ASABE EP621 JUN2017
Guidelines for Calibrating, Validating, and Evaluating Hydrologic and Water Quality (H/WQ) Models

American Society of
Agricultural and Biological Engineers

STANDARD

ASABE is a professional and technical organization, of members worldwide, who are dedicated to advancement of engineering applicable to agricultural, food, and biological systems. ASABE Standards are consensus documents developed and adopted by the American Society of Agricultural and Biological Engineers to meet standardization needs within the scope of the Society; principally agricultural field equipment, farmstead equipment, structures, soil and water resource management, turf and landscape equipment, forest engineering, food and process engineering, electric power applications, plant and animal environment, and waste management.

NOTE: ASABE Standards, Engineering Practices, and Data are informational and advisory only. Their use by anyone engaged in industry or trade is entirely voluntary. The ASABE assumes no responsibility for results attributable to the application of ASABE Standards, Engineering Practices, and Data. Conformity does not ensure compliance with applicable ordinances, laws and regulations. Prospective users are responsible for protecting themselves against liability for infringement of patents.

ASABE Standards, Engineering Practices, and Data initially approved prior to the society name change in July of 2005 are designated as "ASAE", regardless of the revision approval date. Newly developed Standards, Engineering Practices and Data approved after July of 2005 are designated as "ASABE".

Standards designated as "ANSI" are American National Standards as are all ISO adoptions published by ASABE. Adoption as an American National Standard requires verification by ANSI that the requirements for due process, consensus, and other criteria for approval have been met by ASABE.

Consensus is established when, in the judgment of the ANSI Board of Standards Review, substantial agreement has been reached by directly and materially affected interests. Substantial agreement means much more than a simple majority, but not necessarily unanimity. Consensus requires that all views and objections be considered, and that a concerted effort be made toward their resolution.

CAUTION NOTICE: ASABE and ANSI standards may be revised or withdrawn at any time. Additionally, procedures of ASABE require that action be taken periodically to reaffirm, revise, or withdraw each standard.

ASABE, 2950 Niles Road, St. Joseph, MI 49085-9659, USA, phone 269-429-0300, fax 269-429-3852, hq@asabe.org

Guidelines for Calibrating, Validating, and Evaluating Hydrologic and Water Quality (H/WQ) Models

Developed by the NRES-21 Hydrology committee of ASABE. Approved as an ASABE Engineering Practice June 2017.

Keywords: Hydrologic processes, Hydrology, Modeling, Model calibration, Model validation, Water quality.

1 Purpose and Scope

Hydrologic and water quality (H/WQ) models consist of interrelated assemblages of mathematical equations that represent the processes governing the movement, fate, and transport of water, sediments, nutrients, pesticides, bacteria, and other constituents in and on the surface of the soil and through groundwater, stream, river, and reservoir systems. Reliable models provide information for making sound management, policy, and regulatory decisions. While calibration, validation, and evaluation (CV&E) are essential for ensuring model accuracy and precision, there is little consistency in how model practitioners conduct and document this process and communicate the results. Therefore, this Engineering Practice was designed to contribute to good modeling practice recommendations to improve modeling methodology and application of H/WQ models through improved model performance as well as enhanced communication of the processes utilized and results to modelers, decision-makers, and other modeling stakeholders.

It builds on recommended protocols recognizing the need for formal model development, application, and communication guidelines rather than ad hoc approaches because of the often-serious human health, regulatory, ecological, socio-economic, and policy implications of modeling applications. This Engineering Practice is not meant to serve as a comprehensive guide but rather to compile and enhance guidance from foundational manuscripts such as Beck (1987), ASCE (1993), Haan et al. (1995), Refsgaard et al. (2005), Jakeman et al. (2006), Engel et al. (2007), Moriasi et al. (2007), Biondi et al. (2012), Bennett et al. (2013), Ritter and Muñoz-Carpena (2013) and Black et al. (2014), and advances described in recent compilations (e.g., Muñoz-Carpena et al., 2006; Moriasi et al., 2012, 2015a).

1.1 This Engineering Practice aims to enhance calibration, validation, and evaluation of H/WQ models through:

- establishment of consistent terminology and definitions,
- selection of model based on intended use, processes simulated, available data, and temporal and spatial scale,
- compilation and processing of both hard and soft input data and calibration data,
- model parameterization and calibration,
- evaluation of model performance,
- re-examination of input and calibration data and possible model refinement,
- re-evaluation of model results and performance considering intended use, and
- documentation of the modeling process and model results.

2 Definitions

Speaking a common language is an essential first step in improving H/WQ modeling. The following definitions were taken largely from Zeckoski et al. (2015) except where indicated.

2.1 model intended use: the purpose for which the model results will be used. Model intended use affects the required accuracy and precision of the model results (Harmel et al., 2014).

2.1.1 regulatory/legal use: model results will be used for regulatory or legal purposes or have human health implications.

2.1.2 planning use: model results will be used for planning purposes, conservation implementation, or policy formulation.

2.1.3 exploratory use: model results will be used for beta testing, model development, or for initial and approximate comparisons.

2.2 modeling objectives: the specific goals of the modeling study (e.g., determine the contributions from existing pollutant sources to a stream pollutant load or determine the impacts of climate change on drought frequency and duration).

2.3 parameter: quantifiable characteristic of a feature or process represented in a model (Malone et al., 2015).

2.3.1 parameterization: the process of assigning values to model parameters. Parameterization includes calibration and validation, which both require model performance evaluation.

2.3.2 calibration: the process of adjusting selected input parameter values and initial conditions to obtain simulated values that match measured observations with the desired accuracy.

2.3.2.1 manual calibration: the sequence of parameter adjustments is left entirely to the modeler.

2.3.2.2 automatic calibration: a search algorithm is used to decide what parameters are adjusted and by what amount based on the value of an objective function.

2.3.2.3 calibrated model: a model applied to a particular physical setting through appropriate parameterization and calibrated to measured data.

2.3.3 validation: the process of verifying that a calibrated model reproduces measured observations for conditions different than were used for the model calibration.

2.3.4 evaluation: the process of using graphical, quantitative, and/or statistical techniques, along with performance ratings and model intended use to judge the quality of model predictions (Harmel et al., 2014).

2.3.5 verification: the process of determining that the model code is correctly implemented and reflects the conceptual model.

2.4 state variable: time series or function used as input for the model or calculated by the model.

2.5 hard data: data measured within the study area (Arnold et al., 2015).

2.6 soft data: information on individual processes that may not be directly measured in the study area, including temporal or spatial averages, estimated quantities using other models, or qualitative knowledge from experimentalists (Arnold et al., 2015).

2.7 initial conditions: values taken by variables (e.g. soil moisture, snow depth, stage of vegetation) that describe the study area at the beginning of the simulation.

2.8 warm-up period: a simulation period that precedes the time period of interest and is long enough for variables such as soil moisture, perennial vegetation, or reservoir levels to reach values that are independent from initial conditions.

2.9 sensitivity: the relative change of a model to a change in a parameter or input variable.

2.10 sensitivity analysis: the process of determining how sensitive the model output is to selected input parameters.

2.11 uncertainty analysis: the process of estimating the effect of input data and model structure uncertainty on model output variables.

2.12 model performance measures: tools and techniques that evaluate how well simulated values represent measured observations over a specified time period.

3 Consider the intended model use

Upon initiation of modeling projects, the model's intended use (e.g., Exploratory, Planning and Regulatory/Legal) should be taken into consideration (Harmel et al., 2014). Although these intended use categories are not mutually exclusive and may not cover the entire spectrum of modeling applications, they represent general categories that warrant different expectations related to model CV&E.

3.1 Determine the model intended use (e.g., legal/regulatory, planning, or exploratory).

3.2 Determine the modeling objectives.

3.3 Given the intended uses and modeling objectives, determine:

3.3.1 Processes to be simulated.

3.3.2 Spatial and temporal scale of the simulation.

3.3.3 Expectations for accuracy and precision of model predictions.

4 Select model

Select the model that best meets the intended use, modeling objectives, and performance expectations. Moriasi et al. (2012) summarized more than 20 commonly applied H/WQ models; however, that collection is not meant to be exhaustive.

4.1 Select a model that is appropriate given:

- processes that need to be simulated,
- required outputs,
- required spatial and temporal resolution,
- available input and calibration data,
- modeling expertise, and
- computing resources.

4.2 If multiple spatial scales must be considered, the following options should be taken into account (Baffaut et al., 2015):

4.2.1 Simplify by breaking the project into smaller questions using appropriate scales and models. Link two models when addressing interactions of processes that operate over differing spatial scales.

4.2.2 If multiple spatial extents are considered within the same model, perform calibration and validation for smaller areas first, increasing progressively to larger areas. Alternatively, multiple scales can be considered simultaneously.

4.2.3 In the absence of calibration and validation data at the relevant spatial or temporal scale, calibrated parameter values from a model calibrated at a different temporal or spatial scale may not be representative and

