Improving Aquatic Modeling: Evaluating Watershed Model Results Breakout Group 2 – Report Out June 30, 2016

Breakout Group 2
Evaluation of Aquatic Exposure Modeling Estimates

- CHARGE QUESTION (1):
  - **In what ways are a "multiple lines of evidence" approach appropriate for evaluating the results from a watershed model?**
  - **What would be the "lines of evidence" and sources of information?**

Charge Question 1 - Responses

- Yes, multiple lines of evidence are appropriate and add confidence to the conclusions.
- Monitoring data is useful as a line of evidence (must have adequate quality and detail; but does represent an integration of what is happening in the watershed)
- Consider applying known frequency distributions to estimate unmeasured peaks
- Consider Monte Carlo distributions/simulations – pick upper bound of simulations rather than compound high-end inputs
- Look at Monte Carlo distributions of applications / use
  - Use Monte Carlo simulations for sensitivity analysis
  - Look at whole distribution instead of high-end
  - Consider a decision point based on the entire distribution rather than compound high-end inputs
- Use a robust data set to validate watershed modeling

Charge Question 1 - Responses

- Compare process models with statistical models (WARP, SPARROW); they should feed into each other
  - Process models take info and “roll it up;” statistical models identify factors
- Runoff/baseflow: ensure that the hydrology is accurate
- Compare annual measured load loss vs modeled annual load loss
  - literature typically reports annual percent pesticide loss (i.e., load) between 1% to 5%. (See Capel et al., 2001)
- Tiered approach for evaluation: hydrology (stream gauge), sediment (watershed-level), field scale pesticide monitoring; integration at watershed. Evaluate components of a model as well as outputs
- Other models should be considered (e.g. SWAT, WARP, etc.)

WARP – Watershed Regression for Pesticides. SPARROW - SPAtially Referenced Regressions On Watershed attributes. SWAT – Soil and Watershed Assessment Tool

Charge Question 1 - Responses

- What does it mean to evaluate watershed model results with monitoring data?
  - Use as a bounding estimate (upper or lower) in comparison to modeling / validation
  - Use an estimate of exposure provided there is sufficient supporting information
  - Estimate annual loads
  - Statistically compare with flow/parameters where we would expect to see relationships
  - If models don't match expected relationships, need to reevaluate model
- Use data is one of most important issues; it is difficult to obtain but needed to characterize results
- Compare results with other existing models ("well-validated") that simulate processes not captured by PRZM (or current model)
  - Consider guidance on evaluation of models when weighing the strengths and weakness of the model (e.g. https://www.epa.gov/sites/production/files/2015-04/documents/cred_guidance_0309.pdf)
- Evaluate appropriateness of modeled habitat to actual species habitat

PRZM – Pesticide Root Zone Model

Charge Question 1 - Responses

- Fate properties relative to downstream drop-off (e.g., dispersivity) should be considered
- As the spatial scale of the model increases, the amount of detail decreases
  - A larger watershed can be characterized by more averaging of information
  - Knowledge of the watershed factors (e.g. use, flow) is still important but in a more general fashion

Breakout Group 2
Evaluation of Aquatic Exposure Modeling Estimates

- CHARGE QUESTION (2):
  - **How can different types of monitoring data be distinguished?**
  - **What metadata requirements (*e.g.*, use info, sample frequency, etc.) can be used to distinguish types of monitoring data?**

Charge Question 2 - Responses

- Field scale: typically collected with sampler at edge of field. Typically know the following
  - Specific amount of pesticide applied
  - Measured rainfall that occurred
  - Known field characteristics
  - Measured runoff loads (as described in NAS report, p. 40)
  - Historically used to validate PRZM (See FEMVTF Report http://femvtf.com/femvtf/index.htm & W Warren-Hicks, et al. Environ. Tox. Chem, 21, pp.1570-1577, 2002)
- Ambient/general: not targeted to a specific pesticide; supporting data not known
  - Context may be "answerable" by analysis
  - Frequency, flow rate, where collected; flow gage data if available
  - How we use the data is dependent on available metadata
- Field vs "general" monitoring. Is there something in between?
  - Focused/Targeted monitoring is between field and general monitoring
  - Need information on application timing and loading
  - Monitoring can be used to evaluate model components (e.g., runoff or drift)
- Consider the temporal scale of monitoring

NAS – National Academy of Sciences. FEMVTF – FIFRA Exposure Model Validation Task Force.
FIFRA – Federal Insecticide, Fungicide, and Rodenticide Act

Charge Question 2 - Responses

- Metadata and purpose of monitoring program matters
- If we have the monitoring location, we can delineate the watershed to obtain soils, weather, and land cover data
- What scale of use information is needed?
  - Use information needs to be appropriate for the scale of the model (e.g., field vs watershed)
- Use of average application rate and/or application window across watershed can inform at larger scales
- Application information can then be used to evaluate how well the model captures pesticide concentrations from this load, and how the information can be applied to other watersheds
- Analytical information and level of detection need to be considered
- Chesapeake Bay workshop looked at multiple lines of evidence and what to do with them (http://www.chesapeake.org/stac/workshop.php?activity_id=222)

Breakout Group 2
Evaluation of Aquatic Exposure Modeling Estimates

- CHARGE QUESTION (3):
  - **What roles can the various types of monitoring data play in the evaluation of results from a watershed model (e.g., general monitoring doesn't predict maximum but has other roles)?**

Charge Question 3 - Responses

- Similar issues discussed in responses to Question #2
- Use monitoring studies to evaluate the components (e.g., flow) of a watershed model (i.e. SWAT, WARP, SAM)
- Use statistical analysis to predict maximum estimated exposure from general monitoring data
  - Related to size of watershed/basin
  - Scale matters in terms of what we need to know and how frequently we need to sample
  - Error bounds on predictions from monitoring data should be considered
  - Understanding of basin characteristics is needed where data are collected
- Space AND time are both important for endangered species (timing of exposure as well as where they reside)
- More confidence in use of monitoring where species occur vs monitoring elsewhere (part of Weight of Evidence approach)
- Purpose of the monitoring program (whether general or focused on a chemical/event) also plays a role in how it can be used (e.g., trend analysis)

SAM – Spatial Aquatic Model

Breakout Group 2
Evaluation of Aquatic Exposure Modeling Estimates

- CHARGE QUESTION (4):
  - **What other approaches are available for evaluating results from watershed models?**

Charge Question 4 - Responses

- Covered in other question responses (particularly in #1, non-monitoring options)
- A variety of papers are available (multi-model analysis provided in response to Question #2 above)
- Also talked about comparisons with other models (vs statistical models and other types of models)
- Qualitative methods (dispersivity)
- Loading is a reflection of use in the watershed

Breakout Group 2
Evaluation of Aquatic Exposure Modeling Estimates

- CHARGE QUESTION (5):
  - **To what extent can we rely on historical monitoring data when product labeling has changed and application-specific information is lacking?**

Charge Question 5 - Responses

- Uncertainty in the use of historical data is minor compared to other uncertainties (e.g., pesticide use and timing)
- Previous pesticide use could be modeled and compared to monitoring data
- Some mitigation (such as buffers) may result in substantial changes
- Need to be aware of the changes and impacts of those changes
- All things being equal, the newer data would be more reliable
- Change in analytical method/Level of Detection can also have an effect
- Could apply scaling factor if we know what the change (e.g., use) is
- Look at trend analysis (did we see change in concentrations?)
- Label changes and use changes (i.e., market share) must be understood
- “You must understand what you are doing”

Breakout Group 2
Evaluation of Aquatic Exposure Modeling Estimates

- CHARGE QUESTION (6):
  - **Are there new or different types of monitoring that could be employed to further our understanding of aquatic modeling estimates?**

Charge Question 6 - Responses

- Keep in mind that the focus is on the process for any pesticide and not just three organophosphate pesticides
- Bioaccumulation studies/fish tissue? [maybe not for current test chemicals]
- Sediment
- Passive samplers [in conjunction w/ daily/weekly samplers] – these still work for interpretation
  - NAWQA Stream quality assessment program daily vs passive analysis
- More robust use reporting for model inputs and model evaluation
- Monitoring targeted specifically for model validation
- Edge of field runoff (more diversity of chemicals)
- Continuous turbidity (has been useful for phosphorus) – surrogate models using continuous measurements (as a surrogate for pesticide concentration)
- Under-represented habitats (off-channel, low flow, etc)

Long Term Recommendations

- Species-specific habitat modeling
- Using EPA's Spatial Aquatic Model (SAM) for every NHDplus catchment
- Model multiple soils and weather combinations within the watershed

NHD – National Hydrography Dataset