response thresholds. References 13 and 14, and references therein may be useful in this regard.

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, and particularly in an adaptive management context. 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.

References

1. RECOVER, 2010. CERP Adaptive Management Integration Guide. Restoration Coordination and Verification, C/O U.S. Army Corps of Engineers, Jacksonville District, Jacksonville, FL and South Florida Water Management District, West Palm Beach, FL.
2. National Research Council. 2003. Adaptive monitoring and assessment for the comprehensive Everglades restoration plan. Committee on Restoration of the Greater Everglades Ecosystem. National Academy of Sciences. National Academies Press.
3. National Research Council. 2007. Models in environmental regulatory decision making. The National Academies Press.
4. US Environmental Protection Agency. 2000. Toward integrated environmental decision-making. Science Advisory Board Integrated Risk Project. August
5. National Research Council. 2009. Science and decisions: Advancing risk assessment. The National Academies Press.
6. IOM (Institute of Medicine). 2013. Environmental decisions in the face of uncertainty. Washington, DC: The National Academies Press.
7. US Environmental Protection Agency. 2009. Guidance on the development, evaluation, and application of environmental models. Washington (DC): USEPA. EPA/100/K-09/003.
8. US Environmental Protection Agency. 2006. Guidance on systematic planning using the data quality objectives process. Washington (DC): USEPA. EPA QA/G-4. EPA 240/B-06/001.
9. Hakanson, L. 2003. Propagation and analysis of uncertainty in ecosystem models. in "Models in Ecosystem Science", Canham, CD, Cole, JJ, Lauenroth, WK, eds. Princeton University Press, Princeton, NJ.

10. Gardner, RH, and Urban, DL. 2003. Model validation and testing: past lessons, present concerns, future prospects. in “Models in Ecosystem Science”, Canham, CD, Cole, JJ, Lauenroth, WK, eds. Princeton University Press, Princeton, NJ.
11. US Environmental Protection Agency. 2010. Using stressor-response relationships to derive numeric nutrient criteria. Washington (DC): USEPA. EPA-820-S-10-001.
12. US Environmental Protection Agency. April 27, 2010. Letter to EPA Administrator re: SAB Review of Empirical Approaches for Nutrient Criteria Derivation. EPA-SAB-10-006.
13. Krzanowski WJ, Hand, DJ. 2009. ROC curves for continuous data. Boca Raton (FL): Chapman & Hall/CRC. 241 p.
14. McLaughlin DB. 2012. Assessing the predictive performance of risk-based water quality criteria using decision error estimates from ROC analysis. Integr Environ Assess Manag 8:674-684.

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.

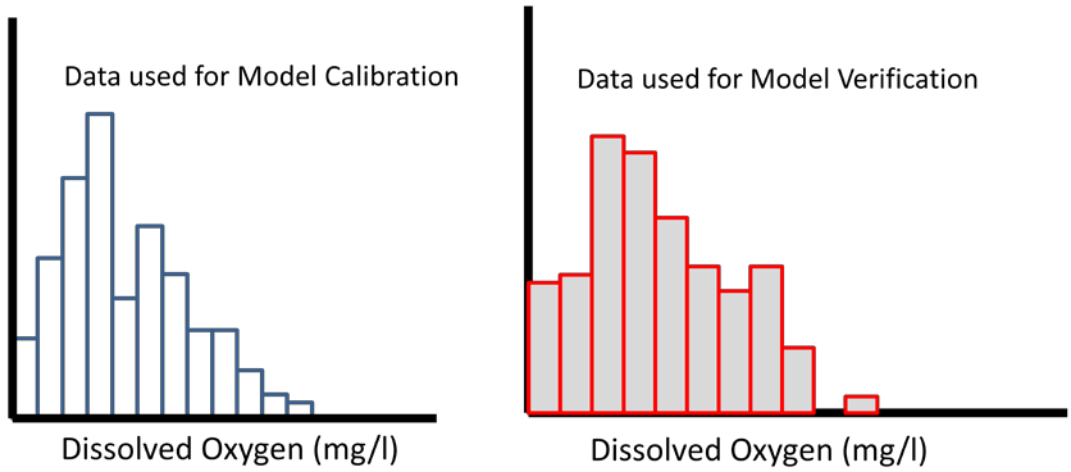
5. **Model Prediction Error** – I find it interesting that the important issue of prediction uncertainty for *Cladophora* growth is of concern, but there is virtually no comment about the prediction uncertainty for the 2D and 3D models. I realize that overparameterization of these models limits the use 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.

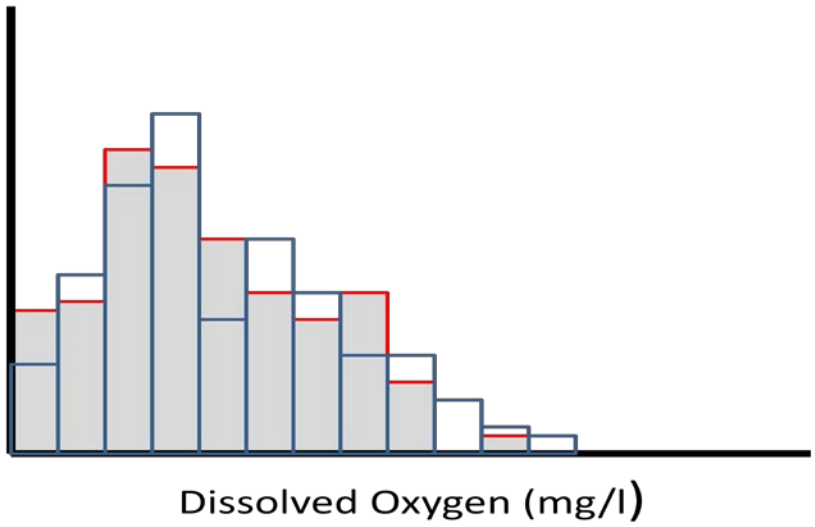
   So, conventional water quality model verification has become basically a charade. This situation is not the fault of modelers; rather, it is simply the consequence of limited available data. Nonetheless, water quality modelers who employ this approach to model verification need to be more candid about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



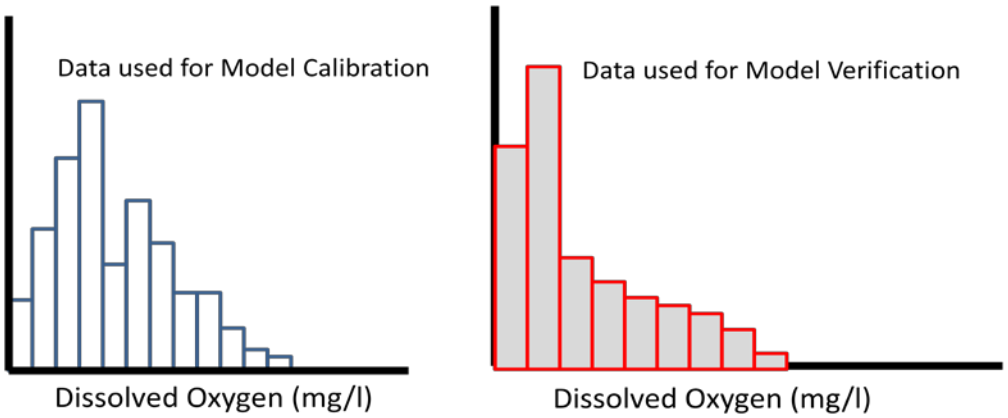
The next figure overlays the calibration and verification histograms for Case 1; notice how similar they are. The lack of difference between these two data sets indicates that "verification" lacks rigor; essentially, the model is being re-assessed with calibration-like data.

Calibration and Verification Data Comparison: Case 1



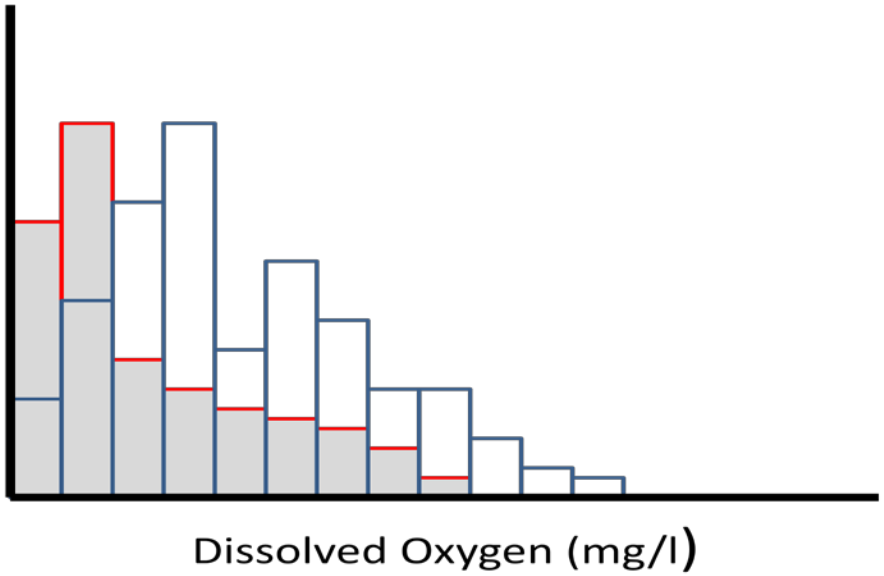
Now consider Case 2 below:

Evaluation of the Strength and Importance of Model Verification
Case 2



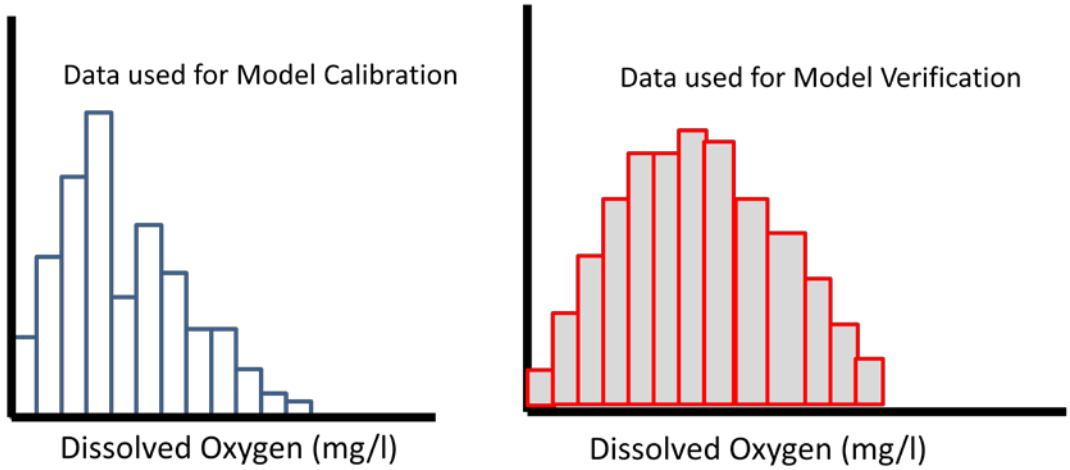
An overlay of the two histograms, shown below, indicates that the calibration and verification data sets are different, which suggests that verification is more rigorous than in case 1. However, note that the verification data in case 2 show DO to be lower than for model calibration. Since model applications 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.

Calibration and Verification Data Comparison: Case 2

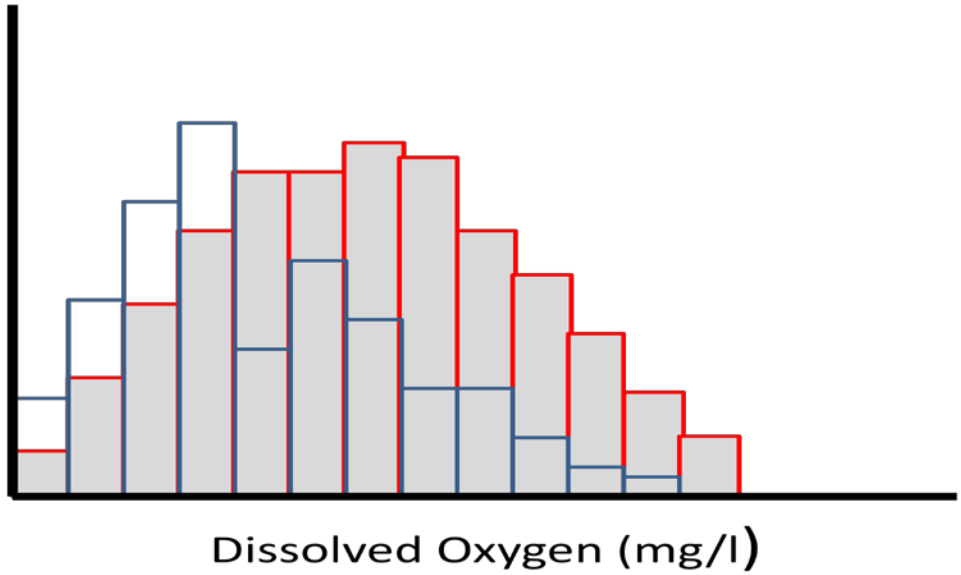


Now consider Case 3 below:

Evaluation of the Strength and Importance of Model Verification
Case 3



Calibration and Verification Data Comparison: Case 3



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. Reddy

Phase 1 Consultation Questions:

In freshwater ecosystems an increase in external nutrient input resulting from anthropogenic activities is frequently the major cause of eutrophication. The main paths of anthropogenic eutrophication (also called cultural eutrophication) in lakes are: increase in input of nutrients (mainly nitrogen and phosphorus), increase of the phytoplankton biomass, loss of biological diversity, dominance by cyanobacteria, diatom, and unicellular green algae, occurrence of algae 'blooms' (high biomass production of certain species of algae at the water surface), reduction in light and oxygen availability, change in heterotroph community composition, death of fish, and production of toxins by certain groups cyanobacterial communities that may pose threat to human or ecosystem health. All these alterations will lead to an ecosystem change, loss of species diversity and decrease in water quality. Hence, lakes with different trophic states (oligotrophic: low productivity, mesotrophic: medium productivity, eutrophic: high productivity and hyper eutrophic: very high productivity) will have distinctive physical, chemical and biological characteristics. To address critical lake management issues, the Lake Ecosystem Objectives (LEO) presented in the Draft Technical Approach for Lake Erie Phosphorus Load-Response Modeling are on target. However, approaches used to accomplish these objectives may need some modifications and refinements. Any strategies used to phosphorus load reduction to Lake Erie must take into consideration of both external sources and internal dynamics associated physical, chemical, and biological processes.

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.***

Ecosystem impairments such as eutrophication of lakes can be attributed to (i) increased external inputs of nutrients from point and nonpoint sources or (ii) accelerated nutrient cycling within the lake associated with change in environmental conditions of the sediment and water column. Typically, eutrophication is linked to increased external inputs of nutrients. However, internal nutrient sources are equally important, especially in eutrophic lakes or lakes that have large reserves of organic and inorganic bound nutrients.

An indicator is a sign or signal that relays a complex message, potentially from numerous sources, in a simplified and useful manner. The eutrophication response indicators meet the criteria of an ecological indicator is defined as a measure, an index, or a model that characterizes an ecosystem or one of its critical components (Jackson et al., 2000). An indicator may reflect biological, chemical, or physical attributes of lake's ecological conditions. The primary uses of indicators are to characterize current ecosystem status and to track and predict future significant changes. The U.S. Environmental Protection Agency provides the following guidelines to develop ecosystem indicators (Jackson et al., 2000). Conceptual relevance- is the indicator relevant to the assessment question (management concern) and to the ecological resource or function at risk? Feasibility of implementation - Are the methods for sampling and measuring the environmental variables technically feasible, appropriate, and efficient for

use in a monitoring program? Response variability - are human errors of measurements and natural variability over time and space sufficiently understood and documented? Interpretation and utility - will the indicator convey information on ecological conditions that is meaningful to environmental decision-making?

For Lake Erie phytoplankton biomass (as represented by chlorophyll a) is selected as lake-wide universal Eutrophication Response Indicator (ERI). This is a traditional indicator and has been used effectively used to evaluate tropic status of lakes. However, it was surprising to note that total phosphorus (TP) concentrations are used as a supportive indicator. No justification was provided for including this important chemical indicator. Document does not provide any clear justification or rationale for selecting phytoplankton biomass as an indicator. Was this decided based on available data for Lake Erie? Was there a mechanistic basis for selection of this indicator? For example, did the data showed a significant correlation between phytoplankton biomass and water column total P concentration. Phytoplankton phosphatase activity of phytoplankton biomass was shown to be sensitive indicator of phosphorus enrichment. How sensitive this indicator is to capture spatial and temporal patterns. Use of satellite imagery based algal bloom monitoring may provide more comprehensive information on the spatial and temporal distribution of bloom occurrence. This approach has been found to successful in a number of central and north Florida lakes. However, validation of this method requires ground-truth measurements.

Often a combination of indicators may help to address ecosystem change than any one single indicator. For example, in addition to phytoplankton biomass other potential indicators are: water column P concentrations, N/P ratios of plankton biomass, and phosphatase enzymes. These indicators have been used in other water bodies including lakes. Is there a sufficient range of values for ERIs so that they may serve as sensitive indicators of nutrient impact/recovery of Lake Erie? Any indicator (simple or complex) selection should be based and supported by complimentary foundational science. Indicators can be viewed at different levels, with Level I indicators (low cost, easily measurable, and less sensitive); Level II indicators (medium cost, moderate complexity, and moderately sensitive); and Level II indicators (high cost, very complex, and highly sensitive). Indicators selected for Lake Erie fall into the category of Level I. Selection of these indicators should be based on strong relationship between response variable (in this case Level I indicator(s) and causal variable such as internal and external drivers. While Level II and Level III indicators provide foundational science to support the selection of Level I indicator. Data can be mined from the peer-reviewed literature and reports from studies conducted on Lake Erie and other lakes to develop strong justification for selection of ERI.

Western Basin ERI – Cyanobacterial biomass can serve as response variable as its production is strongly correlated to phosphorus loads (Anderson et al., 2002; Michalak et al., 2013). Michalak et al. (2013) suggested that severe spring precipitation and long-term agricultural land use associated legacy P in the basin produced a pulse of dissolved reactive P (DRP) load to the western basin of Lake Erie. Understanding the influence of external and internal drivers (potentially serve as indicators) including P concentration and loads that regulate the biomass production and water quality is critical. It is important to document how these blooms have responded to external P load reductions. Are there any other nutrients affecting these blooms? What are the threshold values for P concentrations that influence cyanobacterial biomass? In a recent study by Improvement of water quality of the Western Basin should take high priority.

Central Basin ERI - Hypolimnion hypoxia is an appropriate indicator for this basin. Quantifying hypoxia, especially spatial and temporal patterns may be often complex. In addition to depletion of oxygen, long residence time (approximately 2 years), hypoxia can have significant influence on redox gradients and P mobilization between sediment and water column and increase the bioavailable P and internal load. Models may be able to capture the spatial and temporal linkages between hypoxia and P solubility.
Eastern Basin ERI- *Cladophora* in the nearshore areas as indicator on nutrient impacts. Higgins et al., (2008) concluded that the return of algal blooms in the eastern basin of Lake Erie is linked to ecosystem level changes including: substrate availability, water clarity, and P cycling associated with the invasive dreissenid mussels. No relationship was noted between water column P concentration and *Cladophora* blooms. *Cladophora* is shown to grow on surfaces of substrates such plant material, rock surfaces, and surfaces of animals (Higgins et al., 2008). The ERI is based on the total P content (storage) of the *Cladophora* tissue as a response variable and DRP as a causal variable. This is a very simplistic approach. Dissolved reactive P is very dynamic chemical parameter and it is highly subjected physic-chemical environment in the water and highly variable at spatial and temporal scales. This relationship is valid only if source of DRP for *Cladophora* is only from the water column. However, it should *Cladophora* and filamentous algae grown on substrates can obtain their P needs from the substrate itself. For example, P leakage from plant material can provide P needs of algae growing on this surface. Similarly, substrates such as limestone material can precipitate P on their surfaces. Algae growing on these surfaces can solubilize some of these P to support their needs. In addition, algae can effectively use organic P through enzymatic (phosphatases) hydrolysis (Reddy et al., 2005). Models should consider the internal dynamics and spatial and temporal variability that regulate DRP availability to *Cladophora.*

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?

Several modelling approaches are presented in the literature to evaluate indicators. These may include: (1 Descriptive statistics to define frequency distribution and central tendency of indicators; (2) Multivariate analyses to evaluate relationships between indicators and ecological condition; (3 Geostatistical analyses to evaluate spatial patterns and spatial structure of indicators; and (4) Predictive modeling to forecast evolution of indicators and ecological integrity.

Models chosen to evaluate ERI seem to be appropriate. It was difficult to evaluate these models without proper explanation on input parameters and rate coefficients to be used. I expect this will be done, once appropriate model(s) is selected.

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?

See comments in Question 1 and 2

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?

In large-scale ecosystem management and restoration, there is a pressing need for sensitive, reliable, rapid, and inexpensive indicators of ecological integrity. An appropriate term for indicators that meet all these criteria would be "efficient indicators. In the development of efficient indicators it is critical to determine what portions of the landscape contain the best record of overall ecosystem integrity. If the integrity of a watershed area is compromised, it is likely that it will soon be reflected in the integrity of the associated waterbody such as lakes. Conversely, if the integrity of the lake is compromised it may not be immediately reflected in the condition of the uplands, since materials transfer is largely towards the lake. Eutrophication response indicators selected for monitoring a lake directly respond to changes in land management practices in the watershed. Thus, it is critical to use models that will couple watershed land use practices to in lake processes to determine overall influence of P loads on lake biotic communities.

Improved land use management practices in the watershed can reduce overall P load to Lake Erie. Earlier modeling studies have shown the relationship between P loading and eutrophication, which has led to establishment of 11,000 metric tons P/year as target load for Lake Erie (Baker and Richards, 2002). During the period 1997-2007, external P loads from watershed to western basin of the lake were at target load of 11,000 metric tons per year for 20 out of 30 years. This is probably due to change in land management practices. However, during this period, farmers continue to apply fertilizer P to grow crops. This probably resulted in accumulation of P in soils. Although P application rates have decreased (Baker and Richards, 2002), loads to lake has not changed. This is probably due to the presence of legacy P stored in the watershed that can potentially serve as source of P to the lake. The key questions are: (i) will lake respond to P load reduction? (ii) if so, how long will it take for the systems to recover and reach background conditions? (iii) what are the economically feasible management options in the watershed to reduce P loads hasten the recovery process?, and (5) to what extent the internal load from benthic sediments contribute to the overall eutrophication process.

It is always difficult to evaluate the effectiveness BMPs to reduce overall P load. Phosphorus release from soils is a slow process and characterizing the P release potential (P memory) from the reactive P pool in soils and sediments of the watershed and the lake requires consideration of both biotic and abiotic processes. Biotic processes in the watershed soils and lake sediments include assimilation by vegetation, algae, microorganisms; abiotic processes include sedimentation, adsorption by soils, precipitation, and exchange processes between soil and the overlying water column (Reddy et al., 2005, 2011). This “P memory” can extend the time required for restoration and recovery to a lower nutrient status that historically supported the lake to maintain oligotrophic conditions. This lag time for

recovery should be considered in developing restoration and management strategies for reducing P loads.

Phosphorus concentration in the lake – what drives the TP and forms of P in the water column? Phosphorus concentration in the lake is regulated by: external loading influenced watershed management practices and internal loading from benthic sediments and water column.

External P loads to the lake are influenced by: agricultural practices including nutrient management, phosphorus application beyond the crop needs, accumulation of P in soils and tributary sediments; reactivity and mobility of P in the watershed. To determine the influence of legacy P on loads to the lake, it is important to develop total P budget including P stored in soils and sediment. Questions needed to be addressed include: How much P is stored in the watershed soils and what proportion is present in stable pool and permanently buried in soil. What adaptive management practices can be used in the watershed to reduce mobility of P.? Earlier monitoring studies (1975-1995) in the watershed addressed some of these issues. A series of papers were published on this topic (see Richard et al., 2002; Baker and Richards, 2002). Much can be learned from these studies as new adaptive management strategies are implemented.

Internal P loads are regulated by several biogeochemical processes in stratified water column and in benthic sediments (Reddy et al., 2011). Internal loads can increase the lag time for recovery of lakes, even after external loads are reduced. The processes affecting P exchange at the sediment–water interface include (i) diffusion and advection due to wind-driven currents, (ii) diffusion and advection due to flow and bioturbation, (iii) processes within the water column (mineralization, sorption by particulate matter, and biotic uptake and release), (iv) diagenetic processes (mineralization, sorption, and precipitation dissolution) in benthic sediments, and (v) redox conditions at the sediment–water interface. During windy periods, resuspension and deposition of sediments may be an important mode of phosphorus transfer to the water column. Because sediment resuspension events are transitory, P flux by this process may occur at shorter timescales but at more rapid rates compared to diffusive flux. Future research should focus on improving the understanding of the role/importance of internal load in regulating eutrophication.

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 Lake Ecosystem Objectives. The phosphorus loading targets could be converted to concentration targets, particularly for river mouths/nearshore zones.

References

Auer, M.T., Tomlinson, L.M., Higgins, S.N., Malkin, S.Y., Howell, E.T., and Bootsma, H.A. 2010. Great Lakes *Cladophora* in the 21st Century: Same algae – different ecosystem, J. Great Lakes Research, 36:248-255.

Baker, D. B and R. P. Richards. 2002. Phosphorus budgets and riverine phosphorus export in Northwestern Ohio watersheds. J. Environ. Qual. 31:96–108.

Higgins, S. N., S. Y. Malkin, E. T. Howell, S. J. Guildford, L. Campbell, V. Hiriart-Baer and R. E. Hecky. 2008. An ecological review of *Cladophora glomerata* (Chlorophyta) in the laurentian Great Lakes. J. Phycology. 44:839-854.

Jackson, L. E., J. C. Kurtz, and W. S. Fisher (eds.). 2000. *Evaluation Guidelines for Ecological Indicators*. EPA/620/R-99/005. U.S. Environmental Protection Agency, Office of Research and Development, Research Triangle Park, NC. 107 pp.

Michalaka, A. M; Eric J. Anderson; Dmitry Beletsky; Steven Boland; Nathan S. Bosch; Thomas B. Bridgeman; Justin D. Chaffin; Kyunghwa Cho; Rem Confesor; Irem Dalolug; Joseph V. DePinto; Mary Anne Evans; Gary L. Fahnenstiel; Lingli He; Jeff C. Hol; Liza Jenkins; Thomas H. Johengen; Kevin C. Kuo; Elizabeth LaPort; Xiaojian Liu; Michael R. McWilliams; Michael R. Moore; Derek J. Possel; R. Peter Richard; Donald Scavia; Allison L. Steiner; Ed Verhamme; David M. Wright; and Melissa A. Zagorski. 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. Proc. National Academy Sciences. 110:6448–6452.

Reddy, K. R., R. G. Wetzel, and R. H. Kadlec. 2005. Biogeochemistry of Phosphorus in Wetlands. In Phosphorus: Agriculture and the Environment J. T. Sims and A. N. Sharpley (eds), Soil Science Society of America. pp. 263-316

Reddy, K. R., S. Newman, T. Z. Osborne, J. R. White, and H. C. Fitz. 2011. Phosphorus cycling in the Everglades ecosystem: Legacy phosphorus implications for management and restoration. Crit. Rev. Env. Sci. Technol., 41:149–186.

Richard, F. G, D. B. Baker, and D. J. Eckert. 2002. Trends in agriculture in the LEASEQ watersheds, 1975–1995. J. Environ. Qual. 31:17–24.

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 a 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. In addition to these indicators, it might be informative to actually measure microcystin concentrations.

The document would benefit from further clarification for the reasoning behind segregating the eutrophication indicators in space. The three areas of Lake Erie apparently function somewhat independently and this should be described more thoroughly in the Technical Approach document; on pages 5-6 a further description of this approach would be helpful. In addition, the text clarifying why the metrics are useful in different sections of the lake would be a useful.

The technical document would be strengthened by indicating 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. The point of adaptive management is to do experiments and learn from these experiments. Is there an opportunity for learning in this plan? This should be specifically described in the document. It is also 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 am impressed with the team that EPA has put together to tackle this problem. The team has a strong background on the lake, modeling of the lake metrics and that data that is available. The team has a difficult task in that that the lake dynamics seems to have changed over time and this may make the modeling effort more difficult. In addition, having multiple metrics at different scales makes the process time consuming.

I am still a bit confused about the data that are available, and the scale for modeling and assessment. If the modeling is on an annual scale at a single location, then it seems that a lot less can be accomplished with the models as there are likely to be more model parameters than there are observations to evaluate or fit the parameters. Modeling on a finer scale is more likely to be productive especially if there is adequate data and the model parameters do not increase with the finer scale. There obviously has to be some matching of the scale of the data with the scale of the model.

With regard to the ensemble, I was expecting to see more on the combination of the model predictions. It seems to me that the approach is to fit a number of models then just compare them with the data. If possible, combining the predictions of the relevant metric using a weighted combination with weights connected to model predicted ability should lead to better prediction. There is ample evidence of this in the meteorology literature (Using Bayesian Model Averaging to Calibrate Forecast Ensembles Raftery, A.E., et al., Monthly Weather Review, 2005, 133:1155-1174) and some evidence in other areas (for example, Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) III: Scenario analysis. J.A. Huismana, ,et al., Advances in Water Resources Volume 32, Issue 2, February 2009, Pages 159–170 – there are four articles on this approach by the authors in the issue). A difficulty is the scale of the different models although one might smooth the predictions or aggregate predictions. Thus, unless there is adequate data the approach may not be rewarding. 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 can envision a regression like model that would have the metric as the response variable and input from the model or models as explanatory variable. Other metrics might also be used as input. This would allow for an ensemble approach

For model verification/validation, 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, Rose, K.A., Roth, B.M., Smith, E.P. Skill assessment of spatial maps. (2009). Journal of Marine Science 76:34-48 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 assessment 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?