13 Section 13 Review

13.1 Introduction

The Phase 6 Watershed Model was developed by the Chesapeake Bay Program (CBP) partnership over the period 2012 to 2017. The Modeling Team at the Chesapeake Bay Program Office (CBPO) carries out the directions of the CBP partnership as shown in Figure 13-1. An organizational chart is available on the CBP web site to see how these groups are related within the broader CBP context.

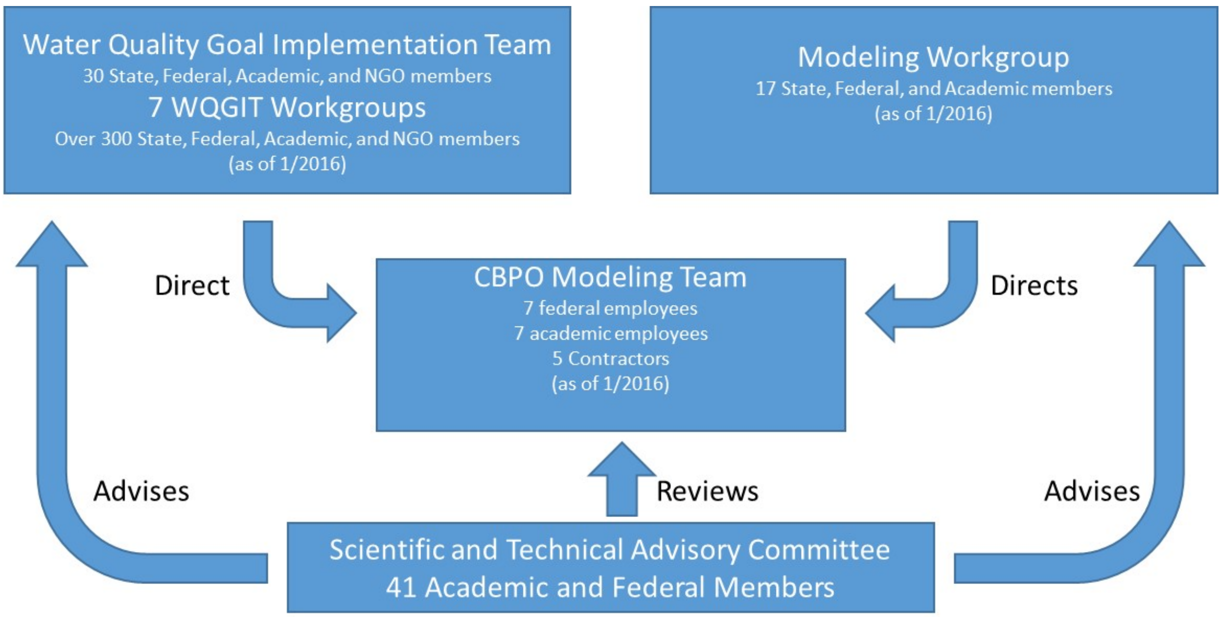

Figure 13-1: CBP Modeling Governance Structure

The CBPO Modeling Team is a cross-disciplinary and multi-organizational group tasked with developing and operating models under the direction of the CBP partnership. Employee numbers in the above figure are approximate as staff at the CBPO are directed to CBP priorities as needed. The Modeling Workgroup (MWG) reports to the Scientific and Technical Analysis and Reporting (STAR) and is primarily tasked with issues related to modeling of the physical environment, calibration, issues that cross sectors, and scientific integrity. The Water Quality Goal Implementation Team (WQGIT) reports to the Management Board and directs the Modeling Team on issues related to model inputs, the effectiveness of BMPs, the extent of BMP implementation, and how the models are used to inform policy. The WQGIT has seven workgroups that direct the Modeling Team or advise the WQGIT, depending on the issue. The Scientific and Technical Advisory Committee (STAC) is an independent scientific body that advises the partnership through recommendations from workshops and reviews, and through letters to the CBP partnership. The MWG, the WQGIT and its workgroups, and STAC also receive considerable input from stakeholders and other interested parties that participate in regular meetings. Expert panels recommending effectiveness values for BMPs report to the WQGIT and its workgroups.

13.2 Partnership review

The Principal Staff Committee (PSC) of the Chesapeake Bay Program (CBP) directed development of the 2017 modeling tools in order to 1) help guide implementation of CBP Watershed Implementation Plans (WIPs) from 2018 to 2025, 2) assess the implications of Conowingo Dam infill on tidal water quality, and 3) assess the influence of future projected increased temperatures and watershed flows on tidal water quality. The CBP partnership have built the suite of 2017 modeling tools over the past half-decade through a multitude of decisions made by the PSC, the Water Quality Goal Implementation Team (WQGIT) and its sub-groups, and the Modeling Workgroup of Scientific Technical Assessment Reporting (STAR). The Phase 6 Model represents the work of the hundreds of partners from dozens of organizations within these groups, as well as in expert panels, and through other modes of participation.

13.2.1 Ongoing review

The WQGIT, MWG, and WQGIT workgroups established priorities in early 2012 for the 2017 Midpoint Assessment (MPA). These are listed on the CBP's MPA website: https://mpa.chesapeakebay.net/Details.html. The WQGIT met on October 22 and 23, 2012 to prioritize these workplans for the partnership's development of the Phase 6 Watershed Model. During the period 2012 to 2017 review and direction of the Phase 6 Watershed Model has occurred in the hundreds of meetings where different aspects of the modeling were discussed and decided upon. A full accounting of meetings where the Phase 6 Watershed Model was discussed is beyond the scope of this documentation. One illustration is given in Table 2-1. Participants in these meetings directed CBPO staff in developing the values determining the differences in loading rates of land uses within a category. The results of these meetings are documented in section 2.2.4.1.

Table 2-1: Partnership Input to relative loading rates

Chesapeake Bay Program committees, goal implementation team, and workgroups	Meeting Date
Modeling Workgroup	9/30/2014, 1/29/2015, 3/26/2015, 4/23/2015, 6/11/2015, 7/21/2015, 10/5/2015
Modeling Team Meeting	9/15/2014, 1/20/2015, ongoing weekly
Land Use Workgroup	9/25/2014; 2/26/2015
Watershed Technical Workgroup	10/2/2014, 3/5/2015
Forestry Workgroup	10/1/2014, 3/4/2015, 3/20/2015
Wetlands Expert Panel	11/12/2014
Urban Stormwater Workgroup	10/21/2014, 12/16/2014, 3/3/2015
Agricultural Workgroup	10/9/2014, 10/22/2014, 2/19/2015, 3/25/2015, and multiple other meetings with updates from the Subgroup
Agricultural Modeling Subcommittee	9/16/2014,12/16/2014, 2/12/2015, 2/18/2015
Agricultural Loading Rate Review Subgroup	3/25/2015, 8/26/2015

The modeling team also participated in meetings with non-CBP scientific, management, and stakeholder groups on numerous occasions. A particularly prominent example is the 'Building a Better Bay Model' workshop that was held with agricultural stakeholders on May 22 and 23, 2013, with over 100 attendees. Recommendations from this workshop were used by the Agriculture Workgroup to prioritize improvements in input data for the Phase 6 Watershed Model.

13.2.2 Fatal Flaw Review

During June and July 2017 a fatal flaw review was conducted based on the Draft Phase 6 (6/1/2017) version of the documentation with the potential for changes to be made during this time. All fatal flaws identified through the partnership review were resolved through the procedures described below. The Modeling Workgroup and the WQGIT met in late summer for final approval of the Partnership's suite of Phase 6 models to be used for the 2017 Midpoint Assessment.

A fatal flaw may be the basis for the implementation of changes to the draft Phase 6 models. A fatal flaw is defined as a ***significant*** impediment, based on a weight of evidence approach, of the ability of the partnership to establish reasonable planning targets or evaluate progress toward achieving the planning targets or meet the conditions of EPA's "Interim Expectations for the Phase III Watershed Implementations Plans," dated January 19, 2017 (Expectations Document) due to:

- A calculation or method that does not follow the documented final decisions of the CBP partnership
- A calculation or method, or combinations thereof, that produce illogical results that result in significant impediment
- The omission of data submitted by the CBP partnership subject to established deadlines
- The overall failure of the model calibration to match observed flows and loads when compared to the level of performance in previous models

Disagreement with a final decision that has been made by the partnership or with a scientific or technical method or product in favor of another method or product are not fatal flaws. Nor are failures to match loads for particular monitoring stations or constituent nutrients fatal flaws unless they create a significant impediment to planning target development or progress evaluation. With respect to fatal flaws, model performance should be judged primarily at the State/Basin scale because the planning target development process provides jurisdiction allocations at that scale. Consideration should also be given to model performance at the impaired segment scale because the Phase 6 suite of modeling tools are intended to be used to determine pollutant reductions necessary to attain water quality standards in all impaired segments. Comments regarding model performance at scales smaller than State/Basin or impaired segment are welcomed when the reviewer perceives potential adverse impacts to local targeting, implementation or crediting. If reviewers find examples of where the Watershed Model's performance is mismatched with monitoring data on smaller/local scales these observations should be noted, not necessarily as fatal flaws, but to inform the Partnership of scale limitations, particularly as related to establishing local goals.

Comments were received from the partnership. CBPO staff and WQGIT leadership developed responses and initial suggestions of whether issues were sufficiently damaging to warrant a change in the model. Issues were resolved through personal communication, and discussion in the Modeling Workgroup, the Water Quality GIT, or its workgroups. The WQGIT discussed the fatal flaw review at each meeting in June, July, and August 2017. A spreadsheet was developed to contain the comments and responses. The spreadsheet presented at the August 28, 2017 WQGIT meeting is the final version and is available on that meeting page or by direct link. Decisions were made by the partnership on the fatal flaw review in the following meetings

- August 14, 2017 – WQGIT discussed fatal flaw review in depth
- August 28, 2017 – WQGIT approved resolution of all review comments except for issues related to soil phosphorus, which was referred to the management board
- September 21, 2017 – Management Board found no fatal flaws with the soil phosphorus

13.2.3 Phase 6 Webinars

The Modeling Workgroup scheduled and conducted a series of webinars to provide partners and stakeholders with a summary of the Draft Phase 6 Watershed Model in considerable detail.

- Phase 6 Sediment Simulation Webinar, Gary Shenk lead. April 19, 2017.
The overall simulation of sediment from estimated target load, land surface detachment, and transport and fate in the watershed will be reviewed.

- Overview of the Integrated Air Watershed and Bay Models, the Midpoint Assessment Decisions the Models Were Designed to Address, and the Decision Framework of Standards, Models, and Planning Target Method, Lew Linker and Gary Shenk Leads. May 9, 2017, 1-3pm.
A description of the integrated models and the decisions the PSC and other groups directed the models to address including an improved approach to get to the Phase III WIPs, Conowingo infill impacts to water quality and an assessment of the influence climate change has on tidal water quality standards.

-Phase 6 Chesapeake Bay Model Poultry Data, Matt Johnston lead. May 24, 2017, 1-3pm.
Webinar on the Phase 6 Chesapeake Bay Model, with an emphasis on poultry data.

- Phase 6 Inputs Webinar, Matt Johnson Lead. May 25, 2017, 1-3pm.
Overall review of Phase 6 inputs.

- Phase 6 Loads Webinar, Gary Shenk Lead. June 1, 2017, 1-3pm.
The nutrient loading rates, how the rates are calculated and how the loading rates are modified due to sensitivities, and sediment.

- Phase 6 Physical Transport Webinar, Gopal Bhatt and Gary Shenk lead. June 20, 2017, 1-3pm.
The processes of riverine and small stream transport and attenuation of nutrient and sediment loads will be reviewed in detail. Included also are the land to water factors for nitrogen and phosphorus loads.

13.3 STAC review and guidance

The CBP's Scientific and Technical Advisory Committee (STAC) provides independent scientific guidance to the CBP partnership through workshops, reviews, and direct communications. Many aspects of the Phase 6 Watershed Model were developed in direct response to recommendations from STAC.

13.3.1 Workshops

Through sponsored workshops, STAC provides a forum for scientists and managers to collaboratively deliberate findings and recommendations based on the latest scientific understanding directly applied to management priorities. STAC reviews and approves workshop reports that provide well-supported recommendations directly to the CBP partnership. STAC typically holds between 4 and 6 workshops each year, many of which have informed development of the Phase 6 Watershed Model. Some of the more important workshops are listed here. Many additional STAC workshops contributed ideas and recommendations to the modeling effort.

Full recommendations and documentation are available from the STAC website http://www.chesapeake.org/stac/. Descriptions and summarized recommendations below are generally quoted directly from letters from STAC to the CBP partnership, also available on the STAC website.

13.3.1.1 Natural landscapes

The Role of Natural Landscape Features in the Fate and Transport of Nutrient and Sediment – March 7-8, 2012. Reference (STAC 2012)

The purpose of the workshop was to understand the scientific basis for nutrient and sediment loading rates from natural landscape features in the Chesapeake Bay Program Watershed Model and how loading rates could be based on ecological health, management status, and landscape position.

Recommendations:

The CBP should pursue upgrades to the Chesapeake Bay Program Watershed Model to provide more accurate estimates of nutrient and sediment loading rates based on the considerations discussed in the report. The recommendations include:

1. Adding three new land use classifications including riparian forest, forested floodplains (in general, these are wider and closer to water table than riparian forests), and other wetlands.
2. Adjusting loading rates for the new land use classes based on spatially explicit landscape attributes, including directional connectivity, multi-direction flow fields, and flow path analysis.
3. Adjusting loading rates for the new land use classes based on landscape feature attributes including type, condition, and possibly forest age.

The CBP should use scenario analysis to identify the most effective landscape features and the most effective configurations of features.

The CBP should consider investing in watershed models for use at the local scale.

Effect on the Phase 6 Watershed Model:

The Forestry Workgroup and wetlands BMP panel invested considerable effort in attempting to map and describe the loading rates for different types of natural landscapes. For the Draft Phase 6 model, the available data support separating true forest from harvested forest and from tree canopy over developed areas. Wetlands are simulated for the first time and separated into non-tidal floodplain, tidal, and isolated categories as discussed in section 2.2 of the Phase 6 documentation.

Recommendations from this workshop lead to the calculation of sediment losses on a fine scale as discussed in section 2.3 and in calculation of spatially-detailed sediment delivery ratios discussed in section 7. Section 10 shows the dramatic improvement in sediment calibration due in large part to these two fine-scale calculations.

13.3.1.2 Multiple Models

Multiple Models for Management in the Chesapeake Bay Watershed. – February 25-26, 2013. Weller and others 2013

The purpose of this workshop was to understand the benefits and methods of implementing multiple model methods.

Recommendations:

The CBP should implement a multiple modeling strategy for each major decision-making model of the Bay (airshed, land use, watershed, and estuary) and analyze the output to quantify skill, advance knowledge, and inform adaptive management.

The CBP should exercise the multiple model systems developed under Recommendation 1 to quantify model uncertainty and confidence in key predictions used in decision-making.

The CBP should estimate and better communicate the appropriate levels of spending on monitoring, modeling, and research relative to the costs of implementation and the cost of poor decision-making.

The CBP should implement ways to better communicate modeling, uncertainty, and multiple model results to partners, decision makers, and the public.

Effect on the Phase 6 Watershed Model:

As discussed in section 1 of the Phase 6 documentation, the structure of the Phase 6 Watershed Model was directly inspired by the recommendations from this workshop to allow for the inclusion of multiple models and multiple lines of evidence. Where available, multiple models are used to estimate coefficients throughout the Phase 6 model. See particularly sections 2 and 4 of the Phase 6 documentation.

13.3.1.3 Conowingo Infill

Conowingo Reservoir Infill and Its Influence on Chesapeake Bay Water Quality - January 13-14, 2016 Linker and other, 2016a

The purpose of this workshop was to address the state of the science on the influence of Conowingo Reservoir infill on tidal Chesapeake Bay water quality.

Recommendations:

Efforts to model the effects of Conowingo on net accumulation or release of nutrients and sediment from the reservoir should be evaluated based on its ability to "hindcast" data from water quality observations and statistical analyses.

In order to quantify the influence Conowingo infill has on Chesapeake water quality, three primary issues should be considered for modeling:

1. Address biogeochemical processes related to sediment scour and nutrient cycling that may influence bioavailability in reservoir sediments, under variable flow ranges in the Conowingo Reservoir.

2. Ensure representation of effects of Conowingo inputs to Chesapeake Bay for the full range of flow conditions including 'extreme' high-flow events.
3. Improve representation of reactivity of particulate organic matter in Conowingo outflow.

Moving forward, an effort should be made to link the sediment transport and biogeochemical models in the 2010 Water Quality and Sediment Transport Model (WQSTM) to enhance modeling of the transport and fate of organic nutrients in the tidal Bay.

Effect on the Phase 6 Watershed Model:

The three primary issues were addressed through the use of statistical and simulation models. The full discussion of this process is in section 10 of this documentation under a Conowingo heading.

13.3.1.4 Climate Change

The Development of Climate Projections for Use in Chesapeake Bay Program Assessments – March 7-8, 2016. Johnson and other, 2016

The main goals of this workshop were to assess the applicability of available climate data, downscaling techniques, projections, and scenarios to establish an approach for climate analysis in Chesapeake Bay Program assessments.

Recommendations:

For the 2017 Midpoint Assessment, use historical (~100 years) trends to project precipitation to 2025.

The Partnership should carefully consider the representation of evapotranspiration in Watershed Model calibration and scenarios, due to its strong influence on future water balance change.

A 2050 timeframe is more appropriate for selecting and incorporating a suite of global climate scenarios and simulations to provide long-term projections for the management community, and an ongoing adaptive process to incorporate climate change into decision-making.

For any 2050 assessment, use an ensemble or multiple global climate model approach, selecting model outputs that bound the range of key climate variables for the Chesapeake Bay region. Additionally, use multiple scenarios covering a range of projected emissions

Beyond the 2017 Midpoint Assessment, it is recommended that the CBP use 2050 projections for best management practice (BMP) design, implementation and performance, given that many BMPs implemented now could be in use beyond 2050.

Effect on the Phase 6 Watershed Model:

As discussed extensively in section 12.2 of this documentation, climate change estimates for the Midpoint Assessment use historical data to project precipitation to 2025 and an ensemble of models to predict temperature.

13.3.2 Reviews

In addition to holding workshop, STAC also organizes formal reviews of CBP products when requested. Review procedures follow the recommendations of the EPA's Science Advisory Board that a review includes both a written review from the review panel and a written response from the CBP.

13.3.2.1 Phosphorus in Phase 5.3.2

Staver, K., P. Kleinman, S. Ator, A. Buda, Q. Ketterings, J.T. Sims, and J. Meisinger. 2014. A review of agricultural P-dynamics in the Chesapeake Bay Watershed Model. STAC Publication Number 14-005, Edgewater, MD. 20 pp.

This review was initiated by STAC following the September 2011 STAC meeting in order to investigate how phosphorus transport from cropland is simulated in the Chesapeake Bay Watershed Model. The following are the recommendations in the report and the responses from the CBP

13.3.2.1.1 Modeling Recommendations Related to Soil P:

- Account for soil P reservoirs as a source of P to runoff on a segment-by-segment basis.
- Track segment P balances to determine whether soil P reservoirs are increasing or decreasing.
- Describe the temporal dynamics of the effects of drawdown/build-up of soil P reservoirs on P losses in runoff.

Response from the CBP

The CBP's Modeling Workgroup (MWG), Water Quality Goal Implementation Team (WQGIT), and the WQGIT's workgroups are generally addressing these three recommendations from STAC's phosphorus report, as well as many other priorities, through the restructuring of the CBWM. As described in the STAC report, the Phase 5 CBWM was dependent on Hydrologic Simulation Program – Fortran (HSPF) for the simulation of phosphorus dynamics. Following recommendations in STAC's report on multiple models, the Phase 6 CBWM has a modular design that allows the use of the most appropriate model or suite of models for each major process. Discussions within the CBP workgroups and with the STAC report authors have led the CBP partners to investigate the Annual Phosphorus Loss Estimator (APLE) as an appropriate model for phosphorus losses that incorporates most of the factors identified by STAC. As of late 2014, the CBPO has recoded APLE into a more convenient format to run multiple large-scale simulations. The MWG will be evaluating APLE for use in the Phase 6 CBWM during 2015.

As identified in the STAC report, APLE requires accurate soil phosphorus level data. The CBP has made arrangements to acquire county-scale data sets from testing labs. This will allow the CBP to better simulate the phosphorus balances on the management scale of the Chesapeake Bay Total Maximum Daily Load (TMDL).

The STAC phosphorus report, combined with the previous STAC report on lag times (STAC 2013), identifies the temporal response of phosphorus draw down as a critical issue. It is anticipated that APLE will adequately simulate this lag and allow the MWG to consider this lag in the calibration of the CBWM. This additional knowledge of lag times will be useful to the CBP's Explaining Trends in Water Quality Team as it investigates the physical processes driving observed changes in water quality. The WQGIT will determine how these descriptions of lag times will be used in running and evaluating scenarios and supporting collaborative management decision making by the larger partnership.

13.3.2.1.2 Modeling Recommendations Related to Management of P Inputs:

- Account for different P application methods, including whether manure is left on the soil surface, incorporated by tillage or incorporated with low soil disturbance full-width applicators, or injected in bands.
- Apply manure at rates and times based on watershed or regional information.

- Improve representation of practices aimed at reducing P runoff potential by adjusting the timing of P applications.
- Account for P stratification that develops in soils in continuous no-till.
- Account for interaction between tillage and manure application on potential for P losses as particulate and dissolved P fractions in overland and sub-surface flow.

Response from the CBP

The above management recommendations will not be modeled explicitly by a mechanistic simulation in the Phase 6 CBWM. The CBP partnership has determined that the most appropriate method to account for these and other conservation practices is to consider multiple models and multiple lines of evidence within an expert panel. The Manure Injection/Incorporation Panel has yet to be to be convened, but its creation has been approved by the Partnership by the WQGIT. The guiding documents are not available, but the Partnership's Nutrient Management Expert Panel has recommended that the Manure Injection/Incorporation Panel develop recommendations related to phosphorus stratification, speciation, and loss related to manure spreading practices and tillage including continuous no-till, covering recommendations 1, 4, and 5 above.

The Nutrient Management Expert Panel extensively discussed recommendations 1 and 3 above on the timing and method of phosphorus application. Due to the Partnership's desire to focus on near term milestone goals and recommendations for the Phase 5.3.2 CBWM, these topics were left out of the Panel's draft final report. However, the Partnership will convene a Nutrient Management Expert Panel to make recommendations for the Phase 6 CBWM that considers these topics.

The Agricultural Modeling Subcommittee (AMS) of the CBP Agriculture Workgroup (AgWG) is charged with developing the rules and data for application of manure, including the rates and times called for in recommendation 2 above. Discussions are currently underway within the AMS and the best available information will be used for the Phase 6 CBWM.

13.3.2.1.3 General Modeling Recommendations:

The following general recommendations from STAC are valuable guides for future modeling efforts. In some cases, these can be incorporated into the Phase 6 CBWM. Others will require a longer time frame to implement.

Identify the fraction of P losses associated with short versus long-term management strategies.

CBP RESPONSE: Given the recommendations and the strategies to meet the recommendations laid out above, this will be a realizable goal. Most conservation practices discussed in Partnership's BMP expert panels will have a short term effect while the mass balance approach and the APLE simulation will allow for calculation of time lags in phosphorus soil drawdowns.

Model functions should be capable of scaling down to provide segment and field guidance on drivers of P.

CBP RESPONSE: The Partnership's BMP expert panels' recommendations and large-scale modeling are necessarily general in nature. While these tools will provide guidance at the segment scale on the relative effectiveness of conservation practices in different regions, field-scale modeling for the

Chesapeake Bay region will likely remain beyond our abilities for the near future although it should remain a long term goal.

Shift away from using model-generated values and proxy data for key input parameters.

CBP RESPONSE: As noted above, the Partnership's Agricultural Modeling Subcommittee will be providing the data and methods on fertilizer and manure applications for the Phase 6 CBWM. The point of view expressed in this report is well represented in the AMS and regional data will likely be used where available.

Consider changing weather patterns associated with climate change.

CBP RESPONSE: The MWG and WQGIT are committed to evaluating the effects of climate change as part of the Chesapeake Bay TMDL 2017 Midpoint Assessment. It is expected that the hydrology and sediment simulation in HSPF, combined with the mechanistic simulation of phosphorus in APLE, will provide the CBP partners with estimates of phosphorus loading responses to climate change.

Better represent and report uncertainty in data sources and model output.

CBP RESPONSE: As noted in the report, this is a recurrent request of STAC and is a long-standing goal of the MWG. The WQGIT requested an evaluation of uncertainty at its October 2014 meeting. Currently, the MWG anticipates working with uncertainty estimation in 2016 and 2017.

Differentiate between surface and sub-surface transport pathways of P loss and account for the role of drainage intensity.

CBP RESPONSE: Subsurface transport of phosphorus, particularly in ditch and tile drain settings, is an area of active research. While currently there is not a mechanistic model available for these processes, the calibration of the Phase 6 CBWM will drive empirical estimates of these loadings.

Future Data Needs to Support Recommended Changes in Modeling Approach:

CBP RESPONSE: The CBP partnership appreciates receiving STAC's list of monitoring priorities. The timing of this list is fortuitous in that the three-stage Building And Sustaining Integrated Networks (BASIN) process is underway in the Integrated Monitoring Networks Work Group, under the oversight of the Partnership's Scientific, Technical, Assessment and Reporting (STAR) Team. It is understood that the list presented here is for assistance in setting future priorities by the CBP partners. Nonetheless, for the record please find updates below where the CBP partners have begun to collect these data sets.

Baseline soil P levels.

CBP RESPONSE: As noted above, work has already begun on gathering this important data set. Additional data access and a longer-term record would improve the modeling.

Information on P application methods.

Spatial and temporal data on manure application.

Inorganic P application rates, including those associated with high-value row and horticulture crops.

CBP RESPONSE: As noted above, the AMS is working to produce the most accurate dataset for the Phase 6 CBWM for recommendations 2, 3, and 4. This work highlights the need for better data availability for manure and fertilizer applications related to the four Rs.

More systematic storm water sampling in predominantly agricultural watersheds for use in model calibration.

CBP RESPONSE: Since 2005, the CBP partnership's watershed water quality monitoring network has required storm samples at all primary monitoring locations, including many that drain watershed which are principally agricultural lands. These stations, now a decade or more into their improved datasets, are beginning to inform load and trend estimates used for model calibration. Additionally, the three USGS-USDA 'showcase' watersheds started in 2010, combining intensive conservation practices and monitoring, are beginning to show results.

Improved mapping of features that restrict water infiltration and promote "saturation excess" runoff.

Improved mapping of drainage intensity as an indicator of hydrologic connectivity and P delivery potential.

CBP RESPONSE: The CBP partnership is now considering comprehensive, consistent basinwide LIDAR mapping that may provide useful elevation data for recommendations 6 and 7 if implemented. The MWG and CBPO Modeling teams are also working as collaborators on NSF-funded research that includes identifying and scaling up hydrologic landscape properties that affect the transport of phosphorus.

13.3.2.2 Nutrient Inputs

Yagow, G., A. Collick, M. Ribaudo, W. Thomason, T. Veith. 2016. Scientific and Technical Advisory Committee Review of Nutrient Input Estimation for the Chesapeake Bay Watershed Model. STAC Publication Number 16-005, Edgewater, MD. 46pp

A STAC review of the nutrient inputs was requested by the CBP in 2016. The review panel was charged with reviewing the appropriateness of methods used to estimate total manure and fertilizer application for each county in the Chesapeake watershed. In keeping with STAC protocol, the reviewers ere also asked to comment on the sufficiency of the documentation and to make suggestions for future work.

Below are the detailed responses from the Chesapeake Bay Program Partnership to STAC's recommendations and comments

Agricultural Modeling Subcommittee Approval: 11/18/2016

Agriculture Workgroup Approval: 11/21/2016

Water Quality Goal Implementation Team Approval: 11/28/2016

Modeling Workgroup Approval: 12/29/2016

The numbered recommendations and comments listed below can be found on pages 2 through 7 of STAC's Review of Nutrient Input Estimates for the Chesapeake Bay Watershed Model (Yagow and others, 2016) **Detailed responses to the STAC peer review are set off from the numbered questions with bullet points and with bold font.**

1. We recommend that a comparative analysis of nutritive content at different stages of decomposition and application be performed across animal types to evaluate volatilization in the storage process and under typical best management practices.
    - **The final documentation will contain detailed figures, with hypothetical nutritive content values, describing the transformation of manure within the barnyard, pasture and field. This will include a description of the volatilization process, and best management practices which can impact volatilization and other transformations.**
2. The current “Scenario Builder” approach assumes zero nitrate nitrogen is available in animal manure. We recommend that this assumption be reviewed. Chastain et al. (2001) found that “the amount of nitrate-N can be significant in poultry litter that has gone through several ‘heats’ in a stacking shed. Manure stored in stacking sheds should be tested for nitrate-N in addition to ammonium-N and organic-N.” Nitrogen can transform speciation quickly, and while ammonium may be the most common form of inorganic N in manure, it can be oxidized rapidly to nitrate.
    - **Many of the values used in the Phase 6 Model are derived from laboratory analyses of manure and litter data. These labs often do not report a value for nitrate-N as it is assumed to be negligible and only measured on request The Partnership will consider including a value for nitrate-N when it is reported by laboratories across the region in the future.**
3. Mineralization of organic nutrients in manure transforms previously unavailable nutrients into forms that can be used for plant uptake. We recommend that the contribution of manure N mineralization from previous years’ applications be credited to the system and the current year crop N demand consequently reduced.
    - **The Phase 6 Model does assume multiple years of manure mineralization beginning in the 1990s and continuing until this day. It is not considered in the 1980s because the Partnership felt that nutrient management plans likely were not considering multiple years of mineralization during that time. The mineralized manure is assumed to be made available to the crop in the year of growth, decreasing the need for additional nutrients from other sources. The exact mineralization rates by year and animal type will be included in the final documentation.**
4. Information for separating manure into areas of deposition is provided by jurisdictions. We recommend that a review of inputs be conducted to ensure relative consistency among the states, and that any difference in values among states be explained and documented.
    - **While the initial information used to separate manure into areas was not thoroughly reviewed, the Partnership has since gathered representatives from state agricultural agencies around the watershed to conduct such a review. This effort resulted in much more consistent information across the states that will be provided in the final documentation.**
5. Direct deposition of manure on pasture is currently assumed to not meet any portion of the nutrient needs of pasture. While this may have made some sense historically prior to the widespread use of nutrient management, this assumption is now suspect. We recommend that some fraction of the direct deposition be counted towards meeting the “crop needs” of the

pasture, perhaps evaluated from recent credits given to this source in nutrient management plans based on P. In this regard, and although the associated documentation was not provided as part of this review, we have observed from results of the alternative Beta3A scenario that many counties already have amounts of direct deposition on pasture in excess of their "crop needs," without any applications from any other source.

- **"Crop need" as referenced in this recommendation is best characterized in the Phase 6 Model as the pounds of nutrients per acre or yield unit that a crop should receive above and beyond any direct deposition. For example, the Maryland Department of Agriculture's Annual Implementation Reports indicate that producers apply an average of 15 pounds of nitrogen per acre of pasture in addition to direct deposition. This value accounts for many pastures receiving nothing in addition to direct deposition while other pastures receive well over 15 pounds of additional nitrogen. The Partnership elected to use this value as the "crop need" for pasture, understanding that direct deposition would occur regardless of the additional applications because this would be approximate the total applications to pasture beyond direct deposition.**
- **The Partnership also feels that much of the direct deposition on pasture would not be available for plant uptake if this calculation was needed for the model because direct deposition of manure and urine on continuously grazed pastures often results in very low per acre N availability values – resulting in N deficient situations for large areas of the pasture.**

6. We recommend that efforts to continually refine the values of manure nutrients in storage be encouraged in order to improve national excreted N and P values.
    - **Agreed. The Partnership is currently collecting manure and nutrient concentration data from swine producers and litter and nutrient concentration data from poultry producers to better characterize the quantity of manure nutrients from these animal industries. The Partnership would also welcome efforts from the cattle and dairy industry to collect similar information in the future.**
7. Since storage and handling losses are some of the most well-documented and understood losses, we recommend calculating storage and handling losses (and perhaps nutrient reductions due to feed additives), before calculating ammonia losses.
    - **The Partnership recently approved this approach for calculating ammonium losses, and the final documentation will reflect this decision. Literature values for total nitrogen losses from excretion to the field (cradle-to-grave) were combined with literature values for ammonium conservation in an effort to characterize ammonium and total nitrogen loss through the system.**
8. The current use of crop application curves to allocate excess manure ignores the fact that for many farms, manure is actually spread first on land operated by the facility. On most farms, vegetables are not likely to be grown in conjunction with those crops most associated with animal operations, such as corn, soybeans and alfalfa. We recommend that the crop application curve procedure be modified with the Economic Research Service's analyses of Census of

Agriculture data to incorporate consideration of crops grown within animal operation systems in its analysis of manure distribution within the Chesapeake Bay watershed.

- **Since the Model operates at the county level, the farm level specificity suggested by the reviewer would be difficult to incorporate. The application curves referenced in this recommendation are used to prioritize applications to certain crops. However, the total manure applications to each crop are based upon a combination of: acres of that crop within a county; average yield for that crop; and percent of application that can be made by organic sources. While the Partnership agreed to prioritize applications to vegetables and grains at a similar level, the combination of these other factors result in very little manure being applied to vegetables. In the most recent Phase 6 Model results, only 3.5 percent of all manure was applied to "specialty crops" – a group that includes vegetables. Conversely, 64 percent of manure was applied to grains.**

9. As a further modification of the crop application curves, we recommend considering BMPs, such as New York's Precision Feed and Forage Management Practice, which prioritizes hay over corn. This procedure minimized farm inputs by balancing crop nutrient needs and dollars spent (Ghebremichael et al. 2007). Implementation and nutrient reduction data from the last 10 years may provide additional allocation mechanisms for excess manure.
   - **Corn is most commonly the highest priority to receive nutrient application throughout the watershed, but individual jurisdictions can indicate other nutrient application priorities for their states as appropriate. Agricultural experts from New York are currently reviewing the latest Phase 6 Model results to determine if applications on hay and corn match their understanding of production systems in their state. Adjustments may be made to application calculations if warranted following this review.**
10. We recommend that the prioritization process of manure applications to crops be revisited and vetted with experts within each state. Transparency should be ensured through the use of written justification for recommended values.
    - **The combination of crop application curves with crop yield and acreage data provides final manure application estimates. Agricultural experts from each state within the watershed are currently reviewing the latest Phase 6 Model results to determine if applications match their understanding of production systems in their state. Adjustments may be made if necessary following review. STAC is encouraged to join in this review. Members of the AMS will provide justification for the prioritization process in the final documentation.**
11. We recommend that the rate of allowable application above the application goal be reviewed by an appropriate group of regional experts. The justification for how these values were established should be documented and the decision substantiated.
    - **The Phase 6 Model has no "rate of allowable application above the application goal." It also has no minimum rate of allowable application other than zero. The tool was designed in such a way that applications could increase or decrease in a smooth, predictable, prioritized fashion across all crops within a county if estimates of**

available manure and fertilizer increased or decreased. If for example, poultry populations increased 10-fold, the tool would not limit the application of poultry litter in that county. As mentioned in previous replies, STAC is encouraged to join in the review of the current Phase 6 Model results to help determine if rates are reasonable.

12. The Association of American Plant Food Control Officials (AAPFCO) data provides information on the fertilizer form (urea, DAP, etc.) and the amount sold (tons). We recommend using these data to calculate the actual ratio of ammonia and nitrate N in the fertilizer sold (e.g., urea=100% ammonia; UAN=50% ammonia and 50% nitrate) as a baseline and/or as a check against the 75:25 assumption.
    - **The Partnership recently completed the analysis suggested by STAC for tons of fertilizer sold in 2012. The results of this analysis indicated that 77% of nitrogen sold was ammonium nitrogen with 23% being nitrate nitrogen. Given the results of this analysis, the Partnership feels the 75/25 split is still appropriate. A description of this analysis along with the assumptions for ammonium and nitrate in each grade will be included in the final documentation.**
13. Although hard data are not available to decisively argue for a plateauing relationship at higher yields versus the linear relationships used in the Phase 6 (P6) equations for crop application goals, it is known that major crops cannot always achieve maximum yield. For example, water stress (even in high-clay soils) may restrict corn growth despite an abundance of nutrients. Thus applying more nutrients under the assumption that maximum yields can be approached is not realistic and may result in increased nutrient loss. We therefore recommend that further research be explored to identify and quantify these plateauing breakpoints to further refine nutrient distribution guidelines and procedures.
    - **The level of crop physiology modeling needed to address the interaction between genetic capacity of the plant and environmental conditions that would likely be difficult to program into the Phase 6 Model. With that said, the Phase 6 Model bases applications not upon maximum yields, but upon average, countywide yields provided by USDA NASS. These yields do not yet approach literature values for maximum yields, so the plateauing of applications is not needed for most scenarios. However, if partners wished to research alternative future scenarios with very high yields, a plateau of applications may be needed.**
14. We recommend that the list of major field crops subject to double-cropping be revisited, as it appears that there are currently too many land use categories included under "Major Field Cropland Harvested Area," whose acreage will never be used to double-crop the crops listed under "Major Field Crops." This would lead to an under-estimation of the amount of identified double-cropped acreage. A suggested list of crop types to be considered under "Major Field Cropland Harvested Area" (for potential double-crop) would only include: i) alfalfa hay; vi) other managed hay; viii) small grain hay; x) wild hay; xiv) field and grass seed crops; xvii) canola; and xviii) popcorn.
    - **The Partnership recently approved a new method to calculate double-cropped acres which follows similar logic as recommended by the reviewers. First, harvested cropland acres are compared to acres of harvested grains, hay, vegetables, orchards and area of greenhouses. The difference becomes the double cropped acres. This difference varies by region with larger differences in the Coastal Plain where double-**

cropping is common, and far fewer acres in New York, where double-cropping is only common about once every few years in an alfalfa or silage cropping routine. Secondly, agricultural experts from each state provided a list of individual crops which would be harvested in the spring and a second list of those harvested in the summer or fall during a double-cropping year. This allows the Phase 6 Model to proportion individual crops into the double crop acreages originally determined. The results again, vary by region. For example, most double-cropped acres in the Coastal Plain are a combination of spring or summer harvesting of small grains and fall harvesting of soybeans. In contrast, most acres in New York are a spring or summer harvesting of small grains haylage with a follow-up planting of alfalfa hay or corn silage. This process will be described in the final documentation.

15. We recommend that the effect of climate change on agricultural forecasting be considered in future forecasting efforts, since climate change may impact crop selection and nutrient inputs.
    - **The Phase 6 Model will be used primarily to project the impact of anthropogenic changes to the landscape (including acreages and yields of crops and numbers of animals produced) through 2025, which is a shorter time period than can typically be addressed with climate prediction models. Through follow-up conversations with STAC review members, it was agreed that the current methods for projecting crop yields, acres and animal numbers by placing a large emphasis upon more recent trends will likely accommodate any changes due to climate over the short term forecasting period out to 2025. However, the Partnership will consult STAC if there is a need to estimate changes in agriculture past 2025.**
16. We recommend that the alpha and beta weighting factors of 0.8 and 0.2 be reevaluated, and that this be based on the agricultural acres and changes in cropland (or individual crop), and not validated using the poultry or cattle data. Trends in crop acreage cannot be expected to follow trends in animal numbers. In fact, under an assumption that crop acres respond to commodity prices, high feed grain prices would be expected to stimulate more feed grain acreages but reduced animal numbers, as animal production is adversely affected by feed grain prices. The hypothetical example in Table 5-3 shows an approximate 30% error rate compared to actual reported acres (Yt) for Aft from 1987 to 2012. If this large discrepancy exists for other crop types/uses, a new strategy is needed. One potential source of forecasting data would be past, current, and future National Agricultural Statistics Service's (NASS) Cropland Data Layer (CDL) maps to look at apparent trends in crop groupings and rotations. Although CDL maps provide less breakdown of specialty crops than the Ag Census, the CDL groupings are at a level more related to the groupings used by Scenario Builder. CDL appears to show year to year fluctuations in overall cropping selection choices (possible relationships to economics, weather and crop damaging insects). Whatever data sources are eventually used to establish the weighting factors, they should be validated, and the hypothesized cause and effect relationship justified.
    - **The Partnership recently conducted an analysis of the beta and alpha weighting factors for the following groups of agricultural data as reported by the Census of Agriculture: Harvested Cropland Area, Pasture, Cattle, Swine and Poultry. The objective of the analysis was to determine which combination of beta and alpha factors would provide: 1) the smallest residual from the actual reported value across the entire watershed in 2012; and 2) the smallest root mean square error across all counties within the watershed in 2012. The analysis resulted in the alpha and beta factors included in the table below. These will be used in future Phase 6 forecasts of agricultural inputs, and the methods for selecting these factors will be described in the final documentation.**

Revised Alpha and Beta Weighting Factors for Agricultural Projections

Category	Alpha Factor	Beta Factor
Harvested Cropland	0.6	0.4
Pasture	0.6	0.4
Cattle	0.6	0.4
Swine	0.8	0.2
Poultry	0.8	0.2

17. We strongly recommend that, in order to improve the transparency of the documentation, references and citations for all of the assumptions and values used in the process (including sources of data used in figures) be included.
    - **The reviewers pointed out many general, and specific, areas where the documentation could be improved. These improvements will be made for the final version. Any decisions made using best professional judgment due to limited information from literature will also be described in the final documentation.**
18. Concerns were previously raised about the relative evaluation of the Revised Universal Soil Loss Equation, Version 2 (RUSLE2) C sub-factors for the "plow" scenario among states by the Ag Loading Rate Steering Committee (ALULRSC) in its December 17, 2015 report: "Although our Sub-group endorsed the use of RUSLE2 for generating sediment loads and relative loading ratios, our review revealed inconsistencies in the range of sub-factor values evaluated between states and crop management zones (CMZs),…[and] our Sub-group strongly felt that these inconsistencies must be addressed before the relative loading rates as used by Scenario Builder will be a valid representation of erosion rates among states and CMZs." An example of this variability was provided in Table 4 from the ALULRSC report for the "crop residue %" sub-factor for the pasture/range land use. While the sub-factor values were independently estimated by National Resources Conservation Service (NRCS) personnel in each state, no attempt was made to normalize them so that they would be comparable across states. The results were highly variable sub-factor values across states. We support and reiterate the previous recommendation by the ALULRSC to provide a cross-state review of all RUSLE2 C sub-factor values which are used to establish relative values of detached sediment by land use, especially noticeable in pasture.
    - **The Partnership is currently providing the RUSLE2 C sub-factors to an independent expert selected by STAC for review. The Partnership requests STAC provide any recommendations for changes to these values that result from that review.**
19. We recommend applying the buffer credit to the same upland land use from which the buffer is taken, rather than applying it to all agricultural land uses in the land-river segment. If the majority of buffers are related to livestock exclusion, it is incorrect and potentially misleading to apply the filtering credit to cropland.
    - **The Partnership's Watershed Technical Workgroup is currently reviewing this recommendation.**
20. The Agricultural Modeling Subcommittee – Pasture Subgroup (PSG) has proposed that all (100%) of manure deposited in riparian pasture areas be directly deposited (DD) to the stream. We recommend, instead, that DD be evaluated as a fraction of the manure load in riparian pasture areas consistent with the P532 loads from degraded riparian pasture areas. The PSG analysis in Appendix B indicates that 74% of the directly deposited total nitrogen (TN) load and 38% of the total phosphorus (TP) load would actually be comparable to the respective P532 loads. If the

PSG's recommended 100% of deposited manure is used, the livestock exclusion BMP in P6 would get bonus reductions of about 35.5% for TN and 163.8% for TP, over that credited under the P532 version of the model.

- **The Partnership reviewed existing TMDL models developed for VA which assessed the impact of riparian area deposition. Each model had an assumption of the amount of manure deposited within the riparian area that was actually deposited in the stream itself, representing 0 attenuation of the manure nutrients by the land. These assumptions varied, with many models assuming 100 percent. The Partnership has chosen to use the average of these models, which was approximately 70 percent, to represent the amount of manure deposited within a riparian area which has 0 attenuation. That leaves 30 percent of manure deposited within nearby riparian areas which should have some attenuation factor applied to it. Butler et. al, 2008 found that of the TN and TP from manure applied to simulated, heavy use riparian areas, approximately 33 percent and 34 percent respectively was exported. These findings and assumptions can be combined in the following manner to estimate manure runoff to streams:**
    - **TN Fraction Runoff from Riparian Pasture = 0.70 X 1 + 0.30 X 0.33, or 0.80**
    - **TP Fraction Runoff from Riparian Pasture = 0.70 X 1 + 0.30 X 0.34, or 0.80**

21. Suggested sources of additional technical data or scientific findings include:
    a. Nutrient management planners and the extensive non-proprietary data contained in nutrient management plans could be engaged by CBPO to refine model characterization of local farming practices, manure/fertilizer application rates and timing, and conservation techniques.
        - **The Partnership is encouraged by the aggregated, countywide results from recent Maryland Annual Implementation Reports (AIRs), in which farmers report** manure and fertilizer applications. Similar, future surveys collected by states across the watershed would provide more data to ground-truth and perhaps adjust modeling assumptions in the future.

    b. Results presented at the 2016 Soil and Water Conservation Society (SWCS) annual meeting which showed that a majority of farmers sought fertilizer rate and timing recommendations from local fertilizer sales companies and agribusinesses (Embertson 2016). These results deserve careful consideration, and recommendations of the sales companies and agribusinesses should be evaluated in relation to other means of estimation. In particular, fertilizer retailers could be surveyed to refine fertilizer sales by crop type.
        - **The Partnership recently began discussions with fertilizer retailers with the objective of improving fertilizer estimates in the future.**

    c. Updated manure production numbers by animal type would help improve manure nutrient load estimates.
        - **The Partnership requests that states provide aggregated laboratory analyses and litter volume analyses for poultry each year in an effort to improve model estimates.**
        - **The Partnership will soon complete an effort to re-characterize nutrients generated by swine and turkeys, and this information will be used in the final Phase 6 Model, and states will be asked to update this data in the future.**
        - **The Partnership would like to expand this effort to the cattle and dairy industries at some point in the future.**

d. Applied academic studies could be used to better support and/or inform nutrient and land management estimates, including a more comprehensive representation of rotational land management.
    - **The Partnership would be interested in developing future STAC workshops that could explore this recommendation further.**

e. Mineralized N from previous year manure applications should be incorporated into future versions of the Watershed Model.
    - **Mineralized nitrogen from previous years' manure applications will be considered in Version 6. Please see the response to question 3 for a more complete description.**

13.3.2.3 Watershed Model

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

The above review was conducted in 2017. The committee used draft documentation and three days of presentation and conversation with the CBP modeling team. The entire response document, which contains the review recommendations is reproduced below. The document was approved by the Modeling Workgroup on May 2 2018.

13.3.2.3.1 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.

13.3.2.3.2 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?*

13.3.2.3.2.1 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.

13.3.2.3.2.2 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.

13.3.2.3.2.3 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.

13.3.2.3.2.4 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.

13.3.2.3.2.5 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.

13.3.2.3.2.6 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.
-

13.3.2.3.2.7 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.

13.3.2.3.2.8 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.

13.3.2.3.2.9 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.

13.3.2.3.3 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?*

13.3.2.3.3.1 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:

13.3.2.3.3.2 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.

13.3.2.3.3.3 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.

13.3.2.3.3.4 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.

13.3.2.3.3.5 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.

13.3.2.3.3.6 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.

13.3.2.3.3.7 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.

13.3.2.3.3.8 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'.

13.3.2.3.3.9 Specific Comments Regarding Model Documentation (c-l)

Various recommendations for clarity

Response from CBP

Thank you! Changes made as suggested.

13.3.2.3.3.10 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.

13.3.2.3.4 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?*

13.3.2.3.4.1 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.

13.3.2.3.4.2 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.

13.3.2.3.4.3 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.

13.3.2.3.4.4 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.

13.3.2.3.5 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?*

13.3.2.3.5.1 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.

13.3.2.3.5.2 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.

13.3.2.3.5.3 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.

13.3.2.3.5.4 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.

13.3.2.3.5.5 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.

13.3.2.3.5.6 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.

13.3.2.3.5.7 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.

13.3.2.3.5.8 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.

13.3.2.3.5.9 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.

13.3.2.3.5.10 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.

13.3.2.3.5.11 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.

13.3.2.3.6 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?*

13.3.2.3.6.1 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.

13.3.2.3.6.2 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.

13.3.2.3.7 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?*

13.3.2.3.7.1 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.

13.3.2.3.7.2 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.

13.3.2.3.8 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).*

13.3.2.3.8.1 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.

13.3.2.3.8.2 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.

13.3.2.3.9 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?*

13.3.2.3.9.1 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.

13.3.2.3.9.2 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.

13.3.2.3.10 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?*

13.3.2.3.10.1 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.

13.3.2.3.10.2 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.

13.3.2.3.10.3 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.

13.3.2.3.10.4 Specific Comments Regarding Model Documentation (a-i)

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

13.3.2.3.11 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.*

13.3.2.3.11.1 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.

13.3.2.3.11.2 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.

13.3.2.3.11.3 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.

13.3.2.3.11.4 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.

13.3.2.3.11.5 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.

13.3.2.3.11.6 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.

13.3.2.3.11.7 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.

13.3.2.3.11.8 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.

13.3.2.3.12 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.*

13.3.2.3.12.1 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.

13.3.2.3.12.2 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.

13.3.2.3.12.3 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.

13.3.2.3.12.4 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.

13.3.2.3.12.5 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.

13.3.2.3.12.6 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.

13.3.2.3.13 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).*

13.3.2.3.13.1 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

13.3.2.3.13.2 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.

13.3.2.3.13.3 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.

13.3.2.3.13.4 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.

13.3.2.3.13.5 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.

13.3.2.3.13.6 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.

13.3.2.3.13.7 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.

13.3.2.3.13.8 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.