Appendix B
Project Quality Assurance Project Plan
Section 8: Model Calibration

This appendix reproduces Section 8.0 of the QAPP which describes model calibration requirements; taken from:

Tetra Tech. (2008) Quality assurance project plan for watershed modeling to evaluate potential impacts of climate and land use change on the hydrology and water quality of major U.S. drainage basins. Prepared for the Office of Research and Development, Global Change Research Program, U.S. Environmental Protection Agency, Washington, DC.

8.0 MODEL SETUP/CALIBRATION

8.1 PROJECT QUALITY OBJECTIVES

EPA emphasizes (USEPA 2000, 2002) a systematic planning process to determine the type and quality of output needed from modeling projects. This begins with a Modeling Needs and Requirements Analysis, which includes the following components:

- Assess the need(s) of the modeling project
- Define the purpose and objectives of the model and the model output specifications
- Define the quality objectives to be associated with model outputs

The first item (needs assessment) is covered in EPA's task order. In essence, simulation models are needed to predict future responses to changes in climate and land use. The existing simulation models HSPF and SWAT are believed to be sufficient to this purpose, and creation of new models is not required.

The second item (define purpose and objectives) is the subject of EPA's Draft Analysis Plan. This proposes both the purpose of the modeling and the specific endpoints to be evaluated as a result of the modeling. At a general level, the objective of this modeling project is to assess the potential effects of climate and land use change on the hydrology and water quality of major U.S. drainage basins; however, this general objective will need to be made more specific to guide development of the modeling effort. The Tt team is tasked with reviewing and commenting on the Analysis Plan as part of this work—and revisions to the existing Analysis Plan could arise as a result of these recommendations. At the end of this review, the Tt team and the EPA COR must agree on the principal study questions to be addressed through the modeling.

The quality objectives for the model(s) follow directly from the purposes and objectives—and can be refined in conjunction with the review of the Analysis Plan. In general, the modeling effort needs to be designed to achieve an appropriate level of accuracy and certainty in answering the principal study questions. This process takes into account the following elements:

- The accuracy and precision needed for the models to predict a given quantity at the application site of interest to satisfy study questions
- The appropriate criteria for making a determination of whether the models are accurate and precise enough on the basis of past general experience combined with site-specific knowledge and completeness of the conceptual models
- How the appropriate criteria would be used to determine whether model outputs achieve the needed quality

EPA's Draft Analysis Plan suggests that the principal study questions to be addressed by the models are changes in (defined on the basis of modeling at a daily time step): (1) the 100-year flood, (2) 7Q10 low flow, (3) runoff center of mass, (4) monthly sediment loads, (5) monthly total nitrogen loads, and (6) monthly total phosphorus loads. This list could be expanded or modified on the basis of the review of the Draft Analysis Plan.

The models will be calibrated and validated to existing (1970–2000) data to establish their credibility for use in forecasting responses to future change. Specific calibration and validation targets for model acceptability (see Sections 8.2 and 8.3) will be selected in light of the intended uses of the model, as identified in the final revisions to the Analysis Plan.

8.2 MODEL CALIBRATION AND VALIDATION

Model calibration is the process of adjusting model inputs in acceptable limits until the resulting predictions give good correlation with observed data. Commonly, calibration begins with the best estimates for model input on the basis of measurements and subsequent data analysis. Results from initial simulations are then used to improve the concepts of the system or to modify the values of the model input parameters. The use of calibrated models, the scientific veracity of which is well defined, is of paramount importance to this project. Because the goal is to be able to assess the potential effects of climate and land use change on the hydrology and water quality of major U.S. drainage basins, model calibration and validation should strive to minimize errors (deviations between model predictions and observed measurement data.).

The Tt Co-TOLs or lead modeler will direct the model calibration efforts. Models are often calibrated through a subjective trial-and-error adjustment of model input data because a large number of interrelated factors influence model output. However, the experience and judgment of the modeler are a major factor in calibrating a model accurately and efficiently. Further, the model should meet pre-specified quantitative measures of accuracy to establish its acceptability in answering the principal study questions.

The model calibration process proceeds through both qualitative and quantitative analyses. Qualitative measures of calibration progress are commonly based on the following:

- Graphical time-series plots of observed and predicted data

- Graphical transect plots of observed and predicted data at a given time interval
- Scatter plots of observed versus predicted values in which the deviation of points from a 45-degree straight line gives a sense of fit
- Tabulation of measured and predicted values and their deviations

After initially configuring the modeling systems, the Tt team will perform model calibration and validation. The watershed models will be calibrated to the best available data, including literature values, and interpolated or extrapolated values using existing field data. If multiple data sets are available, an appropriate time period and corresponding data set will be chosen on the basis of factors characterizing the data set, such as corresponding weather conditions, amount of data, and temporal and spatial variability of data.

A model is considered calibrated when it reproduces data within an acceptable level of accuracy, as described in Section 8.3 and itemized in Table 4 (quantitative measures). A set of parameters used in a calibrated model might not accurately represent field values, and the calibrated parameters might not represent the system under a different set of boundary conditions or hydrologic stresses. Therefore, a model validation period helps establish greater confidence in the calibration and the predictive capabilities of the model. A site-specific model is considered validated if its accuracy and predictive capability have been proven to be within acceptable limits of error independently of the calibration data.

Table 4. General percent error calibration/validation targets for watershed models (applicable to monthly, annual, and cumulative values)

	Relative percent error		
	Very good	**Good**	**Fair**
Hydrology/Flow	< 10	10–15	15–25
Sediment	< 20	20–30	30–45
Water Quality/Nutrients	< 15	15–25	25–35

In general, model validation is performed using a data set separate from the calibration data. If only a single time series is available, the series could be split into two subseries, one for calibration and another for validation. If the model parameters are changed during the validation, this exercise becomes a second calibration, and the first calibration needs to be repeated to account for any changes. Representative stations will be used to guide parameter adjustment to get an accurate representation of the conditions of the individual subwatersheds and streams. The calibration and validation process will be documented for inclusion in the technical reports.

8.3 SPECIFIED PERFORMANCE AND ACCEPTANCE CRITERIA

Model Testing

Model testing includes calibration, verification, and validation. The previous section described model calibration and validation. Model verification is the process of testing the model code, including program debugging, to ensure that the model implementation has been done correctly. Testing usually begins with the best estimates for model input on the basis of measurements and subsequent data analyses. Results from initial simulations are then used to improve the concepts of the system or to modify the values of the model input parameters.

For this project, existing tested model code will be used (HSPF and SWAT). Therefore, model verification is required only for new *bridge* code, such as that required to translate climate scenarios into model input.

The Tt team will calibrate the project models using the best available data, including literature values and interpolated or extrapolated existing field data. The model will be considered successfully tested when it reproduces data at an acceptable level of accuracy.

The work proposed for this project, as defined in the Draft Analysis Plan, differs from other, more common applications of watershed models (e.g., for TMDLs) in several ways that affect the calibration strategy:

- Models will be developed at a very large spatial scale (i.e., HUC4 scale) and will be calibrated at a limited number of points, most of which will likely be at the HUC8 scale.
- Models will be developed for multiple watersheds, and calibration will be done by multiple teams of modelers. The different teams should all apply the same calibration metrics.
- Two separate models (HSPF and SWAT) will be developed for some or all the watersheds. A common set of calibration criteria should be applied to both models to facilitate comparison.
- Models are proposed to be developed using a daily time step (based on the scale of the analysis), which will limit the ability to resolve extreme flows.
- Model application is not for regulatory purposes but to inform possible long-term effects of different change scenarios. While calibration to establish model credibility is essential, the ability to correctly simulate relative changes is most important.
- Comparison of observed and predicted values on a frequency-duration plot.

Quantitative acceptance criteria for the models will be selected to reflect the final set of principal study questions in the revised Analysis Plan and incorporated into the QAPP. Given the considerations listed above, quantitative acceptance criteria will be expressed in relative, rather than absolute form. That is, relevant calibration outputs will be ranked on a scale ranging from *poor* to *very good*. Calibration will strive to obtain the best fit possible; however, specific values of quantitative measures will not be proposed to define whether results should be accepted or rejected. Rather, the level of uncertainty determined in calibration and validation will be documented to decision makers to aid in interpretation of results.

The current Draft Analysis Plan references only three measures related to hydrology (100-year flood, 7Q10 low flow, and runoff center of mass); however, accurate representation of the general water balance is required to demonstrate that the model provides a reasonable representation of reality that can serve as a foundation for water quality simulation. Therefore, commonly accepted measures of model hydrologic fit will be applied.

Model simulation of water quality is, in general, more difficult than simulating hydrology, in part because any uncertainty in the hydrologic simulation will propagate into the water quality

simulation. In addition, the principal study questions related to water quality contained in the Draft Analysis Plan address loads. Loads are not directly observed but are inferred from point-in-time concentration data and continuous flow data. As a result, *observed* load estimates are subject to considerable uncertainty.

Quantitative measures, sometimes referred to as calibration criteria, include the relative percent error between model predictions and observations as defined generally below:

$$E_{rel} = \frac{\sum |O - P|}{\sum O} \times 100,$$

where E_{rel} = relative error in percent. The relative error is the ratio of the absolute mean error to the mean of the observations and is expressed as a percent. A relative error of zero is ideal. Additional statistics that will be applied include the correlation coefficient (R) and its squared value, the coefficient of determination (R^2), where

$$R = \frac{\sum \left(O_i - \overline{O}\right) \cdot \left(P_i - \overline{P}\right)}{\sqrt{\sum \left(O_i - \overline{O}\right)^2 \cdot \left(P_i - \overline{P}\right)^2}} = \frac{\sum O_i \cdot P_i}{\sqrt{\sum O_i^{\,2} \cdot \sum P_i^{\,2}}},$$

where the overbar indicates the sample mean.

For hydrology and the water balance, percent error tests will be applied to the following components:

- Total flow volume
- 10 percent high flows
- 50 percent low flows
- Seasonal flow volumes

For water quality, the outcomes of interest defined in the current Draft Analysis Plan are monthly loads. Therefore, similar calculations of relative percent error will be applied to the series of predicted and *observed* monthly loads (where the observed monthly loads will need to be estimated from observed flow and concentration data using an appropriate estimation technique, such as those described in Preston et al. 1989).

These tests are relevant to monthly and annual values. General calibration/validation targets for percent error consistent with current best modeling practices (Donigian 2000) are shown in Table 4.

For hydrology, there is also an interest in extreme high and low flows. Answering this study question requires calibration to daily flows, rather than just monthly and annual values. Figure 3 (also from Donigian 2000) summarizes R and R^2 ranges for the evaluation of daily and monthly flows:

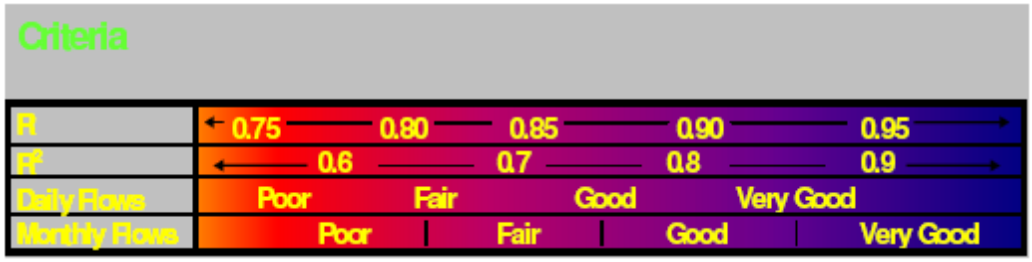

Figure 3. R and R^2 value ranges for model performance

In addition, the Nash-Sutcliffe coefficient of model fit efficiency (COE) will be reported for all calibration and validation runs—although no specific criteria are proposed. This is calculated as

$$COE = 1 - \frac{\sum [O_i - P_i]^2}{\sum [O_i - \overline{O}]^2}.$$

A COE value of one indicates a perfect fit between measured and predicted values for all events. A value of zero indicates that the model fit is not better than using the average value of all the measured data.

Following model calibration, model validation will be conducted using separate, independent portions of the available time series at the calibration stations. Because the Analysis Plan calls for simulating the period 1970–2000, while land use will be based on 2001 NLCD information, the 10-year period from 1991 through 2000 will generally be proposed for calibration, while an earlier period (dependent on data availability) will be used for validation tests. Because the land use distribution during the 1970–1991 period could be different in some regions than during the 1991–2000 period, it is important to note that validation results might not achieve the same quantitative acceptance levels as for calibration.

The Tt team will document model performance over both the calibration and validation period in the technical reports, using the quantitative measures of accuracy documented above (or any additional measures that could be identified in modifications to this QAPP). In addition to measures of accuracy, additional acceptance criteria will include modeling result precision and representativeness:

- Precision of model results: Precision of generated data produced by the model will be examined by performing replicate runs. By confirming that an identical data set is generated when a replicate of the previous model run will rule out numerical instability issues and verify the precision of the model.
- Representativeness of model results: The Tt team technical staff will compare the loadings data and measured environmental concentrations to examine sources and sinks of materials.

An overall assessment of the success of the calibration can be expressed using calibration levels.

Level 1: Quantitative performance measures fall within the *very good* range (highest degree of calibration).

Level 2: Quantitative performance measures fall within the *good* range.

Level 3: Quantitative performance measures fall within the *fair* range.

Level 4: Quantitative performance measures fall within the *poor* range (lowest degree of calibration).

Model Sensitivity Analysis

The sensitivity to variations or uncertainty in input parameters is an important characteristic of a model. Sensitivity analysis is used to identify the most influential parameters in determining the accuracy and precision of model predictions. This information is of importance to the user who must establish required accuracy and precision in model application as a function of data quantity and quality. Sensitivity analysis quantitatively or semi-quantitatively defines the dependence of the model's performance assessment measure on a specific parameter or set of parameters. Sensitivity analysis can also be used to decide how to simplify the model simulation and to improve the efficiency of the calibration process.

Model sensitivity can be expressed as the relative rate of change of selected output caused by a unit change in the input. If the change in the input causes a large change in the output, the model is considered to be sensitive to that input parameter. Sensitivity analysis methods are mostly nonstatistical or even intuitive by nature. Sensitivity analysis is typically performed by changing one input parameter at a time and evaluating the effects on the distribution of the dependent variable. Nominal, minimum, and maximum values are specified for the selected input parameter.

Sensitivity analysis is performed at the beginning of the calibration process to design a calibration strategy. After calibration is completed, a more elaborate sensitivity analysis is performed to quantify the uncertainty in the calibrated model caused by uncertainty in the estimates of the model input parameters.

Informal sensitivity analyses (iterative parameter adjustments) are generally performed during model calibration to ensure that reasonable values for model parameters will be obtained, resulting in acceptable model results. The degree of allowable adjustment of any parameter is usually directly proportional to the uncertainty of its value and is limited to its expected range of values.

8.4 ASSESSMENT AND RESPONSE ACTIONS

The ability of computer code to represent model theory accurately will be ensured by following rigorous programming protocols, including documentation within the source code. Specific tests will be required of all model revisions to ensure that fundamental operations are verified to the extent possible, including testing numerical stability and convergence properties of the model code algorithms, if appropriate. Model results will generally be checked by comparing results to those obtained by other models or by comparing them to hand calculations. Visualization of model results will help determine whether model simulations are realistic. Model calculations

will be compared to field data. If adjustments to model parameters are made to obtain a fit to the data, the modelers will provide an explanation and justification that must agree with scientific knowledge and fit within reasonable ranges of process rates as found in the literature.

As described in Section 5.1, non-project-generated data will be used for model development and calibration. The model calibration procedure is discussed in Section 8.2. The DQOs were discussed in Section 7.0 and 8.0 of this document. Modelers will cross-check data for bias, outliers, normality, completeness, precision, accuracy, and other potential problems.

Data generated outside the project will be obtained primarily from quality assured databases maintained by EPA, USGS, and other entities. Additional data may be obtained from either published or nonpublished sources. The published data will have some degree or form of peer review. Typically, modelers examine these data as part of a data quality assessment. Unpublished databases are also examined in light of a data quality assessment. Data provided by EPA or other sources will be assumed to meet precision objectives established by those entities.

The QA program under which this task order will operate includes surveillance, with independent checks of the data obtained from sampling, analysis, and data gathering activities. This process is illustrated in Figure 4.

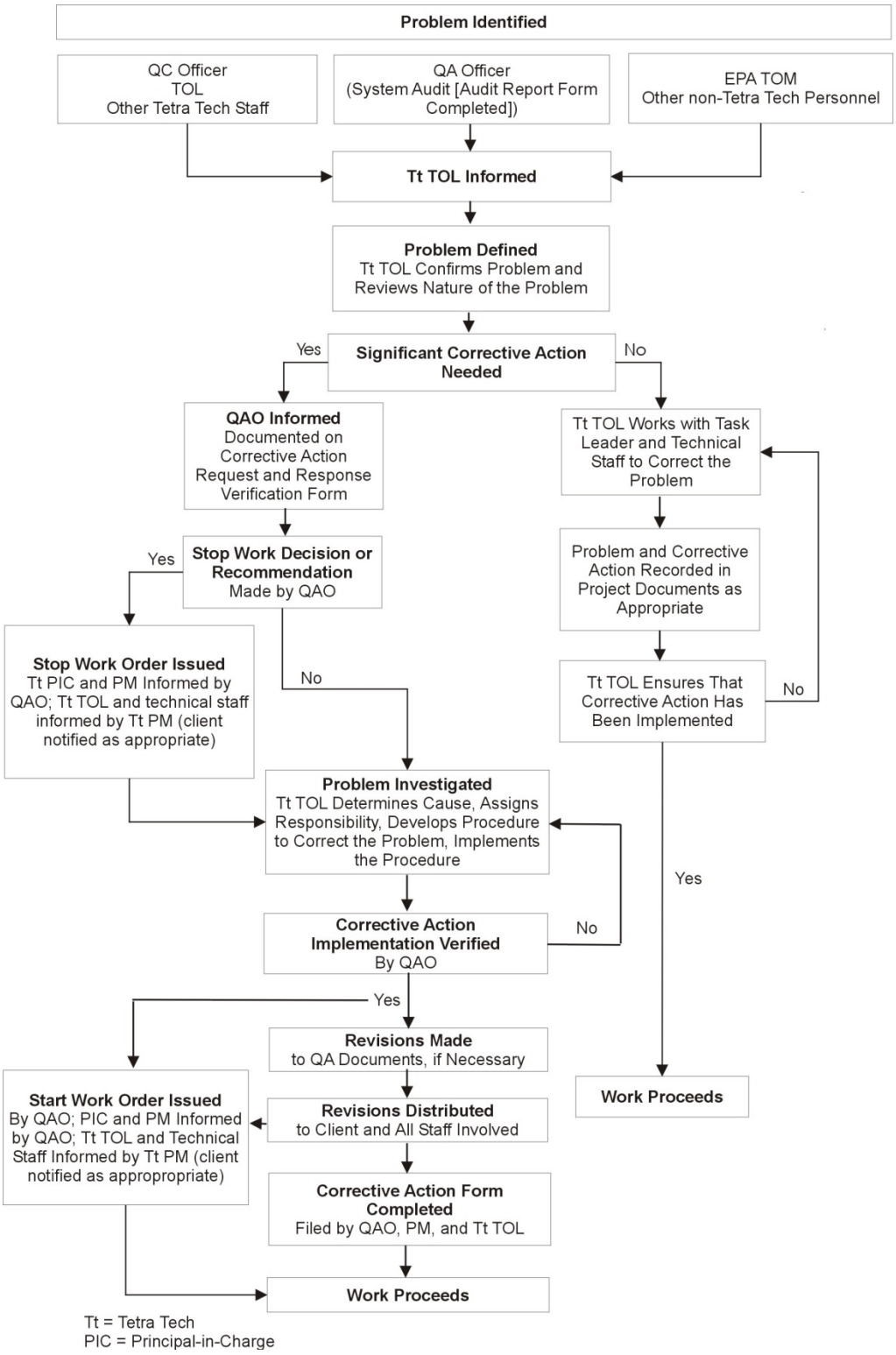

Figure 4. Problem assessment and correction operations

The essential steps in the QA program are as follows:

- Identify and define the problem
- Assign responsibility for investigating the problem
- Investigate and determine the cause of the problem
- Assign and accept responsibility for implementing appropriate corrective action
- Establish the effectiveness of and implement the corrective action
- Verify that the corrective action has eliminated the problem

Many of the possible technical problems can be solved on the spot by staff, for example, by modifying the Initial Technical Approach memorandum or correcting errors or deficiencies in implementation of the approach. Immediate corrective actions are considered SOPs, and they are noted in records for the project. Problems that cannot be solved in this way require more formalized, long-term corrective action.

If quality problems that require attention are identified, Tt will determine whether attaining acceptable quality requires either short- or long-term actions. If a failure in an analytical system occurs (e.g., performance requirements are not met), the Tt team modeling QC officers will be responsible for corrective action and will immediately inform the Tt Co-TOLs or the QAO, as appropriate. Subsequent steps taken will depend on the nature and significance of the problem, as illustrated in Figure 4.

Contract (name) _____

Date of Assessment _____ Request No. _____

Title (of project or other) _____

Project Leader _____ TC# _____

Other Responsible Personnel _____

Auditor or Initiator of This Corrective Action Request _____

Problem Description:

Recommended Action: **Date to Be Completed:**

_____ _____
Quality Assurance Officer Date

_____ _____
Principal-in-Charge or Program Manager Date

Action Taken: **Date:**

Verification of Completion of Corrective Action:

_____ _____
Quality Assurance Officer Date

_____ _____
Principal-in-Charge or Program Manager Date

Original form to be filed in QAO File; one copy to be filed in Project File and one copy in Contract File (if corrective action pertains to a project), or one copy to be filed in Contract File (if corrective action pertains to a contract).

Figure 5. Example corrective action request and response verification form

The Tt Co-TOLs have primary responsibility for monitoring the activities of this project and identifying or confirming any quality problems. The Co-TOLs will also bring these problems to the attention of the Tt QAO, who will initiate the corrective action system described above, document the nature of the problem (using a form such as that shown in Figure 5), and ensure that the recommended corrective action is carried out. The Tt QAO has the authority to stop work on the project if problems affecting data quality that will require extensive effort to resolve are identified.

The EPA COR, Tt PGM and Tt Co-TOLs will be notified of major corrective actions and stop work orders. Corrective actions can include the following:

- Reemphasizing to staff the project objectives, the limitations in scope, the need to adhere to the agreed-upon schedule and procedures, and the need to document QC and QA activities
- Securing additional commitment of staff time to devote to the project
- Retaining outside consultants to review problems in specialized technical areas
- Changing procedures

The Tt Co-TOLs may replace a staff member, as appropriate, if it is in the best interest of the project to do so.

Performance audits are quantitative checks on different segments of project activities; they are most appropriate for sampling, analysis, and data-processing activities. The Tt modeling QC officer is responsible for overseeing work as it is performed and periodically conducting internal assessments during the data entry and analysis phases of the project. As data entries, model codes, calculations, or other activities are checked, the Tt modeling QC officer will sign and date a hard copy of the material or complete Tt's standard Technical/Editorial Review Form, as appropriate, and provide it to the Tt Co-TOLs for inclusion in the administrative record. Performance audits will consist of comparisons of model results with observed historical data. Performing control calculations and post-simulation validation of predictions are major components of the QA framework.

The Tt Co-TOLs will perform or oversee the following qualitative and quantitative assessments of model performance periodically to ensure that the model is performing the required task while meeting the quality objectives:

- Data acquisition assessments
- Model calibration studies
- Sensitivity analyses
- Uncertainty analyses
- Data quality assessments
- Model evaluations
- Internal peer reviews

Internal peer reviews will be documented in the project and QAPP files. Documentation will include the names, titles, and positions of the peer reviewers; their report findings; and the project management's documented responses to their findings.

The Tt Co-TOLs will perform surveillance activities throughout the duration of the project to ensure that management and technical aspects are being properly implemented according to the

schedule and quality requirements specified in this QAPP. These surveillance activities will include assessing how project milestones are achieved and documented; corrective actions implemented; budgets adhered to; peer reviews performed; data managed; and whether computers, software, and data are acquired in a timely manner.

System audits are qualitative reviews of project activity to check that the overall quality program is functioning and that the appropriate QC measures identified in the QAPP are being implemented. If requested by the EPA COR, and EPA provides additional funding, the Tt QAO or designee will conduct an internal system audit of the project and report the results to the EPA COR and the Tt Co-TOLs.

8.5 DOCUMENTATION AND RECORDS

Thorough documentation of all modeling activities is necessary for interpreting study results. Tt will prepare monthly progress reports that will address task and subtask milestones, deliverables, adherence to schedule, and financial progression at the end of each full month while the task order for this project is still open. Data needs and deadlines for Tt's receipt of information needed to meet the project schedule will also be included in the progress reports and Gantt chart. The progress in meeting modeling QA targets (QA reports) will also be included in the progress reports. Other deliverables will be distributed to project participants as indicated by the EPA COR. Data tables, assumptions and analyses used to develop the models will be recorded and provided to EPA as a separate deliverable. The format of the raw data to be used for model parameters, model input, model calibration, and model output will be converted to the appropriate units, as necessary.

The Tt team will save on an external hard drive all modeling output data from all 20 watersheds as digital computer files in a file directory using a file-naming convention specified by the EPA COR. In addition, the Tt team will save on an external hard drive all scripts, project files, calibration data, and other information used to conduct watershed modeling at each of the 20 study watersheds. Tt will deliver these external hard drives to EPA within 2 weeks of the EPA COR's approval of the final report presenting and discussing the goals, methods, results and conclusions of watershed modeling in all 20 study watersheds (see the schedule in Table 2). Tt will maintain a copy of the project files at the Cincinnati, Ohio and/or Fairfax, Virginia, office for at least 3 years (unless otherwise directed by the EPA COR). The EPA COR and Tt Co-TOLs will maintain files, as appropriate, as repositories for information and data used in models and for preparing any reports and documents during the project. Electronic project files are maintained on network computers and are backed up periodically. The Tt Co-TOLs will supervise the use of project materials. The following information will be included in the electronic project files within Tt and on the external hard drives:

- Any reports and documents prepared
- Contract and task order information
- Electronic copies of model input/output (for model calibration and allocation scenarios)
- Results of technical reviews, model tests, data quality assessments of output data, and audits

- Documentation of response actions during the project to correct model development or implementation problems
- Assessment reports for acquired data
- Statistical goodness-of-fit methods and other rationale used to decide which statistical distributions should be used to characterize the uncertainty or variability of model input parameters
- Communications (memoranda; internal notes; telephone conversation records; letters; meeting minutes; and all written correspondence among the project team personnel, subcontractors, suppliers, or others)
- Maps, photographs, and drawings
- Studies, reports, documents, and newspaper articles pertaining to the project
- Spreadsheet data files including physical measurements, analytical chemistry data, and microbiological data (hard copy and on diskette)

The model application will include complete record keeping of each step of the modeling process. The documentation will consist of reports and files addressing the following items:

- Selection of study watersheds and model calibration points
- Assumptions
- Adjustments
- Parameter values and sources
- Nature of grid, network design, or subwatershed delineation
- Changes and verification of changes made in code
- Actual input used
- Output of model runs and interpretation
- Sensitivity analyses results
- Calibration and validation of the models

Formal reports submitted to EPA that are generated from the data will be maintained in the central file (diskette and hard copy) at Tt's Cincinnati, Ohio, and Fairfax, Virginia, offices. The data reports will include a summary of the types of data collected, sampling dates, and any problems or anomalies observed during sample collection.

8.6 OUTPUT ASSESSMENT AND MODEL USABILITY

Tt team technical staff will review model predictions for reasonableness, relevance, and consistency with the requirements of the model development process through model calibration as described in Section 8.0 of this QAPP. Tt team modeling experts will also determine consistency with the acceptance criteria described in Sections 7.0 and 8.0 of this QAPP. The Tt modeling QC officer will ensure that all steps of the modeling process are performed correctly.

Electronic copies of model input/output for model calibration, data quality assessments of output data, and QA reports will be maintained as part of the project files.

LITERATURE CITED

Donigian, A.S., Jr. 2000. HSPF Training Workshop Handbook and CD. Lecture #19. Calibration and Verification Issues, Slide #L19-22. EPA Headquarters, Washington Information Center, January 10–14, 2000. Presented and prepared for U.S. Environmental Protection Agency, Office of Water, Office of Science and Technology, Washington, DC.

Preston, S.D., V.J. Bierman, Jr., and S.E. Silliman. 1989. An evaluation of methods for the estimation of tributary mass loads. *Water Resources Research*, 25(6):1379–1389.

USEPA (U.S. Environmental Protection Agency). 2000. *Guidance for the Data Quality Objectives Process (G-4)*. EPA 600-R-96-055. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.

USEPA (U.S. Environmental Protection Agency). 2002. *Guidance for Quality Assurance Project Plans for Modeling (G-5M)*. EPA 240-R-02-007. U.S. Environmental Protection Agency, Office of Environmental Information, Washington, DC.