

TECHNICAL REVIEW OF LAKE CHAMPLAIN  
TMDL MODELING TOOLS  
(with Tetra Tech responses submitted for the  
Record in BLUE text )

APRIL 2015

Prepared for:

U.S. Environmental Protection Agency, Region I  
5 Post Office Square, Suite 100, Boston, MA

Prepared by:

Peter Shanahan, Ph.D., P.E., Bruce Jacobs, Ph.D., P.E., and Ken Hickey,  
HydroAnalysis, Inc.

Kenneth J. Wagner, Ph.D., CLM, Water Resources Services, Inc.

William H. Frost, P.E., D.WRE, KCI Technologies, Inc.

Paul H. Kirshen, Ph.D., Civil and Environmental Engineering Department and  
Institute for the Study of Earth, Oceans and Space, University of New Hampshire



481 Great Road, Suite 3  
Acton, Massachusetts 01720  
(978) 263-1092

## Contents

Task 1.1	BATHTUB Lake Model.....	3
Task 1.2	SWAT Watershed Model .....	6
Task 1.3	BMP Scenario Tool .....	8
Task 1.4	Climate Change Analyses.....	13
APPENDIX A – Detailed review of BATHTUB and Missisquoi Bay reports .....		15
	Review of BATHTUB Model report.....	15
	Review of Missisquoi Bay Model report.....	20
APPENDIX B – Detailed review of SWAT Model Configuration, Calibration and Validation report .....		22
APPENDIX C – Detailed review of NPS Scenario Tool .....		33
APPENDIX D – Detailed review of climate change analysis .....		38
	Review of LaPlatte River Watershed Pilot.....	38
	Lake Champlain Report.....	40

## **TECHNICAL REVIEW OF LAKE CHAMPLAIN TMDL MODELING TOOLS**

### **Background and limitations of this review**

WaterVision has completed a technical review of the following four modeling and analysis tools applied to support development of the Lake Champlain TMDL:

- BATHTUB Lake Model
- SWAT Watershed Model
- BMP Scenario Tool
- Climate Change Analyses

This technical review was completed by a panel of experts assembled specifically to address the four components of the review. Kenneth J. Wagner, Ph.D., CLM, of Water Resource Services, Inc. completed the review of the BATHTUB lake model; Peter Shanahan, Ph.D., P.E., and Bruce Jacobs, Ph.D., P.E., completed the review of the SWAT watershed model; William Frost, P.E., D.WRE, of KCI Technologies, Inc. completed the review of the BMP Scenario Tool; and Prof. Paul H. Kirshen, Ph.D., of the University of New Hampshire completed the review of the climate change analysis. Peter Shanahan and Ken Hickey coordinated the overall effort and compiled and edited this final report.

This review is not intended to serve as a “peer review” as defined by EPA guidelines. Rather, EPA has requested a “technical review.” This limited our purview and tasks somewhat. As a predicate for the review, we accepted the previously made choices of modeling tools and simply evaluated whether those tools had been applied soundly. For example, it would have been outside the scope of this review to evaluate whether the HSPF model should have been used instead of the SWAT model. Instead, we accepted the choice of the SWAT model and evaluated if the SWAT model was used in a manner consistent with the current state of knowledge and practice. We also did not review the details of the inputs into the models or of the steps taken to apply the models. Thus, we did not have the actual model input and output files to evaluate as a part of this review. Our review implicitly assumes that the mechanics of creating inputs, transferring files, etc., was correct and we did not “check” those or similar aspects of the modeling effort. We considered the data sources used and the nature of the information used in developing inputs, but did not do detailed review of the values of input data and parameters. We

concentrated instead on evaluating assumptions and procedures against a standard of scientific soundness consistent with the current state of knowledge and practice.

For each of the four modeling analysis tool reviews, a section is provided below that provides responses to specific inquiries posed by EPA. Also, detailed technical reviews for each of the four tools are provided in Appendices A through D. Lastly, and as a separate submittal, we provide marked-up versions of the BATHTUB lake model, SWAT watershed model, and climate change reports that capture additional observations and comments.

## Task 1.1 BATHTUB Lake Model

The BATHTUB model appears to have been appropriately set up and calibrated, albeit with assumptions and limitations, as is true of all modeling efforts. We specifically reviewed the formulation of the model and its calibration as presented in the calibration report. While knowledge of how the model is intended to be used informed our review, we did not review the model's application or the interpretation of the model's results. Responses to the specific inquiries posed by EPA are provided below and Appendix A provides a more detailed review of the BATHTUB and Missisquoi Bay reports. Marked-up versions of the reports will also be provided.

- 1. Review and comment on the approach used to set and calibrate diffusion and sedimentation rates. Are the calibrated rates reasonable values in relation to other successful modeling efforts and/or literature information?*

Overall, the approach appears reasonable, but this aspect of the model is probably its weakest link. The applied values for diffusion and sedimentation are believable and consistent with literature and other model applications, but the process by which diffusion values were selected was not well explained and variation suggested by the report is high enough to indicate that this could be a significant source of error.

Lake Champlain is a complex system which includes a broad variety of hydrodynamic environments ranging from narrow causeway channels where exchange flows are constricted, to wide and deep open water situations where extensive mixing would be expected. The BATHTUB model, which is a simple steady-state spatially segmented model, requires assignment of diffusive transport rates between segments of the lake.

Initially diffusive exchange was evaluated using the Fischer Equation coupled with segment-by-segment calibration factors. Optimization of the diffusion factors to match observed chloride data yielded extremely variable results even when the factors were constrained within the range of 0.01 to 10.0

(varied over 3-orders of magnitude). The extreme variability in coefficients was interpreted to suggest that the Fischer approach was not appropriate for portions of this lake as segmented. Hence, the exchange rates between lake segments were evaluated using a direct mass balance for chloride.

The calculated exchange values were quite sensitive to different inputs. One of the biggest reasons for the variability in exchange rates is the small differences in chloride concentrations between adjoining lake segments (which are represented in BATHTUB as control volume boxes of varying sizes, which only have one monitoring location within each) and the sensitivity of the calculated exchange rates to these concentration gradients. In addition it should be noted that the chloride concentrations were not at steady state in the lake between 1990 to 2010, with some significant increases occurring along the main stem of the lake. Of all the approaches evaluated, the calibration of the BATHTUB model using segment-by-segment direct estimation of the exchange rates provided the most direct approach and gave the best calibration, measured in terms of mean percent error and RMSE.

Despite uncertainty in estimating the exchange rates, this does not appear to be a significant source of uncertainty relative to the seasonally averaged nutrient concentrations that are the key management output of the model. Application of the calibrated exchange rates to the validation period resulted in a mean percent error for chlorides of less than 5% on a total lake basis. The spatial variability of chloride and TP concentrations in the lake is more strongly determined by advective fluxes relative to the locations of point source inputs than by diffusive exchanges between segments.

- 2. Review and comment on the model validation process conducted. Given the existing data set, comment on the model's capability to represent varying water quality conditions (i.e., annual average phosphorus concentrations) in response to varying hydrologic and pollutant loading conditions. Are the differences between predicted and observed average annual phosphorus concentrations adequately explained? How do these differences compare with other successful and similar lake modeling efforts (i.e., annual water and nutrient budget models)?*

Model validation appears successful. We did note a problem with the validation in that the predicted TP is substantially higher than observed TP. This may be an issue of the SWAT model not adequately simulating washout of TP from the watershed during a very wet period. Results are similar to other well-constructed BATHTUB model applications.

The results from the SWAT model were only used to configure the direct drainage areas. All other inputs for major/minor tributaries were based on observed data.

This discrepancy in TP during the validation period is largely during water years 1997 - 1998, which was a very wet period, and much of the TP entering

the lake from its tributaries was in particulate form. The global sedimentation rate calibration may not be wholly appropriate to this period. Following discharge to each lake segment, a larger fraction of this particulate phosphorus was likely lost to benthic sediments than predicted by the model, resulting in an attenuation of the response of observed lake TP levels to tributary TP loads. Again the overall percent error for the validation period was found to be ok i.e. <10% if we exclude this anomalous period. In addition it should be noted that although we evaluated the model for the validation period, only the calibration period was used for the basis of the TMDL.

3. *Was the LimnoTech phosphorus model for Missisquoi Bay appropriately substituted for the BATHTUB model structure for that lake segment? (Note that this was done by EPA/DEC after the Tetra Tech report was completed.)*

None of the reviewed documents indicate that the LimnoTech model results for Missisquoi Bay were included in any version of BATHTUB, but the results could indeed be used instead of the SWAT inputs applied in the Tetra Tech report on BATHTUB application. Whether BATHTUB or the LimnoTech model would be used to process in-lake loads in that lake segment may require further consideration, but either should be adequate.

No Tt response, this process will be explained in the TMDL report.

4. *Comment on the model's usefulness as structured and calibrated for determining phosphorus total loading capacities for Lake Champlain segments.*

The model should be quite useful for determining loading capacity and options for achieving needed reductions. There is enough variability in some aspects of the model (e.g., exchange rates, nonpoint source load estimates) that it will be important to quantify uncertainty for any actual values obtained, such as phosphorus concentrations resulting from specific management options, but the model appears to represent reality adequately and seems ready for scenario testing.

Error statistics and CV values on the discrepancies between modeled and observed results are stated along with calibration results (see Tables 15-21). These statistics provide an indication of the net uncertainty in predictions, which can be taken into account with respect to implementation.

5. *Does the final report provide adequate documentation of data analysis steps, modeling decisions, and model output to support its use in establishing phosphorus TMDLs for all of the lake segments in Lake Champlain? Identify any additional documentation needed in the final report.*

Model development and calibration steps appear justified, but the explanation is a bit thin in some places, most notably with regard to exchange rates among

lake segments. Some additional explanation has been provided in responses to questions, but these should be incorporated in the final report.

[Please refer to comment 1; additional explanations have been included in the report](#)

## Task 1.2 SWAT Watershed Model

As the responses below to the specific questions posed by EPA indicate, the SWAT model appears to have been appropriately set up and calibrated. That said, truly accurate prediction of phosphorus loads is impossible and the model predictions appear to be accurate only within an order of magnitude. This is consistent with other similar models. The report does not do a good job of acknowledging the potential errors in model predictions and how those potential errors should be considered by decision makers. We recommend additions to the report to address this issue in Appendix B, a detailed review of the SWAT Model Configuration, Calibration and Validation report. A marked-up version of the report is also being provided.

### *6. Review and comment on the approach used to set up and calibrate the SWAT model for the Lake Champlain Basin.*

The approach used to set up and calibrate the SWAT model appears to be basically sound. The model was constructed by relying on a wide variety of data from many sources. The reliance on actual and site-specific data rather than on generic or literature values is sound procedure and appears to have been done carefully and thoughtfully. The hydrologic component of the model was calibrated against observed stream flows and appears to have matched daily and monthly observed stream flows satisfactorily. Monthly averages of the model's predicted suspended solid and phosphorus loads were calibrated against estimates of monthly loads derived from stratified regression curves of load versus flow for each major watershed. All of the watershed areas with the exception of the Saranac River watershed exhibit a bias of overpredicting phosphorus and suspended solids loads at low flows and underpredicting phosphorus and suspended solid loads at high flows. This problem is not described in the report. Given the uncertainty in measuring loads and difficulty in predicting loads, perfect calibration is impossible. However, the report should be clear that biases exist and discuss the implications of such biases for decision making about NPS management. For example, many NPS management measures focus on containing the runoff from smaller storms. The model's high bias for smaller events indicates the potential to overpredict the effectiveness of such measures in reducing loads to the lake.

[Tetra Tech acknowledges that biases are likely present in the simulated loads. Steps were taken during the calibration/validation process to minimize these biases by examination of loads at various flow regimes. However, it is impossible](#)

to completely eliminate biases. The presence of bias is expected to have some impact on effectiveness of BMPs for NPS management. However, since the reduction strategies are based upon average annual loads over a long period of time, these biases are expected to have minimal impact on NPS management decision making.

7. *Review and comment on the model validation process conducted.*

The model validation process appears to have been conducted appropriately, but could have been extended further. Consistent with the approach advocated by Hassan (2004), we recommend that the report should cite all facets of the model evaluation that build confidence in the model predictions. This opportunity appears to have been missed in a number of places as discussed in the detailed review of the SWAT Calibration report.

Hassan (2004) identifies a number of important considerations for model validation including, prediction uncertainty, data and evaluation test diversity, reliance on objective measures and subjective judgment, and testing sub-models individually and in connection to one another.

While Tetra Tech did not account for all of the above considerations, objective and subjective evaluation of the model performance were performed. The model validation approaches adopted are widely accepted and have been implemented in similar watershed modeling studies.

8. *Comment on the model's capability to estimate annual phosphorus loadings from the identified source sectors at the HUC-8 scale within the Lake Champlain Basin in response to varying hydrologic and climatic conditions. Is the model as structured and calibrated appropriate for use in determining phosphorus loading rates from each major source category?*

Direct calibration of phosphorus loadings from particular source sectors is not possible since there are no practical means of attributing fractions of the measured phosphorus back to particular source categories within a given watershed. Given that limitation, addressing this issue is a question of whether the model developers have first of all successfully calibrated the phosphorus concentration for each watershed and secondly whether they have employed standard practices to make the best possible approximation of phosphorus derived from each source category. The first question, on calibration of the watersheds as a whole is answered by inspection of the individual appendices that in part describe the success of the model in simulating the phosphorus load. This question is addressed in our discussion of the individual appendices contained elsewhere in this review document.

On the question as to whether they have made every effort to accurately estimate phosphorus loadings, we note that they have categorized the watershed land uses in 28 categories and estimated the phosphorus loading from each land use category based on numerous literature references. Further, they document efforts



to refine those initial estimates based on consultation with stakeholders and agencies such as the NRCS, the Vermont AAFM, and the University of Vermont Extension Service. In particular, they have expended effort in refining the agricultural loads based on crop practices, soil drainability, and livestock production. It is beyond the scope of this review to comment on phosphorus loading rates associated with each land use, however we are for the most part impressed with the level of effort to refine the standard loading rates based on local conditions. Based on this observation, we believe the model's phosphorus loading rates from individual land uses are as good or better than standard practice for water quality models of this type.

No action necessary

9. *Does the final report provide adequate documentation of data analysis steps, modeling decisions, and model output to support the model's use in estimating annual phosphorus loads from the various source sectors within the Lake Champlain Basin?*

Overall, we thought the modeling work behind the SWAT Model Configuration, Calibration and Validation report was much better than the report itself. While we eventually reached a level of reasonable confidence in the model, the report made that far more difficult than it should have been. The report is often disorganized and much of the writing is difficult to follow. The level of detail is uneven: some important aspects get only cursory description while others have ample documentation. Even after follow-up responses from Tetra Tech, we are not fully certain of some aspects of the model construction. We have provided detailed review comments on the report and recommend that the report be revised, critically reviewed, and carefully edited before it is finalized.

Based on the reviewers' recommendations, the calibration report has been carefully reviewed and revised to address the concerns raised. Additional details have been added to parts of the document needing more explanation as noted by the reviewers. The overall structure of the document has also been revised to enhance readability.

### Task 1.3 BMP Scenario Tool

The BMP Scenario Tool has been designed to be a means for planners to test a suite of BMPs and determine if pollutant reduction targets can be met. There are two problems that should be addressed in order to meet this goal: complexity and lack of transparency.

**Complexity:** The BMP Scenario Tool as developed does not appear feasible for general use by watershed planners. It cannot be used without a significant amount of training and background understanding. In particular, there are far too many options and BMP choices for practical use. At a minimum, detail on slopes and

soils should be averaged so this can be used as a planning-level tool. Further, data entry is very cumbersome. Unless there is a method for data entry via an input table or batch processing, selecting BMPs one by one through drop-down menus is an infeasible approach. It would take hours or days to replicate the scenarios provided.

As a final check on the BMP Scenario Tool, we ran a test scenario, working through the lake TMDL and existing load summary spreadsheets as recommended in the Introduction worksheet. Our goal was to determine where in the watershed new BMPs should be focused, but the Scenario Tool was not effective in answering that question. The default scenarios provided in the Scenario Tool achieved much higher reductions than any alternatives we tested.

**Lack of Transparency:** All formulas, algorithms, and macros are undocumented. There is no way for the user to know how the calculations are done or to carry out quality control on the results. This is not unusual for models—in most cases the model calculations are complex and it is not feasible to deconstruct them completely to determine if results are reasonable. However, for quality-control purposes, this can be overcome by using models that are well established and have been checked and verified by the model developer (e.g. HSPF, SWMM). With a one-off spreadsheet model like the BMP Scenario Tool, this level of testing and confidence is not available.

[The care and thoroughness undertaken by the reviewers is appreciated. In particular, we appreciate the mostly positive and supportive comments regarding the Tool and its underlying functions. In addition, we think that some more problematic comments are largely driven by a misinterpretation of the purpose and audience of the Tool. While the Scenario Tool would seem to be an obvious application that could be used by local and regional planners to evaluate implementation activities under the TMDL, the objective of the Scenario Tool was in reality defined fairly narrowly by EPA as a tool that would assist EPA in determining whether there is Reasonable Assurance that the Lake Champlain TMDL allocations are likely to be met. Subsequent use by watershed planners is obviously an important ancillary benefit of the Tool but it was not developed as a tool for planners per se.](#)

[The intended audience for the Tool was also a small and targeted group of EPA staff and VT agency staff involved in evaluating BMP implementation scenarios](#)

against TMDL targets for the lake. This means that there was only a fairly limited user group who were targeted with one-on-one instruction as to use of the Tool. Due to the stated objectives of the Tool, EPA did not require broad scale user education.

The Tool is certainly complex as it includes landuse loading rates and totals from the SWAT watershed model of the entire Champlain Basin by segment and drainage area and it includes a large number of BMPs that were selected on the basis of two criteria. The first criteria was that the BMPs should be those that are expected to be widely utilized by managers across the Basin during implementation. The second criteria was that the BMP efficiency should be universally accepted by stakeholders and agencies. While the list of included BMPs is large it does not include all BMPs that can and will potentially be implemented in the basin to achieve the TMDL allocations. Note that the Tool does allow for applying BMPs to average soil and slope conditions as the reviewers indicate would be useful to planners, by selecting the 'ALL' option under soil and slope groups. To new users of the Tool, creating detailed implementation scenarios can be cumbersome; however with familiarity and practice, the speed of the process improves. For the purposes for which EPA created the Tool—to assist in the Reasonable Assurance Demonstration and identification of one or more workable allocation scenarios—it has proven adequate and useful.

We also feel that rather than there being as the reviewers deem, a lack of transparency, it is more fitting to acknowledge lack of documentation with respect to a broad audience of end users. The Tool itself is actually fully transparent and all data and calculations can be seen and reviewed by end users. The project scope did not include resources for a detailed user guide; rather the Design Document that was reviewed served as documentation of functions and operational details throughout development.

10. *Were the BMP efficiencies and other calculation assumptions appropriately described and documented for each sector of nonpoint sources (developed land, agricultural land, streambanks, unpaved roads, forest land)?*

For developed land, BMPs were listed in Table 4 along with the land uses, slopes, and soils to which they could be applied. Table 9 shows the efficiencies for the structural BMPs and Table 6, 7, and 8 showed efficiencies for nonstructural measures. These were appropriately documented.

Table 12 showed agricultural BMPs and their applicability to land uses, slopes, and soils. Efficiencies (Appendix D) were calculated using SWAT and there was no documentation on how they were derived. The procedure for estimating streambank erosion efficiencies was described in detail.

The reduction efficiency for unpaved roads was provided with no documentation, other than the description that the value of 50% “is being used as a placeholder efficiency until information becomes available from a University of Vermont/Lake Champlain Basin Program study currently under way.”

The reduction efficiency for forest practices was given as 5%, with an acknowledgement that there is limited monitoring data to support this or any other value. There is essentially no documentation for this number.

The Tool is being updated to include different levels of controls for harvest area and forest road BMP types with varying efficiency numbers from 5% to 70%. The documentation of which BMPs were linked to selected reduction percentages is being developed by EPA and is anticipated to be included in an appendix to the TMDL document. The Design Document was revised to explain this. The agriculture BMP derivation process is described on page 21 of the Scenario Tool Design Document (i.e., a baseline condition was run in SWAT along with multiple scenarios representing implementation of the specified BMP in order to obtain an estimate of the efficiency of the BMP relative to a baseline load).

11. *Does the final report and accompanying spreadsheet tool provide adequate documentation of data analysis steps and results?*

The documentation for the Scenario Tool describes the structure of the spreadsheet, analysis procedures, and some information on key parameters and their sources. In that respect, there is sufficient information on how to perform an analysis. There is no documentation on formulas, algorithms, or macros which are used to make the calculations.

The math behind the data analysis is pretty simple and the Tool uses standard Excel formulas such as lookup, addition, subtraction, and multiplication functions to process the watershed loading data based on the user’s selected watershed of interest and BMP information based on the selected BMP type. The macros were developed to automate the data processing steps under different user selected options. Any user with basic knowledge of Excel macros can read and understand the code developed for this Tool.

The backend calculations using those formulas, algorithms, or macros were verified to be accurate through a formal EPA review process of the Tool.

12. *Does the BMP Scenario Tool provide a reasonable means of estimating potential phosphorus reductions at the HUC-8 scale associated with a range of BMP scenarios?*

With some revisions as described in the summary, the Scenario Tool could be a reasonable means for assessing pollutant reductions with BMP scenarios. As it stands, data entry is cumbersome and there are too many options which may not be significant at the planning level of detail. With respect to the question of whether the tool represents BMPs accurately, the lack of transparency noted above makes that question unanswerable.

[The overall response to the initial comments also applies to this comment.](#)

10. *Were the BMP efficiencies and other calculation assumptions appropriately described and documented for each sector of nonpoint sources (developed land, agricultural land, streambanks, unpaved roads, forest land)?*

For developed land, BMPs were listed in Table 4 along with the land uses, slopes, and soils to which they could be applied. Table 9 shows the efficiencies for the structural BMPs and Table 6, 7, and 8 showed efficiencies for nonstructural measures. These were appropriately documented.

Table 12 showed agricultural BMPs and their applicability to land uses, slopes, and soils. Efficiencies (Appendix D) were calculated using SWAT and there was no documentation on how they were derived. The procedure for estimating streambank erosion efficiencies was described in detail.

[The agriculture BMP derivation process is described on page 21 of the Scenario Tool Design Document \(i.e., a baseline condition was run in SWAT along with multiple scenarios representing implementation of the specified BMP in order to obtain an estimate of the efficiency of the BMP relative to a baseline load\).](#)

The reduction efficiency for unpaved roads was provided with no documentation, other than the description that the value of 50% “is being used as a placeholder efficiency until information becomes available from a University of Vermont/Lake Champlain Basin Program study currently under way.”

[Comment noted.](#)

The reduction efficiency for forest practices was given as 5%, with an acknowledgement that there is limited monitoring data to support this or any other value. There is essentially no documentation for this number.

[Comment noted. Additional efficiencies are being added to the Tool for forest practices, and EPA is conducting an additional analysis of forest practice reduction efficiencies. Documentation of this analysis will be provided separately in an appendix to the TMDL document.](#)

13. *Does the final report and accompanying spreadsheet tool provide adequate documentation of data analysis steps and results?*

The documentation for the Scenario Tool describes the structure of the spreadsheet, analysis procedures, and some information on key parameters

and their sources. In that respect, there is sufficient information on how to perform an analysis. There is no documentation on formulas, algorithms, or macros which are used to make the calculations.

14. *Does the BMP Scenario Tool provide a reasonable means of estimating potential phosphorus reductions at the HUC-8 scale associated with a range of BMP scenarios?*

With some revisions as described in the summary, the Scenario Tool could be a reasonable means for assessing pollutant reductions with BMP scenarios. As it stands, data entry is cumbersome and there are too many options which may not be significant at the planning level of detail. With respect to the question of whether the tool represents BMPs accurately, the lack of transparency noted above makes that question unanswerable.

## Task 1.4 Climate Change Analyses

15. *Please comment on the overall organization, clarity, and general effectiveness of both reports. Is it clear what was done, why it was done, and what was learned? If not, how can the organization of the reports be improved?*

Both the LaPlatte River and the Lake Champlain reports follow the same format. It is a rather standard, acceptable format for reports on the impacts of climate change on streamflow and water quality and is very reasonable. With the few exceptions noted in the detailed analysis of each report, the reports were clear in terms of what was done, why, and the results. Two of the major exceptions on clarity are the description of how the weather generator functions and the description of the procedure to attempt to capture intensification of precipitation under climate change. No re-organization of the report is necessary.

No comment

*Please comment on the technical quality of the reports. Are the methods appropriate to the goals of the reports? Are the results and conclusions justified and adequately qualified where necessary?*

Overall, the technical quality is high and the level of detail in downscaling and applying the results to the models is appropriate. As noted in the detailed comments on each report, there are some areas of possible methodological improvement and some areas where the results are not as expected. One major area of possible improvement is the use of more emission scenarios to define better the range of possible climate changes because estimated impacts of climate change from different models markedly diverge after or even before 2050. This is particularly important if an adaptation program is to be developed. A second area of possible improvement is that more data could have been used for the SWAT calibration and verification. Some results that require more explanation include why there was not more of shift in monthly flow variation of the LaPlatte River (e.g. Fig 4-12 of LaPlatte River report) and why there are

differences between Table 12 of the Lake Champlain report and Table 4-6 of LaPlatte Report on impacts on the LaPlatte River.

These differences are attributable to the different versions of the SWAT model that were used for each effort. See more detailed explanation in the response the detailed comments on the climate change reports, in Appendix D, below.

14. *Do you have other comments, concerns or suggestions for improving the quality of these reports?*

See detailed comments in Appendix D.

### **Cited Reference**

Hassan, A. E., 2004. Validation of Numerical Ground Water Models Used to Guide Decision Making. *Ground Water*. Vol. 42, No. 2, Pg. 277-290. March-April 2004.

## APPENDIX A – Detailed review of BATHTUB and Missisquoi Bay reports

This detailed review was provided by Kenneth J. Wagner, Ph.D., CLM, Water Resources Manager at Water Resources Services, Inc., 144 Crane Hill Road, Wilbraham, MA 01095.

### Review of BATHTUB Model report

The BATHTUB model appears appropriately set up and calibrated, albeit with assumptions and limitations, which is true of all modeling efforts. The key lies in understanding the constraints introduced by those assumptions and limitations, and interpreting the results of the modeling within a management context that does not go beyond what the model can support. This review covers only the calibration of the model, not its application and interpretation of results, but knowledge of how the model is intended to be used informs this review. The following issues are noted, along with any resolution through my analysis or responses from the modelers.

1. The approach to splitting point and non-point source inputs while estimating total loading from drainage areas with both source types was to subtract known point source loads upstream of monitoring points on tributaries, scale the remaining tributary nonpoint source loads to the whole drainage area, then add point source inputs back in. This induces two sources of error: lack of attenuation of point source loading and assumption of nonpoint source inputs at the same per unit area level between monitored and unmonitored portions of the watersheds. Posed as a question to the modelers, we have been informed that most point sources are near the downstream terminus of the streams and that the water quality data are from the mouth of the stream, minimizing the impact of shortcomings in the point source/nonpoint source loading split and lack of point source attenuation. This appears satisfactory, and in reviewing the available data sources, I see no case in which strongly erroneous estimates would result.

Claims of consistency with the SPARROW model and the policy of Vermont for addressing point source attenuation do not satisfactorily justify lack of attenuation, which can be significant in a free flowing stream. Yet it does appear that the assumptions made in handling point and nonpoint source estimates are reasonable in this case. Considering the relatively smaller magnitude of point source loads, the amount of error induced by assumptions does appear small.

Ok.

2. The approach to point sources assumes no change in loading to WTPs or treatment technology upgrades since 1991, the most recent time of applied point source data. I completed a review for another client in 2003 of the value of further restrictions on P content in possible sources, and had in that instance used the Champlain basin WTPs as an example. It appears that there has been little change in sources or treatment between 1991 and 2004. Further source restrictions were not expected to yield more than a 1% decline in P output. Treatment upgrades were not evident. Another 10 years has passed, and some changes may have occurred, but the applied flows and concentrations appear reasonable.



The existing model uses WWTP data from 1991 through 2010 which is ongoing. Average flows and loads for each of the years based on available data were used and incorporated in the BATHTUB model. Any treatment changes or source upgrades, including during the period from 1991 through 2004 (when many plant upgrades occurred) are reflected in the WWTP data specified in the model.

3. Applied units are not consistent in the report, making it less simple to compare across tables and graphs. I did calculate average TP concentrations in WTP discharges, then found separate data later in Appendix B that agreed within rounding error, which was comforting.

[Comment noted; we have included some clarification in the report related to unusual units used by the model \(e.g.,  \$hm^3\$ \)](#)

4. The time step over which flows or concentrations are averaged can be important when dealing with nonpoint sources. Average flow X average concentration usually underestimates the quantity obtained from a summation of all flows X corresponding concentrations, since the highest flows tend to correspond to the highest concentrations, and this would seem supported by Appendix A. The amount of error induced by using averages, or predicting loads from average flows using regressions, is not immediately apparent in the report. Error terms are developed through comparison of predicted vs. observed values, albeit for no less than biannual periods. As the model is made for steady state conditions, this seems appropriate, and for the purposes of scenario testing, may be adequate. Just bear in mind that the impact of some management actions may not be precisely captured by averaging loads.

[We agree](#)

The relation between flow and concentration shown in virtually all graphs in Appendix A is very strong and indicative of dominance by non-point sources. WTP inputs appear less influential, as the relative loads appear low. If nothing else, this should influence the management actions invoked in the scenarios to be tested. For example, no WTP has both high discharge volume and very high TP concentration; discharge is unlikely to be significantly reduced, and tightening effluent concentrations looks to provide minimal improvement on the scale of loads to each lake segment. An important conclusion of the modeling effort appears to be that nonpoint sources must be addressed to make a difference in the lake.

[Yes, we agree. This lends further support to the notion that implementation and management efforts should be crafted on a segment by segment basis to take into account the specific and unique sources contributing to each.](#)

5. There are a few annoying aspects of the report, including variable spellings of phosphorus and lack of differentiation between observed and predicted values in some text, but it is logically organized and fairly easy to read. There is tendency to describe discrepancies between different approaches as shortcomings of the data, rather than shortcomings of the model, a common issue with model development discussions. Both may be at fault. There is adequate emphasis on the use of the model to look at relative changes, rather than claiming that there will be accurate prediction of values. This may create some issues where a specific target value exists, and it will be important to put error bars around any predictions used to assess compliance with thresholds.

[Comment noted. We have made efforts to increase consistency in the text.](#)

6. Atmospheric inputs of TP and CL are based on limited and old data (2 years, ending in early 1992). Given the success of the Clean Air Act in altering atmospheric deposition, the assumption that inputs from the air remain constant over time seemed questionable. As atmospheric inputs are probably very low compared to nonpoint source runoff, this may be unimportant, but some comparison of loading magnitudes would have been illuminating. In response to this observation, the modelers note that inputs of phosphorus from precipitation represent less than 0.5% of the total, and that an analysis of calcium in air as a surrogate for phosphorus suggests no major change over time. Any related error appears insignificant.

[Comment noted.](#)

7. Table 9 lists the Weed Fish Culture Facility with a “?” after it. The modelers responded that this was an internal note that was resolved, and has no bearing on model review.

[Comment noted.](#)

8. The inclusion of the Burlington CSO as a point source appears appropriate, but there could be complications in the scenario testing since point and non-point elements were combined and SWAT–derived estimates of watershed loading did not include the area that contributes to the CSO. It was uncertain how CSO separation would be modeled in scenario testing, but the modelers responded that the CSO has its own “jurisdiction” in the model, and that BMPs can be tested within the CSO area. It appears that this issue is resolved.

[Yes this is a correct interpretation.](#)

10. Scaling flows or loads based on differential watershed areas introduces error, but the amount of scaling done is not extreme. With regard to flow, it would have been helpful to see a comparison of known water level changes with  $[(\text{Inflow} - \text{outflow})/\text{lake area} + \text{direct precipitation} - \text{evaporation}]$ ; it appears that the data exist to check overall flow estimates against water level changes, and this provides a convenient confirmation of one aspect of the model. With regard to loads, the decision to use loads generated by SWAT for unmonitored drainage areas rather than scaling from monitored watersheds is appropriate and commended, but it would have been helpful to provide a comparison of values from these two alternative methods.

[Comment noted. The comparison values were not presented in the report in an effort to keep the report as clear and concise as possible and we felt it most beneficial to simply include the final setup parameters.](#)

11. The explanation of how exchange rates among lake segments were derived was confusing and inadequate in the calibration report. This is a key part of the model and one that requires more careful explanation. Response to related questions indicates that calibration of diffusion rates based on mass balance analysis of chloride using data from 2001-2010 resulted in negative values for 3 of 13 segments. The modelers suggest that this was a function of shortcomings of the data. Use of the Fisher model necessitated some constraining of values, and the

results were not considered optimal. Rates calculated from 1990-1992 data appeared most believable; at least they contained no negative values. The relationship between calculated exchange rate and the cross sectional area of exchange was relatively consistent among approaches and was considered sufficient to serve as the basis for selecting exchange rates. The explanation and theory behind this approach appear rational, but the degree of variability in the relationship of exchange rate and exchange area (log-log plot in Figure 3 of the calibration report, as much as an order of magnitude difference between comparable estimates in a table supplied in response by the modelers) is high enough to suggest that this could be a significant source of error. Getting the physical exchange rate among segments correct should be easier than accounting for all the variation in loading of non-conservative substances, and the issues encountered in this aspect of model calibration suggest that this may be the weakest part of the model.

[Please refer back to our response to Comment 1 in the main body of this report. Ultimately it should be noted that the exchange rates were verified through the model validation process and were within thresholds required by the modeling QAPP](#)

12. TP/DP ratio can be important to settling rates and there are options for addressing this. Consideration was given to multiple approaches and no major difference was obtained in calibration based on choice. It was appropriate that this was checked and improves confidence in model results.

[We agree.](#)

13. The choice of release rate for TP from sediment in St. Albans Bay is rational and believable. It was a little surprising that this was the only segment that appeared to need such an adjustment, as at least Missisquoi Bay has similar issues as I understand it. Application of the LimnoTech model to Missisquoi Bay may have solved this issue, but is not discussed in the report.

[Subsequent use of the Limnotech model addressed sediment flux issues in Missisquoi Bay](#)

14. Algal vs. non-algal turbidity is an important concept in such modeling, and has been properly handled from my perspective.

[Comment noted.](#)

15. The error terms and R2 values for observed vs. predicted values are quite good. CV values are also low except for St. Albans, which is not really high. There is no indication of variation that will grossly compromise target setting and scenario testing. The only “bad” fit comes in the 1996-97 time period, when the observed values are substantially less than the predicted values for TP. This could be a wash out phenomenon, whereby in reality the available TP on the watershed becomes depleted, but the SWAT model just keeps generating and releasing TP. This problem has been encountered with SWAT before, and can be corrected if that is the problem. Additional effort may not be warranted, and would be part of the SWAT modeling if any adjustment was made.

[Note that FLUX estimates were used to generate these loads as opposed to SWAT. SWAT was only used to configure the direct drainage area loads.](#)

16. Just before the conclusion, the statement is made that “Results suggest that differences between these tributary TP load estimation methods yield relatively small differences in BATHTUB model predicted TP levels, within individual lake segments.” This statement is further described in the conclusions, noting some significant differences in tributary loads among methods employed, but no major differences in the resulting in-lake values when applied to the model. This could be interpreted to mean that lake processes even out those differences in a steady state model or that the model is not very sensitive. This may be important to scenario testing, as huge changes that are unrealistic in a management sense may be needed to reach goals, and it will matter greatly if that is a real situation or a model limitation.

The response to this observation by the modelers indicates that the statement reflects the similarity of results among models representing the same time period and lake segment, not that the model is not sensitive to changes in tributary loading. That the various methods of estimating loads produces similar results in the lake is comforting, if the differences among loading methods are far smaller than projected differences in loading with possible management actions. We are told that the scenario testing evaluation by EPA indicates sensitivity to loading changes, but we do not have any data or results to evaluate in that regard.

[EPA is expecting that the loading changes that will be necessary to reach target criteria in the lake will need to be significant and on a fairly broad scale and the model is sensitive to such changes in loads. In other words, a small change in one part of a watershed \(e.g. implementation of a handful of BMPs in one sub-basin\) will likely not affect the lake conditions. Likewise small changes in modeled loads to small segments of the lake do not affect the BATHTUB model results significantly. However, the model does demonstrate sensitivity to load changes that would be on the scale of necessary changes under the TMDL. For example a 10 percent reduction of incoming loads from the drainages to Otter Creek is predicted to reduce the P concentration in that segment from approximately 16.25 µg/L to 13.5 µg/L.](#)

17. Table 22 of the calibration report lists loading to each lake segment and gives a total. But as each lake segment provides input into another lake segment, it appears that the total given at the bottom of each column is an overestimate of what is going into Lake Champlain. In other words, the table seems to provide the load experienced by each lake segment independently, but those loads cannot be added to yield an accurate total load. It would seem better to express the load from each area (QC, VT, NY) as a percent of the total, rather than an actual load. It also seems that these data would be better represented as a figure, showing the total load for each segment in the spatial arrangement of Figure 1.

In response, the modelers indicate that the loads are listed by segment and jurisdiction and are independent of each other. That is, they do not take into account the load that a segment experiences from another segment. This would seem to verify the appraisal above, and that the listed total loads are overestimates. This summary table remains confusing.

The purpose of the table is to provide information on inputs to the lake segments from the various jurisdictions. Interactions between segments are taken into account in the model. There is no overestimation of inputs.

18. Simplifying the model might be possible, as many influences seem negligible to minor. Inclusion of all influences for which data properly support model development is fine if it doesn't introduce error that interferes with interpretation of results during scenario testing. If certain elements (withdrawals, atmospheric inputs, most WTPs) are of minimal importance to the results, their inclusion might still be necessary for target setting, but would not aid the focus of scenario testing. While there is no obvious reason to alter the approach now, some simplification may be in order once the scenario testing is complete and target concentrations and loads have been established. It would seem that actual management is going to boil down to nonpoint source controls.

Comment noted. EPA is interested in accounting for all source categories and therefore simplification of the model inputs would be counter to that regulatory and policy requirement. In addition, the importance of some smaller sources varies by lake segment; EPA feels it is necessary to be inclusive of all sources..

## Review of Missisquoi Bay Model report

The LimnoTech report on Missisquoi Bay covers separate modeling of Missisquoi Bay. It is a more complicated model than BATHTUB, addressing a smaller area, just one of 13 defined lake segments, but this is one of the most problematic segments and will likely be the focus of considerable management emphasis to meet overall compliance targets for Lake Champlain. Overall, I found it to be a very well-constructed model with appropriate focus on its use in addressing management issues. Observations include the following:

1. EFDC was used for hydrodynamics and RCA for water quality, each modified by LimnoTech from public access versions into the A2EM linked model. Many variables can be applied and modeled, but only a limited set was used. This is a relatively simple version of the model, but appropriate to the questions to be addressed.
2. The model was based on about ten years of data, five for calibration and five for confirmation. Recommendations for more data collection and expansion of the model were made, but the results were suitable for inclusion in the BATHTUB model without further effort.
3. The model was set up to examine reduction needs to reach a TP level of 25 µg/L in Missisquoi Bay (MB). The major conclusion was that a 75% reduction in watershed loading would be needed and will result in a slow decline of sediment P levels over a 30 year model horizon. However, if sediment P inputs could be curtailed (e.g., removal, inactivation treatment or oxygenation), only a 20% watershed input reduction was needed to reach the target TP level. Sediment flux was extensively evaluated and the results are very believable.

4. Point sources represent only about 2% of the load to the Bay, consistent with the impression from the BATHTUB calibration report that nonpoint sources dominate water quality.
5. The report suggests that tributary algae may be important in determining lake algae after high flows; this is an odd conclusion with questionable support, and is not really consistent with lake bloom features. However, it does not detract from TP loading estimation with this model.
6. High flow dynamics are not accurately portrayed; the difficulty in modeling the high end of range was acknowledged, and this may be a function of high variability in water quality in response to large storms.
7. Lack of close daily tracking of observed values by predicted values was noted, but seasonal agreement is much closer, which is fairly typical of complex models. On an annual basis, average difference between predicted and actual values was 12% and the median was 7%. This is an acceptable level of agreement for the intended use of this model.
8. Sediment release of P was estimated at 1-6 mg/m<sup>2</sup>/day under anoxia and 0.02 to 0.4 mg/m<sup>2</sup>/day with oxic conditions. These are very believable estimates, consistent with St Albans data from other work and experience in other lakes.

Put into the same time periods used in the BATHTUB model, the inputs from the LimnoTech model should be useful and appear superior to the SWAT effort associated with the BATHTUB model effort for Missisquoi Bay. The incorporation of sediment loading appears important, and may be needed in other lake segments as well. But given that Missisquoi and St. Albans are the best known locations for sediment P release in Lake Champlain, the LimnoTech model enhances the use of BATHTUB in this case.

## APPENDIX B – Detailed review of SWAT Model Configuration, Calibration and Validation report

This detailed review was prepared by Peter Shanahan, Ph.D., P.E., and Bruce Jacobs, Ph.D., P.E., of HydoAnalysis, Inc., 481 Great Road, Suite 3, Acton, Massachusetts 01720.

### General comments

The reporting of quantities within the report is uneven. Metric units are mostly used in the text and tables, but not always. Frequently, the number of significant digits reported exceeds any reasonable estimate of the actual accuracy of the quantities.

We have made a number of suggestions where we felt additional discussion could build confidence in the model. We do so following the recommendations by Hassan (2004) that confidence building is a realistic interpretation of the notion of “model validation.” Although Hassan’s paper is directed to complex ground-water models, it is equally applicable here.

### Detailed comments

Pg. 9. The report should clarify the elevation datum and explain that the minimum reported elevation (-2 meters) is at the bottom of the lake, not in the land area of the watershed.

[Report text has been clarified.](#)

Pg. 14. Using the percent impervious from the NLCD 2006 seems a reasonable approach.

[No action necessary.](#)

Pg. 14. The reference to a “subsequent analysis (Tetra Tech 2013)” is confusing. Does “subsequent” mean subsequent to the SWAT model development? This paragraph seems to present generic application rates in the first few sentences, then discusses the “subsequent” development of site-specific rates, but never actually says what rate was used in the SWAT model.

[Report text has been clarified.](#)

Pg. 15. The text says that only Codes 2, 3, and 5 were considered to be unpaved. Code 9 for “unknown surface type” as reported in Table 3 is a pretty large fraction and would also seem likely to be unpaved. One would suspect Vermont knows those roads that are paved.

[Explanation has been added to the report justifying the choice of the above surface type codes for classification as unpaved roads.](#)

Pg. 16. The text defines an average total phosphorus concentration in runoff from unpaved roads as 0.681 mg/L. Aside from the fact that stating this to three-digit accuracy implies far more accuracy than justified, giving only the average gives no confidence as to the representativeness (or uncertainty) in the value. The range of values from which the average

is computed should also be stated and the representativeness of the average value justified or qualified as appropriate.

Report text has been revised to say that the average concentration was approximately 0.7 mg/L. A discussion has been added on the range of observed values.

Pg. 14-16 The descriptions of land uses in both the report and subsequent clarifying responses remain confusing with respect to impervious area. Twenty-eight different land uses are listed in Table 2 and more detailed text descriptions are included for a subset of these uses, including for developed land, unpaved roads, and paved roads and driveways. An impervious fraction is said in the report to have been determined for the developed land category. However, there is then a subsequent discussion of roads and driveways, many of which are of course impervious. The report should clarify the criteria by which these seemingly redundant categories of land use are separated. It should also clarify how the impervious fraction of developed lands is adjusted to compensate for any impervious roads and driveways treated as separate land uses.

The section on landcover and land use has been carefully reviewed and revised. The discussion on impervious areas has also been enhanced and it has been clarified in the text that impervious area was not double counted. Tetra Tech has made changes to the model to accommodate the most recent impervious area GIS layer released by the UVM. These changes have also been described in the report. It is important to note that these changes do not impact the current impervious area associated with paved and unpaved roads in the model. Please refer to the revised report for further details.

Pg. 15-16. The land use presentation is also somewhat disorganized with respect to the separate discussions of “Agricultural Lands and Practices” and “Livestock/Manure Production.” These two “land uses” seem to have considerable overlap and the text should explain how these land uses (in fact how all land uses) relate to the categories in Table 2 so there is no ambiguity as to how they are represented in the model. The section on livestock/manure production seems to have been based on a wealth of information and research, but it is presented in a confusing and somewhat unconvincing fashion. See the detailed comments in the report mark-up. While the representation of phosphorus production associated with agricultural land use seems to have been done well, the reporting detracts from the quality of the underlying work.

The discussion on landcover and land use in the report has been carefully reviewed and revised. Tetra Tech would also like to clarify that the section on “Livestock/Manure Production” is intended to provide details on the methods used to estimate manure deposition on pasture land only. The discussion provided in this section has no bearing on the development of manure application rates on croplands. Manure application and rates on croplands in the model are based on input from local experts in the basin as noted in Appendix A of the SWAT Model Calibration Report.

Pg. 19. In initial comments, we had noted that the text did not indicate whether soil properties were missing for any soil horizons and what had been done to complete the representation of soil properties in the event of missing data. The clarifying response from Tetra Tech describes a procedure for filling in missing soil data. This should be added to the report.

Clarifying response has been added to the report. All the parameters listed were available from the cited databases. A small fraction of required data were missing. The approach adopted by Tetra Tech to address missing data is outlined below.



- If values for parameters associated with a given horizon were missing then these were filled using data from an adjacent horizon of the same soil.
- If data for all horizons were missing then the SWAT soils database was used to fill data based upon the name of the soil.

In addition, Official Series Descriptions from USDA (<http://soils.usda.gov/technical/classification/osd/index.html>) were used to guide the gap filling process..

Pg. 21. The reapportionment of HRUs affects a significant portion of the watershed but is not described in detail in the report. The explanations subsequently provided in clarifying responses should be incorporated into the report along with language indicating that it is not expected to significantly affect model predictions.

The reapportionment seems to be more substantial than implied by the clarifying responses. We made a comparison between two sources: 1) the original raster data for land use, slope, and soil type, and 2) the properties of the HRUs in the SWAT output. The two are compared in Table 1. The “Area of generated polygons” in Table 1 refers to polygons of land use, slope, and soil type produced from the original raster data and should equal the entire watershed area for each Model Sub-basin. These were then compared to the HRUs in the SWAT output for matching characteristics and the area found to match was totaled to give the “Area with Matching Properties in SWAT Output.” The “Percent Unmatched” shows the amount of area that was reapportioned and seems to be higher than implied by the tabulated data provided by Tetra Tech.

The reapportionment of areas does not necessarily cause the model to be inaccurate or unrepresentative, but it could have an effect. A more thorough explanation of the potential effect and perhaps a sensitivity analysis for selected subbasins with high reapportionment percentages is recommended.

Table 1 – Analysis of Area Reapportionment

LWC SWAT Model Drainage Sub-basins	Area of Generated Polygons (sq-km)	Area with Matching Properties in SWAT Output (sq-km)	Percent Unmatched
Lamoille River-6	34.5	12.5	64%
Missisquoi River-7	17.8	8.0	55%
Missisquoi River-12	116.1	42.9	63%
Pike River-8	52.3	19.6	63%
Rock River-11	219.1	71.2	68%
Otter Creek-5	33.0	18.2	45%
Otter Creek-32	10.9	4.4	60%
Little Otter Creek-8	22.9	13.4	41%
Little Otter Creek-9	13.6	2.5	82%
Mill River-1	74.5	16.1	78%
Stevens and Jewett Brook-25	59.1	29.8	50%
Lakeshore-6	24.1	0.0	100%

Tetra Tech is cognizant of the fact that every landuse/soil/slope combination may not be available in the models due to effects of the way thresholds are imposed. However, critical sources of phosphorus like developed land and most of the agricultural land were exempt from the landuse threshold implemented in the model. Since the critical landuse categories are expected to produce the bulk of the phosphorus loads and that these are minimally impacted by the imposition of thresholds, the reapportionment of areas is not expected have a significant impact on the accuracy or representativeness of the model. It is important to note that the overarching objectives of this project are to estimate with reasonable confidence the total phosphorus load generated at the HUC8 level and at the landuse level, and to determine possible avenues of phosphorus load reduction. A rigorous calibration process for loads at the mouth of each HUC8 watershed and constraining landuse level loads to literature reported loads in the region provides reasonable confidence on the outputs generated by the models.

It is important to note that the use of thresholds is a standard (and necessary) practice for SWAT. Tetra Tech would also like to point out that the Table presented above is somewhat misleading as it looks at a complete match for landuse-soil-slope category. The re-apportionment process primarily results in aggregation to other nearby soils that have similar properties.

Pg. 21 Some documentation of how the model was run should be included in the report. Was the model run as a single large model domain or watershed-by-watershed? What are the model computation times? Since computation time is the justification given for HRU reapportionment, the times should be given to show the reapportionment was justified and necessary.

Description of how the model was run, model computation times and justification on the use of thresholds has been added to the report.

Pg. 22 For permitted point sources, indicate whether concentrations and loads correspond to permit limits or actual monitored discharges.

Actual monitored discharges were used and this fact has been clarified in the revised report.

Pg. 22 A clearer citation should be provided for the water withdrawal data such as a specific USGS database.

References have been added.

Pg. 23. The reference to “These meteorological time series” is ambiguous. NOAA Summary of the Day data is cited as the source of precipitation and air temperature, but that database included relative humidity, cloud cover, and wind speed as well. Were those time series used or replaced by generated data?

These meteorological time series” refer to precipitation and temperature time-series. The SWAT weather generator was used for the generation of relative humidity, solar radiation and wind speed. The justification for using the weather generator has also been provided in the revised report

Pg. 28. The description of the reservoirs seem to indicate two different options were used to simulate the various reservoirs. Table 12 should include a column to indicate how each was modeled.

Tetra Tech would like to correct its explanation in the previous version of the report. Only the average annual release option was used to model reservoirs explicitly represented in the model.

- Pg. 28. A generic description of the model flow-routing options is given but which option is used is not indicated. The text should indicate which option was used in this model.

The variable storage method was used in the model.

- Pg. 29-32. As indicated in our preliminary comments, the description of channel erosion was difficult to follow and would benefit from a careful revision. A clear and precise definition of “eroding area” should be included. Terminology should be consistent between the text and Figure 7.

Eroding area is defined as the surface area of the channel eroding per unit length of the channel. Tetra Tech has carefully reviewed and revised the section on channel erosion to ensure that the methodology has been adequately explained..

- Pg. 29. The description of code modifications (not “updates”) is a bit confused. Stone Environmental (2011) has a pretty clear description of how they modified the code. Why not simply say you used the algorithm they developed?

The modifications to the code included the modifications made by Stone Environmental and further additions to simulate the settling of sediment bound phosphorus associated with scoured sediment re-settling in a channel.

- Pg. 32-33. SWAT requires as input the values of a number of different parameters that govern phosphorus cycling. The extended discussion provided in the clarifying response, enumerating the variables, their values, and the process of setting their values during calibration should be added to this section and subsequent sections on calibration of the water quality model. As far as the calibrated parameter values, determining those variables by calibration is probably the only practical option. Arnold et al. (2012) indicate others have treated those same variables as calibration parameters, but some discussion of the resulting values could provide greater confidence in the final model. For example, the phosphorus partitioning coefficient (PHOSKD) is the same for most watersheds (value = 200), similar for the Lake Champlain watershed (190), but somewhat different for the Mettawee/Poultney watershed (150). Are there any geologic or other characteristics of this watershed that would give credence to the lower value? Similarly, the phosphorus availability index (PSP) varies across the watersheds. Are there watershed characteristics that provide a rationale for this variation? Finally, the phosphorus enrichment ratio (ERORGP) was treated as a calibration parameter as a function of land use. Our review of the literature (for example, Radcliffe and Cabrera, 2007) indicates that this ratio is more dependent on factors other than land use. The report should provide a rationale for a dependence on land use.

Since the above mentioned parameters are suggested calibration parameters and since their values were well within the suggested ranges, Tetra Tech did not make additional effort to justify the differences between the parameter values across the HUC8 watersheds. The values of these parameters were entirely based on the calibration/validation process.

The comment is correct in saying that ERORGP is dependent on sediment discharge. It is however important to note that phosphorus is preferentially sorbed to clay and silt particles rather than sand. In SWAT, sediment generation at the HRU level is not simulated for the cohesive and non-cohesive fractions separately but rather as one entity that is representative

of the total sediment load. The variable values of ERORGP by landuse were used to account for the uncertainty in cohesive sediment generation at the HRU level.

Pg. 34. The way Lumb et al. (1994) is cited is not accurate. While Lumb et al. (1994, pg. 56) list the criteria they do so only as an example and do not, per se, recommend these specific numerical criteria nor provide any rationale for the criteria.

The comment is correct as regards Lumb et al. (1994). However, the example criteria reported here have subsequently been widely adopted as targets for modeling (see for instance Duda et al. 2012).

Pg. 34. In Table 15, the seasonal volume and summer storm volume errors seem high.

As mentioned in the response to the above comment, these criteria have been widely adopted as targets for modeling.

Pg. 34. The parameter changes needed to achieve total flow balance are trivial indicating that the model was well formulated even before calibration.

No action necessary.

Pg. 35. Calibrated parameter values are given in the report only as ranges, which do not provide much information on the how the model was finally configured. The clarifying response provides a tabulation of values for each watershed which should be added to the report. Most, but not all, of the parameters are commonly used in calibrating SWAT (Arnold et al., 2012). Some of these parameters vary widely from watershed to watershed. A discussion that relates the variation to known characteristics of the watersheds would build confidence in the model. The concentration of organic phosphorus in the channels (CH\_OPCO) is an unusual calibration parameter, all of the values are outside the range indicated in the SWAT IO Documentation (Neitsch et al., 2010), and some are very high. A rationale for these values should be provided.

The additional language provided in the clarifying responses indicates satisfactory formulation of temperature lapse rates. That language should be added to the final report.

The use of higher than suggested values of concentration of organic phosphorus in the channels is based on published literature on work done in the Lake Champlain basin. These studies have been cited in the revised report.

Pg. 36. The calibration discussion notes that “It is evident from the calibration and validation results that the summer flow volumes are consistently over-represented in the SWAT models. A closer investigation into the observed and simulated flows revealed that the bulk of the over-representation is due to a higher-than-observed baseflow (Table 17).” Given that the baseflow is too high, what steps were taken to modify that result? We agree with the explanation that ground water is not an important contributor to phosphorus transport. For one thing, phosphorus is strongly adsorbed by soil and not particularly mobile in ground water. We do not agree with the unequivocal statement that “biases in seasonal flow are not expected to affect the efficiencies of the modeled BMPs.” A more complete discussion of this model deficiency and its potential effect is warranted.

The options provided by SWAT to configure the baseflow component of the model are limited. Notable deficiencies include the lack of ability to address seasonality in the parameters that

control the baseflow component. Given these challenges, Tetra Tech parameterized the baseflow component of the model to minimize the annual average error in observed and simulated baseflow. Since the pollutant of concern here is phosphorus which is minimally impacted by the baseflow component of total flow, the impact of biases in seasonal flow are not expected to affect BMP efficiencies significantly.

Pg. 38. The reference to Moriasi et al. (2007) in the report text indicates that Moriasi had recommended calibrating to monthly loads. We don't find that in the cited paper: Moriasi et al. (2007, pg. 893) consider both monthly and daily loads, and do not recommend one or the other. We concur in the selection of monthly loads since replication of daily load levels is likely to be unimportant to understanding the long-term eutrophication state of a lake of this size.

Table 4 of Moriasi et al. (2007) provides a recommendation on statistical measures for a monthly time-step.

Pg. 38. The reference to Preston et al. (1989) similarly implies that they recommend the specific approach followed for this study. In fact, they evaluated a variety of different approaches and found none to be superior. With respect to regression estimators, the approach used here, they recommend this approach "if the flow-concentration relationship is strong and consistent" (Preston et al., 1989, pg. 1388). Preston et al. also emphasize that this approach can be compromised if there are few data at high flows. The regressions shown in Figure 12 span a good range of flows, show a reasonably strong fit, and indicate  $R^2$  values that would be considered modest for many datasets but are actually fairly good for nonpoint source load estimation. Nonetheless, the curve fit is good only within an order of magnitude and it would be useful to comment on how much error/uncertainty this may introduce in the calibrated model. Moreover, the report text should be more accurate in representing the recommendations by Preston et al.

The comment is correct that Preston et al. (1989) do not recommend a specific method; rather they discuss the strengths and weaknesses of a range of methods. Tetra Tech's selection of method is consistent with the discussion by Preston et al. (1989).

Pg. 39. Even though the regression in Figure 12 is done on  $\ln(\text{TP})$ , it would be preferable to use a version of Figure 12 in which TP is shown on a linear scale. That would give the reader a better sense of the level of accuracy of this regression question for what matters, which is the phosphorus load (and not its log-transform). At whatever scale, the units of TP should be given and the axes label made clearer. TP is the usual abbreviation for the concentration of total phosphorus. The text suggests that Figure 12 plots the TP load but it is not entirely unambiguous.

The regression equations indeed relate TP concentration to flow. The units of TP concentration (and now noted in the revised report) are mg/L.

Pg. 40. Comparison of Figure 13 with Figure 12 suggests that Figure 13 actually plots the residual of the log of TP. That is, the axis label should read " $\ln(\text{TP}_{\text{obs}}) - \ln(\text{TP}_{\text{sim}})$ ." Plotting the logarithms makes the residuals look much less than they actually are. Recognizing that Figure 13 plots the logarithm of the residual indicates discrepancies of roughly plus or minus a factor of ten. This is not surprising—loads are highly variable, subject to very uncertain measurement, and very difficult to predict. But the presentation should be far more forthcoming as to what is being plotted.

The considerable discrepancies between the predictions and observations point to large uncertainty in the predictions. While this is a fact of life in predicting nonpoint-source loads, it would be useful for decision makers to understand the degree of uncertainty. We recommend including some sort of systematic analysis of model error or at least selected sensitivity runs in the report.

The regression approach is described as having been used for both TSS and TP. Presumably, the regression between the log of flow and the log of these independent parameters is evaluated uniquely for each watershed, however only the results for TP within a single watershed are described within the text. The description of the regression approach should be expanded to include a table showing the regression line slope and intercept, flow breakpoints, and the coefficient of determination for each watershed and parameter. Also, all regressions should be fully documented by construction of scatter-plot figures (like Figure 12) and residual magnitude (like Figure 13).

Tetra Tech would first like to point out that the residuals are plotted using natural log and not  $\log_{10}$ . So the discrepancy is not on the order of plus or minus a factor of ten but more likely about 2.7.

Having said that, Tetra Tech does acknowledge that there are uncertainties associated with predictions and observations. While a systematic analysis of model error or sensitivity analysis would be helpful, Tetra Tech is unable to carry out the same due to resource constraints. This may be considered in future enhancements of the model.

Pg. 42. The additional language provided in the clarifying responses indicates satisfactory formulation of phosphorus export rates. That language should be added to the final report.

Report text has been revised to include a discussion on the formulation of phosphorus export rates.

Pg. 43. The cited reference Medalie (2013) is missing.

Reference has been added.

## Appendices C through K

Appendices C through K report the results for the nine major watersheds. Appendix C was reviewed in detail as a prototype of all of the appendices.

Pg. C-5. The mean daily flow is shown in a way that makes it difficult to determine how well the flow is being matched on a day-to-day scale. In Figures 1 and 8, the predicted flow is plotted on top and with a wider line than the observed flow in such a way as to obscure in most cases the observed flow.

Comment noted.

Pg. C-5. The predicted monthly flows appear to match observed monthly flows quite well (Figures 2 and 9). That said, the bias in the summer flows is evident from a comparison of the median observed and median modeled results as seen in Figures 5 and 12.

Comment noted.

Pg. C-20 – C-22. The time-histories of regressed and simulated TSS values shown in Figures 19 and 20 are consistent with respect to the general trajectory and the timing of peak values; however the simulated values have less variability than the regressed values. That is, the simulated high values are consistently less than the regressed high values and the simulated low values are consistently greater than the regressed low values. This is particularly evident on examination of the scatter plots shown in Figure 23 and 24. The points in these curves fail to scatter evenly around the equal-fit line and show pronounced bias in the simulated TSS concentrations for Observed TSS less than 10 tons per day and greater than approximately 500 tons per day.

Comment noted.

Pg. C-20. The use of log axes in Figures 19 and 20 obscures the magnitude of errors. It would be useful to also present comparisons in the format of Figures 19 and 20 between the predicted loads and the actual measured loads using linear axes.

Comment noted.

Pg. C-22. In Figures 19 and 20, the loads are shown in tons per month and the measured values are referred to as “Regression Loads”. Elsewhere, as in Figures 23, 24, 29, and 30 for instance, loads are presented in tons per day and the measured loads are referred to as “Observed.” We suggest that the charts consistently use either tons per month or tons per day. We also recommend consistency in description of the measured loads. We favor use of the term “Regression Loads” in favor of “Observed Loads,” since the presented monthly loads are only indirectly “observed” based on the regressions of TP and TSS versus Q.

Figures 23, 24, 29 and 30 show instantaneous simulated daily loads against instantaneous observed daily loads. The observed loads are indeed actual loads determined from the product of TSS or TP concentration and observed flow for days that water quality samples were available. The purpose of these figures is to evaluate the ability of the model to produce instantaneous loads.

Pg. C-23. The total phosphorus calibration represented in Figures 25, 26, 29, and 30 is qualitatively similar to the analogous TSS calibration plots described above. The timing of peaks seems to be favorably reproduced in the simulated values, however as with the TSS values the low peaks in the simulated loads are greater than the low peaks in the regression loads and the high peaks in the simulated loads are less than the high peaks in the regression loads. This is most obvious on inspection of Figures 29 and 30, which illustrate considerable variability about the “equal-fit” line and pronounced biases for observed high- and low-load values. The report should be clear that such biases exist and discuss the measures taken to address those. We recognize that it may be impossible to calibrate the model so as to eliminate these biases. In that case, the report should discuss the implications of such biases for decision making about NPS management. For example, many NPS management measures focus on containing the runoff from smaller storms. The model’s high bias for smaller events indicates the potential to over-predict the effectiveness of such measures in reducing loads to the lake.

Tetra Tech would like to point out that the regression statistics provided (including the median and average absolute error) indicate that biases do exist in the simulated loads. In addition, the paired daily load and power plots are also intended to show the presence of probable biases in the model outputs. The presence of bias is expected to have some impact on

effectiveness of BMPs. However, since the reduction strategies are based upon average annual loads over a long period of time, these biases are expected to have minimal impact on NPS management decision making

Pg. C-24. Presumably Figures 27 and 28 would be the analog of Figure 12 in the main text for this watershed, but examination of the similar figure in Appendix G for the Missisquoi Watershed indicates otherwise. Whereas Figure 12 shows a flow-stratified regression line for the Missisquoi, Appendix G does not. The robustness of the regression is indicated by Preston et al. (1989) to be a key factor in evaluating whether regression-based calibration is an appropriate approach. Thus, the equivalent of Figure 12 should be shown in the appendices for all watersheds. That is, the regression line actually used for model calibration should be shown for all watersheds. Also, the text should be modified to explain the utility of the so-called power plots shown in Figures 27 and 28 in measuring the effectiveness of the calibration.

The reviewer is mistaken in assuming that these plots are analogous to the Figure 12 in main text. These figures show the simulated daily loads and observed daily loads on the same plot. The purpose of these plots is to show the trend and magnitudes of daily simulated and observed loads with respect to each other.

Appendices D through J. The biases noted above for TSS and TP are apparent in all other watersheds other than the Saranac River (in Appendix J). That is, the simulated lows are too high and the simulated highs are too low. It is perhaps noteworthy that the error for the summer seasonal volume is markedly less for the Saranac than for the other watersheds.

Comment noted.

Appendices I through K. There is no basin map included.

Maps been included for Appendices I through J.

## Cited references

Only references not cited in the SWAT Model Configuration, Calibration and Validation report are included here.

Arnold, J. G., D. N. Moriasi, P. W. Gassman, K. C. Abbaspour, M. J. White, R. Srinivasan, C. Santhi, R. D. Harmel, A. v. Griensven, M. W. V. Liew, N. Kannan, and M. K. Jha, 2012. SWAT: Model Use, Calibration, and Validation. *Transactions of the ASABE*. Vol. 55, No. 4, Pg. 1491-1508.

Hassan, A. E., 2004. Validation of Numerical Ground Water Models Used to Guide Decision Making. *Ground Water*. Vol. 42, No. 2, Pg. 277-290. March-April 2004.

Moriasi, D.N., J.G. Arnold, M.W. Van Liew, R.L. Bingner, R.D. Harmel, and T.L. Veith, 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the American Society of Agricultural and Biological Engineers*. Vol. 50, No. 3, Pg. 885-900.

Radcliffe, D.E., and M.L. Cabrera, eds., 2007. *Modeling Phosphorus in the Environment*. CRC Press, Boca Raton, Florida.



Neitsch, S. L., J. G. Arnold, J. R. Kiniry, R. Srinivasan, and J. R. Williams, 2010. Soil and Water Assessment Tool, Input/Output File Documentation, Version 2009. Report No. TR-365. Texas Water Resources Institute, Texas A&M University, College Station, Texas. May 2010.

## APPENDIX C – Detailed review of NPS Scenario Tool

This detailed review was provided by William Frost, P.E., D.WRE, Senior Water Resources Engineer at KCI Technologies, Inc., 936 Ridgebrook Rd., Sparks, Maryland 21151.

Mr. Frost has extensive experience with similar tools implemented in Maryland—particularly the Maryland Assessment and Scenario Tool (MAST) (Devereux and Rigelman, undated) and Maryland Department of the Environment (MDE) stormwater guidance (MDE, 2011, 2014). The review below includes numerous comparisons of the BMP Scenario Tool with MAST and MDE guidance as a check on the assumptions underlying the BMP Scenario Tool.

### Detailed comments

Pg. 1. The overall approach is a reasonable means of replicating complex models in a spreadsheet tool that can be used by several engineers and planners to assess improvements. It is similar to the approach used in MAST.

[No response.](#)

Pg. 6. Making data from the Existing Load Summary tab available to the user is a useful element of the Tool which allows a level of feasibility testing while developing scenarios.

[We agree and included the tab to provide information to users as they were creating scenarios. We feel it is useful for presenting the results of the modeling for existing conditions throughout the basin.](#)

Pg. 7. Rather than requiring the user to specify percentages of land use to address, MAST currently allows the user to select acres or square miles, which was an upgrade from early versions that used percentages. This was found to be easier for the user than dealing in percentages. This would be a good improvement.

[If future enhancements are made, this can be considered. Note that the intention of this Tool is to assist initial planning activities to explore what percentage of available treatment areas need to be treated to achieve a reasonable assurance that the target load reduction can be met. In later stages, for tracking and accounting purposes it may become more necessary to know the exact amount of acreage to be treated by a certain BMP type.](#)

Pg. 8. Computing a percent reduction as described in the text box is a reasonable approach. It may be the only feasible method when input data for baseline conditions (historic land use or BMPs) are not available.

[No comment](#)

Pg. 11. Regarding the optional Optimization Tool, optimization is complex and can have a large effect on decision making. Use of an optimization algorithm that is not transparent and understood is not a good practice. It is difficult or impossible to provide an adequate QC check in this situation.

[The optimization feature was included preliminarily at VTDEC's request but ultimately was not and is not being utilized in determining loading capacities.](#)

Pg. 15. For calculating reductions, the tool allows the user to select a target HRU in an MS4 area and apply BMPs to a fraction of those areas. Consider using absolute area measures (ha/ac/sq mi) instead of fractions. These are easier for the user to understand.

[If future enhancements are made, this can be considered.](#)

Table 4. Consider adding a column that shows acceptable practices for new development as listed in the Vermont Stormwater Management Manual (VSMM) (Vermont ANR, 2002) cross referenced with the Scenario Tool BMPs listed in the table. Table 4 does not include the following BMPs listed in the Scenario Tool: surface sand filter, underground sand filter, perimeter sand filter, organic filter, open channel/dry swale, wet swale, grass channel. The open channel and dry swale are really two different types of conveyance BMPs and should be separated per Vermont ANR (2002). The following additional non-structural BMPs are listed by MDE (2014) and should be considered for inclusion in the BMP Scenario Tool: sheet flow to conservation area, rainwater harvesting, tree planting, stream restoration, and shoreline restoration.

[Thank you for these detailed suggestions. EPA is considering ways to transition the Scenario Tool into a Tracking and Accounting System for TMDL Implementation, and that tool will have the ability to consider additional BMPs. The Scenario Tool included a set of BMPs for which there is reasonably consistent agreement on P reduction efficiencies and expected use. Because the VT Stormwater manual is currently undergoing an update and the new version is expected to contain many significant changes and additional measures, it was not used as the basis for the Scenario Tool's developed landuse BMPs.](#)

Pg. 18. Because they require the determination of rainfall depth, the curves from USEPA (2010) may be too complicated for scenario analysis. We recommend a default value that the planner can change if enough information is available. A possible default value would be an equivalent of the Water Quality Volume (WEF and ASCE, 2012), say 1.0 to 0.9 inches/impervious acre.

[EPA's BMP Performance Curves were used for this exercise because the rigorous nature of their derivation lends confidence to the predictions. The range of various values used were selected purposely to support the potential analysis needs of Lake Champlain TMDL implementers. EPA believes it is important to be able to represent smaller rainfall depths than the Water Quality Volume, because many BMP retrofit sites will not have the space for a practice designed for a 1 inch storm. Remember, the tool was not designed for site planning purposes, it was designed to simulate the effects of implementing certain categories of BMPs across wide areas. As currently configured, the user can easily select, for example, the 0.5 inch rainfall depth, and simulate the effect of practices designed to treat that amount of rainfall across a large HUC8 watershed. This allows for the simulation of potentially more realistic scenarios for tightly built-up areas.](#)

Pg. 19. Substituting the Chesapeake Bay Program efficiencies is reasonable. Removals from the performance curves in EPA (2010) were significantly lower than CBP and not in line with most published studies of wet ponds.

[Comment noted.](#)

Pg. 19. For Non-structural BMPs (fertilizer ban, street sweeping, etc.) there are citations to unspecified "EPA Analysis." A more definitive reference should be provided.

[The text has been revised to indicate appropriate citations.](#)

Pg. 21. In Table 7, the TP efficiencies are lower than in MDE (2011) or MDE (2014). For the IAR-MDR BMP, 0.89 is a reasonable TP efficiency.

These efficiency rates were developed through a modeling evaluation that simulated the movement of runoff from pervious onto impervious area, using New England regional soil type and climate data. If newer credible performance information for this or other practices becomes available, EPA will consider updating recommendations on the applicable efficiency rates, particularly in the context of tracking BMP implementation going forward.

Pg. 21. In Table 8, the TP efficiency of 0.5 attributed to Ban on P fertilizer use on turf may be high based on the rates used in MAST.

The EPA approach used to calculate phosphorus reductions from restrictions on fertilizer use is based on the Chesapeake Bay Program approach, referenced in the report. The 0.5 efficiency value derives from the difference between 0.4 mg/l P in runoff from lawns where P fertilizer is applied, and 0.2 mg/l P in runoff from lawns without P fertilizer application. The efficiency values used in MAST may differ from the Lake Champlain values due to different assumptions about compliance rates and other factors. As described in Appendix A of the report, a 100% compliance rate was used in the Lake Champlain analysis, as it was considered appropriate to assume that eventually the VT law will be fully complied with.

Pg. 22. We spot-checked Table 9 against USEPA (2010). The runoff depths for wet ponds are much higher than in USEPA (2010). Other values check with USEPA (2010).

The runoff depths are the same for all measures. If the reviewer meant to refer to the efficiencies, those correspond to values adopted by the Chesapeake Bay Program, which are used by the Tool instead of efficiencies predicted by the performance curves. Performance curve wetland efficiencies, as the reviewers noted in a previous comment, are lower than most studies suggest and therefore are not used in the Tool.

Pg. 29. Choosing actions on a reach-specific basis is a reasonable approach considering the variability and coarse resolution of the data. For comparison, both MAST and MDE 2014 treat this as a load reduction BMP, at 0.068 lb/LF of restoration.

We concur that given the resolution of data, a reach specific basis was most reasonable.

Pg. 30. The procedure for assigning CEM data to HUC12 basins needs to be explained for the situation in which the streams are close to equally divided among CEM stages. Overall, the procedure seems a coarse analysis. It entails considerable data aggregation and simplification. Results should be checked against other estimates of stream erosion.

The approach made use of model results, stream erosion study data from the Missisquoi River watershed, and detailed geomorphic assessment data on stream channels throughout the basin (although geomorphic data were not available for 100 % of the basin). The approach was developed in consultation with leading stream geomorphology experts in Vermont. Further explanation is provided in the SWAT model calibration report.

Pg. 32. The report should provide additional information on the conversion of stream sediment into phosphorus loads including the actual conversion factors used.

We have provided a table with the values of stream phosphorus concentration by each HUC8 watershed in the SWAT calibration report.

Pg. 35. The generic 5% reduction factor for forested areas seems entirely arbitrary. In the absence of any data on loads, opportunity for BMPs, and reduction efficiencies, perhaps this option should not be included. Are there any data to support 5% vs 10% or 1% as conservative default values?

[Forest BMP representation is being updated in the tool, and a separate analysis is being conducted by EPA to augment information on potential phosphorus reductions from forest lands. The new analysis will be included as an appendix to the Lake Champlain TMDL document.](#)

Pg. 36. The different approach used for the Missisquoi Bay calculations is a reasonable way to approximate the model results.

[Comment noted.](#)

Pg. 42. The method of using GIS to estimate area of turf is good.

[Comment noted.](#)

Pg. 43. The Confidence Intervals in Table A-2 vary widely. A larger sample may have improved these. The mean turf estimate for low density residential seems to be off. With 20% to 50% impervious, a mean turf value of 29% implies that approximately 50% to 20%, respectively, of the land in this category is neither turf nor impervious. What is the remaining land cover? Woods? This number appears to be low. The values for medium- and high-density residential and commercial/industrial appear reasonable.

[The analysis was done at the parcel level, and the amount of turf was determined as a percentage of the pervious area within a parcel. Low density residential non-turf pervious land cover included shrubs, trees, barren land along roads, meadows, etc. It is not uncommon in VT low density residential areas for significant parcel percentages to be in these categories.](#)

Pg. 45. The values in Table A-5 appear to be reasonable.

[Comment noted.](#)

Table A-6. The units of MG/ha/yr is a strange mix of US Customary and SI units.

[Comment noted; Will consider revising.](#)

Pg. 50. Impervious disconnection is unlikely for commercial/industrial land. There is rarely sufficient pervious area downgrade and in many cases roof drainage is internal and not through downspouts. The same is true to a lesser extent for high-density residential.

[While impervious disconnection may not be implemented widely on commercial/industrial lands, it was nevertheless included as an option for these land uses, as it is a low-cost option that can be effective when site conditions allow.](#)

Pg. 54. The Agricultural BMPs in the Scenario Tool are very detailed and as a result there were not a lot of data available in the literature for comparison. The Tetra Tech QA documentation includes an extensive table of BMP efficiencies for agricultural land use and was used to spot check the Scenario Tool and Table D-1. All three are in agreement. The SWAT derived efficiencies in Table D-1 appear to be reasonable.

[Comment noted.](#)

## Cited references

- Devereux, O.H., and J.R. Rigelman, undated. Maryland Assessment Scenario Tool, About MAST. Maryland Department of the Environment, Baltimore, Maryland. <http://www.mastonline.org/About.aspx>. Accessed April 11, 2015.
- MDE, 2011. Accounting for Stormwater Wasteload Allocations and Impervious Acres Treated, Guidance for National Pollutant Discharge Elimination System Stormwater Permits. Maryland Department of the Environment, Baltimore, Maryland. June (DRAFT) 2011. <http://www.mde.state.md.us/programs/Water/StormwaterManagementProgram/Documents/NPDES%20Draft%20Guidance%206-14.pdf>
- MDE, 2014. Accounting for Stormwater Wasteload Allocations and Impervious Acres Treated, Guidance for National Pollutant Discharge Elimination System Stormwater Permits. Maryland Department of the Environment, Baltimore, Maryland. August 2014. <http://www.mde.state.md.us/programs/Water/StormwaterManagementProgram/Documents/NPDES%20MS4%20Guidance%20August%2018%202014.pdf>
- Vermont ANR, 2002. The Vermont Stormwater Management Manual. Vermont Agency of Natural Resources, Montpelier, Vermont. April 2002. [http://www.vtwaterquality.org/stormwater/docs/sw\\_manual-vol1.pdf](http://www.vtwaterquality.org/stormwater/docs/sw_manual-vol1.pdf).
- WEF and ASCE, 2012. *Design and construction of urban stormwater management systems*. WEF Manual of Practice No. 23, ASCE/EWRI Manuals and Reports on Engineering Practice No. 87. Water Environment Federation, Alexandria, Virginia and American Society of Civil Engineers, Reston, Virginia.

## APPENDIX D – Detailed review of climate change analysis

This detailed review was provided by Paul H. Kirshen, Ph.D., Research Professor in the Civil and Environmental Engineering Department and the Institute for the Study of Earth, Oceans and Space at the University of New Hampshire, 248 Gregg Hall, Durham, NH 03824.

### Review of LaPlatte River Watershed Pilot Report

#### Overall summary:

Reasonably sound procedures were carried out and applied. The writing could be improved: some aspects were difficult to follow.

There are no major flaws and only two possible areas of improvement (below). But in spite of these, the results still adequately show the impacts of climate change on flows and water quality. If an adaptation program was to be developed, more emission scenarios would be needed to better define the range of possible changes.

The following are two areas of possible improvement:

1. A longer period of record could have been assembled to allow for both calibration and verification of the SWAT model. (Page 11)
2. Impacts of climate change markedly diverge after or even before 2050. More emission scenarios could have been used to clarify this behavior. (Page 25)

#### Detailed comments:

Pg 3, Third paragraph. The text should indicate the time step used in SWAT.

Pg. 7, last paragraph. The text states “the magnitude of the response to CO<sub>2</sub> levels predicted by the mid-21st century appears to be on the order of a 10 percent reduction in ET response.” At least in terms of crops and trees, the U.S. National Climate Assessment (Melillo et al., 2014, page 157) reports that effects of CO<sub>2</sub> enrichment may be offset by other factors. Were these offset factors considered in this analysis as well?

Pg. 11, first paragraph of Section 2.4. The text indicates that the record of 1990 to 2004 allowed only for calibration and not validation. Ideally, the met data could have been expanded to 2010 or so, and then calibration done on the first 10 years, verification on the second 10 years. We presume this was not done because the GCRP data were taken from BASINS which ends in 2004. Still, it would have been good to obtain an extended record.

Pg. 12, first paragraph. It would be useful to describe how the SWAT weather generator works and how reasonable it is to use in climate change studies. At minimum, the appropriate sections of Neitsch et al. (2011) should be cited. As described by Neitsch et al., the generator creates daily values. The report should explain how these were disaggregated to hourly or less if that was necessary.

Pg. 14, third paragraph. MM should be defined.

Pg. 15, third paragraph. It is not clear that the sentence beginning “The results presented in Figure 3-1...” is consistent with Figure 3-1. It appears in Figure 3-1 that in many cases the modeled flow exceeds the observed flow, particularly at the end of the calibration period. Fig 3-2, however, shows that the model underpredicts the highest observed flows.

Pg. 16, first paragraph. Table 3-1 reports that the modeled storm volume is ~20% less than observed (~1 inch). The report should discuss the significance of this discrepancy. Also, the report should explain how error is calculated in Table 3-1.

Figure 3-2. The right panel needs further explanation. Explain how these graphs support the statement in the text that there is “reasonably good agreement between total water volumes in the observed and simulated scenarios.”

Figure 3-3. The “seasons” referenced in the figure caption should be defined.

Pg. 20, last paragraph. The text ascribes under-prediction of loads to “uncertainty in the simulation of reservoir trapping.” Might this also be due to the model underestimating storm flows?

Pg. 24, last sentence of second paragraph under Section 4.1. The description of scenario use is well written.

Pg. 25, second paragraph. Regarding the use of a single storyline through mid-century, it is certainly the case that the storylines do not diverge up to around 2040 or 2050, but after that, there is significance divergence. This should be noted in the analysis.

Pg. 26, first paragraph. It would be useful to generally compare these four GCMS individually as, for example, drier or warmer than others, and also to compare to some of the other approximately twenty GCMs for North America so the reader will know where these scenarios fit compared to others.

Pg. 26, second paragraph. It may be possible to compare the HadCM3 and CCSM outputs with the others by comparing statistical values over long time periods such as 30 years or more.

Pg. 27, first paragraph. Although there are alternatives, we regard it acceptable to evaluate changes in other met variables by varying the SWAT weather generator.

Pg. 28, bottom paragraph. Procedure for creating temperature files is OK.

Pg. 29, last paragraph under Precipitation Scenarios. It is reasonable to attempt to simulate the intensification of the precipitation under climate change. The descriptions of Approach 1 and 2 are very difficult to understand and follow. For example, parameter  $q$  is referred to on the top of page 30, but not defined until later.

Pg. 32, PET implementation. The report should explain if the PET simulation includes the impacts of CO<sub>2</sub> enrichment mentioned above.

Table 4-4. Clarify for DEWPT1 what the Adjustment Applied does to possible future changes in relative humidity.



Pg. 32, last paragraph. It was good procedure to check the probability of a rainy day being followed by a rainy day.

Pg. 33, first paragraph. Regarding the assumption that statistical parameters remain unchanged, did you check the sensitivity of SWAT output to, for example, dew point temp, to show the relative impact of this assumption?

Table 4-1. Does this show the difference from the present climate? Why were only the six NARCCAP scenarios considered?

Pg. 35, Section 4.2.1. We suggest mentioning that looking over a 30-year period is necessary as a climate state is defined by ~30 years and that a time period of this length negates the influence of shorter time cycles like ENSO, NAO, etc.

Pg. 35, Section 4.2.2, first paragraph. Good to mention that SWAT is not as accurate at shorter time scales.

Pg. 35, Section 4.3. Is "Annual Runoff" in Table 4-5 the averages of each of the three groups (i.e., NARCCAP, GCM, BCSD)?

Table 4-5. Define 1-day flow shown under flow-duration curve zone.

Figure 4-12. The figure shows no significant change in the timing of the monthly peak flow. This seems a surprising and worthwhile finding to mention in the text.

Table 5-3. Why combine several watersheds into one (e.g. 40 contains 40, 42, 43 as described in the text)?

Figure 5-4. Define the "Aggregate" scenario.

Table 5-4. The meaning of the "Change" column is unclear. Is the percent change over all models in 2041-2070?

## Lake Champlain Report

### Overall summary:

The same methodology comments apply to this report as the LaPlatte report. There is one additional concern.

1. The results for the LaPlatte River seem to significantly differ between this report and the LaPlatte report.

### Detailed comments:

Pg. 5, second paragraph. There needs to be a citation to the SWAT Model Configuration, Calibration and Validation report here.

Pg. 5. It needs to be stated here that actually 14 different combinations of GCMS and downscaling methods were used. Otherwise Table 2 makes little sense.

Pg. 12. A more complete description of the weather generator must be given. Inputs and outputs should be described and the algorithms summarized.

Pg. 13. If SWAT operates at a daily time step, the report should explain why hourly met data are needed.

Table 5. The text should explain why the TSS minimum is negative here and in some other watersheds.

Figure 1. The text should define GDFL-slice.

Figure 5. The text should explain why there is such a large decrease in peak monthly flow under the climate change scenarios.

Table 12. The text should explain why annual flow decreases for LaPlatte and these other rivers. This has very different results compared to Table 4-6 of the LaPlatte report and other results there. These discrepancies should be explained.

SWAT2005 was used for the development of the LaPlatte River model, while SWAT2009 was used for the development of the Lake Champlain watershed model. Significant changes to the algorithms have been made between SWAT2005 and SWAT2009, including but not limited to snow simulation and in-stream sediment dynamics. Some of the differences in the climate change results may be attributed to the differences in the model computational code. It is also important to note that the baseline 30 year period for the two models differ which could be also be a reason for the differences seen in the climate change analysis.

### **Cited references:**

Melillo, Jerry M., Terese (T.C.) Richmond, and Gary W. Yohe, Eds., 2014: Climate Change Impacts in the United States: The Third National Climate Assessment. U.S. Global Change Research Program, 841 pp. doi:10.7930/J0Z31WJ2.

Neitsch, S. L., J. G. Arnold, J. R. Kiniry, and J. R. Williams, 2011. Soil and Water Assessment Tool, Theoretical Documentation, Version 2009. Report No. TR-406. Texas Water Resources Institute, Texas A&M University, College Station, Texas. September 2001.