Environmental Modelling & Software 105 (2018) 286–295

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

On the practical usefulness of least squares for assessing uncertainty in hydrologic and water quality predictions

D. Del Giudice [a, b, *], R. Logsdon Muenich [c], M. McCahon Kalcic [d], N.S. Bosch [e], D. Scavia [f], A.M. Michalak [a]

[a] Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305, USA
[b] Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Raleigh, NC 27695, USA
[c] School of Sustainable Engineering & the Built Environment, Arizona State University, Tempe, AZ 85287, USA
[d] Department of Food, Agricultural and Biological Engineering, The Ohio State University, Columbus, OH 43210, USA
[e] Lilly Center for Lakes & Streams, Grace College, Winona Lake, IN 46590, USA
[f] University of Michigan, Ann Arbor, MI 48104, USA

ARTICLE INFO

Article history:
Received 17 January 2018
Received in revised form
28 February 2018
Accepted 15 March 2018

Keywords:
Uncertainty assessment
Mechanistic modeling
Surface hydrology
Water quality
Least squares
Statistical inference

ABSTRACT

Sophisticated methods for uncertainty quantification have been proposed for overcoming the pitfalls of simple statistical inference in hydrology. The implementation of such methods is conceptually and computationally challenging, however, especially for large-scale models. Here, we explore whether there are circumstances in which simple approaches, such as least squares, produce comparably accurate and reliable predictions. We do so using three case studies, with two involving a small sewer catchment with limited calibration data, and one an agricultural river basin with rich calibration data. We also review additional published case studies. We find that least squares performs similarly to more sophisticated approaches such as a Bayesian autoregressive error model in terms of both accuracy and reliability if calibration periods are long or if the input data and the model have minimal bias. Overall, we find that, when mindfully applied, simple statistical methods such as LS can still be useful for uncertainty quantification.

1. Introduction

Statistical model calibration and uncertainty quantification (UQ) have recently received substantial attention in surface hydrology and water quality research. Several studies have stressed the importance of more realistically describing the behavior of calibration errors, a.k.a. residuals, and thus moving beyond least squares (LS) calibration assumption of independent and normally distributed residuals (e.g., Reichert and Mieleitner, 2009; Renard et al., 2011; Honti et al., 2013). In particular, it has been suggested that, by using error models that explicitly consider the heteroscedasticity (i.e. non-constant variance) and autocorrelation of the calibration errors, parameter estimation and subsequent predictive uncertainty assessment can be improved in a relatively straightforward manner (Sorooshian and Dracup, 1980; Yang et al., 2007b; Schoups and Vrugt, 2010; Del Giudice et al., 2015a). However, while such sophisticated approaches have been shown to be helpful in the specific situations where they were tested, this does not necessarily imply that simpler statistical techniques such as least squares calibration are never useful. Therefore, there is a critical knowledge gap in hydrologic and environmental modeling regarding when simple calibration approaches are acceptable versus when more sophisticated ones are needed. Understanding the domain of applicability of simple least squares along with its limitations is essential. Indeed, we argue that the presupposition that statistical inference always has to be conducted with conceptually and computationally burdensome methods might be inducing modelers to eschew UQ altogether (e.g., Bosch et al., 2011; Coutu et al., 2012; Razavi and Tolson, 2013) or use "pseudo-statistical" methods with unclear probabilistic interpretation (e.g., Freni et al., 2009; Beven and Smith, 2015). In the context of predictive UQ, we therefore address an important yet so far unanswered question: Are there cases in which a simple method such as least squares yields predictions with precision and accuracy that are on par with state-of-the-art approaches that account for error autocorrelation

* Corresponding author. Department of Global Ecology, Carnegie Institution for Science, Stanford, CA, USA.
E-mail address: ddelgiudice@carnegiescience.edu (D. Del Giudice).

https://doi.org/10.1016/j.envsoft.2018.03.009
1364-8152/

and parameter uncertainty? If so, are there features that make a particular case study a better candidate for simple approaches to uncertainty quantification? Here, we contrast three case studies with different degrees of data availability and model discrepancy to answer these questions. By also drawing on published hydrologic and water quality case studies we argue that LS can be used, provided that some criteria are met and that the method is applied with some caution. We additionally shed light upon the specific cases in which more sophisticated statistical methods are needed to deliver useful parameter estimates and uncertainty intervals.

2. Study areas and models

We investigate the suitability of LS calibration and subsequent uncertainty propagation in two catchments that differ substantially in terms of geographic domain, availability of calibration data, quality of input data, type and complexity of hydrological model, variables predicted, and types of systematic errors. For one catchment we consider two cases, one with low systematic deviations between model results and output data and one with high bias induced by forcing the model with less accurate input estimates. In each of the three cases we split the recorded time series into a calibration period, where output data are used for parameter estimation, and a validation period, where output data are used to corroborate model predictive abilities.

2.1. Case studies 1 (CS1 and CS1'): watershed with limited data

CS1 and CS1′ involve the same small, partially combined sewer network located in Adliswil, Zurich Canton, Switzerland (Fig. S1). The watershed has an area of 28.6 ha, only a fraction of which contributes to the sewer outflow. The effective contributing area of the watershed is indeed a calibration parameter (see below). The area is characterized by medium density residential development and a slope of about 8.7%. The site was monitored in 2013 to quantify the occurrence of sewer overflows and to understand the impact of the location of precipitation measurements on discharge predictions. For calibration we use a discharge Q [l/s] time series of an event including 97 observations recorded every 4 min (Fig. 1). This calibration period includes two storm events of duration greater than the catchment response time, which is on the order of minutes. For validation we use a subsequent event that occurred 80 days later and included 179 observations. Such short time series are typical in hydrological modeling of urban catchments (e.g., Freni et al., 2009; Coutu et al., 2012). For CS1, input data were recorded by a pluviometer from the Swiss meteorological office[1] located circa 7.5 km Northeast of the catchment (Fig. S1). The second version of this case study, CS1', uses more accurate input obtained by averaging data from two pluviometers located within the catchment area itself. A comparison of the precipitation records from these pluviometers reveals that the precipitation input data used in CS1 has substantial systematic errors (Fig. S3). The time series of sewer runoff at the outlet of the catchment is modeled using a lumped linear reservoir model with a harmonic function describing the wastewater oscillations (see Del Giudice et al. (2016) for further details about the catchment and the model). In this investigation, we calibrate the three model parameters related to rainfall-runoff, namely A [m^2], the area contributing to the stormwater outflow, k [hr], the mean residence time in the virtual reservoir representing the catchment, and x_{gw} [l/s], the baseflow.

[1] www.hw.zh.ch/hochwasser/foto/DB%20SMA.pdf.

2.2. Case study 2 (CS2): watershed with abundant data

CS2 is the River Raisin basin, which has an area of 2784 km^2 and is primarily rural (72%) and forested (16%). The variables of interest are river discharge Q [m^3/s] and soluble reactive phosphorus load *SRP*, [kg/d]. The calibration period contains 1095 discharge observations and 1095 *SRP* load observations at daily resolution (Fig. 3). This calibration period includes numerous storm events of duration longer than the catchment response time, which is on the order of days. The validation period immediately follows the calibration period and includes 366 discharge observations and 335 *SRP* load observations. The watershed dynamics are simulated using the Soil and Water Assessment Tool (SWAT) (Arnold et al., 1998). SWAT is a hydrologic transport model that operates at catchment scale. It is both more complex, due to its more explicit representation of spatial heterogeneity and watershed processes, and more computationally-demanding than the simple reservoir model used in CS1 and CS1'. The River Raisin basin and model are well studied in the context of furthering the understanding of the dynamics of nutrient loading from agricultural areas (Bosch et al., 2011). The SWAT model used here includes all the same process parameterizations, inputs, and management details as in Muenich et al. (2017). The model is driven by daily precipitation and temperature observations from nine NOAA GHCN land surface stations (Menne et al., 2012), most of which are located within the catchment area (Fig. S2). Daily discharge and *SRP* observations used for calibration and validation are obtained from Heidelberg University NCWQR (2015). In the current application, we calibrate three model parameters: *CN*2 [-], the runoff curve number for moisture condition II, *SMTMP* [°C], the snow melt base temperature, and *PHOSKD* [-], the phosphorus soil partitioning coefficient. These parameters are selected because they are primary controls on three key processes, namely rainfall-runoff, snowmelt, and biochemical reaction, and the output variables of interest are sensitive to them.

3. Methods

3.1. Simple method: frequentist least squares (LS)

The least squares method, LS, is a classic statistical approach for calibrating model parameters, estimating model output errors, and thus producing prediction intervals (Wooldridge, 2015). LS is generally adopted as the basic technique against which new methods for uncertainty quantification are tested (Sorooshian and Dracup, 1980; Schoups and Vrugt, 2010; Renard et al., 2011; Honti et al., 2013; Del Giudice et al., 2016). The simplest application of LS is within a frequentist framework, in which model parameters are assumed to have one true yet unknown value. Consequently, model parameters are estimated by minimizing an objective function and neither prior nor posterior model parameter uncertainties are explicitly considered. Because model residuals in hydrology are typically heteroskedastic and non-normal (Wang et al., 2012; Del Giudice et al., 2013), here we apply LS after having transformed the observed $\mathbf{y}_o$ and modeled $\mathbf{y}$ output using a non-linear monotonic function g (see Supporting Material). The objective function used for calibration is the sum of the squares of the errors:

$$SSE = \sum^{n} (\tilde{\mathbf{y}}_o - \tilde{\mathbf{y}})^2 \quad (1)$$

where tilde represents the transformed output and n is the number of data points in the calibration dataset, i.e. the length of $\mathbf{y}_o$, a vector possibly including multiple outputs. Numerically, we use an adaptive Markov chain Monte Carlo algorithm (as in Del Giudice