A RECOMMENDED CALIBRATION AND VALIDATION STRATEGY FOR HYDROLOGIC AND WATER QUALITY MODELS

P. Daggupati, N. Pai, S. Ale, K. R. Douglas-Mankin, R. W. Zeckoski, J. Jeong, P. B. Parajuli, D. Saraswat, M. A. Youssef

ABSTRACT. *Hydrologic and water quality (H/WQ) models are widely used to support site-specific environmental assessment, design, planning, and decision making. Calibration and validation (C/V) are fundamental processes used to demonstrate that an H/WQ model can produce suitable results in a particular application. However, the lack of comprehensive guidelines has led to the use of ad hoc, inconsistent, and incomplete C/V processes, which have made it difficult to interpret the myriad of published modeling studies, reduced the utility of many modeling applications, and slowed the advancement of H/WQ modeling. The objective of this article is to provide a generalized structure and process to assist modelers in developing a C/V strategy for H/WQ modeling applications. These best practice recommendations were developed based on an expansive review of the modeling literature, including a special collection of articles on H/WQ model calibration, validation, and use, as well as extensive discussion and debate among the authors. The model C/V recommendations include careful consideration, execution, and documentation of the following elements: (1) goals of model use, (2) data and parameters used in C/V, and (3) model C/V processes. Considerations in element 3 include the warm-up period, C/V strategy complexity, C/V process staging, spatiotemporal allocation of C/V comparison data, manual vs. automatic C/V, and additional diagnostics. Notable examples from the literature are provided for each strategy element. The comprehensive C/V strategy described herein will allow for better interpretation of future modeling studies, improved utility of modeling applications, and more systematic advancement of H/WQ models.*

Keywords. *Calibration, Guidelines, Hydrologic modeling, Strategy, Validation.*

More than half a century ago, the Stanford Watershed Model (Crawford and Linsely, 1962) applied digital computer technology to quantitatively describe hydrologic processes in a watershed. Since then, computer power and scientific understanding of hydrologic and biogeochemical processes have evolved significantly, resulting in a great number and variety of hydrologic and water quality (H/WQ) models. Because models are by definition "simplifications of the real world," model developers and practitioners have the responsibility to ensure that the essential characteristics and processes of the real world are simulated appropriately and that the model performs adequately for a given purpose. One important step in model applications is the comparison of model results to observed data through calibration and validation (C/V) (Refsgaard and Storm, 1996; Refsgaard, 1997; Moriasi et al., 2012).

Model calibration is a process in which a generalized model is adjusted so that the model predictions better represent site-specific H/WQ processes and conditions. During calibration, model parameters are optimized in an effort to increase accuracy and reduce model prediction uncertainty. Calibration is performed by carefully selecting model parameter values, adjusting them within their recommended ranges, and comparing predicted output variables with observed data for a given set of conditions (Arnold et al., 2012). Since the crucial goal of model calibration is to optimize unknown parameter values in the model, this process is also called parameter optimization (Šimůnek et al., 2012). A model is considered to be successfully calibrated when it replicates observed data within an adequate level of accuracy and precision (James and Burges, 1982; Konikow and Bredehoeft, 1992; Moriasi et al., 2007, 2015). Valida-

Submitted for review in April 2014 as manuscript number NRES 10712; approved for publication by the Natural Resources & Environmental Systems Community of ASABE in January 2015.

Mention of company or trade names is for description only and does not imply endorsement by the USDA. The USDA is an equal opportunity provider and employer.

The authors are **Prasad Daggupati, ASABE Member,** Assistant Research Scientist, Department of Ecosystem Science and Management, Texas A&M University, College Station, Texas; **Naresh Pai, ASABE Member,** Environmental Modeler, Stone Environmental, Inc., Montpelier, Vermont; **Srinivasulu Ale, ASABE Member,** Assistant Professor, Texas AgriLife Research, Vernon, Texas; **Kyle R. Douglas-Mankin, ASABE Member,** Senior Hydrologist, U.S. Fish and Wildlife Service, Boynton Beach, Florida; **Rebecca W. Zeckoski, ASABE Member,** Owner, Zeckoski Engineering, Charlotte, North Carolina; **Jaehak Jeong, ASABE Member,** Assistant Professor, Texas AgriLife Research, Temple, Texas; **Prem B. Parajuli, ASABE Member,** Associate Professor, Department of Agricultural and Biological Engineering, Mississippi State University, Mississippi State, Mississippi; **Dharmendra Saraswat, ASABE Member,** Associate Professor, Purdue University, West Lafayette, Indiana; **Mohamed A. Youssef, ASABE Member,** Associate Professor, Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, North Carolina. **Corresponding author:** Prasad Daggupati, 1500 Research Parkway, Suite B221, 2120 TAMU, College Station, TX 77843; phone: 785-214-2861; e-mail: pdaggupati@tamu.edu.

tion is the process of demonstrating that a given site-specific calibrated model can make sufficiently accurate simulations in a new modeling situation, although “sufficiently accurate” can vary based on project goals (Refsgaard, 1997).

Despite the importance of C/V to improve H/WQ model results, the topic of C/V strategy development has received relatively little attention. A few examples were identified in the literature describing formal comparisons of multiple C/V strategies to determine the most suitable strategy. Wallner et al. (2012) compared a variety of manual (lumped, one-factor, distributed, and regionalization) and automatic (parameter estimation [PEST], dynamically dimensioned search [DDS], and shuffled complex evolution [SCE]) calibration strategies. They found that the DDS automatic algorithm gave the best results overall. Other studies have assessed different strategies for parameter optimization (Blasone et al., 2007), automatic calibration (Kim et al., 2007), and parameterization (Refsgaard, 1997; Pokhrel and Gupta, 2010). In all cases, these studies provided an improved understanding of specific elements of an overall C/V strategy, but they fell short of providing modelers with guidance in developing a complete C/V strategy that can support the needs of a specific modeling application. Thus, the C/V approaches used by many H/WQ modelers are ad hoc, piecemeal, incomplete, and often inadequate for the task of achieving or demonstrating that model performance meets the needs of the modeling application.

In an effort to develop a comprehensive H/WQ modeling protocol, Engel et al. (2007) suggested the following steps: (1) problem definition and background; (2) model application goals, objectives, and hypothesis; (3) model selection; (4) model sensitivity analysis; (5) available data assessment; (6) data gap assessment; (7) model representation issues; (8) model calibration; (9) model validation; (10) model scenario prediction; and (11) results interpretation and hypothesis testing. However, Engel et al. (2007) provided only cursory guidance on the myriad of issues faced during C/V strategy development and implementation (steps 8 and 9).

Most modelers would agree that a comprehensive model C/V strategy should include consideration and documentation of the goals of model use, data used to compare with model output, parameters and output variables selected for C/V, sequencing of specific steps used in the C/V processes, and the measures and criteria used to characterize model performance. However, comprehensive guidelines are currently unavailable. The objective of this article is to provide a generalized structure and process to assist modelers in developing a C/V strategy for H/WQ modeling applications. Examples of each element of the C/V strategy are drawn from a special collection of articles on H/WQ model calibration, validation, and use (Moriasi et al., 2012) as well as other literature.

The article hereafter is organized into three sections describing each of the three major C/V strategy elements (fig. 1). In the first section, we discuss the influence of modeling goals on developing an appropriate C/V strategy (Element 1). Next, we discuss considerations in using measured data to compare with predicted values and in selecting model parameters and output variables for C/V (Element 2). In the third section, we provide an in-depth discussion of the specifics of various model C/V processes (Element 3), namely the model warm-up period, C/V strategy complexity, staging of the C/V process, spatiotemporal allocation of C/V comparison data, manual vs. automatic C/V, and additional diagnostics. In the conclusion, we emphasize key points and make recommendations for future work.

Element 1. Goals of Model Use

Calibration and validation may differ widely, but they

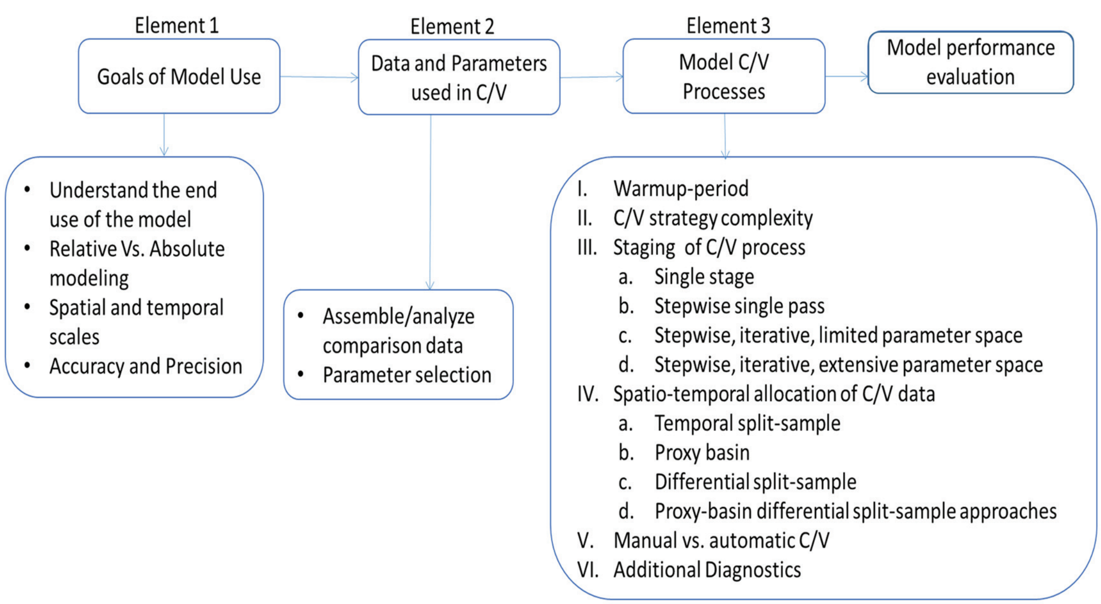

Figure 1. Elements of a comprehensive calibration and validation (C/V) strategy for hydrologic and water quality modeling.

should reflect the modeling application's goals. To develop an appropriate C/V strategy, modelers must know the needs and constraints of the end use and the associated considerations, including absolute or relative predictions, spatial and temporal scales, and levels of accuracy and precision required.

UNDERSTANDING THE END USE

Modeling applications can be conducted for numerous purposes. In developing a C/V strategy, it is important to consider the goal of the modeling application in determining the level of model performance necessary. Model results could be used to guide planning or action (such as addressing impacts of land use or climate change), support regulatory requirements (e.g., developing a permit), or design and evaluate structures or management practices for water quality improvements (e.g., constructed wetlands, water detention ponds, riparian buffers, drainage water management, etc.). The required modeling accuracy may differ for differing applications. Harmel et al. (2014) discussed three general categories of intended model use (exploratory, planning, and regulatory/legal) and emphasized that each category warrants differing expectations related to model performance. Required accuracy may also be determined based on the risk (e.g., health, safety, environmental, economic, and legal) associated with actions that follow from model recommendations or the degree to which such risk can be moderated by adaptive management. High-risk situations may require more accurate results, but application of an adaptive management strategy may reduce such risk and allow action based on less accurate results. Existing guidelines for assessing model performance (e.g., Moriasi et al., 2007) present simple statistical threshold values for given levels of model performance (e.g., unsatisfactory, satisfactory, good, very good) but still leave it to the modeler to determine what level of performance is adequate and consistent with a given end use. We recommend that the model and end-use client have an open discussion about expectations for model accuracy and precision prior to project initiation (as suggested by Refsgaard et al., 2005; Jakeman et al., 2006; Bottcher et al., 2012).

RELATIVE VS. ABSOLUTE RESULTS

Assessment of relative and absolute results may require different model C/V approaches. The focus of relative modeling exercises is to accurately represent changes in H/WQ responses between scenarios. Examples of relevant assessment include climate scenarios (e.g., Parajuli, 2010; Sheshukov et al., 2011; Ahmadi et al., 2014; Mehtha et al., 2015), land-management scenarios, including spatial targeting (e.g., Qi et al., 2009; Woznicki et al., 2011; Daggupati et al., 2011; Pai et al., 2011; Ahmadi et al., 2013; Dile et al., 2015), and land use scenarios, including comparisons among management/mitigation efforts or spatial locations (e.g., Srinivasan et al., 2010; Knisel and Douglas-Mankin, 2012; Deb et al., 2015; Daggupati et al., 2015). Recent ASABE collections provide examples of each of these scenarios using the SWAT model (Douglas-Mankin et al., 2010; Tuppad et al., 2011; Arnold et al., 2012). On the other hand, assessing absolute model performance is needed if the goal is to compare model output to a given criterion or a threshold, such as developing a total maximum daily load (TMDL) for nutrient loading to a lake (Borah et al., 2006). In both relative and absolute modeling situations, adequate model performance must be demonstrated, but differences in expectations of accuracy in relative vs. absolute results should be reflected in the C/V strategy. For example, applications with greater interest in simulating relative results may place more emphasis on assessing performance using differences in observed data relative to a baseline (such as annual differences vs. a baseline year, or daily differences for a field under two management conditions) rather than on the absolute measured values.

SPATIAL AND TEMPORAL SCALES

Spatial and temporal scales are intrinsic to the simulation of model response variables, and these scales must be considered in developing a C/V strategy. Spatial and temporal scales affect many aspects of H/WQ modeling, including selection of an appropriate model or models (Arnold et al., 2015) and selection and allocation of C/V comparison data (discussed subsequently), as well as establishment of the appropriate model assessment criteria and targets. Model C/V strategies may be developed for a variety of spatial (point, field, watershed) and temporal (subdaily to decadal) scales (Moriasi et al., 2012). Scale and temporal considerations of H/WQ modeling are discussed in detail by Baffaut et al. (2015).

Issues of scale (both spatial and temporal) may influence the appropriate use of observed comparison data for a given modeling application. If different scales between observed comparison data and model output variables must be used, these constraints must be resolved, either by upscaling (aggregating data from smaller scale to represent a larger scale) or downscaling (disaggregating data at a larger scale to represent a smaller scale). However, error and uncertainty may result from scaling. Use of C/V comparison data at scales that are not consistent with the model should be carefully justified.

ACCURACY AND PRECISION

Inaccuracy (or error) and uncertainty (or imprecision) in both measured data and modeled predictions are inherent to H/WQ modeling. A clear understanding of the model application will allow modelers to be clear about the accuracy (i.e., the degree to which the model correctly simulates the response variable) and precision (i.e., the smallest difference in response variables that the model can correctly differentiate) required. For example, some planning decisions may require considerably less accuracy than determination of regulatory compliance (see Harmel et al., 2014, for a discussion of model evaluation based on intended use). Likewise, modeling less-understood physical and biogeochemical processes affecting the fate and transport of sediment and nutrients is expected to have coarser model precision compared to modeling better-understood hydrological processes.

A study by Douglas-Mankin et al. (2013) demonstrated the consideration of model use. Their model goal was to encourage farmers to implement specific field-level soil conservation practices by paying them in proportion to the associated SWAT-modeled average annual sediment-yield reduction. This application dictated that "relative" results (between conservation practices) were critical and that model results were needed at the "field" scale, but that coarse temporal averaged results (i.e., average annual sediment-yield reductions) were acceptable. Interestingly, although spatial differences among field locations within the watershed were evident, the project team together with farmers in the watershed found communication of these differences to be somewhat confusing and distracting, and modeled watershed-wide spatially averaged sediment-yield reductions were used as the basis for payment. Because of this, the study dedicated more effort to validating relative model results among practices than among the farmers' fields. This study also demonstrates that modeling strategy should be flexible enough to adapt to emerging situations.

ELEMENT 2. DATA AND PARAMETERS USED IN MODEL C/V

An important step in developing a successful C/V strategy is the selection of appropriate measured data to compare with model predictions and appropriate model parameters to calibrate. We recommend that modelers know the methods used to collect and analyze data and the uncertainty associated with these data and use appropriate methods to select model parameters and their respective value ranges for C/V.

Although not incorporated into most published modeling applications, an analysis of the accuracy and precision of the measured data used for model comparison is central to the development of an effective C/V strategy (Shirmohammadi et al., 2006; Guzman et al., 2015; Harmel et al., 2014). If the measured data used to compare with model predictions have a high degree of uncertainty in representing reality, but are assumed to be accurate and precise, then "poor" C/V statistics for a given model use may be inappropriately interpreted as poor model performance (Harmel et al., 2010, 2014). In addition, "good" C/V statistics may produce inappropriate confidence in model output that has been calibrated to data with a high degree of uncertainty.

COMPARISON DATA

Before starting C/V, an inventory of comparison data is essential. Comparison data are the data used to compare with model outputs to assess model performance. These may be raw measured data or data that are subject to some level of processing to make the data more useful for a given C/V application, such as simple unit conversion, stage-discharge conversion (e.g., Rantz et al., 1982), regression-based constituent load estimation (e.g., LOADEST; Runkel et al., 2004), or spatial or temporal interpolation, aggregation, disaggregation, or re-scaling techniques. Occasionally, comparison data are developed by use of a separate model or modeling process. Comparison data may be collected specifically for the modeling application or obtained from other local, state, and/or federal datasets of measured data.

It is important that modelers know the methods used for comparison data collection, processing, and analysis, which may be included in metadata associated with the dataset, documented in prior literature, or known from personal knowledge of the data collection process. Such information includes sample collection methods, timing, and frequency; compositing methods; handling and storage protocols; analysis methods; etc. With these details, modelers can assess measured data uncertainty and ensure that it is appropriate to meet model C/V goals. When comparison data from multiple sources are used, consistency among these data sources must be assessed to ensure that they are applied appropriately in model C/V. Harmel and King (2005) compared several different sampling strategies and reported the expected ranges of uncertainty associated with various strategies. Further analysis, such as visualization of the data in a time-series graph or other qualitative or quantitative assessments (Moriasi et al., 2015) could be used to ensure that the temporal trends are reasonable and to identify (and address) any anomalies (e.g., outliers) in measured data.

In many cases, modelers have little control over the uncertainty in observed comparison data used for C/V, particularly for historical data sets wherein the data are available but the collection method and associated uncertainty are unreported. As Silberstein (2006) pointed out, the improvement in water resource management does not come with improved models in the absence of improved data. Most model applications assume that measured comparison data are "correct" (i.e., free of error or uncertainty), and parameters are adjusted to best match observed data (Harmel and Smith, 2007; Yen et al., 2014). This approach disregards the presence of error and uncertainty in measured data. Harmel et al. (2006) assessed the expected ranges of uncertainty in various data collection steps for common methods of measurements of streamflow and water quality parameters based on several "data-quality scenario" classifications; these results can be used in the absence of project-specific uncertainty estimates. Cumulative uncertainties for typical scenarios were estimated for streamflow measurement (6% to 19%), sample collection (4% to 48%), sample preservation and storage (2% to 16%), and laboratory analysis (5% to 21%). Harmel et al. (2006) also found large ranges in storm load errors measured under typical conditions for dissolved nutrients (8% to 104%), total N and P (8% to 110%), and total suspended solids (7% to 53%).

We recommend that modelers understand and evaluate various sources of uncertainty when formulating a C/V strategy and interpreting its performance. Good and unsatisfactory model performance should be assessed relative to model and measured data uncertainty (Harmel et al., 2010). For instance, if good performance measures are achieved after model calibration and the measured comparison data uncertainty is high, then the model accuracy is not definitive. In addition, new performance assessment techniques must be developed to differentiate errors in the measured comparison data and model results (Harmel and Smith, 2007; Harmel et al., 2010, did this for model C/V). Until then, we recommend understanding and reporting uncertainty estimates from other studies (e.g., Harmel et al., 2006, 2009), adjusting them based on existing knowledge of data collection techniques for the study area, and considering the impacts of these uncertainties on the performance measures that result from the C/V process (e.g., Harmel and Smith, 2007). Readers are also encouraged to refer to other publications on uncertainty analysis in H/WQ modeling (Shirmohammadi et al., 2006; Harmel et al., 2010; Guzman et al., 2015).

While the quality of comparison data is an important consideration for C/V strategy development, a discussion on quantity is also pertinent. The appropriate amount of monitoring data required for calibration of a continuous H/WQ model continues to be an open question. For instance, it is common to have a period of measured data that is considered too short for adequate C/V. Often this results from having short-term funding for the monitoring portion of a project or from the necessity to conduct modeling in an area with little available data (this is often the reason that the modeling is needed in the first place). In a later section, we discuss "additional diagnostics" that can be used to augment available C/V data to improve the confidence in model results. In contrast, in many watersheds across the globe, long periods (>30 years) of data records for hydrologic parameters are available. Calibrating an H/WQ model to such long periods, although robust, can be computationally expensive. Razavi and Tolson (2013) provide a statistical approach to extracting a short period of "surrogate" data that embeds the information content in the parent dataset.

Parameter Selection

Appropriate selection of model C/V parameters should consider model responsiveness to the candidate parameters. Both formal (e.g., sensitivity analysis) and informal (e.g., experience with model, knowledge of study area, literature review, etc.) approaches have been used to identify parameters (White and Chaubey, 2005; Wang et al., 2012). Sensitivity analysis is the process of determining the rate of change in model output with respect to changes in model inputs or parameters (Arnold et al., 2012; Zeckoski et al., 2015). Sensitivity analyses can be global, which attempts to assess all combinations of all parameter values, or local, which evaluates a specific set of parameters one at a time while the remaining parameters are fixed. Global and local methods may yield different results because of how they select and prioritize parameters (Arnold et al., 2012) and address interactions among parameters (Wang et al., 2005; Tian et al., 2014). It is essential to identify key parameters and define their precision for effective model calibration (Ma et al., 2000; DeJonge et al., 2012). A companion article (Yuan et al., 2015) discusses sensitivity analysis in detail.

In addition to parameter sensitivity, we recommend that modelers also consider the degree to which the candidate parameters can be accurately measured or defined (e.g., based on values published in the literature, prior model calibration in a similar setting, or expert judgment). For example, if the value of a parameter is known accurately for the conditions being modeled, any further adjustment of the parameter may be inappropriate, or at least the range of its adjustment values must be appropriately limited. Similarly, model parameters that cannot be accurately measured or that are used to describe processes or phenomena that are not physically based may be good candidates for calibration.

Not all sensitive parameters need to be calibrated, and not all parameters that are calibrated are adjusted from their original values. During calibration, modelers should clearly document which parameters: are assigned model "default values"; are "baseline values" typically set *a priori* by the model based on expertise, prior modeling work, or literature values and are not changed during calibration; or are "calibration values" based on a calibration process that includes comparison to observed data. Differentiation of these three methods used to set model parameters is critical to the interpretation and replication of model C/V results.

Element 3: Model C/V Processes

Design of an effective C/V strategy must include careful consideration, execution, and documentation of the model warm-up period, C/V strategy complexity, staging of the C/V process, spatiotemporal allocation of C/V comparison data, manual vs. automatic C/V, and additional diagnostics.

Before considering the details of C/V strategy design, the use of uncalibrated models is worth mentioning. In simple cases (e.g., one-dimensional simulation of heat flow, water flow, or solute transport through a soil column), and particularly if the model parameters are based on accepted or measured values and the model processes are mechanistic, it may be appropriate to use a model without C/V. In such cases, model parameters based on default values and/or knowledge of the study site (either from measurements, commonly accepted parameter values, or prior modeling experience) would be used without modification. For example, developers of the Simultaneous Heat and Water (SHAW) model suggest that calibration is only necessary if available data are insufficient to estimate the physically based model parameters (Flerchinger et al., 2012). The SWAT model (Arnold et al., 2012) has also been applied without calibration to simulate hydrologic processes in the upper Mississippi River basin (Srinivasan et al., 2010) and the Smoky Hill River watershed in the Missouri River basin (Tuppad et al., 2010a, 2010b). However, caution should be exercised in the application of model results without calibration, as it is difficult to determine the adequacy of a model without comparison to observed data. Since performance metrics are not used to evaluate the model output in comparison to observed data, the model output should be carefully inspected to confirm that it is within an expected range and follows expected trends. The use of other model diagnostics to supplement calibration or substitute for calibration, when necessary, is discussed in a later section.

Warm-up Period

A warm-up period is used to allow an H/WQ model to run for a sufficient period prior to the simulation period to initialize important model variables or allow important processes to reach a dynamic equilibrium. Use of warm-up periods that are too short may result in biased simulated responses, especially in the initial years when model results may be dominated by uncertainty in characterization of the initial state rather than uncertainty in the model or parameters (Huard and Mailhot, 2008). For example, Muthuwatta et al. (2009) cited an insufficient warm-up period for causing reduced model performance, particularly in the initial years of simulation.

Length of the warm-up period may vary for different watershed-scale processes. Warm-up periods in H/WQ studies may range from months to decades, with one to four years being common for watershed-scale modeling (e.g., Douglas-Mankin et al., 2010). The time needed to reach a realistic initial state is likely related to the temporal and spatial scale of the governing processes, with necessary warm-up times, for example, being shorter to initialize soil moisture than ground-water table, soluble nutrient pools than stable nutrient pools, and surface residue than soil organic carbon. The structural complexity of the model will also impact the length of the warm-up period. Finally, the time needed to approach an equilibrium state will increase the deviation of initial state values relative to their actual values.

Given the complexity of watershed-scale processes, a comprehensive guideline cannot be provided for warm-up period. However, model developers recommend using warm-up periods of two to three years for hydrological processes and five to ten years for sediment and nutrient-related processes (Raghavan Srinivasan, Texas A&M University; Jeffrey Arnold, USDA-ARS; James Almendinger, St. Croix Watershed Research Station, Minnesota, personnel communication, 20 January 2014). In some cases, such as simulation of water impoundments in watershed models, modelers should adjust warm-up period lengths by comparison with observed data before full calibration. Long warm-up periods require measured weather data prior to the start of the simulation period, which may be a limitation for some applications. In such cases, we recommend preparing surrogate weather data for the warm-up period using actual weather data that resemble average weather conditions to ensure that an equilibrium state is attained during the warm-up period.

Strategy Complexity

We classify model C/V strategy as either simple or complex depending on the goals of model use, availability of comparison data, and model parameters selected for C/V.

Simple Strategy

A simple C/V strategy optimizes either a single model output variable at a single site or spatiotemporal scale or, in some cases, multiple model output variables that are simulated independently by the model and are not interactive. In either case, calibration of each output variable can be carried out independently, so conflicting "optimal" parameter sets cannot arise. Examples of simple C/V strategies include optimizing model performance using a time series of streamflow data at a watershed outlet, groundwater flow velocity at a specific location, or soil organic carbon at a specific location.

This simple C/V strategy is the most appropriate for areas with uniform characteristics (e.g., soil, slope, vegetation, meteorology) across the entire modeled area. In these cases, spatial variation in biophysiochemical processes, such as hydrological processes, sediment transport, plant uptake, and nutrient transformations, are assumed to be minimal, such that C/V to data collected at one location can be considered representative of the entire area. There are many examples of single-site C/V in the literature, e.g., ADAPT (Gowda et al., 2012), WEPP (Flanagan et al., 2012), DRAINMOD (Skaggs et al., 2012), RZWQM (Ma et al., 2012), and DAISY (Hansen et al., 2012).

The simple C/V strategy is not recommended for large areas (or watersheds) with highly variable, complex physical characteristics because the contributions of subareas with unique hydrologic and/or water quality characteristics may not be suitably represented. The resulting calibrated parameters must represent some average of the characteristics, or the set of calibrated parameters may represent a combination of overestimated and underestimated calibration values that produce an "adequate" result. In either case, the calibration process may fail to provide accurate values for individual parameters, and the resulting model may not perform well in other areas, thus defeating the purpose of a robust C/V process. This difficulty in representing individual parameters may also be an issue when the desired model use involves scenarios conducted for time periods or circumstances outside those of the calibration, as those scenarios rely on the accurate representation of individual mechanisms or characteristics in the model (e.g., a future land use change scenario).

Complex Strategy

A complex C/V strategy uses a single model output variable at multiple sites or spatiotemporal scales, or different model output variables at either a single or multiple sites or spatiotemporal scales. In cases of complex C/V, there are multiple "optimal" parameter sets that must be resolved to determine a unified "calibrated" parameter set. Examples of complex C/V strategies include calibration using:

- Single variable, multi-site, e.g., time series of streamflow discharge data at multiple, similarly sized subwatershed outlets.
- Multi-variable, single site, e.g., time series of streamflow discharge, suspended sediment yield, and nitrogen yield data at a single stream gauging station.
- Single variable, multi-spatial-scale, e.g., multiple years of crop yield data within a watershed at both field scale and county scale.
- Multi-variable, multi-site, multi-temporal-scale, e.g., multiple years of field-scale erosion yield data within a watershed from three different tillage practices at two different field stations, one at daily scale and one at weekly scale.

A complex C/V strategy may be appropriate for a model that represents an area with greater variability in characteristics and/or when observed data for a given process are available at multiple locations within the study area. The use of data at multiple locations is most beneficial if the data capture the range of physical variations in the study area and allow for a robust model calibration to observed data that represent a greater diversity of characteristics. The complex calibration strategy, which can better account for spatial biophysiochemical variations, may also help reduce the problem of equifinality because fewer model parameter sets will satisfy the calibration criteria at all calibration sites. Multi-site calibration, moving from upstream to downstream locations, was demonstrated for WARMF (Herr and Chen, 2012) and SWAT (White and Chaubey, 2005; Arnold et al., 2012). Variations of multi-site calibration have been used in most watershed scale models, e.g., BASINS/HSPF (Duda et al., 2012), MIKE SHE (Refsgaard, 1997; Jaber and Shukla, 2012), EPIC and APEX (Wang et al., 2012), and KINEROS/AGWA (Goodrich et al., 2012).

Data for multi-site calibration may be of similar scale (e.g., several sites with crop yield at the field scale) or at multiple scales (e.g., surface runoff measured at a field-scale flume at one site and using a streamflow gauge at another site). This may also be an appropriate method if smaller-scale processes need to be modeled accurately (e.g., crop yield, tile drainage) to improve simulation of an aggregate output (e.g., watershed streamflow). Consideration of multiple scales in calibration was recommended for WAM (Bottcher et al., 2012) and KINEROS/AGWA (Goodrich et al., 2012).

A complex strategy may also be warranted when there is more than one output variable of interest. For example, surface runoff is affected by antecedent soil moisture, which in turn is influenced by plant water uptake and transpiration, which is influenced by plant growth. Therefore, in areas of intensive crop production, prediction of runoff without calibration of crop growth could lead to erroneous parameter estimation. Other examples of cases in which it would be beneficial to calibrate multiple, dependent processes include calibrating runoff before sediment yield (Maski et al., 2008; Pai et al., 2011, Douglas-Mankin et al., 2013), nitrate-nitrogen yield (Knisel and Douglas-Mankin, 2012; Ma et al., 2012), or total phosphorus yield (Bottcher et al., 2012) or calibrating sediment yield before total nitrogen or phosphorus yield (Douglas-Mankin et al., 2010; Pai et al., 2011). Calibration of multiple output variables requires consideration of the interactions among the parameters used to calibrate each separate output variable. However, the complexity of a C/V strategy may be limited by the availability of comparison data. In such cases, other model diagnostics should be used to supplement or substitute for calibration, as discussed in a later section.

STAGING

The staging design refers to the systematic approach used by modelers to adjust parameters and assess output variables. The C/V process may use one or more parameters to calibrate a given model, and the model performance may be assessed using one or more output variables (or objectives) to represent system response. The many possible permutations may be distilled to just a few, distinct calibration philosophies: single-stage; stepwise single-pass; stepwise, iterative, limited parameter space; and stepwise, iterative, extensive parameter space (fig. 2).

Single-Stage

Also referred to as the Pareto optimal approach (Fenicia et al., 2007), the single-stage approach adjusts each model parameter across a range of values, and the model response is assessed using a single output variable. The single-stage approach may adjust single or multiple parameters depending on the model and process complexity. In the case of multiple parameters, each combination of model parameter values constitutes a single calibration run, and various combinations of the parameters are tested until an optimal solution is achieved. The optimal solution may be comprised of a single value or possibly a range of equally performing values of the model parameter used for calibration. The single-stage, single-parameter case is a simple subset of the multi-parameter approach. This approach may also use either a single objective or some combination of model output variables that are weighted to produce a single objective. The single-stage approach has been applied in numerous H/WQ models, e.g., ADAPT (Gowda et al., 2012), DRAINMOD (Skaggs et al., 2012), GLEAMS (Knisel and Douglas-Mankin, 2012), and SWAT (Maski et al., 2008).

Stepwise Single-Pass

The stepwise single-pass approach may be referred to as the "pseudo-optimal" approach and implies the calibration of multiple output variables (ET, soil moisture, streamflow, tile-drainage flow, sediment yield, nutrient yield, etc.). In the pseudo-optimal approach, the model is calibrated for multiple output variables in sequence (stepwise fashion). Once a model parameter (or set of parameters) is optimized

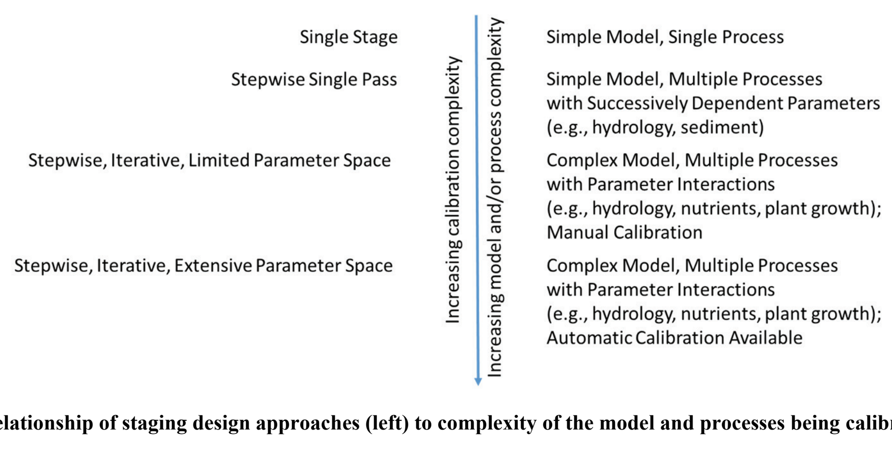

Figure 2. Relationship of staging design approaches (left) to complexity of the model and processes being calibrated (right).

for a given output variable, it is not revised further in calibrating any subsequent output variable.

It is important to consider the sequence in which output variables are calibrated due to interactions among processes. For the SWAT model, Santhi et al. (2001) provided a useful flowchart that guides users on the sequence to be followed for C/V for multiple output variables. Interactions among processes may dictate the most appropriate sequence in which to stage the multiple output variables being calibrated. For example, it is often important to calibrate hydrological processes before other processes that depend on those hydrological processes at each stage to ensure that all relevant influences on a process have already been calibrated. Many studies have applied this stepwise approach with the sequential calibration of processes, e.g., DAISY (Hansen et al., 2012), DRAINMOD-NII (Youssef et al., 2006), and GLEAMS (Knisel and Douglas-Mankin, 2012).

Stepwise, Iterative, Limited Parameter Space

This approach may be referred to as a "local optimal" approach. The local optimal approach applies both to the use of multiple parameters to calibrate a single process output variable and multiple overlapping parameters to calibrate multiple output variables. The latter case is relevant for situations in which there is an interaction between two (or more) model processes. For example, for a given rainfall event, the amount of runoff influences the amount of rainfall available to infiltrate, which influences the amount of soil moisture, which may influence crop growth and the density of plant roots and canopy, which influences plant water uptake, which influences soil moisture, which influences antecedent soil moisture conditions, which (finally) influences runoff. Because of this interaction between crop growth and runoff, calibration of these two output variables would benefit from iterative calibration.

When applied to the use of multiple parameters, this approach suggests that each parameter is optimized in sequence, and after optimizing each successive parameter, modelers readjust prior parameters to ensure that changes to subsequent parameters have not shifted the optimal value. When applied to the calibration of multiple output variables, the approach suggests that modelers calibrate the first output variable and then the second variable (and subsequent variables, if appropriate) in stepwise fashion. However, after calibrating the second output variable (or each subsequent variable), the earlier parameters are revisited to ensure that any interactions between the parameters used to model the output variables have not changed the optimal set of calibrations parameters for the earlier output variable(s). In either case, this is a tedious process when performed manually, and thus it is often accompanied by consideration of a limited parameter space. Most examples in the literature apply a stepwise, iterative process to the calibration of a single output variable such as in ADAPT (Anand et al., 2007), BASINS/HSPF (Duda et al., 2012), RZWQM2 (Ma et al., 2012), and SWAT (Maski et al., 2008; Sheshukov et al., 2015). Calibration of the integrated system model versions of DRAINMOD, i.e., DRAINMOD-FOREST (Tian et al., 2012) and DRAINMOD-DSSAT (Negm et al., 2014), is conducted using a stepwise, iterative approach applied to calibrate multiple variables. For example, the hydrologic, soil carbon and nitrogen transformation, and plant growth parameters of DRAINMOD-DSSAT are calibrated to predict water table fluctuation, drainage flow, nitrogen losses with drain flow, and crop yield. The structure of the GLEAMS model, for example, separates hydrologic, erosion, nutrient, and pesticide processes into separate modules that are intended to be calibrated sequentially (Knisel and Douglas-Mankin, 2012), which minimizes the need for iterative calibration of output variables. The sequential calibration of total nitrogen and total phosphorus in SWAT (Douglas-Mankin et al., 2010) presented an opportunity to apply the multiple output variable approach at the watershed scale because of the influence of a single parameter (BIOMIX) on both nutrients; however, an independent calibration of the two nutrients produced the same optimal value for BIOMIX, making the iteration unnecessary.

Stepwise, Iterative, Extensive Parameter Space

This approach may be referred to as a "global optimal" approach. This approach is similar to the previous approach except that the parameter space is expanded to include a greater number and more extensive range of parameters. This method typically requires an automation procedure and is discussed in greater detail by Duan et al. (1992, 1994), Zhang et al. (2008), and Malone et al. (2015).

SPATIOTEMPORAL DATA ALLOCATIONS (DATA SPLITTING)

The C/V strategy must consider the allocation of spatiotemporally distributed comparison data for the purpose of either calibration or validation. Calibration of an H/WQ model at a single site (or outlet of a watershed) remains a widely used C/V strategy. This common temporal split-sample method, in which the measured comparison data are split into two periods for C/V, is criticized by Rosso (1994) and Qi and Grundwald (2005) because this strategy does not account for the spatial variability of important H/WQ factors within the watershed. Alternative data allocation methods include proxy basin, differential split-sample, and proxy basin differential split-sample. A flowchart describing when to use each approach is presented in figure 3.

Temporal Split-Sample

The temporal split-sample approach is the most commonly used application of single-site comparison data for C/V. It is applicable to cases in which catchment conditions are stationary and sufficient data are available for model testing (Refsgaard, 2000). In this method, the available observed comparison data are split into two parts: one for model calibration and the other for model validation.

The proportion of comparison data allocated to calibration or validation varies by modeling goals and among modelers. Gowda et al. (2012) ran ADAPT for the entire period of available record and then used data from alternate years for calibration or validation, while Bottcher et al. (2012) suggested using no more than one-third of the dataset for the calibration of WAM so that at least two-thirds of the data period would be available for model validation.

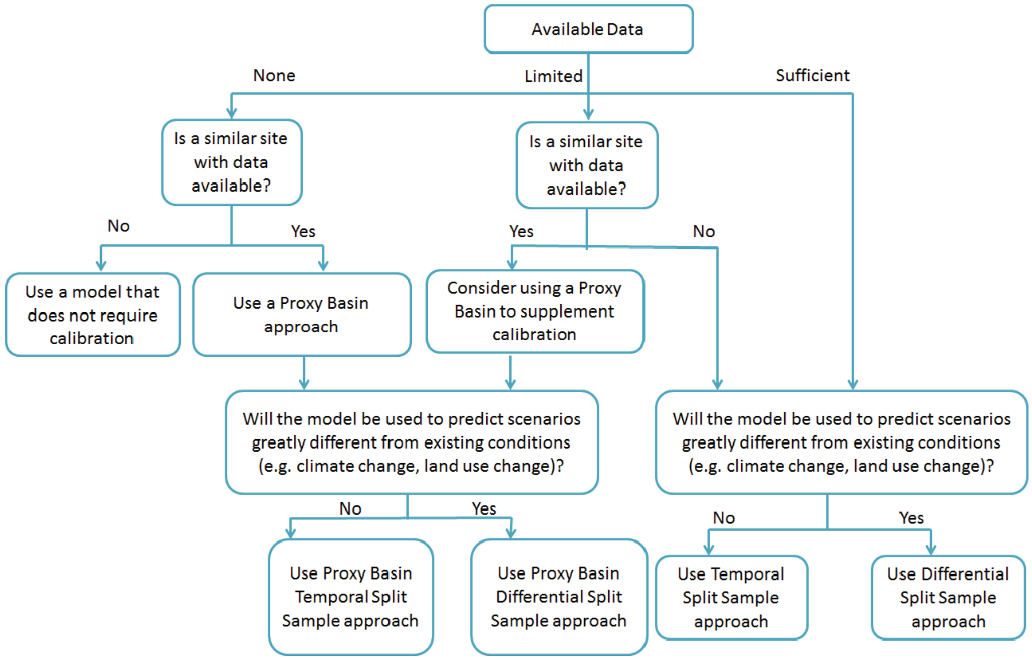

Figure 3. Selection among allocation of spatiotemporal comparison data (data-splitting approaches).

SWAT developers (Arnold et al., 2012) recommend the calibration period to be long enough to contain a range of conditions to be expected in the watershed, such as wet and dry weather extremes over multiple years. Van Liew et al. (2007) followed a different approach and subdivided the validation period (1975-1993) into three periods: wet (1975-1979), dry (1980-1988), and average (1989-1993) climatic conditions. Bennett et al. (2013) presented several structured approaches, such as bootstrapping, cross-validation, hold-out, and *k*-fold partitioning that seeks to randomize the data splitting, so that the model performance evaluation is not biased by the allocation of data.

The disparity in data splitting in previous modeling studies indicates that there is no single rule applicable to all models or applications; however, comparison data could be grouped based on model use. A calibrated model could be applied under conditions that are "similar to" (e.g., when targeting pollutant hotspots or quantifying BMP effectiveness) or "different from" (e.g., climate change, or land use change or ungauged watersheds) those encountered during the calibration period. When using a calibrated model under "similar" conditions, such as applying a model for the same period of record as used in C/V under conditions of shifting (but similar) land use change, it is recommended to split the comparison data in such a way that climatic conditions during the calibration and validation periods are similar. In contrast, when using the model under "different" conditions, such as for climate or land use change impact studies, splitting the comparison data such that the validation data are diverse and deviate greatly from the calibration data is suggested.

Other splits between calibration and validation periods have also been used when comparison data are limited or when modeling goals require greater emphasis on calibration (e.g., when parameter characterization over a wide range of conditions is essential) or validation (e.g., when validation or confirmation of model precision at independent sites or time frames is critical). In some situations, it may also be important to allocate critical time frames to each of the calibration and validation datasets. For example, it might be important for each comparison dataset to contain a proportionally similar number of periods meeting some criteria, such as growing season rainfall, annual net precipitation excess (precipitation minus evapotranspiration), growing degree-days, or antecedent soil moisture conditions prior to major storm events.

Proxy Basin

The spatial proxy basin approach should be used when there are insufficient comparison data for using the temporal split-sample approach or when calibrating a model for ungauged watersheds (Klemes, 1986). For example, if streamflow from an ungauged watershed is to be predicted, two gauged "proxy" watersheds (e.g., A and B) within a similar ecoregion should be selected. The model should be calibrated on watershed A and validated on watershed B, and vice versa. This will result in two sets of calibration parameters. After ascertaining that acceptable and similar results are obtained for both proxy watersheds, one of the calibrated models should be used for predicting streamflow for the ungauged watershed. Model performance statistics may be used as a guideline for deciding which of the two sets of calibrated parameters is best for use in the ungauged watershed. If any comparison data are available for the watershed, they should be used for additional validation of the proxy watershed calibrated model.

Dinicola (1992) performed a proxy basin validation effort to help evaluate rainfall-runoff relations for the Puget Sound area in Washington using the HSPF model. Dinicola (1992) initially calibrated the HSPF model concurrently for 21 stream gauge sites and then validated the model by collecting precipitation and streamflow data from 11 additional drainage basins that are physiographically similar to those used for model calibration. In another study, Parajuli et al. (2009) used the proxy basin approach to calibrate and validate the AnnAGNPS and SWAT models for a USDA Conservation Efforts Assessment Project (USDA-CEAP) agricultural watershed in south central Kansas. They calibrated the models using comparison data from the Red Rock Creek watershed and validated the models using the Goose Creek watershed, both being subwatersheds of the Cheney Lake watershed with similar soil types, land use characteristics, climate, and history of water quality data.

Differential Split-Sample

The differential split-sample approach is recommended whenever a model is used to simulate flows in a given gauged basin under conditions different from those corresponding to the available flow record (e.g., climate change scenarios, land use change studies, or effects of groundwater abstraction). While simulating the effects of climate change, two periods with different values of climate variables of interest, such as high and low average precipitation, should be selected (Refsgaard, 2000). If the model is used to simulate streamflow in a wet climate scenario, it should be calibrated using comparison data from the dry segment of the historical record and validated with data from the wet segment to verify the model's ability to perform in the transition from drier to wetter conditions. Similarly, if the model is used to predict streamflow for a dry climate period, it should be calibrated using data from the wet segment and validated using data from the dry segment. Klemes (1986) demonstrated the potential danger of using a simulation model for climate change studies without subjecting it to a differential split-sample evaluation.

Proxy Basin Differential Split-Sample

This approach is a combination of the two previously described approaches. It is the most difficult evaluation for an H/WQ model, as it is used in situations for which there are no comparison data available for model calibration and when the model is directed to predict non-stationary conditions (i.e., those that vary with time), such as climate change or incremental management practice adoption (Refsgaard, 2000). This approach should be applied in cases for which the model is intended to be both climatically and geographically (or land use-wise) transposable. As described by Klemes (1986), if a model is intended for assessment of impact of climatic change in an ungauged watershed C, this approach should have the following form: two gauged watersheds (A and B), with characteristics similar to those of watershed C, are selected, and segments with different climatic parameters (e.g., w for wet and d for dry) are identified in the historic records of both. For assessment of the impact of a dry climate scenario, the model is first calibrated on Aw and validated on Bd, and then calibrated on Bw and validated on Ad. Calibration is judged adequate if the model performances in both validation runs (Ad and Bd) are acceptable and not significantly different. In the same way, if the model is used for assessment of the impact of a wet climate scenario, it would be calibrated and validated on Ad/Bw and Bd/Aw and judged adequate if the model performances for Aw and Bw are adequate and similar. Once again, the model performance statistics may be used as a guideline for deciding which of the two sets of calibrated parameters is best for use in the ungauged watershed C.

Donnelly (1997) used the four different spatiotemporal approaches discussed to this point to compare the performance of three different rainfall-runoff models (two lumped models and a quasi-distributed model) in two forested catchments in Canada. The author demonstrated that the use of a standard model-testing framework alone could be misleading and recommended conducting statistical analysis in combination with the standard model testing. The author's statistical analysis showed that when split-sample and proxy basin approaches were used, no statistical difference in model performance was found, indicating that no significant benefit would be achieved in applying the quasi-distributed model compared to the simpler lumped models. When a differential split-sample approach was used, the quasi-distributed model performed significantly better than both lumped models in both catchments. Finally, the statistical analysis with the proxy basin differential split-sample approach indicated that the quasi-distributed model performed better than lumped models in one catchment but not in the other (Donnelly, 1997). Further research is warranted to determine if these results hold for various models and settings.

Additional Considerations and Diagnostics

Manual vs. Automatic C/V

The above-mentioned C/V approaches may be performed using either manual or automatic calibration techniques. All models in the 2012 collection (Moriasi et al., 2012) support manual calibration, while nine models (COUP, EPIC/APEX, HYDRUS, KINEROS/AGWA, MIKE SHE, MT3DMS, SHAW, SWAT, and WARMF) were specifically noted to support automatic calibration. However, some tools, such as PEST, allow automatic calibration of any H/WQ model that can be used from a command line; this has been done with HSPF (e.g., Doherty and Johnston, 2003; Kim et al., 2007).

Manual calibration is often time-consuming and labor-intensive when the number of parameters used in the manual calibration is large, especially for complex hydrologic models (Balascio et al., 1998). It is often difficult for modelers to know how sensitive simulation outputs are to parameter adjustments due to non-linear model algorithms (Gupta et al., 1999); however, this can be lessened by conducting a sensitivity analysis, as discussed earlier and by Yuan et al. (2015). These shortcomings of manual calibration have led to the development of automatic calibration techniques that use high-speed computers and various search algorithms to determine best-fit parameters for matching model response to observed data. Automatic calibration techniques have been significantly refined over the past three decades by coupling them with sensitivity and uncertainty analyses to improve parameter specification and estimation.

Automated calibration techniques may save time and labor; however, there is a possibility that the resulting calibrated parameters do not realistically reflect watershed characteristics (Boyle et al., 2000; Arnold et al., 2015). Knowledge of the simulated site is essential for providing initial estimates of model parameters and for evaluating the final set of automatically calibrated parameters. Modelers should take care when selecting the upper and lower ranges for parameters to ensure that they are representative of site conditions. Manual adjustments following an automatic calibration may be necessary, and are recommended, to maintain the overall mass balance and adequate representation of range and magnitude in output variables (Van Liew et al., 2005). More information on parameter estimation and automatic calibration techniques is provided by Malone et al. (2015).

Use of Diagnostics to Supplement or Substitute for Calibration

Comparison data are not always available or at the appropriate spatial or temporal scales for robust model calibration and validation. In these cases, an analysis of model diagnostics may be needed to supplement (or substitute for) calibration and validation efforts to improve confidence in model performance. For example, comparison data may only be available to calibrate daily streamflow at a watershed outlet, but diagnostic analyses could also be performed to demonstrate that crop growth or phenology was modeled reasonably for the climatic conditions each year, that modeled rates of evapotranspiration were appropriate for a given crop, and that relative rates of evapotranspiration or growth among the various crops were reasonable.

The appropriate type and extent of diagnostic analyses must consider the goals of the application, critical or sensitive locations or processes within the study area, and robustness of the calibration and validation process, particularly considering the locations and/or processes with limited or no calibration. Several examples are discussed below, but the suite of diagnostics that are sufficient to demonstrate model performance may vary widely among applications.

- Sensitivity analyses of critical model parameters and output variables can be used to demonstrate reasonable model responses to parameters that are known to be influential.
- Time-series graphs of critical model responses at multiple locations within the study area may demonstrate appropriate spatial and temporal model performance.
- Graphical and/or statistical comparisons among model responses may demonstrate appropriate interactions among model elements.
- Graphical and/or statistical comparisons between model responses and other independent, but potentially related, phenomena may demonstrate appropriate relationships between related processes. For example, a model might not simulate algae production, but recorded data on lake algae blooms may provide important correlation with other modeled outputs that could contribute to the development of algae blooms.
- Comparisons with prior modeling studies (preferably calibrated and validated) under similar conditions may demonstrate consistency (if not accuracy) of the model application.

Conclusions and Recommendations

This article describes the development of a comprehensive model C/V strategy that includes three main elements: (1) goals of model use, (2) data and parameters used in C/V, and (3) model C/V processes. Element 3 includes careful consideration, execution, and documentation of the model warm-up period, C/V strategy complexity, staging of the C/V process, spatiotemporal allocation of C/V comparison data, manual vs. automatic C/V, and diagnostics to supplement calibration. Notable examples from the literature are provided for each C/V element, including guidance presented in 22 model-specific articles on H/WQ model use, calibration, and validation published in a 2012 special collection as well as additional published information.

We expect that use of this comprehensive C/V framework will enhance the quality, consistency, and repeatability of published model results. The comprehensive C/V strategy presented herein will also provide a consistent platform for development and application of model, site, and application specific C/V strategies. Experience gained from the systematic application of this C/V strategy to numerous models by a broad base of modelers will further refine, enhance, and improve the robustness and specificity of the guidance. We encourage future forums to include further discussion and development of these C/V strategies.

Development of a single, consistent, coherent C/V strategy that is applicable across all H/WQ models and biogeochemical, topographic, and climatic settings has ranged from challenging to daunting. Considerable expert judgment was liberally mixed with the technical information gleaned from the published literature in the development of the strategy and specific elements presented herein. Within each element of the C/V strategy, the need exists for further research to provide systematic comparisons of different options and methods across a range of biogeochemical settings for a range of H/WQ models. Examples include studies that advance methods for spatiotemporal allocation ("splitting") of comparison data for calibration or validation, for determining minimum periods of record that are needed to adequately calibrate and validate a given model, and for upscaling or downscaling C/V comparison data or model results. Research is also needed to compile, synthesize, and interpret the myriad of H/WQ modeling C/V studies to document the calibrated parameters for each H/WQ model as well as enhance our understanding of how model-specific C/V strategies and results vary with spatiotemporal setting. The use by modelers of the consistent, comprehensive C/V strategy described herein will allow better interpretation of future published modeling studies, improve the utility of model applications, and allow more systematic advancement of H/WQ models.

Acknowledgements

The authors would like to thank Dr. Jeff Arnold, Dr. Raghavan Srinivasan, Dr. James Almendinger, and Dr. Mehdi Ahmadi for their valuable comments and for sharing their knowledge. We would also like to thank the three anonymous reviewers for their helpful comments and suggestions. The findings and conclusions in this article are

those of the authors and do not necessarily represent the views of the U.S. Fish and Wildlife Service, the USDA Agricultural Research Service, and other respective universities and organizations.

References

Ahmadi, M., Arabi, M., Hoag, D. L., & Engel, B. A. (2013). A mixed discrete-continuous variable multiobjective genetic algorithm for targeted implementation of nonpoint-source pollution control practices. *Water Resources Res., 49*(12), 8344-8356. http://dx.doi.org/10.1002/2013WR013656.

Ahmadi, M., Records, R., & Arabi, M. (2014). Impact of climate change on diffuse pollutant fluxes at the watershed scale. *Hydrol. Proc., 28*(4), 1962-1972. http://dx.doi.org/10.1002/hyp.9723.

Anand, S., Mankin, K. R., McVay, K. A., Janssen, K. A., Barnes, P. L., & Pierzynski, G. M. (2007). Calibration and validation of ADAPT and SWAT for field-scale runoff prediction. *JAWRA, 43*(4), 899-910. http://dx.doi.org/10.1111/j.1752-1688.2007.00061.x.

Arnold, J. G., Moriasi, D. N., Gassman, P. W., Abbaspour, K. C., White, M. J., Srinivasan, R., Santhi, C., Harmel, R. D., van Griensven, A., Van Liew, M. W., Kannan, N., Jha, M. K. (2012). SWAT: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1494-1508. http://dx.doi.org/10.13031/2013.42256.

Arnold, J. G., Youssef, M. A., Yen, H., White, M. J., Sheshukov, A. Y., Sadeghi, A. M., Moriasi, D. N., Steiner, J. L., Amatya, D. M., Skaggs, R. W., Haney, E. B., Jeong, J., Arabi, M., & Gowda, P. H. (2015). Hydrological processes and model representation: Impact of soft data on calibration. *Trans. ASABE, 58*(6), 1637-1660. http://dx.doi.org/10.13031/trans.58.10726.

Baffaut, C., Dabney, S. M., Smolen, M. D., Youssef, M. A., Bonta, J. V., Chu, M. L., Guzman, J. A., Shedekar, V., Jha, M. K., & Arnold, J. G. (2015). Hydrologic and water quality modeling: Spatial and temporal considerations. *Trans. ASABE, 58*(6), 1661-1680. http://dx.doi.org/10.13031/trans.58.10714.

Balascio, C. C., Palmeri, D. J., & Gao, H. (1998). Use of a genetic algorithm and multi-objective programming for calibration of a hydrologic model. *Trans. ASAE, 41*(3), 615-619. http://dx.doi.org/10.13031/2013.17229.

Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D., & Andreassian, V. (2013). Characterising performance of environmental models. *Environ. Model. Software, 40*, 1-20. http://dx.doi.org/10.1016/j.envsoft.2012.09.011.

Blasone, R., Madsen, H., & Rosbjerg, D. (2007). Parameter estimation in distributed hydrological modelling: Comparison of global and local optimisation techniques. *Nordic Hydrol., 38*(4-5), 451-476. http://dx.doi.org/10.2166/nh.2007.024.

Borah, D. K., Yagow, G., Saleh, A., Barnes, P. L., Rosenthal, W., Krug, E. C., & Hauck, L. M. (2006). Sediment and nutrient modeling for TMDL development and implementation. *Trans. ASABE, 49*(4), 967-986. http://dx.doi.org/10.13031/2013.21742.

Bottcher, A. B., Whiteley, B. J., James, A. I., & Hiscock, J. G. (2012). Watershed assessment model (WAM): Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1367-1383. http://dx.doi.org/10.13031/2013.42248.

Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Toward improved streamflow forecasts: Value of semidistributed modeling. *Water Resources Res., 37*(11), 2749-2759. http://dx.doi.org/10.1029/2000WR000207.

Crawford, N. H., & Linsley, R. K. (1962). The synthesis of continuous streamflow hydrographs on a digital computer. Tech. Report No. 12. Palo Alto, Cal.: Stanford University, Department of Civil Engineering.

Daggupati, P., Douglas-Mankin, K. R., Sheshukov, A. Y., Barnes, P. L., & Devlin, D. L. (2011). Field-level targeting using SWAT: Mapping output from HRUs to fields and assessing limitations of GIS input data. *Trans. ASABE, 54*(2), 501-514. http://dx.doi.org/10.13031/2013.36453.

Daggupati, P., Yen, H., White, M. J., Srinivasan, R., Arnold, J. G., Keitzer, C. S., & Sowa, S. P. (2015). Impact of model development decision on hydrological processes and streamflow. *Hydrol. Proc.,* doi: 10.1002/hyp.10536.

Deb, D., Tuppad, P., Daggupati, P., Srinivasan, R., & Varma, D. (2015). Spatio-temporal impacts of biofuel production and climate variability on water quantity and quality in upper Mississippi River basin. *Water, 7*(7), 3283-3305. http://dx.doi.org/10.3390/w7073283.

DeJonge, K., Ascough II, J. C., Ahmadi, M., Andales, A. A., & Arabi, M. (2012). Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. *Ecol. Model., 231*, 113-125. http://dx.doi.org/10.1016/j.ecolmodel.2012.01.024.

Dile, Y. T., Karlberg, L., Daggupati, P., Srinivasan, R., Wiberg, D., & Rockström, J. (2015). Assessing the implications of water harvesting intensification on upstream-downstream social-ecological systems: A case study in the Lake Tana basin of Africa. *J. Sci. Total Environ*., 542(part A), 22-35, doi: 10.1016/j.scitotenv.2015.10.065.

Dinicola, R. S. (1992). Rainfall-runoff relations for the Puget Sound area. USGS Staff-Published Research Paper 121. Reston, Va.: U.S. Geological Survey.

Doherty, J., & Johnston, J. J. M. (2003). Methodologies for calibration and predictive analysis of a watershed model. *JAWRA, 39*(2), 251-265. http://dx.doi.org/10.1111/j.1752-1688.2003.tb04381.x.

Donnelly, L. M. (1997). Comparison of rainfall-runoff modeling techniques in small forested catchments. MS thesis. Burnaby, British Columbia, Canada: Simon Fraser University, Department of Geography.

Douglas-Mankin, K. R., Maski, D., Janssen, K. A., Tuppad, P., & Pierzynski, G. M. (2010). Modeling nutrient runoff yields from combined in-field crop practices using SWAT. *Trans. ASABE, 53*(5), 1557-1568. http://dx.doi.org/10.13031/2013.34914.

Douglas-Mankin, K. R., Daggupati, P., Sheshukov, A. Y., & Barnes, P. L. (2013). Paying for sediment: Field-scale conservation practice targeting, funding, and assessment using SWAT. *J. Soil Water Cons., 68*(1), 41-51.

Duan, Q., Sorooshian, S., & Gupta, V. K. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Res., 28*(4), 1015-1031. http://dx.doi.org/10.1029/91WR02985.

Duan, Q., Sorooshian, S., & Gupta, V. K. (1994). Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol., 158*(3-4), 265-284. http://dx.doi.org/10.1016/0022-1694(94)90057-4.

Duda, P. B., Hummel, P. R., Donigian Jr., A. S., & Imhoff, J. C. (2012). BASINS/HSPF: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1523-1547. http://dx.doi.org/10.13031/2013.42261.

Engel, B., Storm, D., White, M., Arnold, J., & Arabi, M. (2007). A hydrological/water quality model application protocol. *JAWRA, 43*(5), 1223-1236. http://dx.doi.org/10.1111/j.1752-1688.2007.00105.x.

Fenicia, F., Savenije, H. H. G., Matgen, P., & Pfister, L. (2007). A comparison of alternative multiobjective calibration strategies for hydrological modeling. *Water Resources Res. 43*(3), W03434. http://dx.doi.org/10.1029/2006WR005098.

Flanagan, D. C., Frankenberger, J. R., & Ascough II, J. C. (2012). WEPP: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1463-1477. http://dx.doi.org/10.13031/2013.42254.

Flerchinger, G. N., Caldwell, T. G., Cho, J., & Hardegree, S. P. (2012). Simultaneous heat and water (SHAW) model: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1395-1411. http://dx.doi.org/10.13031/2013.42250.

Goodrich, D. C., Burns, I. S., Unkrich, C. L., Semmens, D. J., Guertin, D. P., Hernandez, M., Yatheendradas, S., Kennedy, J. R., & Levick, L. R. (2012). KINEROS2/AGWA: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1561-1574. http://dx.doi.org/10.13031/2013.42264.

Gowda, P. H., Mulla, D. J., Desmond, E. D., Ward, A. D., & Moriasi, D. N. (2012). ADAPT: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1345-1352. http://dx.doi.org/10.13031/2013.42246.

Gupta, H. V. H., Sorooshian, S., & Yapo, P. O. (1999). Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrol. Eng., 4*(2), 135-143. http://dx.doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135).

Guzman, J. A., Shirmohammadi, A., Sadeghi, A. M., Wang, X., Chu, M. L., Jha, M. K., Parajuli, P. B., Harmel, R. D., Khare, Y., & Hernandez, J. (2015). Uncertainty considerations in calibration and validation of hydrologic and water quality models. *Trans. ASABE, 58*(6), 1745-1762. http://dx.doi.org/10.13031/trans.58.10710.

Hansen, S., Abrahamsen, P., Petersen, C. T., & Styczen, M. (2012). Daisy: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1315-1333. http://dx.doi.org/10.13031/2013.42244.

Harmel, R. D., & Smith, P. K. (2007). Consideration of measurement uncertainty in the evaluation of goodness-of-fit in hydrologic and water quality modeling. *J. Hydrol., 337*(3-4), 326-336. http://dx.doi.org/10.1016/j.jhydrol.2007.01.043.

Harmel, R. D., & King, K. W. (2005). Uncertainty in measured sediment and nutrient flux from small agricultural watersheds. *Trans. ASAE, 48*(5), 1713-1721. http://dx.doi.org/10.13031/2013.20005.

Harmel, R. D., Cooper, R. J., Slade, R. M., Haney, R. L., & Arnold, J. G. (2006). Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *Trans. ASABE, 49*(3), 689-701. http://dx.doi.org/10.13031/2013.20488.

Harmel, R. D., Smith, D. R., King, K. W., & Slade, R. M. (2009). Estimating storm discharge and water quality data uncertainty: A software tool for monitoring and modeling applications. *Environ. Model. Software, 24*(7), 832-842. http://dx.doi.org/10.1016/j.envsoft.2008.12.006.

Harmel, R. D., Smith, P. K., & Migliaccio, K. W. (2010). Modifying goodness-of-fit indicators to incorporate both measurement and model uncertainty in model calibration and validation. *Trans. ASABE, 53*(1), 55-63. http://dx.doi.org/10.13031/2013.29502.

Harmel, R. D., Smith, P. K., Migliaccio, K. W., Chaubey, I., Douglas-Mankin, K. R., Benham, B., Shukla, S., Muñoz-Carpena, R., & Robson, B. J. (2014). Evaluating, interpreting, and communicating performance of hydrologic/water quality models considering intended use: A review and recommendations. *Environ. Model. Software, 57*, 40-51. http://dx.doi.org/10.1016/j.envsoft.2014.02.013.

Herr, J. W., & Chen, C. W. (2012). WARMF: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1385-1394. http://dx.doi.org/10.13031/2013.42249.

Huard, D., & Mailhot, A. (2008). Calibration of hydrological model GR2M using Bayesian uncertainty analysis. *Water Resources Res., 44*(2), W02424. http://dx.doi.org/10.1029/2007WR005949.

Jaber, F. H., & Shukla, S. (2012). MIKE SHE: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1479-1489. http://dx.doi.org/10.13031/2013.42255.

Jakeman, A. J., Lester, R. A., & Norton, J. P. (2006). Ten interactive steps in development and evaluation of environmental models. *Environ. Model. Software, 21*(5), 602-614. http://dx.doi.org/10.1016/j.envsoft.2006.01.004.

James, L. D., & Burges, S. J. (1982). Selection, calibration, and testing of hydrologic models. In C. T. Haan, H. P. Johnson, & D. L. Brakensiek (Eds.), *Hydrologic Modeling of Small Watersheds* (pp. 437-472). St. Joseph, Mich.: ASAE.

Kim, S. M., Benham, B. L., Brannan, K. M., Zeckoski, R. W., & Doherty, J. (2007). Comparison of hydrologic calibration of HSPF using automatic and manual methods. *Water Resources Res., 43*(1), W01402. http://dx.doi.org/10.1029/2006WR004883.

Klemeš, V. (1986). Operational testing of hydrologic simulation models. *Hydrol. Sci. J., 31*(1), 13-24. http://dx.doi.org/10.1080/02626668609491024.

Knisel, W. G., & Douglas-Mankin, K. R. (2012). CREAMS/GLEAMS: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1291-1302. http://dx.doi.org/10.13031/2013.42241.

Konikow, L. F., & Bredehoeft, J. D. (1992). Groundwater models cannot be validated. *Adv. Water Resources, 15*(1), 75-83. http://dx.doi.org/10.1016/0309-1708(92)90033-X.

Ma, L., Ascough II, J. C., Ahuja, L. R., Shaffer, M. J., Hanson, J. D., & Rojas, K. W. (2000). Root Zone Water Quality Model sensitivity analysis using Monte Carlo simulation. *Trans. ASABE, 43*(4), 883-895. http://dx.doi.org/10.13031/2013.2984.

Ma, L., Ahuja, L. R., Nolan, B. T., Malone, R. W., Trout, T. J., & Qi, Z. (2012). Root Zone Water Quality Model (RZWQM2): Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1425-1446. http://dx.doi.org/10.13031/2013.42252.

Malone, R. W., Yagow, G., Baffaut, C., Gitau, M. W., Qi, Z., Amatya, D. M., Parajuli, P. B., Bonta, J. V., & Green, T. R. (2015). Parameterization guidelines and considerations for hydrologic models. *Trans. ASABE, 58*(6), 1681-1703. http://dx.doi.org/10.13031/trans.58.10709.

Maski, D., Mankin, K. R., Janssen, K. A., Tuppad, P., & Pierzynski, G. M. (2008). Modeling runoff and sediment yields from combined in-field crop practices using SWAT. *J. Soil Water Cons., 63*(4), 193-203. http://dx.doi.org/10.2489/jswc.63.4.193.

Mehta, V. M., Mendoza, K., Daggupati, P., Srinivasan, R., Rosenberg, N. J., & Deb, D. (2015). High-resolution simulations of decadal climate variability impacts on water yield in the Missouri River basin with the Soil and Water Assessment Tool (SWAT). *J. Hydromet*., http://dx.doi.org/10.1175/JHM-D-15-0039.1.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., Veith, T. L., & Van Liew, M. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE, 50*(3), 885-900. http://dx.doi.org/10.13031/2013.23153.

Moriasi, D. N., Wilson, B. N., Douglas-Mankin, K. R., Arnold, J. G., & Gowda, P. H. (2012). Hydrologic and water quality models: Use, calibration, and validation. *Trans. ASABE, 55*(4), 1241-1247. http://dx.doi.org/10.13031/2013.42265.

Moriasi, D. N., Gitau, M. W., Pai, N., & Daggupati, P. (2015). Hydrologic and water quality models: Performance measures and evaluation criteria. *Trans. ASABE, 58*(6), 1763-1785. http://dx.doi.org/10.13031/trans.58.10715.

Muthuwatta, L. P., Booij, M. J., Reintjes, T. H., Bos, M. G., Geiske, A. S., & Ahmad, M. (2009). Calibration of a semi-distributed hydrological model using discharge and remote sensing data. In *New Approaches to Hydrological Prediction in Data-Sparse Regions* (pp. 52-58). IAHS Publication 333. Wallingford, U.K.:

International Association of Hydrological Sciences.
Negm, L. M., & Youssef, M. A., Skaggs, R. W., Chescheir, G. M., Jones, J. (2014). DRAINMOD-DSSAT model for simulating hydrology, soil carbon and nitrogen dynamics, and crop growth for drained agricultural land. *Agric. Water Mgmt., 137*, 30-45. http://dx.doi.org/10.1016/j.agwat.2014.02.001.
Pai, N., Saraswat, D., & Daniels, M. (2011). Identifying priority subwatersheds in the Illinois River drainage area in Arkansas watershed using a distributed modeling approach. *Trans. ASABE, 54*(6), 2181-2196. http://dx.doi.org/10.13031/2013.40657.
Parajuli, P. (2010). Assessing sensitivity of hydrologic responses to climate change from forested watershed in Mississippi. *Hydrol. Proc., 24*(26), 3785-3797. http://dx.doi.org/10.1002/hyp.7793.
Parajuli, P. B., Nelson, N. O., Frees, L. D., & Mankin, K. R. (2009). Comparison of AnnAGNPS and SWAT model simulation results in USDA-CEAP agricultural watersheds in south-central Kansas. *Hydrol. Proc., 23*(5), 748-763. http://dx.doi.org/10.1002/hyp.7174.
Pokhrel, P., & Gupta, H. (2010). On the use of spatial regularization strategies to improve calibration of distributed watershed models. *Water Resources Res., 46*(1), W01505. http://dx.doi.org/10.1029/2009WR008066.
Qi, C., & Grunwald, S. (2005). GIS-based hydrologic modeling in the Sandusky watershed using SWAT. *Trans. ASAE, 48*(1), 169-180. http://dx.doi.org/10.13031/2013.17960.
Qi, S., Sun, G., Wang, Y., McNulty, S. G., & Moore Myers, J. A. (2009). Streamflow response to climate and landuse changes in a coastal watershed in North Carolina. *Trans. ASABE, 52*(3), 739-749. http://dx.doi.org/10.13031/2013.27395.
Rantz, S. E., et al. (1982). *Measurement and Computation of Streamflow: Volume 1. Measurement of Stage and Discharge*. USGS Water Supply Paper 2175. Reston, Va.: U.S. Geological Survey.
Razavi, S., & Tolson, B. A. (2013). An efficient framework for hydrologic model calibration on long data periods. *Water Resources Res., 49*(12), 8418-8431. http://dx.doi.org/10.1002/2012WR013442.
Refsgaard, J. C. (1997). Parameterisation, calibration, and validation of distributed hydrological models. *J. Hydrol., 198*(1), 69-97. http://dx.doi.org/10.1016/S0022-1694(96)03329-X.
Refsgaard, J. C. (2000). Towards a formal approach to calibration and validation of models using spatial data. In *Spatial Patterns in Catchment Hydrology: Observations and Modeling* (pp. 329-354). Cambridge, U.K.: Cambridge University Press.
Refsgaard, J. C., & Storm, B. (1996). Chapter 3: Construction, calibration, and validation of hydrological models. In M. B. Abbot & J. C. Refsgaard (Eds.), *Distributed Hydrological Modelling* (pp. 41-54). Dordrecht, The Netherlands: Kluwer Academic.
Refsgaard, J. C., Henriksen, H. J., Harrar, W. G., Scholte, H., & Kassahun, A. (2005). Quality assurance in model based water management-Review of existing practice and outline of new approaches. *Environ. Model. Software, 20*(10), 1201-1215. http://dx.doi.org/10.1016/j.envsoft.2004.07.006.
Rosso, R. (1994). An introduction to spatially distributed modelling of basin response. In R. Rosso, A. Peano, I. Becchi, & G. A. Bemporad (Eds.), *Advances in Distributed Hydrology* (pp. 3-30). Highlands Ranch, Colo.: Water Resources Publications.
Runkel, R. L., Crawford, C. G., & Cohn, T. A. (2004). Load Estimator (LOADEST): A FORTRAN program for estimating constituent loads. Techniques and Methods: Book 4, Chapter A5. Reston, Va.: U.S. Geological Survey.
Santhi, C., Arnold, J. G., Williams, J. R., Dugas, W. A., Srinivasan, R., & Hauck, L. M. (2001). Validation of the SWAT model on a large river basin with point and nonpoint sources. *JAWRA, 37*(5), 1169-1188. http://dx.doi.org/10.1111/j.1752-1688.2001.tb03630.x.
Sheshukov, A. Y., Siebenmorgen, C. B., & Douglas-Mankin, K. R. (2011). Seasonal and annual impacts of climate change on watershed response using an ensemble of global climate models. *Trans. ASABE, 54*(6), 2209-2218. http://dx.doi.org/10.13031/2013.40660.
Sheshukov, A. Y., Douglas-Mankin, K. R., Sinnathamby, S., & Daggupati, P. (2015). Modeling pasture BMP effectiveness using a HRU-based sub-field approach in SWAT. *J. Environ. Mgmt.,* doi: 10.1016/j.jenvman.2015.10.023.
Shirmohammadi, A., Chaubey, I., Harmel, R. D., Bosch, D. D., Muñoz-Carpena, R., Dharmasri, C., Sexton, A., Arabi, M., Wolfe, M. L., Frankenberger, J., Graff, C., & Sohrabi, T. M. (2006). Uncertainty in TMDL models. *Trans. ASABE, 49*(4), 1033-1049. http://dx.doi.org/10.13031/2013.21741.
Silberstein, R. P. (2006). Hydrological models are so good, do we still need data? *Environ. Model. Software 21*(9), 1340-1352. http://dx.doi.org/10.1016/j.envsoft.2005.04.019.
Šimůnek, J., Genuchten, M. T., & Šejna, M. (2012). HYDRUS: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1261-1274. http://dx.doi.org/10.13031/2013.42239.
Skaggs, R. W., Youssef, M. A., & Chescheir, G. M. (2012). DRAINMOD: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1509-1522. http://dx.doi.org/10.13031/2013.42259.
Srinivasan, R., Zhang, X., & Arnold, J. (2010). SWAT ungauged: Hydrological budget and crop yield predictions in the upper Mississippi River basin. *Trans. ASABE, 53*(5), 1533-1546. http://dx.doi.org/10.13031/2013.34903.
Tian, S., Youssef, M. A., Skaggs, R. W., Amatya, D. M., & Chescheir, G. M. (2012). DRAINMOD-FOREST: Integrated modeling of hydrology, soil carbon and nitrogen dynamics, and plant growth for drained forests. *J. Environ. Qual., 41*(3), 764-782. http://dx.doi.org/10.2134/jeq2011.0388.
Tian, S., Youssef, M. A., Amatya, D. M., & Vance, E. D. (2014). Global sensitivity analysis of DRAINMOD-FOREST, an integrated forest ecosystem model. *Hydrol. Proc., 28*(15), 4389-4410. http://dx.doi.org/10.1002/hyp.9948.
Tuppad, P., Douglas-Mankin, K. R., Koelliker, J. K., & Hutchinson, M. S. (2010a). SWAT discharge response to spatial rainfall variability in a Kansas watershed. *Trans. ASABE, 53*(1), 65-74. http://dx.doi.org/10.13031/2013.29503.
Tuppad, P., Douglas-Mankin, K. R., Koelliker, J. K., Hutchinson, J. M., & Knapp, M. C. (2010b). NEXRAD stage III precipitation local bias adjustment for streamflow prediction. *Trans. ASABE, 53*(5), 1511-1520. http://dx.doi.org/10.13031/2013.34910.
Tuppad, P., Douglas-Mankin, K. R., Lee, T., Srinivasan, R., & Arnold, J. G. (2011). Soil and Water Assessment Tool (SWAT) hydrologic/water quality model: Extended capability and wider adoption. *Trans. ASABE, 54*(5), 1677-1684. http://dx.doi.org/10.13031/2013.39856.
Van Liew, M. W., Arnold, J. G., & Bosch, D. D. (2005). Problems and potential of autocalibrating a hydrologic model. *Trans. ASAE, 48*(3), 1025-1040. http://dx.doi.org/10.13031/2013.18514.
Van Liew, M. W., Veith, T. L., Bosch, D. D., & Arngold, J. G. (2007). Suitability of SWAT for the Conservation Effects Assessment Project: Comparison on USDA Agricultural Research Service watersheds. *J. Hydrol. Eng., 12*(2), 173-189. http://dx.doi.org/10.1061/(ASCE)1084-0699(2007)12:2(173).
Wallner, M., Haberlandt, U., & Dietrich, J. (2012). Evaluation of different calibration strategies for large-scale continuous hydrological modeling. *Adv. Geosci., 31*, 67-74. http://dx.doi.org/10.5194/adgeo-31-67-2012.
Wang, X., Youssef, M. A., Skaggs, W. R., Atwood, J. D., &

Frankenberger, J. R. (2005). Sensitivity analysis of the nitrogen simulation model, DRAINMOD-N II. *Trans. ASAE, 48*(6), 2205-2212. http://dx.doi.org/10.13031/2013.20106.
Wang, X., Williams, J. R., Gassman, P. W., Baffaut, C., Izaurralde, R. C., Jeong, J., & Kiniry, J. R. (2012). EPIC and APEX: Model use, calibration, and validation. *Trans. ASABE, 55*(4), 1447-1462. http://dx.doi.org/10.13031/2013.42253.
White, K. L., & Chaubey, I. (2005). Sensitivity analysis, calibration, and validations for a multisite and multivariable SWAT model. *JAWRA, 41*(5), 1077-1089. http://dx.doi.org/10.1111/j.1752-1688.2005.tb03786.x.
Woznicki, S., Nejadhashemi, A., & Smith, C. (2011). Assessing best management practice implementation strategies under climate change scenarios. Trans. ASABE, 54(1), 171-190. http://dx.doi.org/10.13031/2013.36272.
Yen, H., Wang, X., Fontane, D. G., Harmel, R. D., & Arabi, M. (2014). A framework for propagation of uncertainty contributed by input data, parameterization, model structure, and calibration/validation data in watershed modeling. Environ. Model. Software 54, 211-221. http://dx.doi.org/10.1016/j.envsoft.2014.01.004.
Youssef, M. A., Skaggs, R. W., Gilliam, J. W., & Chescheir, G. M. (2006). Field evaluation of a model for predicting nitrogen losses from drained lands. J. Environ. Qual., 35(6), 2026-2042. http://dx.doi.org/10.2134/jeq2005.0249.
Yuan, Y., Khare, Y., Wang, X., Parajuli, P. B., Kisekka, I., & Finsterle, S. (2015). Hydrologic and water quality models: Sensitivity. Trans. ASABE, 58(6), 1721-1744. http://dx.doi.org/10.13031/trans.58.10611.
Zeckoski, R. W., Smolen, M. D., Moriasi, D. N., Frankenberger, J. R., & Feyereisen, G. W. (2015). Hydrologic and water quality terminology as applied to modeling. Trans. ASABE, 58(6), 1619-1635. http://dx.doi.org/10.13031/trans.58.10713.
Zhang, X., Srinivasan, R., Zhao, K., & Van Liew, M. (2009). Evaluation of global optimization algorithms for parameter calibration of a computationally intensive hydrologic model. Hydrol. Proc., 23(3), 430-441. http://dx.doi.org/10.1002/hyp.7152.