Summary of Model Evaluation Group (MEG) Meeting
Monday, September 12, 2005
10:00 AM to 3:00 PM
MIT Sea Grant, Cambridge, MA

Attendance

MEG Members:
Eric Adams (chair), MIT
Jack Kelly, EPA Duluth MN Research Lab
Pierre Lermusiaux, Harvard
Rich Signell, USGS
Steve Chapra, Tufts
Jim Kremer, U. Connecticut
John Paul, EPA Research Triangle, NC
Huijie Xue, U. Maine.

Other Attendees:
Martin Dowgert, FDA
Ferdi Hellweger, Northeastern U.
Mingshun Jiang, U. Mass Boston
Wendy Leo, MWRA
Mike Mickelson, MWRA
Andrea Rex, MWRA
Cathy Vakalopoulos, MADEP
Meng Zhou, U. Mass Boston
Jim Fitzpatrick, HydroQual
Carlton Hunt, Battelle
Yong Lao, MWRA
Matt Liebman, EPA
Judy Pederson, MIT, OMSAP
Larry Schafer, retired
Gordon Wallace, U. Mass Boston

Purpose and Process

The Model Evaluation Group (MEG) convened on 9/12/2005. The purpose of this meeting was to review the modeling reports[1] for years 2000-2001 written by the U Mass Boston modeling team and to provide advice on the future direction of the modeling.

The meeting was tape-recorded and transcribed by Cathy Vakalopoulos. An unsorted list of recommendations made by MEG and audience members was culled from the transcripts and is included as an attachment. The following draft summary was compiled by Cathy, Mike Mickelson and Eric Adams.

Background

MEG is a focus group of the Outfall Monitoring Science Advisory Panel (OMSAP) that advises EPA and MADEP on MWRA's outfall monitoring. Since 1992, MEG has provided independent technical advice on the development, maintenance, and running of MWRA's Bays Eutrophication Model (BEM). BEM consists of a hydrodynamic (HD) model[2] coupled to a water quality (WQ) model[3], both adapted to cover Massachusetts Bay and several miles beyond into the Gulf of Maine (GoM).

MEG last convened in 2002[4] and MWRA has since been working on addressing MEG's recommendations. Over the years, many of MEG's recommendations have been implemented. For example:

- Add more monitoring stations near the boundary with GoM
- Add an algal growth carbon/chlorophyll model to simulate the subsurface chlorophyll maximum.
- Add a third algal group to simulate high chlorophyll in fall blooms.
- Increase the horizontal resolution of the WQ model to match that of the HD model.

In addition, MEG recommended that an academic home be found for BEM so it could continue to evolve and be available to other agencies or groups. HydroQual and USGS had used BEM to model years 1990-1999 (with a gap in 1995-1997), and then U Mass Boston modeled 2000-2001.

[1] These are reports # 2004-08 and 2004-09 at http://www.mwra.state.ma.us/harbor/enquad/trlist.html
[2] ECOM-si: Estuarine, Coastal and Ocean Model (semi-implicit). http://woodshole.er.usgs.gov/operations/modeling/ecomsi.html
[3] RCA: Row Column AESOP (Advanced Ecological Systems Operating Program). http://www.hydroqual.com/wr_rca.html
[4] The 2002 MEG report is located at: http://www.epa.gov/region01/omsap/meg0302.html

[Note added after the meeting: MEG comments on the 2000-2001 reports are reflected in improvements in the 2002-2004 reports[5] which were completed after the MEG meeting.]

BEM was developed to model the effects of MWRA's 9.5-mile-long outfall located offshore in Massachusetts Bay. Important modeling results from BEM include

- Harbor DO is improved by secondary treatment; Harbor chlorophyll is improved by outfall relocation.
- Relocation of the outfall has only minor effects on Bay DO and chlorophyll.
- The MWRA outfall contributes only 3% of the total nitrogen load to the Bay.
- Low DO episodes in the Bay are mostly due to inflow from GoM.

Those model results were especially useful as predictions needed before the outfall went on-line, in September 2000. Later, modeling became a requirement in MWRA's National Pollutant Discharge Elimination System (NPDES) permit, with MWRA required to run BEM on an annual basis (hindcasting). MWRA's permit however is currently in the renewal process and MEG's recommendations will help define how to best use the model and assist EPA in determining what (if any) modeling requirements will exist in the new permit. EPA remains interested in whether the model can help discern negative environmental effects of the outfall on Mass Bay.

MEG Review and Recommendations

1 General Conclusions

MEG expressed overall satisfaction with the modeling efforts.

1.1 The model.

A considerable investment has been made in the existing model. While newer models are available, they generally don't differ from the current model in fundamental ways. Improvements can be made to the existing models through increased resolution, improved data, and/or additional processes. While there is little sentiment to change models, there would be merit in increased collaboration with other modelers and in comparing the results of the existing model with results from other models with overlapping domains.

1.2 The modelers.

It is good to have found a home for the model. The baton seems to have been passed successfully, and the modelers are engaged in some interesting studies related to mesoscale processes, sample design, boundary forcing and forecasting that will extend model utility.

1.3 The model results.

The apparent level of agreement between model results and observations seems reasonable and is typical of other models. But the MEG has many questions/comments as indicated below.

2 Questions guiding future model effort

Future modeling effort, and an assessment of how successful this effort is, should be viewed in terms of a number of factors.

2.1 Who is using the model (MWRA, other agencies, other scientists)?

2.2 How will the model be used (hindcasting, nowcasting, forecasting)?

2.3 What is the simulation time frame (average year, interannual variability, monthly variability)?

2.4 How will output be aggregated (predicted concentrations as a function of space and time or aggregated fluxes attributed to boundaries, internal processes, point sources)?

2.5 Will predictions be absolute, e.g., c(x,y,z,t), or relative, e.g., incremental Δc due to outfall?

[5] These are reports # 2006-12 and 2006-13 at http://www.mwra.state.ma.us/harbor/enquad/trlist.html

3 Suggestions to improve model validity and presentation

3.1 Model documentation

The model reports lack information about how the modeling was done objectively, including how boundary data were assimilated. Inclusion of this information would improve confidence in model validity.

Model calibrations, changes, etc. should be cataloged in one place. Differences in model formulations, parameters, and data sources used from year to year should be noted.

Be careful with terminology: don't interchange "calibration" and "validation".

Be consistent and use either calendar or Julian days.

3.2 Model-data comparison

The same filters and depths should be used when comparing model results and observations.

The same forcing functions (irradiance and extinction) should be used when comparing calculated values such as primary production.

Be careful how different phytoplankton groups are defined. It is difficult to defend how the summer, winter and fall groups differ using empirical data, especially field data.

4 Assessment measures

4.1 Summer salinity

The model misses certain aspects of the summer salinity structure. The biological implications could be assessed by plotting predicted bottom as well as surface currents to see how salinity affects transport.

4.2 Interannual comparisons

It is not clear if the model can distinguish different years. To test this capability, important attributes, e.g., yearly minimum dissolved oxygen, should be computed for each year and the years should be tabulated by rank order, for comparison with similar rankings based on observations.

4.3 Effects of outfall

Predictions from post-outfall and pre-outfall years should be compared. If a negative event is observed in a post-outfall year, the outfall should be "subtracted" to see if the event still occurs. Likewise, in a pre-outfall year, one could see how much outfall loading would need to be "added" to trigger concern.

4.4 Focus on boundary conditions

Understanding how open boundary conditions affect interannual variability is critical. Boundary salinities should be displayed to demonstrate whether boundary assimilation is working.

4.5 Algal respiration

Care should be taken to distinguish between net primary production and net community production.

4.6 Additional variables

Model results could be compared with satellite imagery to see if the fronts align.

Observed and predicted extinction coefficients should be compared.

4.7 Other approaches besides simply "observed" vs. "predicted"

Sensitivity to model forcing should be examined. For example as nutrient loading is varied from low to high, do annual productivities, maximum chlorophylls, and sediment deposition totals fit expected patterns?

Look for correlations in related variable (e.g., chlorophyll should be inversely related to nitrogen).

Compare predictions of a climatological year with that of actual years.

Develop a dissolved oxygen budget like the nitrogen budget conducted previously.

5 Can the modeling be improved?

5.1 Use additional data to force the model?

One could obtain atmospheric-ocean flux predictions from weather models, to either drive model or compare with model results. (The model currently computes fluxes from computed surface temperatures, which provides some degree of self-correction.)

The UNH monitoring program (monthly, beginning in 2003) may have useful data to share.

5.2 Nest the model?

Test whether using the Gulf of Maine model for outer boundary conditions is better than the current approach based on limited sampling.

5.3 Improve vertical hydrographic structure?

The thermocline is not sharp enough in many of the modeled vertical sections. Remedial options include a revised turbulence closure model (not strongly endorsed), increased vertical resolution (by a factor of 1.5 to 2 beyond the current 12 layers), and use of 2D atmospheric forcing (as long as it resolves the sea breeze). The need for additional improvement here should be gauged by their effect on the biology.

5.4 Increase the horizontal resolution?

U Mass Boston is evaluating this.

5.5 Increase the complexity of the zooplankton model?

The current model with a simple instantaneous temperature-dependent death rate for phytoplankton is unable to model several observed features, such as the lag between phytoplankton stock and grazing rate. However, zooplankton are very difficult to simulate and augmenting the zooplankton component is not recommended.

6 Future directions

6.1 Model every year?

Yes. An important use of the model is hindcasting: using the model to help determine the cause of observed events. Everyone learns when a model fails, especially for a new reason. We should look at all previous model failures to learn how to make the model better. While not all years are equally "interesting", if you don't model every year, you lose expertise.

6.2 Relationship to monitoring

It's very important to have a monitoring program coupled to the modeling. Having each run continuously allows all the monitoring data to be tied together. The model is available to help answer questions raised by the data, to guide strategic sampling, and to fill in gaps between monitoring.

6.3 Stellwagen Basin

Monitoring data suggest particle transport that focuses in Stellwagen Basin. How much of that carbon is from the outfall or from productivity stimulated by the outfall could be addressed by turning off the MWRA sources of carbon and nitrogen.

6.4 Tighter nutrient management

A sensitivity analysis could be conducted to determine what combinations of effluent/physical conditions promote low DO. This would help if a Mass Bay-wide TMDL is ever developed in the for nutrients or organics.

6.5 Additional output

The state shellfish sanitation program is interested in seeing additional modeling work that involves plume tracking and following a discrete patch of effluent.

MWRA could identify a set of treatment-plant-failure scenarios, feed those into the model, and use the results to determine the scope of ambient monitoring to detect effects on the Bay.

Attachment
Draft attributed unsorted Recommendations

MEG Recommendations

1. Compare interannual variability and minimum dissolved oxygen. (Signell) A table of each year compared side by side would be useful. U Mass Boston has so far modeled 2000-1 and 2003-4 and are looking at the variation between years. The outfall went on-line September 2000.

2. Be careful with terminology: don't interchange "calibration" and "validation". (Paul)

3. Obtain real time forcings or previous analyses from others (e.g. Collier at URI) to calculate proper fluxes. (Lermusiaux) **Counterpoint:** fluxes should be based on model predictions so that if there is some automatic correction in the model. (Adams)

4. Model-data comparisons are best done when data are plentiful. If there is a large gap in field data, don't interpolate in case there is a feature there. (Kremer)

5. Check to see that temperature and salinity are filtered the same way. (Signell)

6. Make sure that there is averaging of the observed temperature field, i.e. confirm smoothing (overhead #24). (Signell)

7. Buoy A is near station N18. Buoy A data are from 5m and N18 data are from the surface. There is a temperature difference between these two locations of 10 degrees. (Xue) Both should be compared at 5 m. (Signell) Could compare with satellite imagery to see if there is a front. (Paul)

8. In general, it would be useful to compare satellite imagery to model results. (Paul)

9. UNH monitoring program may have useful data to share. (Xue)

10. Salinity tends to not be modeled accurately, especially during summer stratification. How good does the model have to be if salinity, for example, does not affect surface currents? (Kelly) The U Mass Boston group noted that data assimilation significantly improved the salinity.

11. It would be helpful to show plots of modeled bottom currents as well as modeled surface currents. The model has a problem with currents at depth because they are affected by salinity. (Kelly)

12. How important is the apparent inability to predict the summer salinities? Are there biological and ecological implications for this in terms of vertical exchanges, i.e. water column stability? (Kremer)

13. The thermocline is too spread out in most of the modeled vertical sections compared to reality. Have there been any trials done on increasing the vertical resolution (currently use 12 layers) in the HD model to see if the stability of the water column improves? Vertical resolution could be increased by 1.5 to 2. Or as a quick improvement, 2D atmospheric forcing could be used (as long as it resolved seabreeze). (Signell & Lermusiaux) **Counterpoint:** Don't necessarily have to go to a more complicated physical structure. If it is decided to improve the vertical physics, keep it in the context of how well the biology is going to work. It is an open question as to whether improving physics will improve the biology. (Kremer)

14. There has been some work done by Bouchard and others on turbulence closures that compare different schemes. Lermusiaux can provide references.

15. To test the model's ability to predict temperature, salinity, and currents, run the model for 2001 (post-outfall) and compare the results to pre-outfall years. (Kelly)

16. Test whether using the Gulf of Maine model for boundary conditions is better than the current approach of using the limited sampling and buoy data available from the boundary. Set up three experiments – model one year, model key events over a year, and use a weighted average of the two. The third approach is potentially the most effective since care must be taken when two models with slightly different scales are forced. It's usually better to use the open boundary model if it captures the mean correctly, instead of manipulating the data. However, problems arise if the mean of the GOM model is slightly inaccurate. (Lermusiaux)

17. If the overall focus is to have an accurate model, then all of these year-by-year questions need to be addressed. If MWRA only needs a model that is representative of the types of features seen in that environment, then in an abstract way, ask the model questions such as the relative influence of the outfall pollutants. If this is adequate, then this becomes less of a challenge. (Kremer)

18. Significant improvement in the WQ model vertical resolution was accomplished when the subsurface chlorophyll maximum could be reproduced. (Paul)

19. To examine how accurate the model needs to be, pick one year and run the WQ model with "what-ifs" on extreme conditions and see how model improvements affect the modeled water quality results (Kelly). If a negative event occurs, figure out how to model it correctly, then "subtract" the outfall and see you the event still occurs. (Adams)

20. One of the important uses of the model is hindcasting: if an event occurs in Mass Bay, use the model to help determine the cause. This may mean that the model does not have to keep running. (Signell) **Counterpoints:** If you don't plan to run the model every year, then you lose expertise. (Paul) If you want the model to guide strategic sampling, then there is value to continually run the model. (Xue)

21. Understanding how open boundary conditions affect interannual variability is critical. (Signell)

22. Now is a good time for MWRA to identify the future direction of the modeling. For example, rapid response studies or trying to integrate the monitoring and the modeling programs better (by trying to improve the design or getting the critical data). Are there any other potential collaborations with other universities or organizations? Would it be advantageous to combine other models? (Lermusiaux)

23. In the context of Mass Bay, the models available are all based on the same physics. Some have slightly better advection schemes or handle vertical coordinates slightly differently, but they all have similar vertical mixing schemes and are capable of the same kinds of resolutions. At the current model resolution of 1 km, one model is not better than another. (Signell) Keep the current model – a lot of work and resources have already been spent on it, and it would take two to three years to get a new model up and running. (Lermusiaux)

24. The modeling field is moving away from competing models and towards using a combination of models. (Lermusiaux) In meteorology, modelers run a suite of models each with their strengths and weaknesses. When they converge, it increases faith in their forecast. There might be some benefit to this approach with oceanic modeling. (Kremer)

25. To help in the evaluation of the modeling results, add a table of observed and predicted extinction coefficients. (Kremer) Need to make sure the same forcing functions are used to make calculations. (Kelly)

26. Be careful when measuring algal respiration. There is a difference between net primary production and net community production. How algal and entire plankton community respiration are calculated makes a significant difference in the observed vs. predicted results. There is a great deal of difficulty in establishing what algal respiration should be. The physiological algal respiration model is based on only a few cultures that cannot possibly represent algal respiration in the field. Our data suggest that algal respiration might reach as high as 50-70% of production at certain times. Algal respiration, grazing, and death should be in the range of 60-80%. It's difficult to evaluate the ecology without some information on the relative balance of those rates and what fraction of those productions go into the various losses seasonally. (Kremer)

27. In reports and presentations, be consistent and either use calendar or Julian days. (Paul)

28. The general approach has been to compare "observed" vs. "predicted" but there are other approaches that can test whether the model is capturing some of the essential features. For example, cross-correlations between related variables, e.g. chlorophyll is inversely related to nitrogen. Cross-system comparisons - load model with very low to very high loading rates, should get annual productivities, maximum chlorophylls, and sediment deposition totals to fit accepted patterns. (Kremer) Could also run a simple baseline model (e.g. a climatology model) to see what additional benefit is gained from the complexity of the model. (Signell)

29. A simpler model doesn't require specifying information that is unknown or can't be measured. (Kremer)

30. Be careful how the different phytoplankton groups are defined. It's difficult to defend how the summer, winter, and fall groups differ using empirical data, especially field data. (Kremer)

31. Zooplankton are very difficult to simulate and so augmenting the zooplankton component is not recommended. Presently, there may not even be enough data available. (Chapra) Progressing with the zooplankton leads up the food chain and predictive value is reduced. Though this would be good research, it wouldn't be helpful in terms of answering questions related to the outfall. (Kremer)

32. The idea that the grazing rate should lag the phytoplankton stock allows for a very simple proxy for a zooplankton and pseudo-predator/prey dynamics. However, this pattern was not shown in the modeling results. Instead, the result was constant high phytoplankton. This may be due to an incorrect magnitude of the grazing. If you increase the magnitude, make sure it is changed all the way through because it may have and effect on sediment deposition and bottom oxygen concentrations. (Kremer)

33. U Mass Boston has shown that there are some interesting and useful uses of the model that investigate some processes, scales, and dynamics that, in combination with the monitoring and buoy data, give some insights on Mass Bay. (Kelly)

34. The question of "is the model good enough?" is difficult to answer because there hasn't been a synthesis shown of how the model and observational data relate. It would also be useful to catalog all calibrations, changes, etc. in one place. We want to make sure that the model produces the right answers based on the right reasons. (Kelly)

35. Conduct a dissolved oxygen budget much like the nitrogen budget that was done a few years ago. (Fitzpatrick) If this is conducted, it should be done looking back at previous years. There are times when the dissolved oxygen minimum was not captured by the model and this would help learn what can be fixed in the model. (Kelly)

36. All models have flaws. We should never take much solace when they appear to work. (Kremer) **Counterpoint:** Because models are based on mass balance, they do bound the problem and do an accounting. So though there are uncertainties, we are not creating something out of nothing. (Chapra)

37. The BEM is typical of other models. (Kremer & Chapra)

38. A sensitivity analysis could be conducted to ask what combinations of effluent and/or physical conditions would promote low dissolved oxygen. This would also help if in the future a Mass Bay-wide total maximum daily limit (TMDL) is developed for nutrients or organics. (Kelly)

39. It's very important to have a monitoring program with a modeling program (Signell). Modeling and data go together. Having monitoring and modeling that runs on a continuous basis from year to year provides a means to tie all the monitoring data together and model is available to help answer questions. (Chapra) Modeling is important to show what is happening when there is no monitoring. (Xue)

40. Everyone learns when the model fails. The only reason to run it every year is if there is an interesting event and the model fails for a new reason. Should make sure have looked at all of the times that the model has failed from the past so that we can learn how to make the model better. (Signell)

41. Overall progress was very good. The future directions identified in the presentations (mesoscale effects, sampling design, open-boundary procedures and schemes, etc) were appropriate. (Lermusiaux)

42. Modeling Investigations (Lermusiaux)
    - Vertical resolution
    - Utilization of SST, SSH and SSC, both for model calibration and data assimilation
    - Turbulence closures
    - Atmospheric forcing (x,y,t)
    - Sensitivity to selected biological processes (e.g. what happens if a given process, or a term in the equations, is removed)
    - Parameter estimation

43. Potential Additional Directions (Lermusiaux)
    - Collaborations among, and integration of, modeling and monitoring/sampling efforts
    - Rapid responses (on demand investigations, events, at-sea accidents, pollutions, homeland security, etc)
    - Seasonal and smaller scales
    - Sampling array design and adaptive sampling
    - Participations in planned academic and business-related research efforts in Mass Bay
    - Diversifications: e.g. nowcasts/forecasts for recreational/tourism activities and their monitoring

Audience Recommendations

44. One thing that was learned from the 1993 *Asterionellopsis* bloom is that there is a lot of variability in the carbon to chlorophyll ratio due to the particulate organic carbon (POC). POC and the respiration that is associated with it should be considered. (Fitzpatrick)

45. We are seeing particle transport that seems to focus in Stellwagen Basin. How much of that carbon is from the outfall and how much is from productivity stimulated by the outfall? This is a chronic, subtle, long term effect. (Pederson & Hunt) Should be able to address the effect of the outfall by turning off the MWRA carbon and nitrogen. This would help answer how much the long term carbon accumulation (and potentially metals) there is in Stellwagen Basin. (Fitzpatrick)

46. U Mass Boston showed phytoplankton biomass, but they should also show how the nutrients were affected. (Fitzpatrick)

47. Dowgert: With respect to the shellfish sanitation program any additional modeling work that involves plume tracing and following a discrete patch of effluent would be of interest. Questions include the following:
    1. What would be the time of travel, dilution and fate of a patch of effluent that represents a 24 hour failure of disinfection and secondary treatment at the upper limit hydraulic flowrate for adequate primary treatment and under several wind, tidal and seasonal regimes?
    2. What if the above disinfection and treatment failure and elevated flowrate persisted for 72 hours?
    3. What would be the time of travel of the leading edge and the dilution of the plume if the plant provides disinfection at the permit upper limit of 14,000 FC and at the upper limit flowrate that provides for full secondary treatment on an ongoing basis? I would suggest that any FC die-off factors used be very conservative to represent the low light conditions encountered at depth and during the winter season.

48. Schafer recommends:
    1. MWRA establish a set of failure episodes to feed into the model
    2. concerned scientists develop a final version of the model, with the goal of
        a. having a working tool for predicting the consequence of any variation in treatment.
        b. define just what on-going monitoring should be set up for the Bay
        c. run the MWRA criteria and write up the results.