Scientific and Technical Advisory Committee
Chesapeake Bay Watershed Model Phase 6 Review

Zach Easton[1], Don Scavia[2], Richard Alexander[3], Lawrence Band[4], Kathleen Boomer[5], Peter Kleinman[6], James Martin[7], Andrew Miller[8], James Pizzuto[9], Douglas Smith[6], Claire Welty[8]

[1]Virginia Tech, [2]University of Michigan, [3]USGS, [4]University of Virginia, [5]The Nature Conservancy, [6]USDA-ARS, [7] Mississippi State University, [8]University of Maryland Baltimore County, [9]University of Delaware

STAC Review Report

September 2017

STAC Publication 17-007

About the Scientific and Technical Advisory Committee

The Scientific and Technical Advisory Committee (STAC) provides scientific and technical guidance to the Chesapeake Bay Program (CBP) on measures to restore and protect the Chesapeake Bay. Since its creation in December 1984, STAC has worked to enhance scientific communication and outreach throughout the Chesapeake Bay Watershed and beyond. STAC provides scientific and technical advice in various ways, including (1) technical reports and papers, (2) discussion groups, (3) assistance in organizing merit reviews of CBP programs and projects, (4) technical workshops, and (5) interaction between STAC members and the CBP. Through professional and academic contacts and organizational networks of its members, STAC ensures close cooperation among and between the various research institutions and management agencies represented in the Watershed. For additional information about STAC, please visit the STAC website at www.chesapeake.org/stac.

Publication Date: September 1, 2017

Publication Number: 17-007

Suggested Citation:

Easton, Z., D. Scavia, R. Alexander, L. Band, K. Boomer, P. Kleinman, J. Martin, A. Miller, J. Pizzuto, D. Smith, C. Welty. 2017. Scientific and Technical Advisory Committee Review of the Phase 6 Chesapeake Bay Watershed Model. STAC Publication Number 17-007, Edgewater, MD. 47 pp.

STAC Administrative Support Provided by:

Chesapeake Research Consortium, Inc.
645 Contees Wharf Road
Edgewater, MD 21037
Telephone: 410-798-1283
Fax: 410-798-0816
http://www.chesapeake.org

Review Team:

Zachary Easton (Panel Co-Chair), Associate Professor, Dept. Biological Systems Engineering, Virginia Tech, Blacksburg VA. STAC Member.

Donald Scavia (Panel Co-Chair), Professor and Director, Graham Sustainability Institute, University of Michigan, Ann Arbor, MI

Richard Alexander, Research Hydrologist, USGS, Reston VA

Larry Band, Professor, Dept. of Environmental Sciences, University of Virginia, Charlottesville, VA. Formerly Director for Institute for the Environment, University of North Carolina at Chapel Hill, Chapel Hill, NC

Kathy Boomer, Watershed Scientist, The Nature Conservancy, Bethesda MD. STAC Member.

Peter Kleinman, Research Leader, USDA-ARS, Pasture Systems and Watershed Management Research Unit, University Park, PA. STAC Member.

James Martin, Professor, Dept. Civil and Environmental Engineering, Mississippi State University

Andrew Miller, Professor, Department of Geography & Environmental Systems, University of Maryland Baltimore County, Baltimore, MD. STAC Member.

Jim Pizzuto, Professor, Dept. of Geological Sciences, University of Delaware, Newark, DE

Doug Smith, Soil Scientist, USDA-ARS, Grasslands Soil and Water Research Unit, Temple, TX

Claire Welty, Professor, Dept. Chemical, Biochemical, and Environmental Engineering, University of Maryland Baltimore County, Baltimore, MD

Table of Contents

Executive Summary

This report documents the conclusions and recommendations reached by a peer review panel convened by the Chesapeake Bay Program's Scientific and Technical Advisory Committee (STAC) to review the Chesapeake Bay Program (CBP) partnership's Phase 6 version of the Chesapeake Bay Watershed Model (P6 WSM). The review was a "responsive review," undertaken at the request of the CBP's Modeling Workgroup and based on questions formed collaboratively by CBP and STAC. The review is intended to provide guidance for the CBP as they refine the P6 WSM for the Mid-Point Assessment in 2017. Overall, the review panel as a whole was favorably impressed with the integrated P6 WSM framework. Recommendations from the review panel focus largely on a longer-term suggestion for the CBP to more fully exploit the multiple model framework and incorporate estimates of uncertainty into the output. Other recommendations are for better justification and documentation of approaches taken.

Background and Review Process

Phase 6 is the most recent of a series of increasingly refined versions of the Chesapeake Bay Watershed Model (WSM) developed since 1982. Different versions of the model have been operational and serving to guide CBP management decisions for more than three decades. However, the P6 WSM is a major departure from previous versions which were largely based on a highly modified Hydrological Simulation Program -FORTRAN (HSPF) framework. While the deterministic HSPF framework is preserved in the P6 WSM for hydrologic and sediment simulations, the approach to water quality simulation is entirely new, relying on integration of multiple models for different biogeochemical processes in the watershed.

The review was conducted in two phases between June 2016 and August 2017. Responses relevant to all questions except nine and ten were finalized in Phase 1 of the review, which was completed in December 2016. Responses to questions nine and ten were delayed because the two issues involved (Conowingo Reservoir modeling and WSM-related aspects of climate change assessment) had not yet been finalized and documented by the CBP. The final documentation relevant to these questions (and a revised STAC-approved question ten) were provided in June 2017 and early July 2017, respectively. Panel responses relevant to questions nine and ten were completed in early August 2017 and were subsequently reviewed by STAC membership.

Summary of Major Recommendations

1. A more detailed and comprehensive description and rationale of model structure and linkages is needed.
2. The precise role that multiple models play and the structure that is used to accommodate multiple models needs to be clarified.
3. An accuracy or skill assessment of the underlying individual models used in the multiple model approach is warranted.

4. The panel encourages the CBP to transition from a multi-level model approach (e.g., several models providing a single point of input to the larger watershed model, which results in a single model realization) to a true ensemble model approach, which would allow for a Bayesian model analysis and a more thorough quantification of uncertainties.
5. Related to summary recommendation 4, above, uncertainty analyses should be developed for each model component; the panel believes this would be a natural extension of the ensemble model approach.
6. Use of expert panels for establishing BMP (best management practices) efficiencies should develop an explicit basis/approach to evaluating uncertainty in the estimates because this information would constitute priors for the Bayesian analysis.
7. The CBP should commit to a process for improving the model's capability to represent processes of particle transport, storage, and reworking in the Chesapeake Bay watershed, as the Revised Universal Soil Loss Equation 2 (RUSLE2) foundation is questionable at the river basin scale.
8. The CBP should encourage the development of sub-models that attempt to down-scale the watershed models while also exploring process-based mechanisms affecting water quality.

Introduction

This report documents the conclusions and recommendations reached by a peer review panel convened by the Chesapeake Bay Program's Scientific and Technical Advisory Committee (STAC) to review the Chesapeake Bay Program (CBP) Phase 6 Watershed Model (P6 WSM). .

Background

Phase 6 is the most recent of a series of increasingly refined versions of the Chesapeake Bay Watershed Model (WSM) developed since 1982. Different versions of the model have been operational for more than three decades and have served throughout this period to help guide management decisions by multiple CBP partners, including the US EPA. However, the P6 WSM is a major departure from previous versions which were largely based on a highly modified Hydrological Simulation Program-FORTRAN (HSPF) framework. While the deterministic HSPF framework is preserved in P6 WSM for hydrologic and sediment simulations, the approach to water quality simulation is entirely new, relying on multiple models.

Review Process

The CBP Office (CBPO), through the Modeling Workgroup, submitted a request for a responsive review of the P6 WSM, with particular emphasis on reviewing the new multiple model aspects of the watershed simulation and obtaining guidance as the CBPO modeling team continued to refine the P6 WSM for the Mid-Point Assessment in 2017. The final set of 12 review questions, determined collaboratively by the CBPO and STAC, are attached herewith in Appendix A.

The review was conducted in two phases. The first phase was conducted between August and December of 2016 and focused on all questions except questions nine and ten, with the understanding that final documentation for these questions was not yet available. The second phase of the review was conducted between June and August of 2017 and focused on questions nine and ten, as well as associated issues relevant to questions 11 and 12.

For Phase 1 of the review, an initial set of questions was posed by CBPO in August 2016 for approval by the STAC membership. After modest revision, the amended questions and associated documentation were received by STAC from the CBPO Modeling Workgroup on September 15, 2016. A review panel consisting of 10 recognized experts in the various components of the P6 WSM was convened and met for initial discussion via a conference call in September 2016. On September 28 2016, the panel met at the CBP office in Annapolis MD where Gary Shenk (USGS) and the CBPO model development team presented a P6 WSM overview and responded to panel questions, after which the panel met to discuss and respond to the questions posed by the CBP.

Phase 2 of the review focused on questions nine and ten, which relate to the P6 WSM handling of the Conowingo Reservoir within the Lower Susquehanna River Reservoir System (LSRRS) and P6 WSM approaches to the assessment of climate change effects. Final approaches and documentation were still under development by the CBP partnership at the time of Phase 1 review, therefore, Phase 2 review efforts were postponed until receipt of final documentation in June 2017. Notably, both the LSRRS and climate change assessment work were each also being subjected to separate independent reviews. In particular, an independent panel review of model enhancements for the LSRRS was coordinated by the Chesapeake Research Consortium (CRC) that was conducted between April 2016 and August 2017. The LSRRS expert panel review (Ball et al. 2017) provided independent review of modeling efforts undertaken by an Exelon-sponsored team in cooperation with the Maryland Department of Natural Resources and Maryland Department of the Environment. One of the reviewers on that effort (James Martin) was added to the WSM team to provide liaison to this prior review, which had helped inform the CBP in their P6 WSM development effort. Second, a separate STAC review of the CBP's broader Climate Change Assessment Framework (CCAF) is underway at the time of this report. Within the context of these other reviews, the CBP asked the P6 WSM review panel not to focus on the LSRRS modeling effort or the broader aspects of the CCAF, but to instead focus principally on the P6 WSM *application* of results from those and other previously peer-reviewed research efforts, as available in published literature.

Final documentation for questions nine and ten and a STAC-approved revision to question ten were provided by the CBP in June 2017 and early July 2017, respectively. Following several conference calls, this aspect of the review was completed in August 2017. Complete responses to all twelve review questions follow.

Panel Responses and Recommendations

The document is structured to respond to each question individually, thus, where there is overlap among questions, some information is presented more than once. For comments regarding the model documentation, specific comments can be found in responses to each question, and general recommendations are made in response to review question 12.

Question 1. Please comment on the overall appropriateness of the approach taken in the Phase 6 structure of a deterministic hydrology and sediment transport management model combined with a nutrient model informed by multiple models and multiple lines of evidence as described in Section 1. Please comment on the multiple model structure of the Phase 6 nutrient simulation particularly to its utility to watershed management in the Chesapeake restoration? How can the Phase 6 multiple model approach be improved going forward?

Overview

Overall, the review panel as a whole was favorably impressed with the integrated P6 WSM framework. The approach represents an exciting opportunity to leverage multiple modeling and ongoing field monitoring efforts to advance adaptive management in ways that should help guide CBP decision making, enhance understanding of watershed processes, and ultimately improve Bay water quality.

Future efforts should continue to focus on recommendations by the Phase 5 WSM review team (Band et al. 2008): to promote development of process-oriented, distributed modeling at the sub-basin scale. Importantly, given the limited resources available for research and model development, these science-based tools should explicitly address decision making needs while also providing a basis to define and explore alternative hypotheses of system dynamics.

Recommendations

(a) Provide a more comprehensive introduction to the modeling conceptual design and structure, with particular emphasis on describing the features of the new "data-driven" approach for the steady state model. This new approach reflects excellent progress in the evolution of the CBP modeling, and is a markedly different approach than the CBP has used before. Therefore, it is important that stakeholders and others understand clearly stated rationales and merits of the approach, with an emphasis on the more prominent role played by monitoring data. The CBP should view the opening section of the documentation as an opportunity to inform stakeholders and others about this evolution in the CBP modeling, consistent with the overall desire for the method to be more transparent and for model predictions to be more consistent with observations.

(b) Clarify the precise role that multiple models play and the structure that is used to accommodate multiple models. For most readers, the reference to multiple models is understood to mean multiple independent models are run and the model outcomes (predictions) combined in some weighted fashion. This is not the framework for the CBP model. Therefore, the rationale for the CBP approach, which only combines selected components from different models (e.g., several models providing a single point of input to the larger watershed model, which results in a single model realization), needs to be clarified. This approach appears to stem from the desire to use a single spatially explicit structure, which does readily accommodate the integration of multiple models, but this needs to be clearly explained. The advantages and the tradeoffs of this approach should also be made clear. For example, please discuss why the CBP chose to expand the model framework by using the SPARROW (Spatially Referenced Regression On Watersheds) model as a basis for estimating the P6 land-to-water delivery and aquatic transport components rather than relying solely upon the core HSPF watershed model.

(c) Cite the results of evaluations of the accuracy of the P6 WSM predictions, process components, and input models (e.g., Agricultural Policy Environmental eXtender [APEX]; Soil and Water Assessment Tool [SWAT]; Annual Phosphorus Loss Estimator [APLE]; and SPARROW) in the first section and present the details of these results in other sections, where appropriate. The shift to a more data-driven modeling approach elevates the need for a more comprehensive discussion of model performance and diagnostics, with some consideration of how the model uncertainties might affect decisions on load allocations. The assessment should especially examine the spatial bias and precision of the model.

I. At a minimum, a rigorous skill assessment is needed for the 66 "calibration" sites (where long-term records allow the use of the Watershed Regression on Time Discharge Season-WRTDS load method). However, an examination of model performance at an additional ~60 sites, which were used to calibrate the SPARROW model, would improve understanding of spatial variability in model bias and precision that could also inform development of more formal estimates of prediction uncertainties. These sites are likely to require the use of other load estimation methods (e.g., Ratio estimator, Loadest) because of their shorter records. A recent USGS comparative analysis of different load estimation methods (Lee et al. 2016) should be consulted to determine which methods might be best suited for use on shorter records.

II. To enhance understanding of possible causes of prediction errors, there's value in developing regression-based models of the model prediction errors, with explanatory factors related to sources/land use, transport properties, and physiography. For example, there is a particular need to investigate the possible causes of the over-predictions in nitrogen that were reported to occur at many of the 66 WRTDS sites.

III. In addition to completing a skill assessment, it would be informative to use Monte Carlo analysis to quantify prediction uncertainties related to the errors in model parameters and process components, and especially the uncertainties associated with BMPs (best management practices), which are likely to be one of the more highly uncertain features of the model. In the future, more sophisticated parameter estimation and error assessment (e.g., Bayesian analysis) should be used (see below).

(d) Consider implementing a formal optimization procedure for the next-generation (Phase 7, or P7) static watershed model in which the land-use export, land-to-water delivery, and aquatic transport components are simultaneously estimated. One concern with the P6 WSM procedure is that by performing an upstream sequential extraction of process effects on loads, based on using the downstream River Input Monitoring (RIM) loads as a constraint or boundary condition, the model doesn't provide a statistically optimal set of predictions for source generation, delivery to streams, and aquatic decay. The procedure treats the downstream monitored loads and the intervening process components (point source loads, aquatic decay), which are used to derive the upstream load constraints, as essentially error free. This does not explicitly isolate the model uncertainties, but allows the errors to be implicitly included as part of the land-use exports and transport components. A more formal optimization procedure, which would simultaneously estimate all of the components, would be desirable for the next-generation (P7) model. This would provide a more statistically rigorous method to allocate nutrient mass inputs and losses over space, while explicitly accounting for model errors. While such optimization may be cumbersome for the dynamic model, the static model is well suited for application of procedures for estimating parameter and state variable uncertainties that account for both input and model uncertainties (e.g., Bayesian estimation). This approach also allows the program to capture the wealth of information (and associated uncertainties) generated through the many stakeholder and expert workshops as formal priors. As such, in the future, the CBP should consider calibrating the static model and testing impacts with the dynamic model – the reverse of the current approach.

(e) Consider additional applications of alternative ensemble modeling approaches. While the integration of multiple models is a significant advancement over the Phase 5 approach, the review team felt more could be done to leverage the multiple model. In reality, the P6 WSM is not a true multi model (e.g., ensemble) approach but rather a multi-level integrated model, with several models providing a single point of input to the larger watershed model – resulting in a single model realization. Instead of averaging the multiple outputs from the input models (e.g., APEX, APLE, SPARROW) to provide a single input to the larger, watershed model, one could envision the multi-level model approach providing an ensemble of end member predictions. Thus, one recommendation would be to use each of the input models to provide a discrete parameter set to the watershed model, which would then be run and output predictions made; then the next set of input models would run and predictions made, etc. Together the models provide an average prediction, and differences among the predictions can be summarized as a range or probability distribution to provide an estimate of the uncertainty in the model average prediction. In addition, or alternatively, a similar ensemble model approach could be applied to sub-model components.

(f) Clarify the logic of the static modeling approach. While the review team appreciates the intent to simplify modeling efforts, we were uncomfortable with organizing and describing the static model as a separate, simpler model than the transient model used to drive the estuary circulation model. The results from the complex sub-models in fact represent critical components of the static model predictions. Presenting the management model as a summary of the TMDL (Total Maximum Daily Load) simulation seems a more tractable and transparent approach than hosting a management vs a TMDL model. Of further concern, the static model logic seems circular; first, observed loads at river outlets are decomposed to predict land use land cover (LULC)-specific loading rates, but then the model structure apparently is applied in reverse to predict river discharge. This *may* be justifiable if each component is viewed as a "composite" parameter of the overall framework, and these back-and-forth adjustments actually describe the calibration process. If this is a correct interpretation, it further highlights a need to reserve a subset of the RIM data exclusively for assessing model performance and to consider future versions of the static model that employ formal optimization procedures (see (d) above).

(g) Provide a more comprehensive assessment of the strengths and weaknesses of individual models and how these affect model performance and model prediction. For example, to what extent have input models like APEX and APLE been calibrated and validated for the Chesapeake Bay watershed? Have the impacts of alterations in SPARROW (e.g., eliminating some P-related functions (see below)) been tested by recalibration of the model to observations? Some of this would be accomplished by implementing the recommendation in (c), above, but additional descriptors of model processes and relationships would strengthen the overall section. Overall, the model descriptions are not clear (for the overall model structure as well as sub-model architectures), particularly the flow charts attempting to summarize relationships among key model components.

(h) Organize the documentation according to traditional modeling steps (i.e., conceptual model description; model implementation; model calibration; model validation, sensitivities, and uncertainties; and model application). This could enhance understanding and transparency significantly. Note that this is not a suggestion to dive deeper into the details of the underlying equations or model parameterizations, which we have detailed above; rather, it is a suggestion to enhance communications by organizing the text parallel to the modeling process.

(i) Commit to a process for improving the model's capability to represent processes of particle transport, storage, and reworking in the Chesapeake Bay watershed (perhaps for Phase 7). The current science upon which the P6 is built – as related to watershed-scale particle storage, residence times, and time scales for sediment delivery – is still evolving. Therefore, management decisions based on the P6 modeling results could be subject to future challenges as the research clarifies the dominant processes moving and transforming sediments. See some closely related comments in response to Questions 4 and 7 below.

Question 2. Please comment on the scientific rigor of the methods used for the average nutrient export rates described in Section 2. Are they calculated appropriately? Is there any additional scientific information that should be included?

Overview

P6 represents a substantial reorganization of the Chesapeake Bay WSM and includes significant updates at most steps of the process of assigning loads to individual land uses. The CBP has clearly responded to input from previous reviews, addressing a number of concerns that arise from evolving scientific understanding of watershed-scale processes, better approaches to simulating nutrient sources and transport, and the availability of new data and models to inform the model results. Although there are some major changes to the way nutrient loads are calculated at the finest scales, including the averaging of several model predictions to obtain so-called sensitivity values, much of the underlying approach to calculating loads builds upon the data and routines used in the P5.3.2 model. Significant updates in the P6 WSM include incorporating the SPARROW model into the calculation of riverine nutrient transport, the use of APEX/SWAT from USDA's Conservation Effects Assessment Program (CEAP) modeling to evaluate unit acre loading rates, and the use of APLE to better differentiate between legacy and applied P loads in developing sensitivity factors.

There are numerous issues of clarity in the P6 WSM documentation, and although many were rectified by presentations to the review panel, a fundamental rewriting is needed for clarity if this document is to serve as an accessible reference. The clarity of descriptions, especially within sub-sections, vary considerably (see prior comments suggesting a revised organizational structure.) Small issues, such as clarifying that "average loads" are actually "average land use loads" (or something along those lines) could help to avoid confusion. Similarly, "land simulation targets" has only contextual meaning and is otherwise jargon. In addition, there is a use of the term "fall line" to differentiate the tidal and non-tidal portions of the Coastal Plain that is not consistent with applications in the literature where it represents the Coastal Plain/Piedmont divide.

The general approach used in the P6 WSM is to calibrate hybrid models to USGS RIM data for large basins above the tidal zone, then to back out the contributions of various processes and sources to nutrient loads along river systems until individual land uses are assigned an average nutrient export rate at small stream outlets. This approach balances the availability of data, availability of appropriate fate and transport routines, and availability of resources (including computational). Given all of these considerations, as well as the need for simplicity, transparency, and consistency, the methods used to calculate nutrient export rates are, for the most part, defensible. Yet there were questions about model circularity as noted in response to Question 1. Again, better and more linear documentation may help to clarify these concerns.

While there is uncertainty at every stage of the modeling process (from RIM station to small stream outlet), some of the larger concerns with past and current approaches involve the assignment of nutrient export rates to land use categories at the finest scale of inference followed by the derivation of so-called "sensitivity" factors to assess the effect of land management. Here, the watershed model is being asked to represent the interactive effects of biogeophysical processes and management actions on nutrient export at approximately the county level. The model relies upon increasingly precise and accurate information on land use at this scale, as well as nutrient reduction efficiencies from expert panels to predict management contributions to nutrient export rates. These panels generally focus upon field and landscape scale studies that are at a much finer scale than the smallest scale of inference used in the watershed model and are asked to derive efficiency factors for single practices that are contributing, in combination with many other factors, to nutrient export.

Nutrient loads are first assigned to sectors by averaging estimates from three models (the P5.3.2 WSM, SPARROW and CEAP's APEX/SWAT). Examples of watershed model averaging (P5.3.2 WSM, CEAP, SPARROW) show considerable deviation in the sector loads estimated by each model. No evaluation of the appropriateness of each model for this application is provided (e.g., to what extent have APEX/SWAT been calibrated and validated for the range of conditions in the Bay watershed?). Rather, they are all given equal weight. In the example provided, SPARROW substantially underestimates nutrient loads from hay relative to the other models. A critical evaluation of the causes of these differences is needed before simply accepting the models as appropriate and weighting them equally.

Following the assignment of nutrient loads to different land use sectors, nutrient loads are further divided into specific land use sub-categories using ratios that quantify deviation from a standard (median) loading rate for that sector. These ratios are provided by expert panels and are largely literature-derived. They are reviewed by the CBP's modeling workgroup for consistency. Until a more distributed, process-driven approach can be employed by the CBP's WSM, this coarse approach seems to balance issues of data availability and computational parsimony and is satisfactory, if not ideal. However, more could be done to capture and propagate the uncertainties imbedded within the assessments by expert panels.

A notable omission, which is a carryover from previous phases of the WSM, is the inability to simulate loads from land in the tidal region. This leaves significant swaths of the Coastal Plain unrepresented in the modeling, including areas that have the greatest hydrologic connectivity to the Bay and are most vulnerable to sea level rise. The model currently assumes that nutrient loads from these regions are consistent with those above the tidal zone.

It seems that RUSLE2 plays a prominent role in estimating P loss from different land management categories. This model, while widely applied, has significant limitations with regard to the prediction of sediment loads (and associated phosphorus) from land uses with low erosion rates (e.g., pasture, no-till). Some versions of RUSLE2 (Foster 2013) have been found to overestimate soil erosion, especially from pastures. This overestimation of sediment was due to low biomass estimates in RUSLE2 crop management routines (Dabney and Yoder 2012). Further concerns regarding the RUSLE2 application are discussed below.

Recommendations

(a) The multiple model approach is new to the P6 WSM and therefore warrants the greatest scrutiny and reflection. At a minimum, the variability in model estimates should be used as a measure of uncertainty in output.

(b) An evaluation of model skill is recommended using the RIM station data. Currently, assessments of skill are based upon loads. It is recommended that model skill is evaluated for estimates of watershed discharge and for estimates of nutrient and sediment concentrations.

Specific Comments Regarding Model Documentation

(a) Consider the number of significant digits to report more carefully. The panel felt that it is not appropriate to report the amount of nutrients lost from an acre of land to the nearest 1/100 of a pound. At least round to the nearest pound if not the nearest 10 pounds.

(b) Figure 2-1: Indicate the outcomes of the model process (i.e., what is the arrow pointing to?)

(c) page 2-3, Section 2.2.1/2.2.1.1: The description of the four adjustments to RIM loads is difficult to follow. Suggested alternative description: Loads are allocated to LULC with consideration of BMP practice effect(s) after adjusting observed loads to account for 1) in-stream and river losses; and 2) additions from sources contributing directly to in-stream loads (e.g., point sources, atmospheric depositions, etc.).

(d) page 2-4, Section 2.2.1.2: Why are septic systems not considered a contributor to Non-Point Source (NPS) pollution?

(e) page 2-5, Figure 2-4: Why focus on comparison of regional factors for forests when other sectors have a much stronger influence on predicted discharges? Additional explanation of how factors were derived for would be helpful.

(f) Page 2-7, last paragraph: Basis for excluding CEAP export rates for developed lands requires further explanation. Its “general assumptions” suggest a simpler model, which might actually be better.

(g) Tables 2-4, 2-5: Basis for deriving average sector export rates are not clear.

(h) Table 2-6: For each sector, highlight the reference/unit LULC class

(i) Equation 2-4: 1) Where does the 1.2M inches come from? 2) How is the runoff estimated? 3) Is there a conversion factor missing (mg to pounds)?

(j) Table 2-9: Why differentiate among palustrine wetland types, rather than floodplain vs headwater wetlands or other hydrogeomorphic classes? Similar to forest, consider differentiating disturbed vs undisturbed wetlands.

(k) Table 2-10: In general, a summary of underlying data rather than references to reports or accountings of who did the research would be preferable. For example, in Table 2-10, in addition to the model parameters, indicate the number of studies (including what proportion were in the Chesapeake watershed) and the range of reported values.

(l) Terminology concerns:

   i. Confusing terms: i) “average load” (rather than loading rate or yield); ii) “target load” rather than sector or sub-sector loading rate (because a target most often refers to an objective)

   ii. Jargon: global variable

   iii. Needs additional explanation: edge-of-?small?-stream, “true” forest, nitrogen species (p 2-20)

Question 3. In Section 4, how justified are the sensitivities of nutrient export from land uses to nutrient inputs, given the approach used and data available? Do the sensitivities to nutrient inputs derived from multiple models reflect our best understanding of the current condition of nutrient load processing and attenuation on the landscape? Is there any additional scientific information that should be included?

Overview

For the most part, the sensitivities to nutrient inputs derived from multiple models seem to be in line with what we think would happen. There are a couple of instances (as noted below) that may or may not make sense.

In regard to the sensitivities of nutrient export, the approach of using multiple models to develop the averages is an improvement from using a single model. There seems to be confusion in whether the section discusses absolute or relative sensitivity, which would clearly make a difference when it comes to interpretation. While the method is improved, relying solely on model output without any documentation of calibration and validation of the sub-models raises concerns that they may lead to erroneous conclusions.

Inclusion of "soft data" (e.g., edge of field) for the purpose of verifying the model output would be an improvement. Using this approach, you will not be able to ground-truth every point of model output; however, by ground-truthing what you can with available data sources a future direction for the modeling can be projected. If the soft data is in general agreement with the group-model output, then one could argue that the models are representative where data are available, so (in theory) should also be trustworthy. If the data sources are outside the bounds of the sensitivity projected by the group-modeling output, then it is possible that the observed data may shed light on areas in all of the models that need improvement. As this is being done, the uncertainty around the observed data should also be considered, as the measurement techniques to collect such data can have great influence on the quality of the data. While this has been done somewhat with the inclusion of work from the CEAP project it would be helpful to have data from a specific land use in a specific county (or soils that also happen to be in the Chesapeake Bay watershed) that can be used to verify that the model output are fitting observational data.

Recommendations

(a) Incorporating "soft data" verification would help lend some validity to the reported results, meaning observed data should be used to ensure the sensitivities are at least in the ballpark of what is reported. Using three or four models is preferable to using only one, but there needs to be some ground-truthing of as much of the modeling as possible.

(b) A strong recommendation of assessing the modeled output to local/regional observed data at plot to watershed scale will help make the case that the models are working or need some improvement in specific areas.

Specific Comments Regarding Model Documentation

(a) There are some inconsistencies throughout the Section 4 draft. For instance, the caption for the equations on P. 4-2 notes "definition of sensitivity and relative sensitivity". The equations are actually for absolute sensitivity and relative sensitivity. Throughout much of the document, it is difficult to know if it is the relative sensitivity (Sr) or the absolute sensitivity (Sa) that is being referred to.

(b) Page 4-3: "In Phase 5, a linear sensitivity was assigned to PQUAL and IQUAL such that reducing all inputs to zero would result in half the calibrated nutrient load." What is the logic behind this? Is this ½ the load to CB or ½ load from land use?

(c) Page 4-6: last line before section 4.2.4 cites Figure 4-4… there is no figure 4-4.

(d) Page 4-8: "Variability in the APEX values are from scenarios and model versions. Variability in the SPARROW output is from model versions found in the literature. Variability in the Phase 5.3.2 output is from the different land uses." So what is being compared here is not exactly 'apples to apples'.

(e) Page 4-9: at the bottom of the page, the units given are pounds export per pounds import… so this is absolute sensitivity?

(f) Tables 4-2 to 4-4: are these relative or absolute sensitivities? These seem to ignore absolute sensitivities. The does not seem to make sense – atmospheric deposition is more sensitive in an [ag]riculture landscape for the sum of N species than on an impervious surface. This signal is likely overwhelmed in agricultural landscapes.

(g) Figure 4-8: The caption mentioned the plot is for "tree canopy over scrub shrub land use." The panel thought that this land use was deleted.

(h) Figure 4-11: what does the x-axis represent on the top graph? What kind of sensitivity?

(i) Page 4-15: is fertilizer really the dominant nutrient source on conventional till with manure land uses?

Question 4. Please comment on the scientific rigor of the methods used in the use of Spatially Referenced Regression On Watersheds (SPARROW) for land to water factors in <u>Section 7</u>. Are they reasonably implemented? Is there any additional scientific information that should be included?

Overview

The land use steady state SPARROW model described here provides a useful modeling structure to inform estimation of the source and transport (terrestrial and aquatic) components of the newly developed steady state P6 WSM. The spatially explicit properties of SPARROW are generally well-suited for the P6 WSM approach, which like SPARROW separates nutrient source generation from terrestrial and aquatic process effects on transport. The development of land-use based sources in the SPARROW model, rather than the mass-based sources that are commonly used, is reasonable to ensure consistency with the overall P6 WSM concepts that emphasize the use of land-use export coefficients from multiple watershed models. The SPARROW land-to-water delivery factors in the P6 WSM should provide an acceptable representation of the major factors that explain the long-term average transport and storage (and loss) of nutrients along surficial and sub-surface flow paths across many Chesapeake Bay watersheds (although see comments in 'b' below). The area-weighted averaging of the SPARROW land-to-water factors within the catchments of the ~2300 river segments in the P6 WSM will greatly reduce the spatial variability in these factors, but provides a reasonable method to link the two different spatial scales present in these models.

The central role that SPARROW now plays in quantifying terrestrial and aquatic transport in the steady state P6 WSM, based on a modified version of a recent USGS model (Ator et al. 2011), underscores the need for the documentation to more clearly describe the details of the methods and performance of the new SPARROW model. Evaluations of the model structure and performance should include the following. First, owing to the change in the structure of the source component (shifting to land use from mass inputs), the possible influence of additional land-water delivery factors should be evaluated as part of the model development phase (the extent to which other factors were examined is unclear, including whether other functions, such as continuous in-stream decay, were tried). Second, information about model performance (e.g., Root Mean Square Error-RMSE, yield, r^2, etc.) and residual diagnostics should be reported and discussed. Of particular importance is a careful review and reporting of the spatial biases in the predictions, which are available for approximately 180 calibration monitoring sites. Spatial biases that are observed at this stage will be important to consider in relation to those observed for the full model, which is currently evaluated for only about 60 Chesapeake Bay monitoring sites with the longest records.

For the phosphorus delivery factor, several statistically significant variables (erodibility, Coastal Plain, and precipitation) were eliminated from the SPARROW P model to avoid potential redundancies with the APLE model, which was reported to include similar properties. However, because of the multiplicative interaction between the sources and land-to-water transport factors in SPARROW, the elimination of important variables without recalibration of the model would be expected to introduce biases to the model predictions, and is an unconventional approach. Thus, it would be informative to evaluate the extent of correlation in the explanatory variables for these factors in the two models (SPARROW and APLE) to determine the extent to which the eliminated factors are accounted for by the other model terms.

During the discussions, concerns were raised about the applicability of the Index of Connectivity (IC) to estimation of sediment delivery from Bay watersheds in view of the limited prior testing of this predictive measure by Cavalli et al. (2013). Questions also centered on whether the spatial variability in IC makes physical sense. The P6 WSM documentation should explain the physical basis for the IC metric. Furthermore, an effort should be made to justify why the metric, originally developed in high relief, small drainage basins of Switzerland, is actually appropriate for use in the Chesapeake Bay watershed. Data should be cited that test the effectiveness of the IC in the Chesapeake Bay watershed, if any exists (we assume that no real evaluations are available). The observations spatial patterns of the IC, for example, might be correlated with various watershed properties in a reasonable way. Lacking this information, however, the use of the IC seems rather *ad hoc*, and poorly justified scientifically (i.e., untested by observations in the watershed).

Recommendations

<u>Short term Recommendations:</u>

(a) Improved evaluations are needed of the SPARROW model performance and diagnostics.

(b) Document specific values of the sediment delivery ratios used in the model, and explain to the reader what these values imply conceptually for sediment movement in the watershed. Document geographic variation in sediment delivery ratios and justify patterns in terms of watershed characteristics.

<u>Long-term Recommendations (See also Question 11.):</u>

(a) Additional evaluations are needed of the sediment components, such as IC. Data should be obtained justifying the use of this metric, which appears to have been developed for conditions very different from the Chesapeake Bay. Its use presently is an extrapolation without support of local data and results are likely to have high uncertainty. Further validation is needed.

(b) The conceptual basis for the entire sediment modeling approach requires further investigation in preparation for the next phase of the watershed model. The basic idea in this model is that sediment is generated from uplands, some of it is stored on the landscape between this upland source and its delivery to small streams, and additional source/sink terms (bank erosion and floodplain deposition) are included for small streams, but not for larger streams. The sediment delivery ratio approach reflects this conceptual framework, but the evidence that the watershed really works this way does not have a strong empirical foundation (i.e., the data supporting it is not extensive). Other sources and sinks should be considered; while some scientists believe upland sources are not important, others consider them very important. Rills and gullies often represent incision of the upland landscape and headward extension of the drainage network, either ephemeral or more permanent, and these may be important sources as we discuss in response to Question 7. If sediment is important to model, then the scientific foundation for doing so really needs to be improved.

Specific Comments Regarding Model Documentation

(a) Given that the introduction to SPARROW in Section 7 (Equation 7-1) applies to both Sections 2, 7, and 9, this material might be more appropriate to locate in the introduction (Section 1), where a more comprehensive treatment could be given to the overall model concepts. The material also might be located within a section that provides an introduction and background for the three modeling approaches that are used for P6.

(b) Table 7-1: The text should indicate that the transport processes for land-to-water delivery are inclusive of selected groundwater effects for nitrogen. This table also includes an overview of transport processes that operate at all scales within the Chesapeake Bay. It is therefore more important to the documentation than simply to provide an explanation for the land-to-water factors. It should be introduced early in the documentation, and its conceptual basis should be explained and justified.

(c) Equation 7-4 should be assigned to the area-weighted equation on p. 7-7, which is missing an equation number.

(d) Fig. 7-1 of the river and stream segments needs more explanation because there was some initial confusion about the spatial domain for the P6 river catchments; displaying these watershed-river segment boundaries may be helpful. In particular, the figure capture should contain specific details describing what the figure illustrates. For example, the yellow shaded area should be identified and explained (the brief acronyms in the legend are not really comprehensible). The more detailed "brownish" catchment boundaries should also be explained – why are these included? They are not defined in the legend.

(e) The documentation should include a chapter on uncertainty and risk analysis. The greatest uncertainty for sediment is that the conceptual basis for the model is not well supported by observational data, and this creates a significant risk associated with using the model for management decisions.

(f) In Section 7.3.1.2 (of the provided model documentation) on 'feeding space', losses of 30% for N and 90% for P are assumed for the nutrient transport to streams rather than using the SPARROW land-to-water delivery factors. Therefore, notation should be added to Table 7-4 to indicate that land-to-water delivery interactions with the pasture land-use source were not allowed in the specification (it's also worth checking that this specification was used by USGS in the updated SPARROW calibrations).

(g) In regard to Section 7.4: This section notes that Sediment Delivery Ratios are a common concept in sediment modeling. While this is likely true, it does not necessarily provide much confidence in this aspect of the modeling approach. The available data regarding how sediment moves from upland landscapes to hillslopes to small streams in the Chesapeake Bay watershed is very limited. The conceptual basis for the sediment delivery approach is not well verified in the Chesapeake Bay region. The documentation provided is very sparse, and in fact almost nothing is said about what specific values of sediment delivery ratios are used or how they vary geographically. This part of the modeling approach is poorly documented.

Question 5. Please comment on the overall appropriateness of the methods used in the application of multiple methods to estimate stream-to-river factors for nutrients in Section 9? Is there additional scientific information that should be included?

Overview

In the P6 WSM, the finer scale of National Hydrography Dataset Plus (NHDplus) catchments is used to generate loads to the edge of small streams, compared to the much coarser Land-River-Segmentation (LRS) used in the P5 WSM. The in-stream losses in these smaller streams prior to loading in the river segments are computed using different methods derived from the NHDplus SPARROW models for nitrogen and phosphorus. Methods to estimate sediment loss or additions at the small stream level are currently being developed by the USGS and by the Center for Watershed Protection but are not yet incorporated into the P6 WSM.

The nitrogen and phosphorus small stream (NHDplus stream) decay are modified from an approach taken by Hoos and McMahon (2009) that was applied to the land-to-water (L2W) delivery, differentiating among the nutrient sources. This approach computes a sector mean delivery, but then differentiates rates at the catchment level based on Delivery Variance Factors (DVF) to increase or decrease decay rates based on local conditions. Similarly, in the P6 WSM, Stream-to-River (S2R) factors, developed in the NHDplus SPARROW model were used, and calculated by general sector category (crop, pasture, developed and natural land). Documentation needs to be improved as the equations on p. 9-2, of the supplied model documentation, do not appear to include sector specific S2R (the S2R on the right hand side of equation 9-2 are not indexed by sector), but figures 9-3 through 9-10 show sector specific S2R values for each Land-River Segment. It is not clear why the sector source of nutrients in the stream are differentiated with S2R, as it is presumed the NHDplus reaches are well mixed, and there is no reason to expect nitrogen or phosphorus from different sectors would be processed differently, unless it is dependent on speciation from different sources, or by the location of edge-of-small stream delivery within the LRS. In this latter case, it may be that a sector nutrient delivery may be in a specific upstream or downstream location in the LRS, and so subject to a greater period of in-stream decay. This should be more clearly described. Note that the NHDplus-scale SPARROW model developed by Ator et al (2011) does not provide sector specific decay rates in the aquatic phase; rather, the first-order rate constants vary according to stream size (i.e., mean discharge).

The method developed by the Center for Watershed Protection, limited to urban catchments, is based on estimating Stream Source Ratios (SSR) to differentiate between upland and in channel sources in urban catchments. The SSR is based on the development of a dataset of upland derived and total loads within a watershed for which total sediment delivery can be estimated using a regression model that predicts SSR as a function of available watershed characteristics. This information can then be used to apportion total load to upland and in-stream sources. A small data set of nine urban watersheds was used for which gauge data of flow levels were available.

Upland sediment load estimates are based on an EMC (Event Mean Concentration) estimated from measurements for storm sewer outfalls available from the National Stormwater Quality Database for the county of the LRS. Hourly flows were taken from P5.3.2 simulations. In this case, EMCs were taken as constants for a county, based on nearest sampled watershed. The errors of these estimates are not described in the documentation, providing no estimates of uncertainty. The method assumes no deposition, so that integrated loads are not balanced or reduced by loss rates in the small urban streams.

For total loads, EMCs on an event basis are co-located with gauged flow measurements. A load-discharge power function relationship for each gauge is then integrated over hourly flow. It is important to note that while the estimates of the upland loads use a constant EMC, the total loads are estimated with an EMC-discharge relationship, which may result in some discharge weighted bias of the SSR. The final regression model predicting SSR is a function of the Hydrologic Soil Group C/D soil and impervious surface proportions. As the estimates of SSR from regression can be below zero or greater than one for some feasible combinations of C/D soil and impervious proportions, SSR's are arbitrarily bounded between 0.05 and 0.95.

Recommendations

(a) This method is strongly limited by available data, as well as potential inconsistencies and high uncertainty in the estimate of upland and total loads. The results of an uncertainty analysis should be reported based on sources of error for the loads and the SSR, as well as an analysis of spatial patterns. If it is possible to test the model with independent observations of in stream derived loads, as an example from the data sets of Noe et al (2015**a,b**), that could provide diagnostics for refining and building more confidence in the methods.

Question 6. Please comment on the scientific appropriateness of the approach taken for Phase 6 lag times described in Section 10 given the current state of information and understanding of groundwater and particulate processes. How can the structure and processes of nutrient lag time simulation on the land be improved in Phase 6 or future watershed model applications? Is the application of the Ranked Storage Selection (rSAS) function for groundwater nitrate and Unit Nutrient Export Curves (UNEC) for all other nutrient species appropriate for the management questions?

Overview

Reviewers felt that it was not reasonable or efficient to address the issue of lag time for dissolved species (where delays occur primarily in groundwater) at the same time as addressing lag times for particulate materials (where delays are predominantly on land surfaces, on banks, and in beds), since the processes and pathways are completely different. In this regard, issues of particulate processes are discussed above under question 4 and below under question 7. Hence, the remainder of reviewer response to this question relates to lag-time modeling for dissolved species.

The CBP modelers have presented new methods they are using in the P6 WSM for calculating nutrient loads based on transit times. These approaches are based on recently published literature and appear to be a promising improvement over approaches used in Phase 5 for load calculations. Given that a great deal of thought and discussion have gone into the new calculations and the approach appears to have merit, the panel recommends that this trajectory should continue to be pursued in the P6 WSM with caveats discussed below. If, as additional new approaches come to light, the CBP can consider adopting them.

Question 6 addresses whether the application of the Ranked Storage Selection (rSAS) function for groundwater nitrate is appropriate for addressing management questions, but no definition of rSAS or reference to the literature is provided in Section 10 of the provided model documentation. Section 10.1.2.2.2 states that "UNEC and rSAS use the water and sediment fluxes along with the nutrient inputs to calculate nutrient budgets for the land uses. Parameters for transit time distribution of nutrients are provided as inputs for both UNEC and rSAS based on estimates of lag times that include models and other lines of evidence". This brief description of *how* this routine is used does not provide any information of what calculations rSAS *actually does*. In Section 10.1.2.1, the reader is referred to Section 5.2 for a definition of rSAS; Section 10.5.2 provides the following definition of rSAS: "Ranked Storage Selection (rSAS) is an approach to simulate transient response of the system. The model was used to simulate response for the groundwater nitrate. Operational details were worked out over the Beta 2 development phase. A full application will be made available for review with the Beta 3 phase." There is no further mention of rSAS in Section 10 and thus makes review of this algorithm impossible based on the written material provided.

However, to complement the written documentation, the description of the basis of rSAS was presented in a briefing to the panel, and although not explicitly mentioned on the slides, the panel was told this concept is based on the work of Harman and others. Harman's work has been published in *Water Resources Research* (Harman 2015) and is generally viewed by the scientific community to be a robust, novel, and elegant approach to evaluating watershed transit times. The CBP modelers have gone to great lengths to try to parameterize the rSAS method with a number of published data sets and results from other groundwater models. This is meritorious; however, the approach needs to be better documented to stand up to public scrutiny. Furthermore, it is not clear that the rSAS approach has been validated in conditions similar to the Chesapeake Bay watershed, as the original publications evaluate data collected from watershed in Ireland.

Whereas the rSAS approach is used for simulating nitrate transit times, a new "Unit Nutrient Export Curves" (UNEC) approach is being used to simulate transport paths for other species (other nitrogen species; phosphorus species). For UNEC, some description of the conceptual and computational basis of UNEC is provided in Section 10 of the P6 WSM documentation. The model is based on previously published approaches used in EPA Q-TRACER. The idea behind the UNEC is that a unit pulse of nutrient application is defined to have an associated empirical concentration response function (breakthrough concentration curve) that characterizes travel time distribution relevant for the location on the landscape to which the nutrient pulse is applied. To obtain a load time series, a time series of unit (pulse) nutrient applications is specified, a composite concentration response function is constructed by adding together the time series of unit concentration response functions, and a final load function is created by multiplying the composite concentration response function by simulated flow.

Details on how UNEC model parameters are chosen are not provided. However, evidence of model performance is provided by comparing aggregated monthly model output with USGS WRTDS-calculated loads. The comparisons show a significant improvement over comparisons of the same data with P5.3.2 WSM output. For total nitrogen, the modeled load output is in agreement with about 87% of USGS WRTDS data using the P6 UNEC approach compared to a 67% agreement for P5.3.2 WSM. In addition, the spread about the median value in the box-plot comparison is greatly reduced for the P6 WSM compared to P5.3.2 WSM. Whereas improved agreement of model output with observational data is documented, it is difficult to prove that this is due to the new method for calculating loads. However, because other components of the model have not changed much since P5.3.2 WSM, it is reasonable to assume that the improvement agreement in model output with data results from implementation of UNEC.

Although there could be criticisms in the use of UNEC owing to the empirical nature of the parameterization, the logical basis for the approach is sound and appears to be an improved path forward to calculating loads using the modeling structure. The panel therefore feels that the approach is reasonable.

An overarching concern of the CBP WSM especially relevant to understanding lag times is its intent to generalize hydrologic processes across the entire Bay watershed. Clearly, the fate and transport of excess nutrients and sediment strongly depends on the physiographic province and the hydrologic connectivity between a ‘contaminant’ source and downstream waterbody. Numerous reports provide general descriptions of how the geology in each province uniquely influences surface- and ground-water interactions. While model segmentation and application may capture related effects explicitly, the model structure abandons a key opportunity to explore how this set of factors influence land management effects. The lag time model component provides a prime opportunity to capture this key set of factors.

Sanford and Pope (2013) note two other important considerations:

- Due to the significant influence of hydrologic connectivity on ground- and surface-water exchange, stream network length and distribution likely influence groundwater transport lag times. This may be especially critical to consider when evaluating effects of seasonal and climate change in addition to drainage/land use management (de Wit and Stankiewicz 2006).
- Extremely flat areas, such as the outer Coastal Plain lowlands, have relatively long (>300 years) residence time. This contradicts assumptions/predicted data presented in Figure 10-27.

Recommendations

Rather than solely using the lag time models to adjust load estimates, additional post-processing of results also could address the following key questions of concern: 1) What is the status of 'legacy' nitrate in our groundwater systems? Where does long-term groundwater storage and discharge possibly outweigh impacts from current land management practices and thus limit the response of down-gradient ecosystems? 2) Does our understanding of lag-times suggest where or which areas of the landscape might be more critical to manage for water quality concerns? For example, are there highly leachable areas (i.e., with no lag-times) that perhaps also incur excessive fertilizer applications (because so much is lost)? These questions provide exciting examples of how data from the CBP WSM can be used beyond evaluating TMDL obtainment, to support management decisions directly.

Question 7. Please comment on the scientific rigor of the methods used in the Phase 6 sediment simulation components using a detailed Revised Universal Soil Loss Equation 2 (RUSLE2) (Section 2), an interconnectivity metric (Section 7), and the inclusion of sediment source/sink estimates from stream banks and flood plains (Section 9).

Overview

The conceptual framework for the sediment modeling assumes that sediment is primarily produced from erosion of upland soils (hence the use of RUSLE2), and routed to small streams with some storage accounted for by the Interconnectivity (IC) Metric and sediment delivery ratios. Along small streams, some sediment is stored in floodplains and additional sediment is generated by bank erosion. Sediment that reaches large streams is transported downstream to the Chesapeake Bay. There is no accounting for the distribution of lag-times, which ideally should be evaluated in relation to sediment texture.

Our comments first address the strengths and weaknesses of each of the three components (RUSLE2, the Interconnectivity Metric, and the streambank/floodplain source/sink estimates) followed by more general comments and recommendations.

RUSLE2: Although it is not entirely clear from the documentation provided, it seems more appropriate to describe the P6 WSM hillslope erosion model as a method to disaggregate average annual erosion RUSLE estimates to more of an event-based time-scale rather than using RUSLE estimates as calibration targets. The latter implies adjusting model parameters to known values, which absolutely is not true. At best, RUSLE estimates can be considered as *qualitative* indicators of where soil erosion may impose greatest impact on in-stream suspended sediment loads (Wu et al. 2005).

Numerous publications highlight concerns regarding watershed-scaled, USLE-based raster applications (e.g., Boomer et al. 2004), providing strong evidence that such estimates should not be used as model calibration data. For one, the mismatch between the field design upon which the USLE is based and the raster-based implementation raises significant concerns with model implementation. More importantly, there is increasing evidence to suggest that hillslope and rill erosion often play a relatively minor role in regulating sediment transport at the watershed-scale, compared to impacts from gully and scour erosion, as well as instream deposition (De Vente et al. 2013) and sediment remobilization. Indeed, pinpointing these small-scale point sources and sinks (i.e., precision conservation) may provide important means to managing downstream water quality (e.g., Tomer et al. 2013; see response to question 8). Accordingly, the reliance on RUSLE to generate calibration data presents major concerns, both from a mechanical/implementation and conceptual basis.

If RUSLE estimates essentially are model endpoints (starting points?) that were disaggregated to a finer timescale, rather than calibration endpoints, this alleviates some concern about the RUSLE usage in the current sediment simulation framework, especially given the lack of tractable alternatives. This also could explain the near perfect (and highly suspect) correlation presented in Figure 10-26. In defense of RUSLE, the model remains conceptually sound at the small field scale, and it still is considered best available sediment modeling technology to identify areas with relatively greater risk for erosion. Especially notable, for example, is the current scarcity of quality data and associated scientific approach for properly representing gully erosion and scour. The qualitative evaluation and forcibly incomplete representation of other sediment transport processes in the CBP P6 WSM, do however, raise major concerns about the accuracy and precision of sediment simulation. The data and science are continuing to evolve in these regards, and the CBP should commit to updating the P7 WSM as new data become available.

In addition to the overarching concern of how RUSLE is being used within the model CBP model structure, the review panel shared concerns regarding the limited description of RUSLE parameters. For example, adjustment factors listed in Table 2-19 report categories that were not evaluated in the original, empirical research (e.g., roads); such assumptions require documentation outlining which and how information was combined for use within the RUSLE framework.

IC Metric: The Interconnectivity (IC) Metric raises many questions. First, it is based on a method that was developed for use in small mountain watersheds in the Alps, which have complex histories of glacial erosion and very different surficial materials and topographic patterns and spatial scales by comparison with locations in the Chesapeake Bay watershed. These watersheds as illustrated in the original paper drain across alluvial fans at their downstream ends. Whether results derived from these watersheds are transferable to a very different landscape in the Chesapeake Bay watershed is, at a minimum, open to question. The logic of the paper seems reasonable and no claim is made against or for the method itself. Although there may be a project currently underway to assess its broader applicability, the review panel is concerned about justifying results derived for locations throughout the Chesapeake Bay watershed on a method whose applicability has not been demonstrated.

From what the panel can see, the direct application of the method generates values that are well outside the anticipated range, which required adjustment and rescaling. The panel was somewhat alarmed by presentation to the panel stating: "Need to convert to scale of 0 to 1 with an average of 0.25; distribution looks reasonable, so assume linear translation." With a large enough data set, it is not particularly surprising that the distribution of IC scores appears normally distributed; beyond that, all of the adjustments are based on other considerations and it is not clear how the Cavalli et al. (2013) method really informs what was done. Essentially the data have been massaged enough to generate sediment delivery ratios that are helpful in matching values recorded by the WRTDS stations, but that does not tell us whether the method is applicable here. If it is indeed applied in the P6 WSM, we would recommend attaching an asterisk to its use indicating that the validity of the approach needs further study and that other approaches will be examined in planning for P7.

Stream banks and floodplains: The USGS floodplain network regression models are anticipated to be applied to account for exchange between streams and floodplains. One of the two documents cited for Noe et al. (2015**b**) is a seminar presentation that is available online and is clearly work in progress with an ambitious agenda to generate regression results for application across different physiographic provinces and across the entire Chesapeake Bay watershed. The availability of field measurements at a network of sites is helpful; this is supplemented with data extracted from light detection and ranging (LiDAR) to characterize channel and floodplain geomorphology. We have not assessed the validity of the tools, the output or the conclusions being drawn and therefore we are not in a position to make any strong recommendations about their use at this point other than to say that the availability of empirical data represents a significant advance and these data should be incorporated into the modeling framework. If sufficient information is collected to develop robust regression models, this may be a useful method of estimating sediment loss or gain in small streams as fine scale LiDAR and spectral imagery is developed to represent channel geomorphology and land cover conditions for the independent regression variables. It is important that the results of this analysis be qualified as a minimum rate of deposition and erosion, as the initial depth of roots below the soil surface is likely not known. Clearly there is exchange between the channel and the floodplain, the floodplain has the ability to serve as an important storage reservoir, and the related processes will influence fluxes and time lags at the watershed scale.

The regression approach used to predict floodplain exchange has some strengths, but also has two significant limitations. One is related to the variables used for the regression model. These are geomorphic variables and watershed characteristics; they may be useful for predicting floodplain processes as they currently operate, but are insufficient to assess future changes. How might changing flow frequency or sediment supply affect floodplain exchange? Relevant variables are not included. The second limitation is the lack of temporal component to the approach. Floodplain storage encompasses long timescales that need to be explicitly considered in watershed scale sediment routing schemes. We hope that these components can be incorporated into the modeling framework, if not for the P6 WSM then potentially for P7.

Summary: The panel had serious scientific questions regarding the conceptual framework used in the P6 WSM sediment modeling. First, observational evidence linking upland soils to downstream sediment delivery is weak; some studies suggest that other sources (e.g., gully and stream bank erosion) may be more important than upland soil erosion. These processes are not directly included in the modeling effort. Second, sediment storage (particularly on floodplains) imposes long timescales (centuries!) on sediment delivery processes that are not accounted for in regression-based estimates of floodplain exchange. Long storage timescales might be neglected when processes are approximately "steady", but the entire point of a management model is to predict and evaluate changing conditions. Furthermore, the long distribution of timescales for sediment delivery, encompassing days to centuries, is not consistent with the use of steady state framework for modeling management decisions in the Chesapeake Bay watershed. The timescales required to reach a steady-state sediment delivery are much longer than any timescale envisioned for management decisions, so the steady-state condition isn't really relevant. What the model needs to address is the extent of sediment delivery following management actions within a reasonable time frame, recognizing that steady-state conditions may not be achieved.

Recommendations

(a) In the short-term (P6), little can be accomplished towards improving the sediment modeling approach – too many changes are needed, especially given the amount of time available. Over the longer term (P7), new model structures should be created that account for the variety of potential sediment sources in the watershed and the wide distribution of timescales for sediment delivery. A coordinated modeling and field research program will be needed to support such an effort. Current scientific understanding is not sufficient to accurately quantify the relevant processes, for example, to make predictions of lag times and delivery rates for sediments at the watershed scale with a reasonable degree of confidence. Therefore the P6 modeling approach should be regarded as an interim solution with the expectation that improved scientific understanding will allow a more comprehensive approach in P7.

    Furthermore, although there is a strong incentive to rely on fall line gaging station measurements of sediment flux as calibration targets, matching calibration targets at the mouth of the watershed for existing conditions does not guarantee that management activities at specific locations upstream will be successful in achieving their stated goals. New approaches should seek to capture the most important sediment sources at distributed locations throughout the watershed and to incorporate new research results on timescales and intensity of sediment exchange processes.

Question 8. Given the fine scale 1m x 1m land use data that's used in Phase 6, what opportunities does this open to the CBP and scientific community in the next phase of watershed model development? What are the advantages in a distributed representation of hydrology, land cover, and sediment? Given the availability of nutrient inputs from Agricultural Censes at the county scale only does a higher resolution of the watershed model make sense?

Overview

With one exception, the panel does not recommend using high resolution products to advance the CBP-HSPF component but rather to embrace the utility of these data in developing sub-models that can inform our understanding of system processes and provide critical information to stakeholders, in particular to drive management decisions. Especially given the costs in terms of processing time, it is essential to report that high resolution land use and topography data, finer than 10 m resolution, appear to provide limited benefits to improving discharge predictions generated by regional watershed models (Zhang and Montgomery 1994, De Vente and Poesen 2005, Bormann et al. 2009, Yang et al. 2014). While the digital elevation model (DEM) grid size and vertical accuracy influence hydrologic modeling performance, model calibrations likely compensate for this effect due to interactions between model parameters and spatial factors (Wu et al. 2008). The current relatively coarse resolution of soils, bedrock, and surficial geology maps likely also limit the utility of detailed data in a watershed modeling context. High resolution weather data presents an exception; specifically, precipitation data finer than 1 km grid, hourly measures has potential to improve watershed model predictions significantly (Bormann et al. 2009).

Detailed topography and land use/land cover data, however, provide great promise to advancing process-based, distributed models at a spatial-scale relevant to identifying where land use/land management practices impose the greatest impact (and present the greatest opportunity) to downstream water quality (e.g., as a downscaling model or as a sub-model to the overarching CBP model framework). While limited access to comparable high-resolution information on agricultural practices will remain a constraint, application of high resolution data to mapping terrain characteristics more accurately has shown value, including watershed boundaries and stream networks (Yang et al. 2014). Additional applications show that detailed data can be used successfully to map surface processes including field erosion (Verachtert et al. 2010), gully formation (Momm et al. 2011), channel head migration (James et al. 2007; Tarolli and Fontana 2009), in- and near-stream incision and deposition (Thoma et al. 2005, Notebaert et al. 2009, Walter and Merritts 2008), and wetland function (Murphy et al. 2007, 2009; Rayburg et al. 2009, Richardson et al. 2010). Further, spatially explicit predictions of biogeochemical processes present an exciting opportunity to link field research and modeling efforts, especially if designed to evaluate/investigate alternative models of predominant system drivers (e.g., verifying variation in field-scaled predictions of evapotranspiration, which also could provide critical information to improving regional watershed models).

Recommendations

(a) Identification of field-scaled opportunities to install practices which will provide the greatest water quality (and habitat) benefits at the least cost remains one of the most frequently-cited information needs among state and federal outreach agents, county planners, and restoration managers. High resolution land-use data combined with other more detailed information, such as LiDAR-derived topography data, present exciting opportunities to address this information gap. Rather than using these data as input for the HSPF-based framework, the CBP should encourage the development of sub-models that attempt to down-scale the watershed models while also exploring process-based mechanisms affecting downstream habitat conditions.

Question 9. Better simulation of the deposition and scour processes in the reservoir reach of the Lower Susquehanna is an important feature of the Phase 6 Model. It is crucial to 2017 Midpoint Assessment decision making to be able to represent the net deposition of sediment, nitrogen, and phosphorus in this reach of the Susquehanna as fully as possible. Does the Phase 6 representation of the dynamics of the reservoir system rely on the best science available at this time? Do the simulations approximately represent the observed changes in storage of sediment, nitrogen and phosphorus as seen in the historical record from the last few decades? How can the representation of Conowingo infill be improved going forward beyond the Phase 6 Model?

Overview

The CBP modelers have presented revised methods being used in the P6 WSM for calculating particulate organic and sediment loads from the Lower Susquehanna River Reservoir System (LSRRS), including Lake Clarke (Safe Harbor Dam), Lake Aldred (Holtwood Dam) and the Conowingo Pool (Conowingo Dam). The Susquehanna contributes 41 percent of the nitrogen, 25 percent of the phosphorus, and 27 percent of the suspended sediment to the tidal Bay (Linker et al., 2016b). However, the infilling has altered the solids retention capacity and nutrient trapping efficiency of the reservoirs over time, which has management implications and necessitated revision of the WSM.

The CBP modeling revisions were, in part, based on recommendations of the STAC Conowingo workshop (Linker et al., 2016a). The workshop resulted in nine issues (five recommendations and four questions), which were broadly summarized in the August 18, 2016 STAC letter to the CBP Management Board as:

1. Efforts to model the effects of Conowingo on net accumulation or release of nutrients and sediment from the reservoir should be evaluated based on its ability to “hindcast” data from water quality observations and statistical analyses.
2. Address biogeochemical processes related to sediment scour and nutrient cycling that may influence bioavailability in reservoir sediments, under variable flow ranges in the Conowingo Reservoir.

3. Ensure representation of effects of Conowingo inputs to Chesapeake Bay for the full range of flow conditions including ‘extreme’ high-flow events.
4. Improve representation of reactivity of particulate organic matter in Conowingo outflow.
5. Moving forward, an effort should be made to link the sediment transport and biogeochemical models in the 2010 Water Quality and Sediment Transport Model (WQSTM) to enhance modeling of the transport and fate of organic nutrients in the tidal Bay.

In addition, the CBP modelers addressed questions from the CBP Modeling Workgroup (February and April 2017) identified as necessary to be resolved in order to calibrate the P6 WSM and address management questions:

1. What is the current state of the Conowingo and the two upper reservoirs with regard to long- term mass balance?
2. What information can be used to estimate the change in scour and deposition over time for the purposes of calibration?
3. Does the trapping efficiency change with different levels of nutrient inputs?
4. How does the availability of organics change with respect to flow?

General Review Comments

The P6 WSM of the LSRRS as described is supported by observations and directly informed by complementary modeling studies including the application of WRTDS and modeling of LSRRS sediment and nutrient processes as part of the "LSRRS Model Enhancements" project by Exelon. The Exelon project included the application of an unsteady-flow HEC-RAS model developed by WEST Consultants and Gomez and Sullivan (GSE) Development of a HEC-RAS model and applied from Marietta, PA to Holtwood Dam. The LSRRS enhancement project also included the development and application of a coupled hydrodynamic sediment transport model (ECOM-SED) to Conowingo Pond and development of the Conowingo Pond Mass Balance Model (CPMBM) by HDR and GSE. The CPMBM also included a sediment flux model which is basically the same model incorporated into the CE-QUAL-ICM model on which the estuarine model WQSTM is based. However, for the CPMBM, the model was modified by HDR to allow simulation of multiple sediment layers and the impact of scour and deposition on the sediment bed and fluxes of particulate organic matter (POM) and dissolved materials to the water column. The POM (POC, PON, POP) is subdivided into separate G-classes representing their reactivity. These supporting studies and observations can be considered to represent the best science presently available.

In Section 10.7.2 of the model documentation provided, the CBP modelers addressed the three CBP Modeling workshop questions.

(a) With regard to question 1 above, the CBP modelers indicated that the weight of evidence suggests that the three reservoirs in the LSRRS are currently in dynamic equilibrium. That conclusion seems to be well supported.

(b) With regard to question 2 above, the CBP modelers, based on recommendations from the CBP Modeling Workgroup, determined that the P6 WSM can be calibrated to WRTDS loads, since WRTDS matches the observed change in the reservoir behavior over time. The CBP modelers calibrated scour and deposition parameters and their change over time based on WRTDS which formed the basis of the methodology discussed in 10.7.3. In using the WRTDS as the basis for scour and deposition parameters the CBP modelers also adequately addressed the issue identified in the above question 2.

(c) With regard to question 3 above, the CBP modelers indicated they are retaining the assumption of constant delivery ratios in the P6 WSM. The assumption is that the trapping efficiency for nutrients does not change in response to changes in nutrient inputs. That assumption also seems well supported by multiple lines of evidence. As indicated, the Exelon-supported model results for nutrients (application of the CPMBM by HDR, Exelon April 2017) supported the idea of a linear behavior in response to decreases or increases in nutrients delivered to Chesapeake Bay from the watershed and Conowingo Pond.

(d) With regard to question 4 above, the CBP modelers address the reactivity of the organics as a function of flow based on results from the Exelon supported Conowingo Pond Mass Balance Model (CPMBM) developed by HDR and GSE. In the Exelon supported study, HDR applied both a stand-alone version of the sediment flux model and the CPMBM which included linked hydrodynamic, water quality and sediment flux models. The CPMBM sediment flux model was modified by HDR to include multiple layers (previous versions were a static 2-layer model) to allow the incorporation of the impacts of scour and deposition. The model could then be used to estimate the reactivity of POM in the sediment bed and resuspended POM. The sediment flux model subdivides particulate organic matter (POM; POP, PON and POC) into G classes, including G1 (labile; half-life of weeks to months), G2 (refractory; half-life on the order of a year), and G3 (inert components). The CBP modelers used relationships developed by the CPMBM to specify the G-class distribution of TP and TN as a function of flow (Table 10-25). The approach taken by the CBP modelers seems reasonable and based on the best information available. The approach taken also addresses STAC questions 2a and 2c. The multi-layer sediment flux model developed by HDR could also be considered for implementation in a future version estuarine model WQSTM in response to the STAC recommendation 3.

In addressing question 2, above, from the CBP Modeling Workgroup, STAC recommendation 2 was also addressed. That is the results from WRTDS were incorporated to ensure representation of the effects of Conowingo inputs to Chesapeake Bay for the full range of flow conditions including ‘extreme’ high-flow events. In addressing question 4, above, from the CBP Modeling Workgroup recommendation 2 the STAC recommendations have also been addressed by incorporating the results from the Exelon supported studies that included simulation of the biogeochemical cycling of POM and the reactivity of the POM exported from the Conowingo Pond. The STAC recommendations were also addressed in more detail and more specifically in 10.7.4. Section 10.7.4 also addressed question 3, above for the next generation of the tidal water quality and sediment transport model. As stated above, the sediment flux model modifications implemented by HDR in the application of the CPMBM can also be considered as a basis for future modifications to the WQSTM model as in STAC recommendation 3.

STAC recommendation 1, above was addressed in 10.7.3. That recommendation was that efforts to model the effects of Conowingo on net accumulation or release of nutrients and sediment from the reservoir should be evaluated based on its ability to “hind cast” data from water quality observations and statistical analyses. The CBP modelers described a 4-step process whereby the WSM was calibrated to WRTDS. For hindcasting, the model parameters representing the 1990s condition were gradually varied to the parameters identified in Steps representing the 2010s condition. The CBP modelers also developed long-term trapping efficiencies for sand silt and clay delivery from the results of the Exelon supported HEC-RAS study of the LSRRS including Lakes Clarke and Lake Aldred. The approach for the LSRRS, including Conowingo Pond, seems reasonable given that the WSM cannot simulate changes in deposition and scour with the change in bathymetry or infill, so that those processes must be parameterized.

The discussion provided in 10.7 suggests that the P6 WSM representation of the Conowingo Pond is based on the best science presently available. The P6 WSM revisions are based largely upon the WRTDS application and the LSRRS models supported by Exelon, including the HEC-RAS application and the Conowingo Pond hydrodynamic, mass balanced and sediment flux models. However, there are few data to support those model applications. In particular, there are few data available within Conowingo Pond and few data available for high flow events within or below the Conowingo Pond. Therefore, any conclusions regarding whether the simulations approximately represent the observed changes in storage of sediment, nitrogen and phosphorus as seen in the historical record from the last few decades (question 9B) are limited by the data in those records, emphasizing the need for additional field data in future studies to support modeling efforts.

Recommendations

(a) Section 10.7 needs stronger organization and a more detailed discussion of the model components, and how these components ultimately tie together to generate WSM model predictions. The CBP modelers responded to STAC reviewer questions regarding Section 10.7 in a memo on 7/13/2017. That response clarified a number of the questions regarding the application and the response should be considered in part or whole for incorporation into Section 10.7.

(b) The final charge question (9C) asked how the representation of Conowingo infill can be improved going forward beyond the P6 WSM. The P6 WSM revisions were largely informed and supported by a powerful set of models of the LSRRS, including WRTDS, HEC-RAS, the sediment-flux model (stand-alone), the 3-dimensional ECOM-SED hydrodynamic and sediment transport model, and the RCA model which was the basis for the CPMBM. A recommended approach would be to mine these models to identify data deficiencies, developed improved data plans and continue to support and develop these models in conjunction with the WSM in an iterative fashion to develop an improved understanding of processes in the LSRRS and how they impact management questions and to support adaptive management.

Specific Comments Regarding Model Documentation

(a) It would be helpful if all of the P6 WSM inputs and outputs to and from Conowingo Pond be tabulated and their source identified. Presently 10.7 discussions are limited to sediments, PON and POP.

(b) In 10.7.2 the CBP modelers described their approach for total suspended solids. However, for the results from HEC-RAS the CBP modelers (7/13/2017 memo) indicated that sand, silt and sand fractions were considered. Discussion should be provided as to how each of those fractions were considered in the simulation of suspended sediments in the Conowingo Pond and its exports.

(c) In the discussion in section 10.7.2 it is not clear how the dynamic steady state assumed for sediments and nutrients will be incorporated into P6 WSM simulations.

(d) The responses to the questions also focused on suspended solids and particulate nitrogen and phosphorus. POC was not discussed nor were dissolved constituents. Per the first comment, all of the constituents simulated to and through the Conowingo Pond should be discussed.

(e) In the 20170712 - Phase 6 WSM Review Responses the CBP modelers indicated that Table 10-25 was derived from figures in a presentation by Qian Zhang et al. on 02/14/2017 (Slides III-7 to III-12). However, the Conowingo Pond mass Balance Model on which the figures were based was not submitted for peer review until April 2017. It is recommended that the finalized results from the CPMBM be reviewed to ensure that the equations in 10-25 are still applicable and the interpretation appropriate.

(f) In 10.7.2.2 there was discussion of concentrations and loads but no explicit indication of concentrations and loads of what. Was the analysis conducted for solids, nitrogen and phosphorus (only SS illustrated in Figure 10-69; note that panels a and b appear to be reversed in that figure)?

(g) Expand the discussion on the 4 steps of the iterative calibration process in 10.7.3 per the review questions and response in the 7/14 memo.

(h) In Table 1 there are columns for the Conowingo Infill Scenarios and for the Climate Change Scenarios. Since these are shown as percent change, we wanted to be sure we understood what the percentages in the Climate Change columns represent – are these percentages of the values achieved AFTER differences associated with the infill scenarios are accounted for?

(i) Figures 5 and 6 illustrate comparisons of predictions to observations for the Susquehanna at Conowingo for suspended solids and for phosphorus. The panels showing the comparison between simulated and observed values and the comparisons of the cumulative distributions are interesting and raise a question. For suspended sediment and for phosphorus the fitted linear trend between simulated and observed departs from the dotted 1:1 line and to a much greater extent for phosphorus than for suspended sediment. This presumably accounts also for the difference in shape of the cumulative distributions for both parameters. However, the comparison of monthly loads for both parameters after the 4-stage model calibration and the Nash-Sutcliffe efficiencies for both annual and monthly loads appear to show a closer fit than one might guess after looking at figures 5 and 6. It would be helpful to have some explanation associated with these figures to clear this up.

Question 10. Please comment on the scientific appropriateness of the methods used in the representation of climate change in watershed nutrient and sediment loads estimated for the 2025 and 2050 time periods.

Overview

The P6 WSM incorporates an assessment of the influence climate change has on Chesapeake Bay water quality, which is necessary as a contribution to the Midpoint Assessment. CBP decision makers need this information to determine if and when climate change impacts should be incorporated into the jurisdictions' Watershed Implementation Plans (WIPs).

The treatment of climate change scenarios in the P6 WSM is a response to recommendations from STAC and others that the Bay Program integrate consideration of climate change into the management framework to embed climate change among partnership goals in decision making, identify and prioritize vulnerabilities of restoration efforts and management actions, and use partners' ongoing research efforts to better assess and evaluate responses to changing climatic conditions.

The modeling team focused primarily on projections of precipitation volume and intensity, temperature, and evapotranspiration as inputs to the watershed model. Other projections, such as future sea level, were incorporated into the Water Quality and Sediment Transport Model (WQSTM) of the tidal Bay, and were not under review by this panel.

Decisions in regard to which scenarios to use were based on recommendations provided by the 2016 STAC workshop on Climate Change in the Chesapeake (Johnson et al. 2016). Projections were generated for the 2025 and 2050 target dates. The Coupled Model Intercomparison Phase 5

(CMIP5) set of Global Climate Models as described in the IPCC Fifth Report (IPCC 2013) were employed in the assessment of temperature trends for both 2025 and 2050 and projections were made for alternative sets of Representative Concentration Pathways (RCP) scenarios.

As is common for climate models, the expected changes in temperature are in much better agreement among the various models than is the case for precipitation projections. For this reason the model projections were used for temperature for both 2025 and 2050. However, for 2025, the 2016 STAC Climate Change Workshop recommended that precipitation scenarios should be based on extension of long-term observed trends in intensity, duration, and frequency of precipitation events, and the modeling team followed this recommendation. For 2050 it was recommended, and the modeling team followed the recommendation, that the P6 WSM utilize the CMIP5 projections of precipitation change.

The delta method was employed using downscaled GCM results, as described in the documentation provided to the review team. Median estimates from the suite of models were chosen as the expected change for the 50th percentile. Estimates were also developed for the 10th and 90th percentiles for both 2025 and 2050.

The P6 WSM makes a significant change in the approach used to estimate potential evapotranspiration (PET), relying on the guidance provided by P.C. Milly in the STAC Climate Change Workshop. The Hargreaves-Samani approach was found to produce a more realistic response to temperatures than the Hamon approach that was used in the P5.3.2 WSM.

Members of the review team assigned to this question examined the documentation and provided questions to the modeling team about several issues, followed by a conference call to follow-up on some of the answers provided in written form. The panel is in agreement with the choices made and recognize that they are consistent with the recommendations of the 2016 STAC Climate Change Workshop, and we agree further that the methodology used consists of the best available science and are appropriate for use in developing scenarios of nutrient and sediment loads for the 2025 and 2050 time periods. However, we have several recommendations outlined below.

Recommendations

(a) **Bias Corrections and Downscaling:** More information is needed about the bias corrections and the results of the hindcast, as well as showing some monthly or annual time series of the output, if available. The documentation should also clarify which data and models were actually used. For example, for the 2050 model estimates, it would be useful to explain more clearly the link from the CCIP5 ensemble projections and BCSD downscaling to the delta-method estimates of precipitation. It should be fairly straightforward but a figure or two and some text would clarify where these estimates came from. Using bias corrections in the projections based on biases between model and historical observation assumes the future bias is the same as the historical bias. Some discussion about why this is not problematic would be useful.

(b) **Precipitation:** The reviewers understand that the STAC Climate Change Workshop recommended using precipitation projections for 2025 based on extending the historical record, and that projections for 2050 should be based on the ensemble model projections. As the 2025 precipitation estimates are based on linear trends developed from the 87-year PRISM records at the county level, it would be helpful to clearly demonstrate how the well the regressions fit the historical data. It would be useful to see one or two examples of the actual trend analysis and projections.

It was shown in recent presentations that the projections from the historical record are not unlike the mean results from the ensemble, *albeit* with the latter having wider error envelopes. Because 2025 is only 8 years out from the present, reviewers are in agreement that it is reasonable to use these projections. However, given this concurrence in the means, it would be useful to carry forward the watershed model results driven from both approaches. This could be especially important in comparing the relative uncertainties.

It appears the transition in the future from snow to rain had significant impacts on evapotranspiration and the time frame of delivery of water and nutrients. This should be explored in more detail.

(c) **Flow and Nutrient Flux:** The projected increases in Flow, N, P, and sediment derived from the projected increases in precipitation clearly are dependent in large part on the way that the precipitation increases are parsed across the intensity deciles and this is in turn defines the runoff response and loads. The summary table 11.2.3 is not enough by itself to explain the nature of how these results are derived for the climate scenarios. Although the description of the change in runoff modeling approach (particularly with regard to evapotranspiration) is clear, it is not clear how the quantitative predictions of increased flow, N, P and sediment are derived from the climate model scenarios. More specifically, how are the increased loads related to projected changes parsed by intensity deciles, and how are the N and P increases parsed in terms of dissolved vs particulate forms and what is the basis for this? Since there is a discussion of an alternative modeling scenario based on assuming uniform increase across all deciles, perhaps there is also some indication of the sensitivity of these results to the distribution of precipitation intensity increase across deciles. It would be helpful to know more about this given its importance to the midpoint assessment.

The discussion of nitrogen sensitivity to flow was somewhat surprising in the level of uncertainty implied. The ratio of percent N change to percent flow change was determined from the model to be 0.7. The USEPA 2013 study provided very different results and much larger ratios for the Susquehanna watershed, and even though that was an outlier, the projections for other watersheds were still mostly larger than 1 and averaged 1.5. The choice of 1.0 to be used here seems like something of a stopgap choice. Given the importance of the ratio chosen, the final statement leaves one wondering what the path forward is on this issue: "Given the wide variability in outcomes a ratio of 1 is selected for initial study with additional input being sought." It would help to have some clarification on what additional input might be of use.

Specific Comments Regarding Model Documentation

(a) In general, the documentation could be more explicit about how exactly it is drawing from sources and showing the basis for projections rather than just citing sources and including a few figures, particularly as they relate to the bias correction and model hindcasting skill.

(b) There are a couple of discrepancies between the text and figures 11.2.2.2 and 11.2.2.4 - in both case the increase for 2050 does not match the value cited in the text

(c) Incorporate a more rigorous analysis of the uncertainty surrounding the use of climate projections, and how those uncertainties propagate through the watershed model to impact management decisions.

(d) Provide additional justification and explanation for the selection of nitrogen to flow ratio, how sensitive are the results to this ratio?

Question 11. For longer term CBP considerations, how can the overall approaches and procedures used in Phase 6 be improved and what alternative approaches and data gathering might you recommend?

Overview

In an effort for continual improvement in the CBP WSM the panel recommendations center around several overarching themes: a) further exploiting the multi model approach to develop a true ensemble model; b) more formalized optimization techniques; c) evaluation of model uncertainty (e.g., via Bayesian techniques); d) development of higher spatial resolution models to inform management; e) further refining the consensus based approach to the BMP expert panels; and f) developing improved modeling strategies for key processes that are not adequately quantifiable based on available scientific knowledge (e.g., identifying and quantifying sediment sources, estimating sediment lag times, etc.). Additional detail and recommendations for each of these points are given below.

Recommendations

(a) Moving forward, the P6 WSM could benefit from several approaches, in particular the development of a true ensemble modeling framework, which is easily accomplished (for the land segments), given the existing multilevel modeling approach. In this approach, rather than using averages from multiple model outputs as input to the single, overall model, each input model (or combinations of input models) are used as input to the overall model. The overall model thus produces several results that can be summarized as averages with error characteristics. This ensemble approach allows one to sample the uncertainties in both the initial conditions and model formulation through the variation of input data, analysis, and methodologies of the ensemble members. As such, this approach will be less likely to result in systematic errors and exhibit less variation than would be expected in single-model prediction systems, ultimately allowing the CBP to develop metric of uncertainty, and perhaps better target land segments acting as critical source areas. The multiple solutions also provide options to select more or less conservative management targets. Weighted averaging and Bayesian methods can improve multi-model ensemble integration.

(b) As mentioned for question 1, there's value in considering the use of a model structure that could accommodate the formal use of optimization techniques in which the source generation, land-to-water delivery, and aquatic transport are simultaneously estimated. This would provide a more statistically rigorous mass balance method than the current approach, and would allow for an explicit accounting of model error.

(c) This type of optimized model structure is also well-suited for the use of Bayesian methods, including their application with hierarchical (nested) model structures, especially for the linear static model. Bayesian methods have several advantages. First, the methods allow an explicit accounting of the uncertainties in stream monitoring load estimates, BMP efficiency estimates, and other model components. As noted in the group discussions, one option is to treat the BMP efficiencies, which are derived from expert panel assessments, as prior information in a Bayesian structure, thereby providing a more precise accounting and evaluation of the BMP uncertainties in the model. Second, a hierarchical Bayesian structure would permit one or more of the model parameters to be treated as random variables that vary spatially. This would allow model processes and predictions to be more sensitive to sub-regional and local variations in water-quality conditions, which in the current P6 WSM may contribute to prediction biases. Bayesian methods are currently being used and refined for SPARROW, which could potentially serve as a guide for their use in future Chesapeake Bay steady state models.

(d) Some additional thought might be given as to whether a higher resolution steady state model should be developed. An important question is whether the current 2,200 segmentation (large-river) stream network for the P6 WSM is sufficiently detailed or should be refined to provide more spatially resolved information on sources and land-water transport, which could inform within-state allocations. In the group discussions, it was acknowledged that there would be value in providing small-scale information to help inform local needs to target conservation and manage inputs, yet it was recognized that the uncertainties in model predictions generally increase with reduced spatial scale. Model accuracy is limited by monitoring that occurs more commonly in large rivers and by county-scale data for certain model inputs. However, accuracy is also potentially enhanced by high resolution land use and climate data that are currently used in all of the watershed models. The P6 river segmentation causes a loss of resolution and spatial variability in the modeled process interactions between sources and transport factors that are currently obtained from the NHD SPARROW model, which operates with ~80,000 reaches. Stakeholders may benefit from having access to predictions from a CB model with a spatial resolution that falls between that of the SPARROW NHD and the P6 WSM segmentation.

(e) With regard to assigning BMP efficiencies, concerns from the previous model persist: "Removal efficiencies of BMPs are known to be dependent on climate, flow rates, hydrogeologic setting, and implementation and maintenance conditions. Within the External Transfer Module (ETM) framework, these efficiencies are currently fixed at constant values. However, they could either be sampled from a distribution function (with form and bounds set from the literature, ideally tied to the hydrogeologic setting or conditioned on flow rates (if appropriate)). This would allow "breakthrough" of sediment and nutrients for a subset of the population of BMPs, which could have important downstream impacts."

(f) There was concern about the consensus-based approach for establishing BMP efficiencies through expert panels without an explicit basis/approach to evaluating uncertainty. Expert panels should be encouraged to incorporate or develop understanding of uncertainty/risk associated with estimated efficiencies. For example, the range of opinions about BMP efficiencies that are reflected in expert panel discussions should be preserved to support uncertainty analyses; unfortunately, this information is lost in the current approach. As mentioned above, a Bayesian estimation framework would enable use of this information in establishing priors and associated uncertainties. There was also concern about the limited evaluation and discussion of uncertainty, in general, and its implications for both management and research.

(g) The BMP expert panels should recognize that that mean retention efficiencies derived from the literature represent a model of expected outcomes. The BMP expert panels represent a model of expected outcomes and should be encouraged to refine these models in a way that describes BMP performance in relation to location and climate/seasonal weather/event condition. This focus on process-oriented, local-scale models may be where we can encourage development of relatively simple models to represent competing hypotheses of system dynamics and best leverage the advantages/opportunities presented through Bayesian modeling.

(h) The CBP should take greater advantage of intermediate modeling products to better understand seasonal dynamics, also to better understand storm-based loads, critical to understanding BMP performance.

(i) As noted in the answer to question 7, many changes should be considered in the future to improve the approach for sediment modeling.

Question 12. Please comment on the Phase 6 documentation. Is it clear, well organized, concise, and complete (taking into account that it is the third Beta out of an expected four Beta versions and about six months ahead of final release)?

Overview

The review team was generally impressed with the documentation, particularly compared to the Phase 5 documentation. While the vast majority of recommendations related to the documentation are encompassed in the specific comments for each section, there are some additional recommendations and comments detailed below.

Recommendations/Comments

(a) Document needs stronger organization that will facilitate a more well-rounded and complete discussion of the model components, and further, how these components ultimately tie together to generate P6 WSM model predictions. Specifically:

I. The opening chapter should be divided into two, with one outlining the decision contexts, the questions the P6 WSM model is designed to address, and bureaucracy associated with model development and use, and the other providing a general overview of the model conceptualization and structure. This second part needs to provide a more comprehensive coverage of method and approach, describing the data-driven methodology, which represents a marked evolution in the CBP approach.

II. The opening chapter also needs to present a clearer conceptual diagram, with all of the source/transport components. There's also value in considering the presentation of a supporting model equation to identify how the various components are linked together and processed.

III. Consider implementing a parallel organization structure for subsequent sections. For example, adopt the 'traditional' modeling framework to describe individual model components: i) system conceptualization (ideally including a "cartoon" or flowchart figure); ii) model selection and 'code' description; iii) model design (e.g., spatial scale, boundary conditions, input data, etc.); iv) calibration; v) sensitivity analysis/uncertainty assessment; vi) verification; and vii) predictions (ideally along with estimates of uncertainties).

IV. Given that the introduction to SPARROW in Section 7 (Equation 7-1) applies to Sections 2, 7, and 9, this material might be more appropriate to locate in the introduction (Section 1), where a more comprehensive treatment could be given to the overall model concepts. The material also might be located within a section that provides an introduction and background for the three modeling approaches that are used for P6.

(b) Additional discussion is needed as to how the steady state model is used to inform the operation of the transient model. For example, the text should clarify whether the biogeochemical process rates in the HSPF transient model are active or whether components of the steady state model (e.g., land-to-water delivery) are used as surrogates for these processes. Additionally, in the case of aquatic decay in streams and reservoirs, it would be helpful to clarify the sensitivity of the two models to different processes. The transient HSPF model simulates time-varying nutrient processes, associated with algal uptake and denitrification, which are then adjusted to be generally consistent with long-term averages of in-stream decay estimated by SPARROW. The SPARROW long-term average decay rates are associated with long-term storage or permanent removal processes, and thus should be acknowledged to differ from those in the transient model.

(c) More discussion is also needed in regard to the transient components for estimating time-lags in the watersheds, based on the two models: UNEC (with exponential decay imposed) and rSAS (gamma distribution to pull from different groundwater layers that reflects application timing as well). It would be informative to include some of the details that were given in the group presentations and discussions, such as how time series of nutrients exports are determined for hourly (or monthly) edge of small stream (EOSS) export using the basic inputs (fertilizer, etc.), and how this is derived to ensure that the sum is equivalent to the steady state mean.

(d) More and clearer discussion is also needed to describe the approach to sediment lag times. During the review panel/CBP modelers group meeting, some specific methods were described to account for some lags between sediment production and delivery, but these are not described adequately in the documentation.

(e) Table 1-2 is unclear and distracting to understanding document layout; needs column headings and caption. Perhaps move to program history section. Also, cross-walk with recommendations from CBP P5 review panel.

Literature Cited

Ator, S.W., J.W. Brakebill, and J.D. Blomquist. 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 pp. (available at http://pubs.usgs.gov/sir/2011/5167)

Ball, W.P., B.D Michael, D.C. Brady, J.L. Martin, S.H. Scott, and P.R. Wilcock. 2017. Lower Susquehanna River Reservoir System Model Enhancements Peer Review. CRC Publication No. 17-173, Chesapeake Research Consortium, Edgewater, MD. 59 pp.

Band, L., T. Dillaha, C. Duffy, K. Reckhow, and C. Welty. 2008. Chesapeake Bay Watershed Model Phase V Review. STAC Publication Number 08-003, Edgewater, MD. 13 pp.

Boomer, K.B., D.E. Weller, and T.E. Jordan. 2004. Empirical Models Based on the Universal Soil Loss Equation Fail to Predict Sediment Discharges from Chesapeake Bay Catchments. Journal of Environmental Quality 37: 79-89.

Bormann, H., L. Breuer, T. Graff, J.A. Huisman, and B. Croke. 2009. Assessing the Impact of Land Use Change on Hydrology by Ensemble Modelling (LUCHEM) IV: Model Sensitivity to Data Aggregation and Spatial (Re-)Distribution. Advances in Water Resources 32: 171-192.

Cavalli, M., S. Trevisani, F. Comiti, and L. Marchi. 2013. Geomorphometric assessment of spatial sediment connectivity in small Alpine catchments. Geomorphology 188: 31-41. DOI: 10.1016/j.geomorph.2012.05.007

Dabney, S.M., and D.C. Yoder. 2012. Improved description of herbaceous perennial growth and residue creation for RUSLE2. Agronomy Journal 104(3): 771-784.

De Vente, J. and J. Poesen. 2005. Predicting Soil Erosion and Sediment Yield at the Basin Scale: Scale Issues and Semi-Quantitative Models. Earth-Science Reviews 71: 95-125.

De Vente, J., J. Poesen, and G. Verstraeten. 2013. Predicting Soil Erosion and Sediment Yield at Regional Scales: Where Do We Stand? Earth-Science Reviews 127: 16-29.

de Wit, M. and J. Stankiewicz. 2006. Changes in Surface Water Supply across Africa with Predicted Climate Change. Science 311: 1917-1921.

Foster, G.R. 2013. Science Documentation User's Guide Version 2 RUSLE2. USDA – Agricultural Research Service, Washington, D.C. (available at http://www.ars.usda.gov/sp2UserFiles/Place/60600505/RUSLE/RUSLE2_Science_Doc.pdf)

James, L.A., D.G. Watson, and W.F. Hansen. 2007. Using LiDAR Data to Map Gullies and Headwater Streams under Forest Canopy: South Carolina, USA. CATENASoil Erosion and Sediment Transport under Different Land Use/land Cover Scenarios 71: 132-144.

Harman, CJ. 2015. Time-variable transit time distributions and transport: Theory and application to storage-dependent transport of chloride in a watershed. Water Resources Research 51: 1-30. DOI: 10.1002/2014WR015707

Hoos, A.B., and G. McMahon. 2009. Spatial analysis of instream nitrogen loads and factors controlling nitrogen delivery to streams in the southeastern United States using spatially referenced regression on watershed attributes (SPARROW) and regional classification frameworks. Hydrological Processes 23(16): 2275-2294.

IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, doi:10.1017/CBO9781107415324.

Johnson, Z., M. Bennett, L. Linker, S. Julius, R. Najjar, M. Mitchell, D. Montali, R. Dixon. 2016. The Development of Climate Projections for Use in Chesapeake Bay Program Assessments. STAC Publication Number 16-006, Edgewater, MD. 52 pp.

Lee, C.J., R.M. Hirsch, G.E. Schwarz, D.J. Holtschlag, S.D. Preston, C.G. Crawford, A.V. Vecchia. 2016. An evaluation of methods for estimating decadal stream loads. Journal of Hydrology 542: 185-203.

Momm, H., R. Bingner, R. Wells, and S. Dabney. 2011. Methods for Gully Characterization in Agricultural Croplands Using Ground-Based Light Detection and Ranging. F. Bhuiyan (Editor). Sediment Transport - Flow and Morphological Processes. InTech.

Murphy, P.N.C., J. Ogilvie, K. Connor, and P.A. Arpl. 2007. Mapping Wetlands: A Comparison of Two Different Approaches for New Brunswick, Canada. Wetlands 27: 846-854.

Murphy, P.N.C., J. Ogilvie, and P. Arp. 2009. Topographic Modelling of Soil Moisture Conditions: A Comparison and Verification of Two Models. European Journal of Soil Science 60: 94-109.

Noe, G.B., C.R. Hupp, and E.R. Schenk. 2015**a**. Measuring and predicting sediment and associated nitrogen and phosphorus fluxes of bank erosion, floodplain deposition, and net balance in stream reaches of the Chesapeake Bay watershed: current results from the USGS Chesapeake Floodplain Network. Draft.

Noe, G., C. Hupp, E. Schenk, and P. Claggett. 2015**b**. How to model sediment and nutrient fluxes of floodplains and streambanks across the Chesapeake watershed. Presentation to the Modeling Workgroup, November, 4 2015. http://www.chesapeakebay.net/channel_files/22984/noe_cfn_cbp_4nov15.pdf

Notebaert, B., G. Verstraeten, G. Govers, and J. Poesen. 2009. Qualitative and Quantitative Applications of LiDAR Imagery in Fluvial Geomorphology. Earth Surface Processes and Landforms 34: 217-231.

Rayburg, S., M. Thoms, and M. Neave. 2009. A Comparison of Digital Elevation Models Generated from Different Data Sources. Geomorphology 106: 261-270.

Richardson, M.C., C.P.J. Mitchell, B.A. Branfireun, and R.K. Kolka. 2010. Analysis of Airborne LiDAR Surveys to Quantify the Characteristic Morphologies of Northern Forested Wetlands. Journal of Geophysical Research-Biogeosciences 115: 16.

Sanford, W.E. and J.P. Pope. 2013. Quantifying Groundwater’ S Role in Delaying Improvements to Chesapeake Bay Water Quality. Environmental Science & Technology 47: 13330-13338.

Tarolli, P. and G.D. Fontana. 2009. Hillslope-to-Valley Transition Morphology: New Opportunities from High Resolution DTMs. Geomorphology 113: 47-56.

Thoma, D.P., S.C. Gupta, M.E. Bauer, and C.E. Kirchoff. 2005. Airborne Laser Scanning for Riverbank Erosion Assessment. Remote Sensing of Environment 95: 493-501.

Tomer, M., S. Porter, D.E. James, K.M.B. Boomer, J.A. Kostel, and E. McLellan. 2013. Combining precision conservation technologies into a flexible framework to facilitate agricultural watershed planning. Journal of Soil and Water Conservation 68(5): 113A-120A; doi:10.2489/jswc.68.5.113A

U.S. EPA (Environmental Protection Agency). 2013. Watershed modeling to assess the sensitivity of streamflow, nutrient, and sediment loads to potential climate change and urban development in 20 U.S. watersheds. National Center for Environmental Assessment, Washington, DC; EPA/600/R-12/058F.

Verachtert, E., M. Van Den Eeckhaut, J. Poesen, and J. Deckers. 2010. Factors Controlling the Spatial Distribution of Soil Piping Erosion on Loess-Derived Soils: A Case Study from Central Belgium. Geomorphology 118: 339-348.

Walter, R.C. and D.J. Merritts. 2008. Natural Streams and the Legacy of Water-Powered Mills. Science 319: 299-304.

Wu, S., J. Li, and G. Huang. 2005. An Evaluation of Grid Size Uncertainty in Empirical Soil Loss Modeling with Digital Elevation Models. Environmental Modeling & Assessment 10: 33-42.

Wu, S., J. Li, and G.H. Huang. 2008. Characterization and Evaluation of Elevation Data Uncertainty in Water Resources Modeling with GIS. Water Resources Management 22: 959-972.

Yang, P., D.P. Ames, A. Fonseca, D. Anderson, R. Shrestha, N.F. Glenn, and Y. Cao. 2014. What Is the Effect of LiDAR-Derived DEM Resolution on Large-Scale Watershed Model Results? Environmental Modelling & Software 58: 48-57.

Zhang, W.H. and D.R. Montgomery. 1994. Digital Elevation Model Grid Size, Landscape Representation, and Hydrologic Simulations. Water Resources Research 30: 1019-1028.

Appendix A

Phase 6 Peer Review Questions:

From the Charge to the Panel: "The Chesapeake Bay Program (CBP) partnership requested a scientific review that directly addresses the following questions. The review committee was also encouraged to make recommendations for future work by the CBP partnership that build on the questions or are related to the scientific or management issues raised in the Phase 6 peer review. The review committee was provided with relevant documentation, a detailed briefing, and access to CBP modeling practitioners. The review committee generated this report for submittal to CBP through the chairs of the Modeling Workgroup and the Modeling Coordinator. The Modeling Workgroup will ensure the STAC peer review record includes responses to the peer review panel's comments."

The charge questions outlined in that document as "Questions/Requests for STAC Review of the Phase 6 Watershed Model" are as follows (note that reference to sections in each question refers to the model documentation provided to the Panel by the CBPO):

1. Please comment on the overall appropriateness of the approach taken in the Phase 6 structure of a deterministic hydrology and sediment transport management model combined with a nutrient model informed by multiple models and multiple lines of evidence as described in Section 1. Please comment on the multiple model structure of the Phase 6 nutrient simulation particularly to its utility to watershed management in the Chesapeake restoration? How can the Phase 6 multiple model approach be improved going forward?
2. Please comment on the scientific rigor of the methods used for the average nutrient export rates described in Section 2. Are they calculated appropriately? Is there any additional scientific information that should be included?
3. In Section 4, how justified are the sensitivities of nutrient export from land uses to nutrient inputs, given the approach used and data available? Do the sensitivities to nutrient inputs derived from multiple models reflect our best understanding of the current condition of nutrient load processing and attenuation on the landscape? Is there any additional scientific information that should be included?
4. Please comment on the scientific rigor of the methods used in the use of Spatially Referenced Regression On Watersheds (SPARROW) for land to water factors in Section 7. Are they reasonably implemented? Is there any additional scientific information that should be included?
5. Please comment on the overall appropriateness of the methods used in the application of multiple methods to estimate stream-to-river factors for nutrients in Section 9? Is there additional scientific information that should be included?
6. Please comment on the scientific appropriateness of the approach taken for Phase 6 lag times described in Section 10 given the current state of information and understanding of groundwater and particulate processes. How can the structure and processes of nutrient lag time simulation on the land be improved in Phase 6 or future watershed model applications?

Is the application of the Ranked Storage Selection (rSAS) function for groundwater nitrate and Unit Nutrient Export Curves (UNEC) for all other nutrient species appropriate for the management questions?

7. Please comment on the scientific rigor of the methods used in the Phase 6 sediment simulation components using a detailed Revised Universal Soil Loss Equation 2 (RUSLE2) (Section 2), an interconnectivity metric (Section 7), and the inclusion of sediment source/sink estimates from stream banks and flood plains (Section 9).
8. Given the fine scale 1m x 1m land use data that's used in Phase 6, what opportunities does this open to the CBP and scientific community in the next phase of watershed model development? What are the advantages in a distributed representation of hydrology, land cover, and sediment? Given the availability of nutrient inputs from Agricultural Censes at the county scale only does a higher resolution of the watershed model make sense?
9. Better simulation of the deposition and scour processes in the reservoir reach of the Lower Susquehanna is an important feature of the Phase 6 Model. It is crucial to 2017 Midpoint Assessment decision making to be able to represent the net deposition of sediment, nitrogen, and phosphorus in this reach of the Susquehanna as fully as possible. Does the Phase 6 representation of the dynamics of the reservoir system rely on the best science available at this time?[1] Do the simulations approximately represent the observed changes in storage of sediment, nitrogen and phosphorus as seen in the historical record from the last few decades? How can the representation of Conowingo infill be improved going forward beyond the Phase 6 Model?
10. Please comment on the scientific appropriateness of the methods used in the representation of climate change in watershed nutrient and sediment loads estimated for the 2025 and 2050 time periods. ~~How well do the models used for producing future climate scenarios show skill in hindcasting the actual climatic and hydrologic changes that have happened over the past several decades?~~ [2]
11. For longer term CBP considerations, how can the overall approaches and procedures used in Phase 6 be improved and what alternative approaches and data gathering might you recommend?
12. Please comment on the Phase 6 documentation. Is it clear, well organized, concise, and complete (taking into account that it is the third Beta out of an expected four Beta versions and about six months ahead of final release)?

[1] Note that the Phase 6 Watershed Model is informed by other models and analyses including WRTDS and the Conowingo Pool Model. The Phase 6 Model reviewers are asked to address specifically the structure and calibration of the Phase 6 Model rather than these complementary tools which have their own ongoing independent peer reviews.

[2] Following deliberation with the CBPO during Phase 2 of the review, it was determined that adequate documentation was not available for reviewers to answer this specific question. It was removed from consideration on July 6, 2017.