Detailed response to “Scientific and Technical Advisory Committee Chesapeake Bay Watershed Model Phase 6 Review”

Chesapeake Bay Program Modeling Workgroup

May 2, 2018

Introduction

The Phase 6 Watershed Model is the most recent in a series of watershed models produced by the Chesapeake Bay Program (CBP) for use in management of nitrogen, phosphorus, and sediment dating to the early 1980s. The Phase 6 Model differs in structure from previous models in that its physical simulation components are greatly simplified. This structure allows for better stakeholder understanding of the processes, speeds up computations, and resulted in a demonstrably better agreement with water quality observations. The Phase 6 Model was built with a stakeholder-driven process through the CBP’s Modeling Workgroup and Water Quality Goal Implementation Team.

In the course of the normal business processes of STAC and the larger CBP, there has been significant opportunity to discuss the major themes, findings, and recommendations from this important review. The reviewers and Phase 6 Model practitioners made presentations to STAC and the Modeling Workgroup where the uncertainty issues were highlighted with possible solutions discussed. For example, the STAC workshops on BMP uncertainty and on the future of CBP modeling were able to build on the content of the Phase 6 Model review.

The CBPO modeling team found the STAC review team to be thoroughly engaged in the review process. The interactions during the review process were positive and respectful, leading to review outcomes that will help guide the CBP partnership in future model development. The panel were extremely clear in their recommendations, both positive and negative. This document will contain responses to the recommendations on a point-by-point basis. The CBP responses reflect the understanding that has arisen from conversations and workshops as of early 2018. However, significant changes in direction could result from future CBP partnership input.

The document presents both the recommendation from the review panel and the responses from the CBP. Responses in this document are relatively concise. Many recommendations prompted larger changes in the model documentation which are referenced here. Other recommendations are for efforts that will take years to implement. Only a short description of a path forward or an indication of the likelihood of CBP enthusiasm for the recommendation is included in the CBP responses.

Many of the recommendations focus on the importance of the estimation of uncertainty and the opportunities presented by the Phase 6 structure to pursue an uncertainty quantification. The Modeling Workgroup supports an uncertainty quantification and appreciates the significant effort by the STAC review team to discuss feasible methods. To begin work on uncertainty in earnest it will be necessary to work with the managers on the Water Quality Goal Implementation Team in order to understand how uncertainty will inform the management process. Only when the management goals are understood can we design an effective method of uncertainty quantification.

Question 1 – Section 1 Overview

Question from CBP: *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?*

Summary of recommendations from Question 1

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 STAC 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.

Response from CBP

Thank you for pointing out areas of the documentation that needed improvement. The first section has been restructured so that the physical setting and the segmentation are in separate sections (11 and 12) so that the overview in Section 1 can be concentrated on describing the motivation for the structure of the Phase 6 Model and how multiple models are used as an integral part. The recommendations for a future extrapolation of this modeling method are exciting to contemplate.

Recommendations 1a and 1b

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.

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.

Response from CBP

We appreciate the comments on documentation and communication made in writing and verbally during the in-person review. It is in the best interest of the CBP and the CBPO modeling team to have clear documentation of the rationale driving change in the watershed model and the structure that arose to meet stakeholder and scientific needs. We have moved discussions on the physical watershed and model segmentation out of the introductory section of the documentation so that it could be devoted to communicating the rationale and structure of the Phase 6 model. Language was added to the introductory section to clarify how multiple parallel models are currently used in Phase 6. Additionally, the language regarding multiple models was strengthened in Section 2 dealing with average loads and section four on model sensitivities. Text has been added to Section 7 on land to water factors to explain the decision to use sparrow rather than HSPF factors.

Recommendation 1c

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.

- 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.
- 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.
- 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).

Response from CBP

We agree that uncertainty estimation is a priority. This work is not currently feasible in the timeframe of this response, but we expect to work on uncertainty quantification over the next several years. Responses to the next few recommendations also deal with uncertainty. Regarding the specifics of this recommendation:

- We did not feel that adding calibration and uncertainty information to the first section was advisable given the need to focus on the model structure in that section.
- Section 4 was amended to address the skill assessment of CEAP and Phase 5.3.2
- Section 7 contains the uncertainty of the coefficients generated by Sparrow.
- The skill assessment of the Phase 6 Model is presented in Section 10 and is also directly available on the CAST documentation page.
- The second and third bullets suggest analyses that are presented more completely in recommendation (d). The CBPO team agrees with this general approach.
- The second bullet contains a recommendation to develop regression-based models to relate prediction errors to explanatory variables. This was done in the calibration in a more *ad hoc* fashion. Residuals were mapped and visually compared to known factors to determine the suitability for inclusion into the model. Different values for some coefficients were tested for improvement of model predictions. We agree that a formal structure would prove valuable.

Recommendation 1d

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.

Response from CBP

This is a particularly exciting recommendation and one that the CBP is likely to take very seriously in the coming years. During the calibration process, there was considerable discussion in the Modeling Workgroup and in the modeling team sharing the reviewers' concern that the calibration was relying on just the last parameter, river delivery, to match observation and that everything else was assumed to be true. This is not a completely accurate characterization as during the calibration process; the modeling team and Modeling Workgroup compared model errors to the spatial variability in alternative upstream factors. For example, land-to-water factors were modified to not include the factor 'coastal plain' after it was demonstrated that the calibration was improved with its exclusion and it was theorized that it was present in sparrow to make up for the lack of consideration of soil storage. However, an ad-hoc iterative calibration process is a far cry from a formal Bayesian analysis system.

Recommendation 1e

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.

Response from CBP

We agree that this is a good approach to estimate the uncertainty of the phase 6 Model. As the CBP and CBPO turn toward model uncertainty quantification, this is a good candidate approach to apply. A related approach would be to use the multiple models to estimate a parameter distribution to use in a Monte Carlo-type approach. No decisions have been made at this time, but conversations will continue through the next few years.

Recommendation 1f

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).

Response from CBP

Thank you for this recommendation as it highlights the need to better explain the purpose of the model, the structure of the static model, and the relationship between the static model and the temporal model. Section 1 of the documentation has been reorganized and clarified.

The CBP is not viewing the static model as a simpler version of the transient model. The static model (CAST) is the watershed model for both management and TMDL purposes. The transient model is used as one of the sources of input parameters for the static model and to temporally disaggregate the loads of the static model to run through the estuarine model.

We disagree with the characterization of the static model logic as 'circular' for the following reasons

- The river outlet loads are used for the mass balance of the overall RIM-shed. They are not used individually to predict spatial differences. The model then predicts spatial differences that are compared against the RIM sites individually.
- The primary prediction of the watershed model is change in anthropogenic load which is not estimated by the total time-averaged observed load at the RIM sites.
- Finally, as pointed out in the recommendation, a circular model may be justifiable if the components are viewed as composite parameters. This is the case and as discussed in recommendation 1d and the response to 1d, this method may be formalized going forward.
-

Recommendation 1g

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.

Response from CBP

The sections of the documentation have been improved as suggested.

Recommendation 1h

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.

Response from CBP

In a way, the documentation is ordered as suggested. Section 1 is the conceptual model description. Sections 2-9 are the model implementation. Section 10 is the calibration and validation, as well as the conceptual description and implementation of the temporal model. Unfortunately, uncertainties remain unaddressed. The CBP would argue that this organization better reflects the modeling process for Phase 6 in that Sections 2-9, the implementation, is where the partnership spent the majority of its development time. These are broken out in separate sections for the distinct parts of the model so that the CBP partners who were part of the development of these sections can access the documentation more easily.

Recommendation 1i

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.

Response from CBP

The CBP agrees with this recommendation in general and understands that it will require a sustained and focused effort to create a new conceptual model for sediment transport and to then create a new numerical model. During the STAC workshop "CBP Modeling in 2025 and Beyond", this idea of a new model for sediment transport and the CBP partnership processes that would be required were more concretely defined.

Question 2 – Section 2 Average Loads

Question from CBP: *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?*

CBP summary of Panel response to question 2

The review panel's response to question 2 on Section 2 of the documentation dealing with average loads by land use was generally favorable with some reservations: "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." There are a number of review comments that invite responses. These are listed below:

Scale issues

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.

Response from CBP

This is a reasonable observation that the CBP is using information that was derived at a small scale and applying it to larger scales. We are generally aware of that issue and take steps to minimize it, but we are limited by available information. The CBP strategy is to use absolute numbers where the scale of the observation is the same as the simulation and use percentages where it is not. For example, absolute concentrations, flows, and loads are used at gauging stations where the data and the model are evaluated at a precise location. Absolute pounds per acre are assigned at the edge-of-stream scale since that is the endpoint of the available calculations described in section 2 of the documentation. There is no estimation of edge-of-field loads since no method to estimate total transport losses was determined. Land-to-water factors are only used as a relative factor. Similarly, differences in loads between land classes and land uses are represented as proportions rather than absolute numbers. The essential untested assumption here is that these proportionalities hold across scale. In other words, if a BMP is estimated to have a 10% reduction at the edge-of-field, it is assumed to have a 10% reduction at the edge-of-stream.

Coastal Plain Loads

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.

Response from CBP

The CBP model development is absolutely hampered by the lack of monitoring data in the Coastal Plain. These are important areas for loads for the reasons noted by the panel and also because of the high loading land uses that tend to be located there. The model assumes that the same rules apply to the coastal plain as the rest of the watershed, but the delivery of nutrients is affected by characteristics that are specific to the Coastal Plain including low surface runoff, low sediment runoff, high recharge rates, and soil hydric state.

RUSLE

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.

Response from CBP

This specific point about pasture and RUSLE came up again during the partnership review of the draft Phase 6 Model that took place in late 2017. Changes were made to the cover factor to address this issue and significantly lower the sediment estimated from pasture. Further review or replacement of RUSLE for sediment estimation will likely be a recommendation from the STAC "CBP Modeling in 2025 and Beyond" workshop.

Recommendation 2a

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.

Response from CBP

We agree with the recommendation. As we move forward with uncertainty quantification we will use the variability in input coefficients in the estimation of model uncertainty.

Recommendation 2b

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.

Response from CBP

We agree with the points and this is how the calibration was done. Calibration is performed to match flows and concentrations while the loads are used for validation. The model documentation was not clear on this point and has been updated. The skill assessments are available for download.

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

Response from CBP

We agree up to a point. It is certainly true that we do not have confidence in the estimates of load to the nearest 1/100th of a pound and that the proper estimate of confidence would be better represented by rounding to the nearest 10 pounds. Rounding effectively communicates the confidence in the results, however it also introduces arbitrary error. The loading rate expert panel estimated through summarizing available literature that the land use 'grain with manure' should load about 20% higher than the land use 'silage without manure'. If we rounded to the nearest 10 pounds, we would show them with the same load which is not reflective of the panel's work.

Standard practice is to round only for the final output and not for intermediate calculations. The final product of this model, planning targets for the Chesapeake Bay TMDL and estimated progress toward them, are reported with two or three significant digits. All of the values presented in this section are intermediate calculations and in theory should not be rounded. However, the point is well taken that others will use these values for other purposes and need to understand the confidence that we have in them. We balance the competing priorities of communication and error reduction by inserting text discussing the level of confidence and by reducing the number of significant figures shown in the tables.

Specific Comments Regarding Model Documentation (b)

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

Response from CBP

Thank you for this comment! This is an obvious improvement once the recommendation is made, but this is the first time that is has been suggested in years of presenting this 'blue arrow'.

Specific Comments Regarding Model Documentation (c-l)

Various recommendations for clarity

Response from CBP

Thank you! Changes made as suggested.

Specific Comments Regarding Model Documentation (j)

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.

Response from CBP

The comment predicted the final decision of the CBP wetlands panel. The panel found no conclusive data to support loading rates differentiation between wetland types, or even between wetlands and forest. The differentiation between floodplain and headwater wetlands are for tracking purposes only.

Question 3 – Section 4: Sensitivities

Question from CBP: *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?*

CBP summary of Panel response to question 3

The review panel's response to question 3 on Section 4 of the documentation dealing with sensitivities of the output loads to inputs was generally favorable. STAC supports the theory and practice of using multiple models to estimate parameters and is interested in improving the confidence in the sensitivities by bringing additional observed data.

Recommendation 3a

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.

Response from CBP

We agree with the recommendation. This ground-truthing could be in the form of literature comparing changes in inputs to changes in outputs. Many of these types of studies were used in the creation of the models that were used to estimate the sensitivities, so it is likely that the models would match the central tendency of the literature. A new sub-section has been added to the end of the Section 4 to discuss literature findings. Another ground-truthing method is to use statistical models such as SPARROW to estimate the effects of changes in inputs. This is essentially what was done in that SPARROW was used as a support for the nitrogen sensitivities calculated from the Phase 5.3.2 Model.

Recommendation 3b

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.

Response from CBP

We agree with the recommendation. Section 10 covers the calibration of the model to observed flows and concentrations at more than 100 sites and the validation of the model against calculated loads at 60 or more sites. The process of calibration was to spatially compare the output of the models against observation and then to re-configure the model as needed to improve the agreement.

Specific Comments Regarding Model Documentation (a-i)

Various recommendations for clarity and correctness

Response from CBP

Thank you! Changes made as suggested. The absolute versus relative sensitivity references have been clarified throughout.

Question 4 – Section 7: Land-to-water factors

Question from CBP: *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 Section 7. Are they reasonably implemented? Is there any additional scientific information that should be included?*

CBP summary of Panel response to question 4

The review panel's response to question 4 on Section 7 indicates that they approved of the idea of using SPARROW to estimate land to water factors. Significant questions were raised on the specific methods of use of these factors, however. For the nitrogen land-to-water factors, concerns were raised that a version of SPARROW that used land use as a source rather than mass inputs required more study of the land use source model. For phosphorus a concern was raised that the dropping of some land-to-water terms would require a re-estimation of parameters. Additionally, the applicability of the interconnectivity metric for sediment was questioned given the limited previous application of this method.

Response from CBP

We appreciate the thoughtful recommendations. The documentation has been strengthened on the point that the nitrogen land-to-water factors come from the published version 4 SPARROW and not from the version that used land use as a source. The land use SPARROW was only used for the average loads in Section 2 rather than for land-to-water factors, but a discussion has been added to the documentation for comparison with the version 4 SPARROW.

We did use only a single land-to-water coefficient from the phosphorus model while we did not use the other three. We did not think that is was appropriate to re-estimate SPARROW without the other land-to-water variables for fear that any spatial co-variance between these variables would cause difficulties in interpreting the coefficient in the remaining variable. As part of the calibration procedures we compared the spatial agreement with data, in both load and yield terms, before and after dropping some of the phosphorus land-to-water terms and found improvement in the final Phase 6 Model from their exclusion.

Comments on the applicability and scientific justification of the interconnectivity metric are acknowledged and understood. It was the experience of the CBPO modelers that the inclusion of this particular variable was the major factor in improving the Nash-Sutcliffe efficiency of the spatial variability in sediment yields from approximately zero in Phase 5.3.2 to well over 0.9 in Phase 6. The success of the interconnectivity metric points to the necessity of understanding the processes better to determine why it resulted in such a significant improvement.

Recommendation Short-term 3a

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

Response from CBP

Additional text has been added to the documentation to discuss the land use-based SPARROW model coefficients.

Recommendation Short-term 3b

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.

Response from CBP

Graphics and text have been added to Section 7 to address this recommendation.

Recommendation Long-term 3a

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.

Response from CBP

We agree with the recommendation. The inclusion of the interconnectivity metric and its linear translation to a sediment delivery ratio was the major factor in improving the Nash-Sutcliffe efficiency of the spatial variability in sediment yields from approximately zero in Phase 5.3.2 to well over 0.9 in Phase 6, which is some level of validation. However, the point is well-taken that a more thorough physical understanding is warranted.

Recommendation Long-term 3b

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.

Response from CBP

We agree with the recommendation. More discussion is given in response to question 7.

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.

Response from CBP

We understand the recommendation and agree that inclusion of the full SPARROW description in Section 7 is somewhat ungainly. A major recommendation regarding documentation from the review panel was to concentrate on the rationale and explanation of the conceptual model in Section 1 so we will not discuss the three models there. Moving the description of all three supporting models to Section 2 where they are first used is a reasonable approach, but we would prefer that Section 2 on average loads be more accessible to managers and would prefer to leave the descriptions in Sections 4 and 7.

Specific Comments Regarding Model Documentation (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.

Response from CBP

The table and supporting text have been updated.

Specific Comments Regarding Model Documentation (c and d)

Comments on captions and explanations.

Response from CBP

Thank you for the comments that improve the documentation. All comments have been addressed.

Specific Comments Regarding Model Documentation (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.

Response from CBP

An uncertainty chapter is not possible within the current time constraints, but we expect to work on uncertainty within the CBP partnership over the next few years. See responses to recommendations 1d and 1e in particular.

Specific Comments Regarding Model Documentation (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).

Response from CBP

Feeding spaces, which are defined in Section 5, are not part of the pasture land use. Pasture still receives the SPARROW pasture land-to-water factor.

Specific Comments Regarding Model Documentation (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.

Response from CBP

We agree that available data for sediment delivery ratios are lacking in the Chesapeake Bay watershed and in general. The approach we took improved the calibration, but we agree that we need to understand the processes better. Please refer to our general response to question 4.

Question 5 – Section 9: Stream to River factors

Question from CBP: *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?*

CBP summary of Panel response to question 5

The review panel commented on two main parts of section – the use of SPARROW stream and reservoir decay and the use of the Stream Source Ratio. The panel requested better documentation as to why stream-to-river factors were specific to land classes since SPARROW did not differentiate attenuation by source.

Response from CBP

The panel is correct that stream-to-river factors are not differentiated by source in the SPARROW models. Estimates of stream and reservoir attenuation in SPARROW are at the NHDplus catchment level. Any spatial correlation of attenuation and land use is taken into account when applying values to the coarser phase 6 land-river segmentation. For example, policies may drive land use upstream of drinking water reservoirs to be higher in forest and lower in developed which would show up as a higher attenuation rate for forests relative to developed in land-river segments containing drinking water reservoirs. The documentation will be strengthened on this point.

Recommendation 5

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 (2015a,b), that could provide diagnostics for refining and building more confidence in the methods.

Response from CBP

We agree with the recommendation that the stream source ratio is strongly data limited. The final Phase 6 Model did not use the stream source ratio except as background information.

Question 6 – Section 10: Lag times

***Question from CBP**: 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?*

CBP summary of Panel response to question 6

The review panel saw the approach as a promising structure for the temporal model and worthy of continued development with the understanding that new modeling techniques may be developed that would improve upon the Phase 6 method. Concerns were expressed about the level of documentation for rSAS and the level of documentation for the parameterization for UNEC. It was pointed out in general that the lag model should be exercised to generate implications for management. The panel also made two technical points about the parameterization of lag time relating to the dependence of stream density on lag times and the exception of the coastal plain to the general relationships found for the watershed describing indicators of lag time.

Response from CBP

Documentation has been updated to better describe the UNEC parameterization. The point about the relationship between stream density and lag times is well taken and this is a good suggestion for future work. The fine-scale work that will be necessary to map stream density and then develop process-based hydrologic models is still in early stages. We agree that the initial relationship used between depth to groundwater recharge and lag times was inappropriate for the coastal plain. This particular issue was resolved for the final version of phase 6 where the coastal plain parameters were generalized from the USGS Eastern Shore groundwater model only. The documentation has been updated for this change.

Recommendation

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 the 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.

Response from CBP

These recommendations are aligned with the intentions of the CBP. There have been discussions at the Modeling Workgroup, the WQGIT, and within the CBPO modeling team on the ability of the lag time model to give the CBP the opportunity to assess the temporal effectiveness of management actions and to spatially target areas that are likely to quickly show results in the Bay. The panel has articulated the management questions well and the recommendation will help to move these questions to the forefront.

Question 7 – Sediment

***Question from CBP**: 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).*

Panel summary of response to question 7

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.

Recommendation 7

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.

Response from CBP

We agree with the panel that the conceptual framework for sediment in the Phase 6 Watershed Model, and watershed models in general, are due for a major revision. We are very encouraged by the discussion with the panel and the yet-unpublished results of the STAC "CBP Modeling in 2025 and Beyond" workshop that the time is approaching when effective conceptual models can be formulated. The CBP will soon be discussing the formation of a Sediment Subcommittee to report to the CBP

Modeling Workgroup. It is expected that the Sediment Subcommittee will make both short and long-term improvements to the model as called for in this recommendation.

We understand the point that matching fall line (and many more upstream) gaging station measurements does not mean that the effect of management practices will be accurately simulated. Nevertheless, the Phase 6 Watershed Model is able to predict spatial variability in sediment load reaching gauging stations much better than previous versions of the CBP watershed model and that does suggest that the Phase 6 Model can provide some guidance on the relative management effectiveness of spatial locations. Understanding how to 'observe' change in sediment load will be important for additional confidence in predictions of management practice effectiveness.

We also agree that there is not an accounting of time scales for sediment delivery within the watershed model, which should be part of any conceptual models that are developed in the future. Furthermore, we agree that the steady-state, meaning equilibrium, sediment delivery is not relevant to management since the time scales required to reach equilibrium sediment delivery are very long. However, the Phase 6 Model is not meant to simulate the equilibrium state, but rather the hydrologic average of current watershed conditions. For example, in developed areas there is a higher sediment export related to the amount of impervious. This higher export is not considered to be an equilibrium state of the channel, but rather the non-equilibrium load from the stream bed and bank that would be expected over a typical 10-year hydrologic period. The documentation has been strengthened on this point. A clear conceptual model that includes time scale will clarify the difference between a non-equilibrium average hydrology condition and an equilibrium condition.

Question 8 – Future high resolution simulation

***Question from CBP**: 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 Census at the county scale only does a higher resolution of the watershed model make sense?*

CBP summary of Panel response to question 8

The panel recommended that the CBP not pursue a high-resolution dynamic model to cover the entire watershed and cited many instances in the literature where higher resolution had not been shown to increase accuracy for dynamic simulation applications. However, the panel pointed out that high-resolution data could be used for developing dynamic process-based models with more limited spatial scope and for the mapping of sediment and biogeochemical processes on fine scales. These products could then be used to inform management decisions watershed-wide.

Recommendation 8

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.

Response from CBP

The CBP appreciates the careful arguments made in this response and recommendation. The recommendation to only use distributed dynamic modeling sparingly rather than a large-scale application was unexpected, but well supported and consistent with conversations in the STAC workshop "CBP Modeling in 2025 and Beyond", which is still to be published.

The panel makes the point that decisions made at the field scale have a great deal of leverage for cost-effective implementation while acknowledging that the models must still work at the large watershed scale. The way forward is not entirely clear. Perhaps a time-average high-resolution model could be built for the entire watershed based on knowledge gained from small-scale process models combined with mapping of geographic features relevant to sediment and biogeochemical processes. Formulating an effective plan to take on the issue of scale will present a significant challenge to the CBP and to STAC over the next several years and may generate appropriate topics for a STAC workshop or synthesis once the CBP has had the opportunity to digest the current set of STAC products.

Question 9 – Lower Susquehanna

***Question from CBP**: 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?*

CBP summary of Panel response to question 9

The panel found that the current Phase 6 simulation of the lower Susquehanna is supported by observations and informed by complementary modeling studies. Specifically, the conclusion that the reservoirs and in dynamic equilibrium seems to be well supported; the use of WRTDS for calibration is reasonable; the linear behavior of the reservoir in response to nutrient input changes is supported by multiple lines of evidence; and the change in reactivity of nutrients was addressed through models and observations. The panel found that the STAC recommendations from the Conowingo workshop were followed to a significant extent, but that the CBP models, complementary models, and statistical estimations of load were all hampered by a lack of observations during high flow events and a lack of observations within the Conowingo pool itself.

Recommendation 9a

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.

Response from CBP

The memo has been incorporated into Section 10.7 of the watershed model documentation.

Recommendation 9b

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, develop 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.

Response from CBP

The CBP community understands the primacy of observation in the modeling process. The WRTDS, the data that support it, and the interpretability of the results will continue to be improved through time. Unfortunately, it is not likely that the funding will be available to continue the modeling of the

Conowingo pool specifically, but gains made in sediment flux and the simulation of organic bioavailability will be used in the construction of future Chesapeake Bay models.

Specific Comments Regarding Model Documentation (a-i)

Changes were made as suggested. Thank you for your thorough review.

Question 10 – Climate Change

***Question from CBP**: 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.*

CBP summary of Panel response to question 10

The panel found that the work of the CBP modeling team had appropriately followed the guidance of the 2016 STAC workshop on climate change and the recommendations of the CBP's Climate Resiliency Workgroup. The panel is in agreement that these procedures represent the best available science.

Recommendation 10a – 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.

Response from CBP

We thank the reviewers for suggesting these much-needed details for describing the inputs used for climate change scenarios. The documentation was updated as follows: (a) Section 12.2.2.3 was added to describe the bias correction process; (b) Sections 12.2.2.2, 12.2.2.3, 12.2.2.4, 12.2.2.5 and 12.2.2.6 now provide details involved in input preparation, providing clarity for the sequential processing steps from CMIP5 model projections, to bias correction and downscaling, to delta-method for generating model inputs. Although known structural uncertainties are present in global circulation models, it is assumed with the bias correction approach that the biases in the hindcast simulation are the only biases carried over to expanded model simulation for the future conditions. This is consistent with practices used in standard climate change scenario applications but it does raise the philosophical question of whether a well calibrated model or one that is not well calibrated but offers better structural foundation is more reliable for prediction.

Recommendation 10b – 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.

Response from CBP

Thank you for those specific suggestions to improve the documentation. In response, Section 12.2.2.1 now provides details for the long-term trends using linear regression of annual rainfall volumes. As suggested, examples for a few land segments were included. We agree that uncertainty for the 2025 rainfall should be considered, but it was not included in the Phase 6 documentation as CBP Climate Resiliency and Modeling Workgroups did not recommend use of CMIP5 rainfall projections. As per suggestion, the snow hydrology is being investigated for inclusion in the documentation.

Recommendation 10c - 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.

Response from CBP

Thank you for this important review comment. New sub-sections 12.3.3.3, 12.3.3.4, 12.3.3.5, and 12.3.3.6 were added to describe the hydrology, sediment, nitrogen, and phosphorus model simulation responses for the climate change assessment. For the land simulation processes the projected changes are evaluated for total nitrogen and total phosphorus using the sensitivities. The estimated change in the total is then divided between the corresponding nutrient species as described in Section 10.5. Model simulation where rainfall volume change was distributed uniformly and non-uniformly across rainfall events (rainfall deciles) resulted in smaller differences in the transport of flow, sediment, and nutrients as compared to that with the volume change. The documentation was updated to include those model results.

The use of nitrogen and phosphorus sensitivities to flow and sediment provided a framework for using findings from numerous modeling studies in the CBP climate assessment. It is a novel step but results from available model results provided the nitrogen sensitives with wider uncertainty. In the current assessment, the available information were reconciled using best professional judgement, and modeling team's recommendations to the Modeling Workgroup. It is noted that a nitrogen sensitivity of 1.0 was applied for land transport processes alone, and does not include downstream responses of streams and rivers. This work along with reviewer's comments emphasize the need for a closer investigation of simulated climate change nitrogen response from multiple models to develop a better understanding.

Specific Comment 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.

Response from CBP

The model documentation was updated as suggested to provide more details on various aspects of climatic inputs for climate change scenarios.

Specific Comment Regarding Model Documentation (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.

Response from CBP

The discrepancies have been corrected.

Specific Comment Regarding Model Documentation (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.

Response from CBP

The documentation was updated as suggested for the uncertainty in climate projections and corresponding model simulated results.

Specific Comment Regarding Model Documentation (d)

Provide additional justification and explanation for the selection of nitrogen to flow ratio, how sensitive are the results to this ratio.

Response from CBP

Section 12.2.3.5 now provides more details on the sensitivity of nitrogen transport to flow.

Question 11 – Future approaches

***Question from CBP**: For longer term CBP consideration, how can the overall approaches and procedures used in Phase 6 be improved and what alternative approaches and data gathering might you recommend.*

Panel overview of response to question 11

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 11a–c

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.

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.

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

Response from CBP

The STAC has long requested that the CBP estimate the uncertainty of the CBP models. In particular, the 2005 and 2008 review of the Phase 5 Watershed Model and the 2014 STAC workshop on multiple models discussed the importance of uncertainty quantification for management decisions. A successful uncertainty quantification is more likely than it was in the past both because of increased management interest and the enhanced ability of the Phase 6 structure to support uncertainty quantification. In the past managers have been reluctant to discuss modeling uncertainty, but recently there is increased awareness and discussion of how it can improve models and the decisions that depend on the models. Additionally, as the reviewers note, the structure of Phase 6 is much more amenable to uncertainty quantification given the short run time and either the availability of multiple inputs or distributions of inputs. The specific suggestion of using models as separate realizations of the inputs may be the best method, but the CBP will also consider other methods, such as using the multiple models to define a distribution of inputs from which samples can be selected.

The specific idea of a Bayesian structure to the entire model neatly combines the issues of calibration and uncertainty quantification together in a single solution. The CBPO has begun the recoding of CAST in a procedural language and the porting of the code to the cloud so that optimizations can be run. Significant parts of the software system required for optimization are the same components that will be required for the Bayesian model and uncertainty quantification. As mentioned in the response to question 1, this idea is a particularly exciting idea that fits very well within the structure of Phase 6.

Recommendation 11d

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.

Response from CBP

The recommendation to investigate running CAST at a finer level of spatial detail is a welcome one for the CBP. As mentioned in the recommendation, since the SPARROW data that generate the land-to-water and stream-to-river factors are available at the NHD scale and CAST is essentially linear, CAST would return similar results whether run at the current land-river segment scale or the NHDplus catchment scale. Running at the NHD catchment scale would give additional nutrient reduction credit for implementing BMPs in areas with less stream attenuation or greater nutrient transport properties. The point that the agricultural input data are limited by the county scale in most cases is a good one, but making a model available to the partnership on a finer scale may incentivize the partners to collect data on a more useful scale. Increasing the resolution would increase run times by the same factor. This would not be a problem for a single run of CAST, but would perhaps inhibit the ability of the CBP to carry out recommendations 11a-11c or to implement cost optimization for a given reduction, which is a priority of the CBP. As with all modeling exercises, there must be a balance between spatial scale and run time.

Recommendations 11e–f

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."

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.

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.

Response from CBP

In response to the 2008 review quoted above, the capability to pull random values from a distribution for each BMP was implemented during the development of the Phase 5 Model in 2009 and 2010. Spatial uncertainty was represented by pulling the effectiveness of a BMP in a land-river segment from a distribution. Temporal uncertainty was estimated by relating the effectiveness of a BMP to flow. The spatial uncertainty method was tested but was never used for management purposes due to the lack of a specific management question. The temporal-flow uncertainty was used in a EPA/RAND study of robust decision making relative to climate change:

Fischbach, Jordan R., Robert J. Lempert, Edmundo Molina-Perez, Abdul Ahad Tariq, Melissa L. Finucane, and Frauke Hoss, Managing Water Quality in the Face of Uncertainty: A Robust Decision Making Demonstration for EPA's National Water Program. Santa Monica, CA: RAND Corporation, 2015. https://www.rand.org/pubs/research_reports/RR720.html.

Aside from the lack of management interest, these methods suffered from a lack of data to parameterize the uncertainty in BMP effectiveness. There are some promising developments in the CBP community, however. STAC has recently held a workshop on BMP uncertainty which will likely lead to changes in the WQGIT protocol for BMP panels that require estimates of uncertainty.

Stephenson, K., C. Hershner, B. Benham, Z. Easton, J. Hanson, S. Julius, E. Hinrichs. (2018). Consideration of BMP Performance Uncertainty in Chesapeake Bay Program Implementation. STAC. http://www.chesapeake.org/pubs/390_Stephenson2018.pdf.

The major recommendations from this workshop were:

1. Systematically document and represent uncertainties throughout the BMP treatment process;
2. Produce information about the distribution of removal effectiveness of each BMP;
3. Develop a method for simply and effectively communicating the degree and type of uncertainty across all approved BMPs; and
4. Provide additional guidance for how to most effectively solicit “best professional judgment” as part of the expert panel process, including best practices for structured literature syntheses, identifying and avoiding potentially inappropriate heuristics (shortcuts) and biases when obtaining expert opinion, and expert elicitation.

At the time of this writing, the process of revising the BMP panel protocol to address these recommendations is beginning. These three recommendations will be provided to support the STAC BMP uncertainty workshop recommendations and to provide specifics for how the panels may best provide uncertainty data.

Recommendation 11h

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.

Response from CBP

The CBP agrees that understanding seasonal dynamics, particularly the expected effects of climate change is crucial for making the linkage between BMP implementation and water quality improvements in the Bay. The Phase 6 Model has an improved seasonal calibration to observation relative to Phase 5 which is important for simulating eutrophication in the Bay. The empiricism of the Phase 6 approach, however, means that we must rely on external and intermediate models to provide expected seasonal changes. Storm loads are a priority in calibration because they generate roughly half of the nitrogen load and a majority of the phosphorus and sediment load, but also, as is stated in the recommendation, a good characterization of storm loads is needed for understanding BMP response.

Recommendation 11i

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

Response from CBP
The CBP agrees as noted in the response to question 7.

Question 12 – Documentation

***Question from CBP**: 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).*

Panel summary response to question 12

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

Recommendation 12a I – Organization

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.

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

Response from CBP
This recommendation led to a reorganization of the introductory section. The segmentation and physical description of the watershed were moved to Section 11. The management description and the model conceptualization were kept within the opening section, but a better division was made between these subsections. Considerable effort went into editing the conceptualization and structure sections so that these are more easily understood.

Recommendation 12a III – parallel organization of model documentation

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).

Response from CBP
The CBPO modeling team has adopted a somewhat parallel structure where the first section of each chapter begins with the conceptual diagram of the Phase 6 model, proceeds to describe the model, multiple models, data, or other analysis that are used for the factor that is the subject of the chapter, and ends with a table or map describing the output. As the model is further developed, we will tend toward the full reorganization of the documentation along the lines suggested. However, since there is

variability in the types of information available, the suggested organization will be a general guideline rather than a required format.

Recommendation 12a IV – parallel organization of model documentation

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.

Response from CBP

(Repeated from question 4) We understand the recommendation and agree that inclusion of the full SPARROW description in Section 7 is somewhat ungainly. A major recommendation regarding documentation from the review panel was to concentrate on the rationale and explanation of the conceptual model in Section 1 so we will not discuss the three models there. Moving the description of all three supporting models to Section 2 where they are first used is a reasonable approach, but we would prefer that Section 2 on average loads be more accessible to managers and would prefer to leave the descriptions in Sections 4 and 7.

Recommendation 12b – Clarity of process simulation

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.

Response from CBP

Text has been added to the opening part of Section 10 to make this clearer. The relationship between the time-averaged model and the transient model can be more easily understood by describing the processes that connect them. The time-averaged model supplies the long-term edge-of-river loads of nitrogen, phosphorus, and sediment to the transient model. The transient model separates these long-term loads into hourly loads into the rivers. The process of temporal disaggregation of the loads is informed by hydrology, seasonality, and lags. The HSPF biogeochemical river process model is then calibrated to observed concentration data in rivers. The resultant temporal simulation is then used to estimate the river delivery coefficient in the time-averaged model.

Recommendation 12c – UNEC and rSAS description

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.

Response from CBP

A significant amount of additional discussion has been added to Section 10 to clarify how UNEC are rSAS are used to generate time series of nutrients to rivers and how these are made to match the time-averaged model.

Recommendation 12d – Sediment lag description

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.

Response from CBP

Sediment lag times are not considered in the time-averaged model. There are coefficients to account for production and loss in the landscape and rivers, but these do not consider the time dimension.

Recommendation 12e – Table 1–2

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.

Response from CBP

Tables 1–2 has been updated to include headers so that is it clearer. We chose not to include the Phase 5 review cross-walk. The most recent Phase 5 review had approximately 50 recommendations and this would expand the table beyond what is reasonable in an overview section.