Environmental Modelling & Software 59 (2014) 135–145

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Toward improved calibration of watershed models: Multisite multiobjective measures of information

Mehdi Ahmadi [a], Mazdak Arabi [b], James C. Ascough II [c,*], Darrell G. Fontane [b], Bernard A. Engel [d]

[a] *Spatial Science Laboratory, Texas A&M University, 1500 Research Parkway, College Station, TX 77843, USA*
[b] *Dept. of Civil and Environmental Engineering, Colorado State University, Fort Collins, CO 80523, USA*
[c] *USDA-ARS, ASRU, 2150 Centre Ave., Bldg. D, Suite 200, Fort Collins, CO 80526, USA*
[d] *Dep. of Agricultural and Biological Engineering, Purdue University, West Lafayette, IN 47907, USA*

ARTICLE INFO

Article history:
Received 27 January 2013
Received in revised form
15 May 2014
Accepted 16 May 2014
Available online

Keywords:
Parameter estimation
Hydrologic modeling
Optimization
SWAT model
Water quality

ABSTRACT

This paper presents a computational framework for incorporation of disparate information from observed hydrologic responses at multiple locations into the calibration of watershed models. The framework consists of three components: 1) an *a-priori* characterization of system behavior; 2) a formal and statistically valid formulation of objective function(s) of model errors; and 3) an optimization engine to determine the Pareto-optimal front for the selected objectives. The proposed framework was applied for calibration of the Soil and Water Assessment Tool (SWAT) in the Eagle Creek Watershed, Indiana, USA using three single objective optimization methods [Shuffled Complex Evolution (SCE), Dynamically Dimensioned Search (DDS), and DiffeRential Evolution Adaptive Metropolis (DREAM)], and one multi-objective optimization method. Solutions were classified into behavioral and non-behavioral using percent bias and Nash–Sutcliffe model efficiency coefficient. The results showed that aggregation of streamflow and NO_x (NO_3-N + NO_2-N) information measured at multiple locations within the watershed into a single measure of weighted errors resulted in faster convergence to a solution with a lower overall objective function value than using multiple measures of information. However, the DREAM method solution was the only one among the three single objective optimization methods considered in this study that satisfied the conditions defined for characterizing system behavior. In particular, aggregation of streamflow and NO_x responses undermined finding "very good" behavioral solutions for NO_x, primarily because of the significantly larger number of observations for streamflow. Aggregation of only NO_x responses into a single measure expedited finding better solutions although aggregation of data from nested sites appeared to be inappropriate because of correlated errors. This study demonstrates the importance of hydrologic and water quality data availability at multiple locations, and also highlights the use of multiobjective approaches for proper calibration of watershed models that are used for pollutant source identification and watershed management.

Published by Elsevier Ltd.

1. Introduction

Watershed models are increasingly embedded in the decision making process to address a wide range of hydrologic and water quality issues. In the United States, federal law requires individual states to develop total maximum daily loads (TMDLs) for impaired water bodies to attain ambient water quality standards through the control of point and nonpoint sources (NRC, 2001). Similarly, the European Water Framework Directive aims to enhance the water quality status of all water bodies within its jurisdiction (Kaika, 2003). Environmental simulation models play a central role in successful implementation of watershed management programs by providing the means to assess the relative contribution of different sources (i.e., stressors) to the impairment (Ahmadi et al., 2014b). Therefore, it is of keen interest to evaluate the performance validity of watershed models according to the past observations of fluxes of water and contaminants at multiple locations on the stream network. In most cases, daily or more frequent discharge measurements are available at watershed outlets on many rivers and

* Corresponding author. Tel.: +1 970 492 7371.
E-mail addresses: ahmadi@tamu.edu (M. Ahmadi), mazdak.arabi@colostate.edu (M. Arabi), jim.ascough@ars.usda.gov (J.C. Ascough), darrell.fontane@colostate.edu (D.G. Fontane), engelb@purdue.edu (B.A. Engel).

http://dx.doi.org/10.1016/j.envsoft.2014.05.012
1364-8152/Published by Elsevier Ltd.

streams. On the other hand, nutrient concentrations are often measured by local watershed groups on less frequent time steps (e.g., weekly or monthly) at the smaller subwatershed level.

Application of models that credibly represent important processes of the natural system presents a scientific challenge (Ahmadi et al., 2014a; Konikow and Bredehoeft, 1992). With the temptation to incorporate more parameters in models to represent a broader range of hydrologic and water quality processes has come an insidious effect: the ever-increasing complexity of model structures. Therefore, efficient and effective use of observed data is vital for calibration of complex spatially distributed process-based models. The literature is replete with application of automated calibration methods to minimize the error between observed and modeled values. However, an important and often neglected issue in calibration of complex models of the environment is that while optimization techniques facilitate the search for solutions with minimum errors, they do not necessarily ascertain model adequacy for mimicking the observed behavior of the system (Matthews et al., 2011). System behavior is often defined according to specific model application goals and characteristics of model and data (Bennett et al., 2013; Moriasi et al., 2007). For example, ambient water quality standards for nutrients and pesticides are expressed in terms of average annual responses. Thus, when models are used to support nutrient or pesticide TMDL, statistical measures such as average annual errors as a percentage of observed responses are used in the selection of final model parameters. Mean, minimum, peaks, variance, distribution, skewness, and trends of data are commonly used indicators of system behavior.

On the other hand, hydrologic and water quality observations are characterized by varying measurement errors and uncertainties, varying sample size, and are typically noncommensurable (Willems, 2009). These considerations must be taken into account when using observed data in construction of the proper likelihood function(s) for calibration purposes (Ahmadi et al., 2014a; Beven and Binley, 1992; Beven and Freer, 2001; Mantovan and Todini, 2006; Sorooshian and Dracup, 1980; Willems, 2009). The effectiveness of parameter estimation techniques depends greatly on the selection of proper likelihood function. There is an ongoing debate in the scientific community on the use of either a formal or informal likelihood function for calibration of hydrologic models (Beven et al., 2012; Clark et al., 2011; Mantovan and Todini, 2006; McMillan and Clark, 2009; Stedinger et al., 2008; Vrugt et al., 2009). Model performance is usually evaluated (based on subjective judgment of the analyst) against informal likelihood measures such as the Nash and Sutcliffe (1970) model efficiency coefficient or percent bias. Alternatively, formal likelihood functions are based on a strict assumptions about the structure of residuals represented by a statistical model (Beven et al., 2012). The use of a formal Bayesian-based likelihood function can provide more acceptable and statistically valid prediction intervals for future observations (Stedinger et al., 2008) and may result in a better coverage of observed data and more acceptable posterior distribution of parameters (Vrugt et al., 2009). However, in the context of watershed management, a statistically valid likelihood function may not exist (Beven, 2006; Gupta et al., 1998). On the other hand, Beven et al. (2007) and Beven (2008) showed that the formal Bayesian identification of models can be considered as a special case of generalized likelihood uncertainty estimation (GLUE) and is applicable if a strong assumption about the nature of the modeling errors can be made.

Overall, both formal and informal likelihood functions have their strengths and weaknesses. Both approaches could result in similar exploration of the parameter space, estimation of parametric uncertainty, and representation of the observed behavior of the system under study (McMillan and Clark, 2009; Vrugt et al., 2009). Automatic calibration procedures are often employed using formal likelihood measures, but the final choice of parameter values will always depend on informal/subjective measures that adequately capture analyst preferences. Thus, a successful calibration approach should identify models that represent behavior of the system in addition to finding the minimum error. The scientific literature contains numerous studies on noncommensurable measures of performance for classification of model parameter sets as behavioral (i.e., good or acceptable) or non-behavioral (i.e., bad or unacceptable) solutions (see, for instance, Beven and Binley, 1992; Blazkova and Beven, 2009; Klepper et al., 1991; Nash and Sutcliffe, 1970; Spear and Hornberger, 1980).

Model calibration at multiple sites and for many responses is inherently a multiobjective problem (Gupta et al., 1998; Madsen, 2003). Multiobjective optimization approaches enable the analyst to assess trade-offs associated with conflicting objectives and determine a set of nondominated solutions that comprise the Pareto-optimal front. The Nondominated Sorted Genetic Algorithm II (NSGA-II) (Deb, 2001) is a commonly used multiobjective approach. The complexity of multiobjective methods increases substantially with an increasing number of objectives in the optimization problems. Typically, these methods require more model simulations than single objective techniques for convergence and are more difficult to implement. Single objective optimization methods are diverse, computationally less intensive, easier to visualize, easier for statistically analyze, and less prone to search process stagnation. Therefore, an analyst may opt to use a single aggregated objective function of weighted errors; however, this can lead to loss of important information from some of the observations (Fenicia et al., 2008). The Shuffled Complex Evolutionary (SCE) algorithm (Duan et al., 1992), Markov Chain Monte Carlo (MCMC) methods such as DiffeRential Evolution Adaptive Metropolis (DREAM) (Vrugt et al., 2009), and Dynamically Dimensioned Search (DDS) (Tolson and Shoemaker, 2007) are among the most popular single objective optimization methods for calibration of hydrological models. Several software applications have been developed to facilitate implementation of single objective and multiobjective optimization algorithms for watershed model analysis. For example, Wu and Liu (2012) developed a framework (R-SWAT-FME) for calibration, sensitivity, and uncertainty analysis of the Soil and Water Assessment Tool (SWAT, Arnold et al., 2011). Zhang et al. (2013) developed a Python-based calibration package (PP-SWAT) for calibration of SWAT using a parallel computing technique and the "A Multi-method Genetically Adaptive Multiobjective Optimization Algorithm" (AMALGAM, Vrugt and Robinson, 2007).

While considerable progress has been made in addressing system behavior, formal vs. informal likelihood functions, and single objective vs. multiobjective optimization approaches individually, there is little if any research in the literature that addresses these issues in an integrated fashion. The primary goal of this study is to present a computational framework for multisite multiobjective calibration of watershed models that integrates the strengths of both formal likelihood functions and informal measures of model performance. The proposed framework is a complementary approach and includes use of system behavior definition for classification of the optimal solution(s) based on the model application objectives. Two specific objectives are examined *en-route* to the overall goal of the study: 1) develop an integrated approach that uses a formal likelihood function to identify models with minimum errors and subjective-informal statistical measures that incorporate user-specified priorities to represent behavior of the system; and 2) demonstrate the utility of the proposed approach using a real-world case study that evaluates the effectiveness of single objective and multiobjective optimization approaches in generating optimal solutions consistent with model application purpose.