JOURNAL OF THE AMERICAN WATER RESOURCES ASSOCIATION
Vol. 43, No. 5 AMERICAN WATER RESOURCES ASSOCIATION October 2007

A HYDROLOGIC/WATER QUALITY MODEL APPLICATION PROTOCOL[1]

Bernard Engel, Dan Storm, Mike White, Jeff Arnold, and Mazdak Arabi[2]

ABSTRACT: This paper presents a procedure for standard application of hydrologic/water quality models. To date, most hydrologic/water quality modeling projects and studies have not utilized formal protocols, but rather have employed *ad hoc* approaches. The procedure proposed is an adaptation and extension of steps identified from relevant literature including guidance provided by the U.S. Environmental Protection Agency. This protocol provides guidance for establishing written plans prior to conducting modeling efforts. Eleven issues that should be addressed in model application plans were identified and discussed in the context of hydrologic/water quality studies. A graded approach for selection of the level of documentation for each item was suggested. The creation and use of environmental modeling plans is increasingly important as the results of modeling projects are used in decision-making processes that have significant implications. Standard modeling application protocols similar to the proposed procedure herein provide modelers with a roadmap to be followed, reduces modelers' bias, enhances the reproducibility of model application studies, and eventually improves acceptance of modeling outcomes.

(KEY TERMS: simulation; quality assurance/quality control; runoff; nonpoint source pollution.)

Engel, Bernard, Dan Storm, Mike White, Jeff Arnold, and Mazdak Arabi, 2007. A Hydrologic/Water Quality Model Application Protocol. *Journal of the American Water Resources Association* (JAWRA) 43(5):1223-1236. DOI: 10.1111/j.1752-1688.2007.00105.x

INTRODUCTION

Planning for modeling projects is just as important as planning traditional environmental measurements for data collection projects. If model predictions are to be used for regulatory purposes, research or design, then the modeling effort should be scientifically sound, robust, and defensible. To ensure this and to lead to confidence in results, the U.S. Environmental Protection Agency (USEPA, 2002) recommends a planning process that incorporates the following elements:

(1) A systematic planning process including identification of assessments and related performance criteria.
(2) Peer-reviewed theory and equations.
(3) Carefully designed life-cycle development processes that minimize errors.
(4) Documentation of changes from the original plan.

[1]Paper No. J05065 of the *Journal of the American Water Resources Association* (JAWRA). Received May 18, 2005; accepted February 5, 2007. **Discussions are open until April 1, 2008**.

[2]Respectively, Professor, Agricultural and Biological Engineering, Purdue University, 225 S. University Street, West Lafayette, Indiana 47907; Professor and Research Associate, 121 Agricultural Hall, Oklahoma State University, Stillwater, Oklahoma 74078-6016; Agricultural Engineer, USDA-ARS, 808 East Blackland Road, Temple, Texas 76502; and Postdoctoral Research Associate, Agricultural and Biological Engineering, Purdue University, 225 S. University Street, West Lafayette, Indiana 47907 (E-mail/Engel: engelb@purdue.edu).

(5) Clear documentation of assumptions, theory, and parameterization that is detailed enough so others can fully understand the model predictions.
(6) Input data and parameters that are accurate and appropriate for the application.
(7) Model prediction data that can be used to help inform decision-making.

A Quality Assurance Project Plan and good project management in modeling projects are closely linked. A good Quality Assurance Project Plan documents all criteria and assumptions in one place for easy review and reference. The plan can be used to guide project personnel through the model development or application process and helps ensure that choices are consistent with project objectives and requirements. However, it should be noted that many assumptions and decisions cannot be made until the modeling effort is underway. A well prepared plan can be helpful in providing guidance in such situations. Assumptions and decisions made during the modeling process should be documented.

Quality assurance (QA) in hydrologic modeling is the procedural and operational framework put in place by the organization managing the modeling study to ensure adequate execution of all project tasks, and to ensure that all modeling-based analyses are verifiable and defensible (Taylor, 1985). The two major elements of QA are quality control (QC) and quality assessment. QC addresses the procedures that ensure the quality of the final product. The procedures include: (1) the use of appropriate methodology in developing and applying computer models; (2) suitable verification, calibration, and validation procedures; and (3) proper use of the methods and model. Quality assessment is applied to monitor the QC procedures (van der Heijde, 1987).

Use of a modeling protocol provides several potential benefits to projects that include a significant modeling component. These include: (1) reduces potential modeler bias, (2) provides a roadmap to be followed, (3) allows others to assess decisions made in modeling the system of interest, (4) allows others to repeat the study, and (5) improves acceptance of model results.

A modeling protocol, preferably written, should be established prior to conducting a modeling study. To date, most hydrologic/water quality modeling projects and studies have not utilized formal modeling protocols, but rather *ad hoc* approaches are typically employed. The goal of this paper is to define the content of a modeling protocol or a modeling QA plan that can be used to help hydrologic/water quality modelers establish such protocols for their modeling projects.

LITERATURE REVIEW

In following the scientific method, steps should be taken to minimize the potential influence of scientists' possible bias. The use of a modeling protocol or a QA plan in modeling projects can provide the documentation needed to assess the project and can be helpful in reducing potential bias. By definition, the scientific method is impartial and the results from the application of the scientific method must be reproducible. Therefore, the modeling protocol and associated documentation must provide enough detail to allow the modeling project to be repeated. It should be noted that models are not hypotheses, but are simply tools that are used to evaluate a hypothesis. As applied to hydrologic modeling, the steps in the scientific method may be given as follows:

(1) Based on existing theory and data, develop a hypothesis that is consistent with the current understanding of the system being modeled.
(2) Based on the hypothesis, make predictions by applying an appropriate hydrologic model.
(3) Test the hypothesis by comparing model predictions with observed data.
(4) Accept or reject the hypothesis based on appropriate criteria.
(5) If needed, modify the hypothesis and repeat Steps 2-5.

Refsgaard (1997) defined a modeling protocol as depicted in Figure 1. Refsgaard makes a distinction between a model and a model code; a model is any hydrologic model established for a particular watershed. Others might refer to Refsgaard's definition of a model as a model "setup" or a "parameterized" model. Refsgaard (1997) defined a model code as a generalized software package, which without changes, can be used to establish a model with the same basic types of equations (but allowing different parameter values) for different watersheds. Refsgaard (1997) defined model validation as the process of demonstrating that a given site-specific model is capable of making "sufficiently accurate" predictions, where "sufficiently accurate" will vary by application and project needs. A model is considered validated if its accuracy and predictive capability in the validation period have been proven to lie within acceptable limits. Again, acceptable limits will vary by application and project requirements. Interestingly, Refsgaard (1997) does not include a model sensitivity analysis in his steps. Sensitivity analyses, discussed in more detail later in the paper, can be helpful for a variety of purposes in modeling projects.

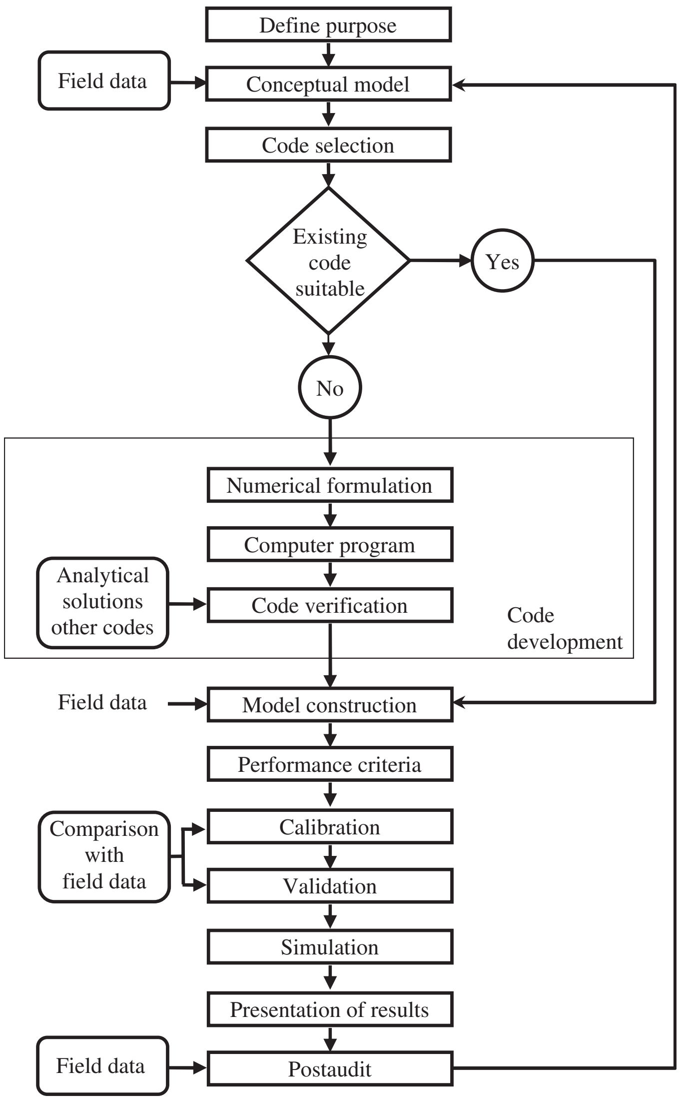

FIGURE 1. A Hydrological Model Protocol as Proposed by Refsgaard (1997).

Developing efficient and reliable hydrologic/water quality models and applying them requires numerous steps, each of which should be taken conscientiously and reviewed carefully. Taking a systematic, well-defined and controlled approach to all steps of the model development and application process is essential for successful model implementation. QA provides the mechanisms and framework to ensure that decisions made during this process are based on the best available data and analyses.

USEPA Quality Assurance

The USEPA uses the Quality Assurance Project Plan to help project managers and planners document the type and quality of data and information needed for making environmental decisions. The USEPA (2002) has developed a document, *Guidance for Quality Assurance Project Plans for Modeling* (EPA QA/G-5 M), to provide recommendations on how to develop a Quality Assurance Project Plan for projects involving modeling (e.g., model development, model application, as well as large projects with a modeling component). A "model" is defined by USEPA as something that creates a prediction. The guidance regarding modeling is based on recommendations and policies from USEPA Quality Assurance Project Plan protocols, but is written specifically for modeling projects. However, modeling projects have different QA concerns than traditional environmental monitoring data collection projects. The structure for the Quality Assurance Project Plan for modeling is consistent with the *EPA Requirements for Quality Assurance Project Plans* (QA/R-5) (USEPA, 2001) and *EPA Guidance for Quality Assurance Project Plans* (QA/G-5) (USEPA, 1998), though for modeling not all elements are included because not all are relevant.

The USEPA Quality System defined in USEPA Order 5360.1 A2 (USEPA, 2000), *Policy and Program Requirements for the Mandatory Agency-Wide Quality System*, includes environmental data produced from models. Environmental data includes any measurements or information that describes environmental processes, location, or conditions, ecological or health effects and consequences, or the performance of environmental technology. As defined by USEPA, environmental data includes information collected directly from measurements, produced from models, or compiled from other sources, such as databases or literature. The USEPA Quality System is based on the American National Standard ANSI/ASQC E4-1994.

Graded Approach to QA Project Plans

USEPA defines the graded approach as "the process of basing the level of application of managerial controls applied to an item or work according to the intended use of the results and degree of confidence needed in the quality of the results" (USEPA, 1998). This allows the application of QA and QC activities to be adapted to meet project specific needs. Models that provide an initial "ballpark" estimate or nonregulatory priorities, for example, would likely not require the same level of QA and planning as would models that will be used to set regulatory requirements. However, USEPA provides no explicit categorizations or other specific guidelines for applying the graded approach (USEPA, 2002).

In applying the graded approach, USEPA suggests two aspects that are important for defining the level of QA that a modeling project needs: (1) intended use of the model and (2) the project scope and magnitude (USEPA, 2002). The intended use of the model is a determining factor because it is an indication of the potential consequences or impacts that might occur because of the QC problems. For example, higher standards might be set for projects that involve potentially large consequences, such as Congressional testimony, development of new laws and regulations, or the support of litigation. More modest levels of defensibility and rigor would often be acceptable for data used for technology assessment or "proof of principle," where no litigation or regulatory action are expected. Still lower levels of defensibility would likely apply to basic exploratory research requiring extremely fast turnaround, or high flexibility and adaptability. In such cases, the work may have to be replicated under more stringent controls or the results carefully reviewed prior to publication. The USEPA (2002) suggests peer review may be substituted, to some extent, for the level of QA. By analyzing the end-use needs, appropriate QA criteria can be established to guide the program or project. The examples presented are for illustration only, and the degree of rigor needed for any particular project should be determined based on an evaluation of the project needs and resources.

Other aspects of the QA effort can be established by considering the scope and magnitude of the project. The scope of the model development and application determines the complexity of the project; more complex models or modeling projects likely need more QA effort. The magnitude of the project defines the resources at risk if quality problems lead to rework and delays.

The QA Project Plan Elements for a Model Application Project

The USEPA (2002) defined the following nine model application tasks and mapped them into Quality Assurance Project Plan elements: (1) needs assessment; (2) purpose, objectives, and output specifications; (3) define quality objectives, desired performance criteria, and documentation needs for model output; (4) select the most appropriate model; (5) data development, model parameterization, and model calibration; (6) determine whether data, models, and parameters for the application meet desired performance criteria; (7) run the computer code; (8) model output testing and peer review; and (9) summarize results and document. Further details on how these modeling tasks fit within a potential modeling QA plan are described in detail in *Guidance for Quality Assurance Project Plans for Modeling* (USEPA, 2002).

In this paper, we develop a standard protocol for conducting modeling efforts with an emphasis on hydrologic/water quality modeling. To this end, the work of USEPA (2002) and other relevant literature were reviewed to arrive at a preliminary list of recommended steps in conducting modeling studies. These steps were further extended based on the authors' experience to include issues, such as representation of nonpoint source (NPS) best management practices (BMPs). A detailed description of the proposed steps follows.

MODEL APPLICATION PROTOCOL STEPS

A hydrologic/water quality model application protocol is proposed based on the authors' experiences and review of the literature, including the USEPA (2002) *Guidance for Quality Assurance Project Plans for Modeling* document. The authors recognize that a "graded" approach in implementing a modeling protocol will be required, and thus not all modeling QA plans will include all sections or issues suggested. Further, the level of detail in such plans will vary greatly depending on the purpose of the model application project. The USEPA (2002) suggests that a graded approach can be used to define the level of QA effort that a modeling project needs based on the intended use of the model and the project scope and magnitude (Table 1).

The following items or sections should be included in a hydrologic/water quality modeling protocol: (1) problem definition/background; (2) model application goals, objectives, and hypothesis; (3) model selection; (4) model sensitivity analysis; (5) available data; (6) data to be collected; (7) model representation issues – data, BMPs, etc.; (8) model calibration; (9) model validation; (10) model scenario prediction; and (11) results interpretation/hypothesis testing.

The proposed modeling protocol steps may be iterative. For example, the scientific literature and a preliminary sensitivity analysis using general data may initially be used to identify the model parameters that are the most sensitive. A more comprehensive sensitivity analysis assessment may be performed later once more detailed location specific data have been collected or obtained.

Decisions made throughout the modeling effort and the rationale for these decisions should be documented. In most instances, it will be necessary to make various assumptions and decisions throughout the modeling project. Many of these assumptions are best made during the modeling project rather than before the modeling starts, as information from prior steps may impact decisions. The amount of documentation that should be created depends on the project

TABLE 1. Examples of Modeling Projects With Differing Intended Uses (adapted from USEPA, 2002).

Purpose for Obtaining Model-Generated Information (intended use)	**Typical QA Issues**	**Level of QA**
Regulatory compliance	Legal defensibility of data sources	
Litigation	Compliance with laws and regulatory mandates applicable	
Congressional testimony	to data gathering	
Regulatory development	Compliance with regulatory guidelines	
State Implementation Plan (SIP) attainment	Existing data obtained under suitable QA program	
Verification of model	Audits and data reviews	
Trends monitoring (nonregulatory)	Use of accepted data-gathering methods	
Technical development	Use of widely accepted models	
"Proof of principle"	Audit and data reviews	
Basic research	QA planning and documentation at the facility level	
Bench-scale testing	Peer review of novel theories and methodology	

goals and the consequences of decisions that will be made as a result of the project findings. Each of the modeling protocol steps is discussed in more detail in the sections that follow.

Item 1. Problem Definition/Background

Background information and preliminary data for the study area should be obtained to help initially define the overall problem that will be addressed by the study. The background information and data collected in this step will be useful to determine whether modeling will be necessary, assist in defining the modeling objectives (if modeling is required) and to select the model or models to be used. More detailed objectives or hypotheses to be examined within the project are defined in the subsequent step. This initial step is similar to the initial observation phase commonly employed within the scientific method.

Questions that may be addressed when defining the problem include the following:

(1) What is the specific problem?
(2) What are the overall goals and objectives of this project that will address this problem?
(3) Why should a modeling approach be used to address the problem?
(4) How will modeling of the problem help to address the overall goals of the project?

It is also important to place the problem in context to provide a sense of the project's purpose relative to other project and program phases and initiatives. Questions that might be addressed include the following:

(1) What information, previous work, or previous data may currently exist that this project can use?
(2) Given that the problem is best solved by a modeling approach, what models currently exist (if any) that can be used to achieve this project's goals and objectives?
(3) What are the advantages and disadvantages of using these models?

The presentation of background information may also include a discussion of initial ideas or potential approaches for model application.

Item 2. Model Application Goals, Objectives, and Hypothesis

The specific objectives and/or hypotheses to be accomplished or tested by the modeling effort are defined based on the background information and data collected in the first step. The objectives or hypotheses should be stated in a manner that they can be tested or evaluated using the model predictions.

In setting the objectives or hypotheses to be tested, one should keep in mind that models are typically more accurate when making relative comparisons rather than making absolute predictions. Thus, an objective or hypothesis might be written to compare expected pollutant losses for different tillage systems rather than examining whether a particular tillage system results in pollutant losses below a given magnitude. Model calibration can help improve the accuracy of absolute predictions, but adequate data for calibration that represent the range of conditions of interest for the location of interest are not always available.

A summary of the work to be performed and the "products" to be created by the model application effort should be identified. These will be described in more detail in subsequent sections.

Item 3. Model Selection

An appropriate model should be selected based on (1) the project goals, objectives or hypotheses; (2) how model results will be used; (3) the characteristics of the hydrologic/water quality system that are important to the objectives or hypotheses; and (4) various other factors including: (a) appropriate level of detail (space and time); (b) important chemical, physical, and biological processes are included; (c) calibration requirements; (d) data requirements and availability; (e) previous applications of the model and acceptance in the scientific, regulatory, and stakeholder communities; (f) ease of use; (g) sensitivity to processes of interest; and (h) Available resources (e.g. modeler expertise, model technical support) and time.

Item 4. Model Sensitivity Analysis

A model sensitivity analysis can be helpful in understanding which model inputs are most important or sensitive and in understanding potential limitations of the model. Additional care should be taken when estimating model parameters that are the most sensitive. Data collection efforts that support the modeling study may focus on obtaining better data for the most sensitive parameters.

The sensitivity analysis can also identify potential limitations of the model. If a model is not sensitive to parameters that are to be varied in testing the project objectives or hypotheses, a different model may need to be selected. Models are abstractions of the systems they simulate and therefore typically represent system components with varying levels of detail. For example, the scientific literature may indicate that differences in tillage practices influence pesticide losses in surface runoff. In such a case, the use of a model that is not sensitive to tillage to examine the impact of switching from conventional tillage to conservation tillage on pesticide losses in surface runoff is likely inappropriate.

The literature and model documentation are often excellent sources of information on model sensitivity. For example, Muttiah and Wurbs (2002) identified the sensitivity of Soil and Water Assessment Tool (SWAT) to various parameters. However, it may be necessary to conduct a sensitivity analysis for the study watershed if its conditions are significantly different than those for model sensitivity analyses reported in the literature, as model sensitivity may be specific to the model setup. Thus, limited data for parameterizing the model may need to be collected prior to conducting a sensitivity analysis. Generally, the sensitivity analysis should be completed using an uncalibrated model setup, as the sensitive parameters and those with the greatest uncertainty are typically used for model calibration. For example, Spruill *et al.* (2000) conducted a SWAT sensitivity analysis to evaluate parameters that were thought to influence stream discharge predictions. During calibration, the average absolute deviation between observed and simulated streamflows was minimized and used to identify optimum values or ranges for each parameter.

Item 5. Available Data

The goal of this step is to select the most appropriate data for the modeling effort. Data available for the modeling effort will likely come from numerous sources. An assessment of available data, its quality, and the time period it covers should be made. The amount of data available for a watershed can vary greatly, as can the quality of the data. For example, flow and water quality data may be available for 1983 through 1988, while land use data might have been developed for conditions in 1995. This may result in a misrepresentation of the land uses that were present during the observed flow and water quality data period, especially for areas experiencing rapid urbanization. In other instances, differences in data collected at different dates may be negligible. For example, soil property data used in modeling runoff from a watershed would not typically change significantly over time, even over periods of tens of years. In instances where data, such as land use, may have changed significantly, it may be necessary to estimate data for the period of interest by interpolating between datasets for different time periods or by adjusting the data from the available time period using other sources of data and information.

The USEPA (2002) indicates that a Quality Assurance Project Plan for modeling should address the following issues regarding information on how nondirect measurements (data and other information that have been previously collected or generated under some effort outside the specific project being addressed by the Quality Assurance Project Plan) are acquired and used in the project:

1. The need and intended use of each type of data or information to be acquired;
2. How the data will be identified or acquired, and expected sources of these data;
3. The method of determining the underlying quality of the data;
4. The criteria established for determining whether the level of quality for a given set of data is acceptable for use in the project.

Water quality and runoff data for the study watershed may be available from federal, state or local government agencies. For example, the U.S. Geological Survey is often an excellent source of streamflow data and the USEPA STORET database may provide useful water quality data. Datasets may also be available from past studies, and often are documented in project reports. In many instances, these data will not be identified by simply conducting a literature search; rather contacts with local universities, state and local agencies, and local watershed groups will likely be necessary.

Well documented and widely used datasets, such as soil properties from the U.S. Department of Agriculture Natural Resources Conservation Service, often have well understood properties and degree of uncertainty. It is useful to understand this degree of uncertainty, the assumptions in the data and how these will likely impact the model results. Spatial or geographic information systems (GIS) data may be available from federal, state and local government agencies. Increasingly, county and local governments are developing detailed spatial datasets. For example, many county governments with urban areas have developed detailed elevation datasets that provide more detail than state and national elevation datasets. Spatial data from these sources should have metadata available that describe the accuracy and other properties of the data that will be helpful in understanding the data quality and limitations.

Remotely sensed datasets from satellites and aerial photography can potentially provide land use and other data needed in hydrologic/water quality modeling studies. In addition, archived satellite data and aerial photography may be useful in creating land use information for the past. Remotely sensed datasets will require interpretation to create the land use or other data that are needed. Accuracy assessments of the interpreted results should be performed to provide information concerning the quality of the land use products created.

The scientific literature may contain some information about the study area. Project reports, however, are more likely to contain the detailed data typically required for a model application project. Scientific papers may also provide insight into transformation of various data into the data or parameters required by the model. In most cases, these data must be transformed into values and formats required by the model.

After identifying the data available and its various properties, including quality and temporal aspects, an assessment of the suitability of the data for use in the model that has been selected must be made. The model data requirements and the sensitivity of the model to various parameters should be considered when evaluating and selecting the data to use. The rationale for the data selected for use in the model should be well documented, as should any required data transformation.

The scientific literature contains numerous studies on the impacts that various data sources and data errors can have on model results. For example, Chaubey *et al.* (1999) explored the assumption of spatial homogeneity of rainfall when parameterizing models and concluded large uncertainty in estimated model parameters can be expected if detailed variations in the input rainfall are not considered. Nandakumar and Mein (1997) examined the levels of uncertainty in rainfall-runoff model predictions as a result of errors in hydrological and climatic data, and considered the implications for prediction of the hydrologic effect of landuse changes. Studies, such as these highlight the importance of understanding the consequences of the data used in the project on the model results and their interpretation.

Item 6. Data to Be Collected

Based on the project objectives and hypotheses, available data and model sensitivity should be considered in deciding what, if any, additional data should be collected. After assessing these issues, the modeler may conclude that additional data should be collected. Following calibration or validation, the modeler may also decide that additional data should be collected in an attempt to improve model performance. The collection of additional data can be expensive, as well as require a significant amount of time. An appropriate QA plan for the collection of additional data should be prepared and followed (USEPA, 1998).

Item 7. Model Representation Issues – Data, Best Management Practices, etc.

Models are abstractions of the systems they are simulating. Therefore, the modeler will be required to make decisions on how to represent the various components of the system being modeled. This may include decisions on representation of components within the model and in the transformation of available data into the formats needed by the model. These decisions should be documented. The expected effect of these assumptions on the results, relative to alternative assumptions that could have been made should also be documented.

One of the data representation issues typically faced is related to pollutant sources. It is typically impossible to include all pollutant sources in the modeling effort. For example, if the amount of phosphorus leaving a watershed is of interest, the modeler may decide not to include phosphorus losses from septic systems, if they are small relative to other sources. Criteria should be established to determine which pollutant sources to include in the model and/or overall analysis. A simple mass balance for water and pollutants of interest may be helpful to identify the most important components of the hydrologic cycle and system to model and the most important sources of pollutants to consider. Another option is to use the selected model to perform a simple preliminary model simulation using limited data. Based on such a model, criteria to exclude pollutant sources that represent less than 5% (or other levels deemed appropriate) of the pollutant might be established. It should be noted, however, that pollutant sources less than this threshold may be included if these data are readily available and easy to incorporate into the model. When potential pollutant sources are not incorporated into the model, care in the interpretation of the final model results is required.

The representation of BMPs within the model may not be well defined. Model documentation and the scientific literature can often provide guidance in BMP representation (Bracmort *et al.*, 2006). However, in most instances, these sources do not fully describe how a specific BMP, such as a grassed waterway, should be represented within a particular model; rather the modeler must exercise judgment in the BMP representation decision. Therefore, the modeler will need to determine how BMPs will be represented in the application of a model to a given location.

The accuracy of hydrologic/water quality models also depends in part on how well model input parameters describe the relevant characteristics of the watershed. Data that are obtained for a watershed will typically require some transformation and interpretation to create the inputs required by the model. For example, soil properties in the SSURGO database are often reported with a range of values, while the model will require a single value for each soil property. The model documentation and scientific literature can often provide guidance in transforming commonly available data into the inputs required by the model. These data were used and the decisions made in data transformations should be documented.

Input parameter aggregation may have a substantial impact on model output. For example, FitzHugh and Mackay (2000) used SWAT to determine how the size or number of subwatersheds used to partition the watershed affect model output and the processes responsible for model behavior. Mankin *et al.* (2002) explored the errors introduced when translating GIS data into model-input data. Watershed modelers using GIS data should be aware of the issues related to appropriate grid cell sizes, generation of land-management practice GIS coverages, accuracy of GIS data, and accuracy of interface algorithms.

Refsgaard and Storm (1996) indicated that a rigorous model parameterization procedure is crucial to avoid methodological problems in subsequent phases of model calibration and validation. They suggest the following points are important to consider in model parameterization:

(1) Parameter classes (soil types, vegetation types, etc.) should be selected so it is easy, in an objective way, to associate parameter values. Thus, when possible parameter values in the classes should be determined based on available field data.
(2) Determine which parameters can be assessed from field data or the literature and which will require calibration. For parameters subject to calibration, the physically acceptable intervals for the parameter values should be estimated and documented.
(3) The number of calibration parameters should be minimized both from practical and methodological points of view. Fixing a pattern for a spatially varying parameter but allowing its value to be modified uniformly throughout the watershed can help minimize the number of calibrated parameters.

Item 8. Model Calibration

The USEPA (2002) indicates that if no nationally recognized calibration standards exist, the basis for the calibration should be documented. Quality Assurance Project Plan guidance indicates that calibration for data collection efforts address calibration of the analytical instruments that will be utilized to generate analytical data. In modeling projects, by analogy, the "instrument" is the predictive tool (the model) that is to be applied (USEPA, 2002). All models, by definition, are a simplification of the processes they are intended to represent. When formulating the mathematical representations of these processes, there are relationships and parameters that need to be defined. Estimating parameters for these relationships is called *calibration*. Some model parameters may need to be estimated for every application of the model using site-specific data. Similar to an analytical instrument, models are calibrated by comparing the predictions (output) for a given set of assumed conditions to observed data for the same conditions. This comparison allows the modeler to evaluate whether the model and its parameters reasonably represent the environment of interest. Statistical methods typically applied when performing model calibrations include regression analyses and goodness-of-fit methods. An acceptable level of model performance should be defined prior to the initiation of model calibration. The details of the model calibration procedure, including statistical analyses that are involved, should be documented.

Calibration Procedures

Model calibration is often important in hydrologic modeling studies, as uncertainty in model predictions can be reduced if models are properly calibrated. Factors contributing to difficulties in model calibration include calibration data with limited metadata, data with measurement errors, and spatial variability of rainfall or watershed properties poorly represented by point measurements. Model calibration can be done manually or by a combination of manual and automatic procedures. Manual calibration can be subjective and time-consuming (Eckhardt and Arnold, 2001). Initial values can be assigned to parameters, which are then optimized by an automatic procedure (Gan *et al.*, 1997). Chanasyk *et al*. (2002) calibrated SWAT until the predicted and observed results were visibly close. Many studies use comparable *ad hoc* approaches in calibration. However, approaches that use only visual comparison should be avoided. One of the advantages of an automated approach to calibration is that it uses a systematic approach in adjusting the model parameters, thereby removing potential modeler bias. With an *ad hoc* calibration approach, the modeler could potentially adjust model parameters during calibration that would create a model setup or parameterization that would be more likely to provide desired results when testing the project objectives or hypotheses.

Santhi *et al.* (2001a) presented a flow chart with the decision criteria used during the calibration of SWAT. This flowchart has been adapted by Arabi *et al.* (2004) and others for calibration of SWAT, and an adapted version is presented in Figure 2. In some instances, this approach is too rigid to be strictly followed because of the interactions between model parameters, and thus the modeler may need to deviate from strictly following such an approach.

The approach that will be followed in calibrating the model should be identified prior to beginning calibration. Performance criteria should also be established prior to beginning model calibration so that the modeler knows when the model has been successfully calibrated. The scientific literature can often provide an idea of the likely performance of the model following calibration. Statistical measures can be used to identify performance criteria for determining whether the model has been calibrated successfully. For some efforts, an *ad hoc* calibration approach may be acceptable, while in other instances it will be desirable to have a specific calibration protocol.

For projects supporting regulatory decision-making, the USEPA (2002) suggests the level of detail on model calibration in the Quality Assurance Project Plan should be sufficient to allow another modeler to duplicate the calibration method, if the modeler is given access to the model and to the data being used in the calibration process. For other projects (e.g., some basic research projects), it may be acceptable to provide less detail on this issue for the Quality Assurance Project Plan. In some instances, projects may use procedures that are somewhat different from standard calibration techniques, such as "benchmarking" procedures, and therefore the level of detail may differ from what is generally portrayed for calibration.

Examples of features that the model calibration portion of the Quality Assurance Project Plan may address include the following:

(1) Objectives of model calibration activities, including acceptance criteria;
(2) Details on the model calibration procedure;
(3) Method of acquiring the input data;
(4) Types of output generated during model calibration;
(5) Method of assessing the goodness-of-fit of the model calibration equation to calibration data;
(6) Method of quantifying variability and uncertainty in the model calibration results; and
(7) Corrective action to be taken if acceptance criteria are not met.

The calibration plan should identify the parameters that will be adjusted, the order in which they will be adjusted, and ranges in which the adjusted parameters must fall. The ranges of parameters used in calibration and the calibration results obtained should be documented during calibration.

Not all models must be calibrated prior to using the model to test the project objectives or hypotheses. However, in most cases, calibration of the model for the study watershed(s) conditions can reduce the uncertainty in model predictions. If models are not calibrated, they should still be validated for the study watershed if data are available.

For hydrologic/water quality models, the hydrologic components are usually calibrated first. In the calibration of the hydrologic components of the model, it may be necessary to separate streamflow into direct or surface runoff and base flow. In balancing surface runoff and base flow volumes for a system of interest, other components associated with the hydrologic component could be ultimately balanced. The

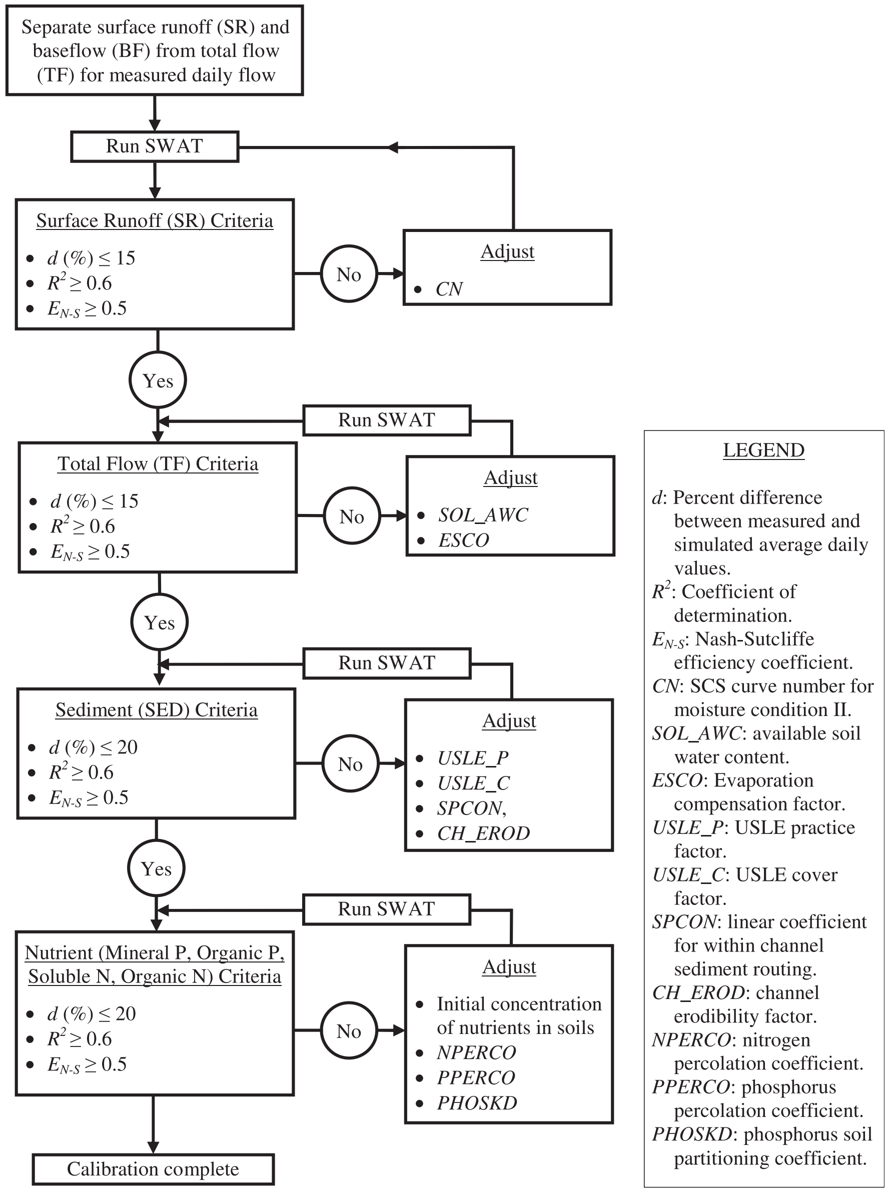

FIGURE 2. Example SWAT Calibration Flowchart (adapted from Santhi *et al.*, 2001a).

model is typically calibrated first to obtain acceptable performance in the hydrologic components, then for sediment, and finally for nutrients, pesticides, bacteria, or other constituents.

Calibration Data

Data that will be used for calibration should be identified. One common method is to split observed data into one dataset for calibration and one for validation. It is important that the calibration and validation datasets each have observed data of approximately the same magnitudes. For example, both calibration and validation datasets should have periods with high and low flows in order to increase the robustness of the model.

Yapo *et al.* (1996) used varying lengths of calibration data and found that approximately eight years of data were needed to obtain calibrations that were insensitive to the calibration period selected for their watershed. Gan *et al.* (1997) indicated that ideally, calibration should use three to five years of data that include average, wet, and dry years so that the data encompass a sufficient range of hydrologic events to activate all the model components during calibration. However, the required amount of calibration data is project specific.

Calibration Statistics

The goodness-of-fit statistics to be used in describing the model's performance relative to the observed data should be selected prior to calibration and validation. The American Society of Civil Engineers (ASCE) Task Committee (1993) recommended graphical and statistical methods useful for evaluating model performance. In most instances, both visual comparisons of predicted and observed data, as well as goodness-of-fit statistics, should be used. Plotting of predicted results and observed results along with the 1:1 line can be helpful in identifying model bias. The percent deviation of predicted values from observed values is one numerical goodness-of-fit criterion. A second basic goodness-of-fit criterion recommended by the ASCE Task Committee (1993) is the Nash-Sutcliffe coefficient or coefficient of simulation efficiency. Legates and McCabe (1999) evaluated various goodness-of-fit measures for hydrologic model validation and suggested that correlation and correlation-based measures (e.g., the coefficient of determination) are oversensitive to extreme values and are insensitive to additive and proportional differences between model estimates and observed values. Thus, correlation-based measures can indicate that a model is a good predictor, even when it is not. Legates and McCabe (1999) concluded that measures, such as the Nash-Sutcliffe coefficient of efficiency and the index of agreement are better measures for hydrologic model assessment than correlation-based measures. Legates and McCabe (1999) suggested a modified Nash-Sutcliffe coefficient that is less sensitive to extreme values may be appropriate in some instances. They also suggested additional evaluation measures, such as summary statistics and absolute error measures (observed and modeled means and standard deviations, mean absolute error and root mean square error) should be reported for model results.

There are no standards or a range of values for goodness-of-fit statistical parameters that will adjudge the model performance as acceptable (Loague and Green, 1991). Ramanarayanan *et al.* (1997) suggested values of goodness-of-fit statistics computed based on monthly computations for determining the acceptable performance of the APEX model. They indicated that values close to zero for the correlation coefficient and/or the Nash-Sutcliffe coefficient indicated the model performance was unacceptable or poor. They judged the model performance as satisfactory if the correlation coefficient was greater than 0.5 and the Nash-Sutcliffe coefficient was greater than 0.4. Santhi *et al.* (2001a) assumed a Nash-Sutcliffe coefficient greater than 0.5 and a goodness of fit (R^2) greater than 0.6 indicated acceptable model performance when calibrating SWAT. However, acceptable statistical measures are project specific.

The literature can provide typical ranges of goodness-of-fit statistics for models. For example, Saleh *et al.* (2000) obtained Nash-Sutcliffe coefficients for average monthly flow, sediment, and nutrient loading at 11 locations with values ranging from 0.65 to 0.99, indicating reasonable SWAT predicted values. SWAT also adequately predicted monthly trends in average daily flow, sediment, and nutrient loading over the validation period with Nash-Sutcliffe coefficients ranging from 0.54 to 0.94, except for NO_3-N, which had a value of 0.27. Fernandez *et al.* (2002) developed a GIS-based, lumped parameter water quality model to estimate the spatial and temporal nitrogen-loading patterns for lower coastal plain watersheds in eastern North Carolina. Predicted nitrogen loads were highly correlated with observed loads (correlation coefficients of 0.99 for nitrate-nitrogen, 0.90 for total Kjeldahl nitrogen, and 0.96 for total nitrogen). However, the limitations of correlation coefficients, as discussed previously, should be considered in interpretation of these results. Spruill *et al.* (2000) evaluated SWAT and its parameter sensitivities for streamflow from a small central Kentucky watershed and concluded the model adequately predicted the trends in daily streamflow, although Nash-Sutcliffe coefficient values were 0.19 for calibration and −0.04 for validation. The Nash-Sutcliffe coefficients for monthly total flows were 0.58 for validation and 0.89 for calibration.

In some instances, model calibration may not yield results that are acceptable based on the predefined model performance criteria. If this occurs, the observed flow and pollutant data as well as the model input data should be examined for potential errors. The poor model performance may be an indication that more detailed model inputs are required. In other cases, this may be an indication that the model is unable to adequately represent the processes of interest for this watershed.

Item 9. Model Validation

When possible, it is important to reserve some observed data (e.g., flow and water quality data) for model validation. Additional discussion of the data for validation and calibration can be found in the Model Calibration section. Prior to beginning model validation, the criteria used to validate, that is, accept, reject, or qualify the model results, should be documented (USEPA, 2002). The same statistics used and reported for model calibration should be used in model validation. Typically, the values of these statistics are lower for validation than calibration. Acceptable levels of performance may be difficult to identify. Acceptable model performance levels that have been proposed are discussed in the Model Calibration section. The scientific literature can provide suggestions for levels of performance that might be anticipated for a given model. The specific purpose of the study, the available data and other factors should be considered when establishing the performance criteria. For example, the time period considered can impact model performance. Typically, model performance is poorer for shorter periods than for longer periods (e.g., daily *vs*. monthly or yearly). For example, Yuan *et al.* (2001) found that AnnAGNPS provided an R^2 of 0.5 for event comparison of predicted and observed sediment yields, while the agreement between monthly data had an R^2 of 0.7.

In some instances, acceptable model performance may not be obtained during the validation step. Note that the utility of the model may not depend on a single performance indicator, and therefore, some judgment will be required by the modeler. The potential uncertainty associated with models and model setups that do not attain the desired level of performance during validation will be greater than those for which model performance is deemed acceptable. Unacceptable model performance for validation can be an indication that the validation period data ranges or conditions are significantly different than those for the calibration period. Therefore, care in the selection of the data for calibration and validation periods is needed. In other cases, poor performance during validation may be an indication that the model has not been adequately or properly calibrated. It is possible that numerous model setups or parameterizations can provide acceptable model results for calibration. However, during validation such setups may provide poor results. In such cases, the model should be re-calibrated and then validation attempted again. In addition, in some cases, the lack of acceptable validation may be the result of inaccurate validation data.

If data are unavailable for validation, other approaches might be used to evaluate the potential performance of the model. The literature on the model may provide an indication of the model's expected performance. However, care should be taken in inferring the model's likely performance for the study watershed based on validation results found in the literature. Data used and model parameterization for studies reported in the literature are not often described with enough detail to allow a good assessment of the model's likely performance in other watersheds. Further, if the model study reported in the literature included calibration, assessment of the model's likely performance in the study watershed will be even more difficult as the model will not be calibrated for the study watershed.

Observed runoff and water quality data from a similar watershed could potentially be used to determine the likely performance of the model for the study watershed. Sogbedji and McIsaac (2002) demonstrated the expected performance of the ADAPT model through calibration of the model using data from a comparable watershed and then applying it to similar watersheds. However, it may be desirable not to calibrate the model for the similar watershed, but rather simply validate the model for such watersheds, as data are unavailable for calibration of the model in the study watershed.

The USEPA (2002) indicates that a model can be evaluated by comparing model predictions of current conditions with similar field or laboratory data not used in the model calibration process, or with comparable predictions from accepted models or by other methods (e.g., uncertainty and sensitivity analyses). The results of a simple mass balance model could be compared with those of the model used in the study to see how well results match. Multiple comprehensive models might also be applied to the study watershed if data are unavailable for calibration and validation. If multiple models provide similar results, confidence in the results that are obtained may be increased. One must be cautious though with the interpretation of results in such a case, especially if the models use similar modeling components or approaches.

If validation is not possible, varying ranges of model inputs might be used in later stages of the modeling effort to determine the sensitivity of the model results to the model inputs. The use of Monte Carlo techniques and other approaches can also be used to identify confidence limits on outputs. "Biasing" the model inputs may also be used in later stages of the modeling effort to determine the sensitivity of the results to assumptions in model inputs. In such a situation, the model inputs would be set to extreme values in their expected ranges. If the same conclusions are reached with these inputs, the confidence in the conclusions reached would be increased since the conclusions are not sensitive to the model input assumptions.

Item 10. Model Scenario Prediction

Once the model has been validated and the results are deemed acceptable, the model is ready to be parameterized to the conditions of interest (e.g., a landuse change, implementation of BMPs). The parameterization of the model and the rationalization for decisions regarding data and representations within the model should be documented to allow others to recreate the model setup (see *Model*

Representation Issues section). These data and representation decisions should be consistent with those used in setting up the model for calibration and validation.

If possible, the uncertainty in model predictions when parameterized for the condition(s) of interest should be explored. The results from the validation stage provide some basis for expected model performance and level of uncertainty. Monte Carlo and other techniques can also be used to place confidence intervals on the expected results. For example, Kuczera and Parent (1998) used two Monte Carlo-based approaches for assessing parameter uncertainty in complex hydrologic models.

An approach that can be helpful in exploring the extremes in the uncertainty of model predictions is to bias model inputs in a direction that would be expected to represent the "worst case." If the model results for such a case result in the same conclusion being reached, the confidence in the conclusion should be high.

Item 11. Results Interpretation/Hypothesis Testing

Model results should be interpreted accounting for the expected uncertainty. Typically, the uncertainty in models cannot be quantified because of complexity of interactions, and thus it will be necessary to qualitatively assess the objectives or hypotheses taking into account the expected uncertainty in the results. The approach to be utilized in testing the objectives or hypotheses should be identified and documented prior to initiating the modeling.

The literature contains numerous examples of interpretation of model results. For example, Kirsch *et al.* (2002) tested the SWAT model within pilot watersheds and then applied it throughout a larger watershed in Wisconsin to quantify impacts from the application of basin-wide BMPs. Modeling results indicated that implementation of improved tillage practices (predominantly conservation tillage) could reduce sediment yields by almost 20%. They deemed this a significant reduction relative to current conditions. Santhi *et al.* (2001b) applied SWAT, which had been validated for flow and sediment and nutrient transport, to a watershed to quantify the effects of BMPs related to dairy manure management and municipal wastewater treatment plant effluent. King and Balogh (2001) used 99-year SWAT simulations for three locations to test hydrologic/water quality impacts of continuous corn, a forested environment, a golf course built in a previously agricultural setting, and a golf course constructed in a previously forested setting. Differences in hydrologic, nitrate-nitrogen, and pesticide impacts were examined using Tukey's pairwise comparison to determine whether differences were statistically different.

SUMMARY AND CONCLUSIONS

Data collection for environmental projects typically follows a QA/QC plan. Likewise, QA planning for environmental modeling projects should follow a standard procedure. A modeling protocol, preferably written, should be established prior to conducting modeling studies. A modeling standard protocol would: (1) reduce potential modelers' bias, (2) provide a roadmap to be followed, (3) allow others to repeat the study, and (4) improve acceptance of model results.

This paper presents a model application protocol for hydrologic/water quality studies. The present work is an adaptation and extension of the guidance available in the literature, including the USEPA, for QA project plans for modeling. Eleven issues that should be addressed in hydrologic/water quality model application plans were identified that include: (1) problem definition/background; (2) model application goals, objectives, and hypothesis; (3) model selection; (4) model sensitivity analysis; (5) available data; (6) data to be collected; (7) model representation issues – data, BMPs, etc.; (8) model calibration; (9) model validation; (10) model scenario prediction; and (11) results interpretation/hypothesis testing.

It is essential to document the decisions made for each of these items and the rationale for these decisions. The extent of documentation that should be prepared depends on various factors, including the purpose of the modeling study. A graded approach was recommended for defining the level of QA effort that is required for a modeling study. A detailed discussion of the above-mentioned steps was provided with an emphasis on hydrologic/water quality studies including considerations relevant to representation of NPS BMPs with watershed models and appropriate criteria for evaluation of the performance hydrologic models.

LITERATURE CITED

Arabi, M., R.S. Govindaraju, and M.M. Hantush, 2004. Source Identification of Sediments and Nutrients in Watersheds – Final Report."EPA/600/R-05/080. National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, Ohio, 120 pp. http://www.epa.gov/nrmrl/pubs/600r05080/600r05080.pdf. *accessed* August 3, 2007.

American Society of Chemical Engineers, Task Committee, 1993. Criteria for Evaluation of Watershed Models. Journal of Irrigation and Drainage Engineering 119(3): 429-442.

Bracmort, K., M. Arabi, J. Frankenberger, and B. Engel, 2006. Modeling Long-Term Water Quality Impacts of Structural BMPs. Transactions of the ASABE 49(2):367-374.

Chanasyk, D.S., E. Mapfumo, and W. Williams, 2002. Quantification and Simulation of Surface Runoff From Fescue Grassland Watersheds. Agricultural Water Management 1782(2002):1-17.

Chaubey, I., C.T. Haan, S. Grunwald, and J.M. Salisbury, 1999. Uncertainty in the Model Parameters Due to Spatial Variability of Rainfall. Journal of Hydrology 220(1999):48-61.

Eckhardt, K. and J.G. Arnold, 2001. Automatic Calibration of a Distributed Catchment Model. Journal of Hydrology 251(2001):103-109.

Fernandez, G.P., G.M. Chescheir, R.W. Skaggs, and D.M. Amatya, 2002. WATGIS: A GIS-Based Lumped Parameter Water Quality Model. Transactions of the ASAE 45(3):593-600.

FitzHugh, T.W. and D.S. Mackay, 2000. Impacts of Input Parameter Spatial Aggregation on an Agricultural Nonpoint Source Pollution Model. Journal of Hydrology 236(2000):35-53.

Gan, T.Y., E.M. Dlamini, and G.F. Biftu, 1997. Effects of Model Complexity and Structure, Data Quality, and Objective Functions on Hydrologic Modeling. Journal of Hydrology 192(1997):81-103.

King, K.W. and J.C. Balogh, 2001. Water Quality Impacts Associated With Converting Farmland and Forests to Turfgrass. Transactions of the ASAE 44(3):569-576.

Kirsch, K., A. Kirsch, and J.G. Arnold, 2002. Predicting Sediment and Phosphorus Loads in the Rock River Basin Using SWAT. Transactions of the ASAE 45(6):1757-1769.

Kuczera, G. and E. Parent, 1998. Monte Carlo Assessment of Parameter Uncertainty in Conceptual Catchment Models: The Metropolis Algorithm. Journal of Hydrology 211(1998):69-85.

Legates, D.R. and G.J. McCabe, Jr., 1999. Evaluating the Use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation. Water Resources Research 35(1):233-241.

Loague, K. and R.E. Green, 1991. Statistical and Graphical Methods for Evaluating Solute Transport Models: Overview and Application. Journal of Contaminant Hydrology 7:51-73.

Mankin, K.R., R.D. DeAusen, and P.L. Barnes, 2002. Assessment of a GIS-AGNPS Interface Model. Transactions of the ASAE 45(5):1375-1383.

Muttiah, R.S. and R.A. Wurbs, 2002. Scale-Dependent Soil and Climate Variability Effects on Watershed Water Balance of the SWAT Model. Journal of Hydrology 256(2002):264-285.

Nandakumar, N. and R.G. Mein, 1997. Uncertainty in Rainfall-Runoff Model Simulations and the Implications for Predicting the Hydrologic Effects of Land-Use Change. Journal of Hydrology 192(1997):211-232.

Ramanarayanan, T.S., J.R. Williams, W.A. Dugas, L.M. Hauck, and A.M.S. McFarland 1997. Using APEX to Identify Alternative Practices for Animal Waste Management. Paper No. 97-2209. ASAE, St. Joseph, Michigan.

Refsgaard, J.C., 1997. Parameterisation, Calibration and Validation of Distributed Hydrological Models. Journal of Hydrology 198(1997):69-97.

Refsgaard, J.C. and B. Storm 1996. Construction, Calibration and Validation of Hydrological Models. *In:* Distributed Hydrological Modeling, M.B. Abbott, and J.C. Refsgaard (Editors). Kluwer Academic Press, Dordrecht, The Netherlands, pp. 41-54.

Saleh, A., J.G. Arnold, P.W. Gassman, L.M. Hauck, W.D. Rosenthal, J.R. Williams, and A.M.S. McFarland, 2000. Application of SWAT for the Upper North Bosque River Watershed. Transactions of the ASAE 43(5):1077-1087.

Santhi, C., J.G. Arnold, J.R. Williams, W.A. Dugas, R. Srinivasan, and L.M. Hauck 2001a. Validation of the SWAT Model on a Large River Basin With Point and Nonpoint Sources. Journal of the American Water Resources Association 37: 1169-1188.

Santhi, C., J.G. Arnold, J.R. Williams, L.M. Hauck, and W.A. Dugas, 2001b. Application of a Watershed Model to Evaluate Management Effects on Point and Nonpoint Source Pollution. Transactions of the ASAE 44(6):1559-1570.

Sogbedji, J.M. and G.F. McIsaac, 2002. Evaluation of the ADAPT Model for Simulating Water Outflow From Agricultural Watersheds With Extensive Tile Drainage. Transactions of the ASAE 45(3):649-659.

Spruill, C.A., S.R. Workman, and J.L. Taraba, 2000. Simulation of Daily and Monthly Stream Discharge From Small Watersheds Using the SWAT Model. Transactions of the ASAE 43(6):1431-1439.

Taylor, J.K. 1985. What is Quality Assurance? *In:* Quality Assurance for Environmental Measurements, J.K. Taylor and T.W. Stanley (Editors). American Society for Testing and Materials Special Technical Publication No. 867, ASTM International, Philadelphia, Pennsylvania.

USEPA (U.S. Environmental Protection Agency), 1998. Guidance for Quality Assurance Project Plans QA/G-5. Report EPA/600/R-98/018, Office of Research and Development, Washington, D.C.

USEPA (U.S. Environmental Protection Agency), 2000. Policy and Program Requirements for the Mandatory Agency-Wide Quality System. EPA Order 5360.1 A2, Washington, D.C. http://www.epa.gov/quality/qs-docs/5360-1.pdf. *Accessed* August 3, 2007.

USEPA (U.S. Environmental Protection Agency), 2001. EPA Requirements for Quality Assurance Project Plans QA/R-5. Report EPA/240/B-01/003, Office of Environmental Information, Washington, D.C.

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

van der Heijde, P.K.M., 1987. Quality Assurance in Computer Simulations of Groundwater Contamination. Environmental Software 2(1):19-28.

Yapo, P.O., H.V. Gupta, and S. Sorooshian, 1996. Automatic Calibration of Conceptual Rainfall-Runoff Models: Sensitivity to Calibration Data. Journal of Hydrology 181(1996):23-48.

Yuan, Y., R.L. Bingner, and R.A. Rebich, 2001. Evaluation of ANNAGNPS on Mississippi Delta MSEA Watersheds. Transactions of the ASAE 44(5):1183-1190.