Ecological Informatics 24 (2014) 107–111

Contents lists available at ScienceDirect

Ecological Informatics

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

# A suggestion for computing objective function in model calibration

Yiping Wu [a,*], Shuguang Liu [b,*]

[a] ASRC Federal, contractor to the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD 57198, United States
[b] U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center, Sioux Falls, SD 57198, United States

ARTICLE INFO

*Article history:*
Received 1 November 2013
Received in revised form 4 August 2014
Accepted 6 August 2014
Available online 13 August 2014

*Keywords:*
Biogeochemical modeling
Hydrological modeling
Model calibration
Objective function
Parameter optimization

ABSTRACT

A parameter-optimization process (model calibration) is usually required for numerical model applications, which involves the use of an objective function to determine the model cost (model-data errors). The sum of square errors (SSR) has been widely adopted as the objective function in various optimization procedures. However, 'square error' calculation was found to be more sensitive to extreme or high values. Thus, we proposed that the sum of absolute errors (SAR) may be a better option than SSR for model calibration. To test this hypothesis, we used two case studies—a hydrological model calibration and a biogeochemical model calibration—to investigate the behavior of a group of potential objective functions: SSR, SAR, sum of squared relative deviation (SSRD), and sum of absolute relative deviation (SARD). Mathematical evaluation of model performance demonstrates that 'absolute error' (SAR and SARD) are superior to 'square error' (SSR and SSRD) in calculating objective function for model calibration, and SAR behaved the best (with the least error and highest efficiency). This study suggests that SSR might be overly used in real applications, and SAR may be a reasonable choice in common optimization implementations without emphasizing either high or low values (e.g., modeling for supporting resources management).

Published by Elsevier B.V.

## 1. Introduction

Numerical models have been widely used in environmental science for understanding the natural processes, predicting impacts of global changes, and decision making for the sustainable management of resources. As knowledge of physical processes grows, models become more sophisticated and more parameters may be introduced (Beck, 1999; Brun et al., 2001; Legates and McCabe, 1999). We can see examples of the continuous developments of process-based models such as Soil and Water Assessment Tool (SWAT) (Arnold et al., 2012; Arnold et al., 1998) in hydrology and Erosion Deposition Carbon Model (EDCM) (Liu et al., 2003), a modified version of CENTURY (Parton et al., 1994), in ecology. These mathematical models include some parameters that need to be calibrated through an optimization procedure, which is to sample the parameter values from the allowable ranges until the value of the objective function (i.e., a function of differences between observations and simulations) is minimized or maximized (Diskin and Simon, 1977; Legates and McCabe, 1999; Nash and Sutcliffe, 1970).

From a literature review, a number of objective functions were used for model calibration in hydrology such as mean squared error, absolute mean/maximum error, residual bias, and Nash objective function (Boyle et al., 2000; Diskin and Simon, 1977; Gupta et al., 1998; Servat and Dezetter, 1991; Yapo et al., 1998). However, the sum of square errors (SSR) is the most commonly used objective function for a variety of optimization processes even in recent years (Confesor and Whittaker, 2007; Diskin and Simon, 1977; Gupta et al., 1998; Van Liew et al., 2005; Zhang et al., 2009). We also observe use of SSR in some popular optimization procedures such as the SWAT Auto-calibration Tool (Green and van Griensven, 2008; van Griensven, 2006; van Griensven et al., 2006), SWAT Calibration and Uncertainty Program (SWAT-CUP) (Abbaspour, 2012), the Flexible Model Environment (FME) R package (Soetaert and Petzoldt, 2010), and reservoir operation optimizations (Jothiprakash and Shanthi, 2006; Momtahen and Dariane, 2007; Raman and Chandramouli, 1996; Reddy and Kumar, 2006).

In evaluating model performances mathematically, studies have illustrated that the correlation-based measures characterized by 'square error' such as square correlation coefficient ($r^2$) and Nash–Sutcliffe Efficiency (NSE) are oversensitive to extreme values (outliers) and insensitive to additive and proportional differences between observations and simulations (Legates and Davis, 1997; Legates and McCabe, 1999; Moore, 1991). Legates and McCabe (1999) proposed a modified NSE (mNSE), which uses the 'absolute error' to replace the 'square error' in the original NSE calculation to evaluate the goodness-of-fit of hydrological models. For a similar purpose (avoiding oversensitivity to extreme values), Krause et al. (2005) revised the NSE based on relative deviations (i.e., replacing the 'square error' by 'square relative deviation'). Using multiple examples, they concluded that both mNSE and rNSE can suppress the oversensitivity to peak values, and the latter is more sensitive to the low values (Krause et al., 2005).

As stimulated by the above findings, we can infer that the widely-used objective function, SSR, also emphasizes the extreme values of a

* Corresponding authors.
*E-mail addresses:* ywu@usgs.gov, rocky.ypwu@gmail.com (Y. Wu), sliu@usgs.gov (S. Liu).

http://dx.doi.org/10.1016/j.ecoinf.2014.08.002
1574-9541/Published by Elsevier B.V.

set of observation data and neglects the low values during model calibration because squaring calculation usually means a relatively larger weight for peak or higher values. Thus, we hypothesize that using the sum of absolute errors (SAR) and the sum of absolute relative deviations (SARD) may be better than SSR and the sum of square relative deviations (SSRD), respectively, for an environmental model calibration without emphasizing either high or low values. The objective of this study is to test this hypothesis by implementing model calibrations using these four different objective functions and evaluating the corresponding model performances. For this purpose, we used two large complex models in different disciplines: the widely-used hydrological model —SWAT—with a case study of monthly streamflow calibration in a headwater area of the East River Basin in South China and the well-established biogeochemical model—EDCM—with a case study of monthly gross primary production (GPP) calibration at a forest flux tower site in the eastern United States.

## 2. Materials and methods

### 2.1. A hydrological model and a biogeochemical model

The hydrological model, SWAT was developed by the U.S. Department of Agriculture (USDA) Agricultural Research Service for exploring the effects of climate and land cover changes on water, sediment, and agricultural chemical yields (Arnold et al., 1998). This physically-based, watershed-scale, continuous model can simulate the hydrological cycle, cycles of plant growth, the transportation of sediment, and agricultural chemical yields on a daily time step (Arnold et al., 1998, 2012; Neitsch et al., 2005). The latest version, SWAT2012, was used in the current study.

The biogeochemical model, EDCM (Liu et al., 2003), is a modified version of CENTURY (version IV) (Parton et al., 1994). EDCM uses up to 10 soil layers to simulate the soil organic carbon (SOC) dynamics in the whole soil profile instead of the one single top-layer structure of CENTURY. EDCM can dynamically keep track of the evolution of the soil profile and carbon storage as influenced by both soil erosion and deposition (Liu et al., 2003). This process-based biogeochemical model is used to simulate carbon and nitrogen cycles in diverse ecosystems at a monthly time step (Liu et al., 2003; Tan et al., 2009). In particular, was used to evaluate carbon dynamics across the entire conterminous United States (Liu et al., 2014; Liu et al., 2012b; Zhu, 2011; Zhu et al., 2010).

### 2.2. Modification of the modeling frameworks

To implement the model calibration procedure for SWAT and EDCM, we used the developed R-SWAT-FME (Wu and Liu, 2012, 2014) and R-EDCM-FME (or EDCM-Auto) (Liu et al., 2012a; Wu et al., 2014), respectively. The two frameworks were developed to provide a variety of functionalities (e.g., parameter identifiability, optimization, and sensitivity and uncertainty analysis) for the two models (SWAT and EDCM), respectively. For the function of parameter optimization, the pseudo-random search algorithm (PseudoOptim) of Price (Price, 1977; Soetaert and Herman, 2009) included in FME was used in the current study, which was successfully tested for SWAT and EDCM calibrations. Because the original FME package uses SSR only as the objective function to compute model cost (Soetaert and Petzoldt, 2010), we modified the related function (modCost) to introduce the other three objective functions (i.e., SAR, SSRD, and SARD) we proposed as alternatives. The corresponding mathematical expressions of the four objective functions are listed in Table A.1 in Appendix **A**.

### 2.3. Scenarios for comparing objective functions

For comparing the four objective functions, we set four scenarios with one objective function being assigned for each scenario while holding the others constant (such as optimization algorithm, input data, and calibration time period) during model calibration. This kind of scenario setting was the same for both hydrological modeling with SWAT and biogeochemical modeling with EDCM.

### 2.4. Criteria to assess model performance

To assess different objective functions for model calibration, it is important to select a uniform and widely-accepted set of evaluation criteria. Because of the drawbacks (e.g., oversensitivity to extreme values) of correlation-based measures (e.g., NSE and $r^2$) (see Introduction), the use of mNSE and mean absolute error (MAE) terms for overall assessment was recommended (Krause et al., 2005; Legates and McCabe, 1999). In this study, we adopted these two terms as the primary criteria to evaluate the model performances, although the other commonly-used terms NSE, $r^2$, and RMSE are also reported for reference. The mathematical expressions of these five terms can be found in Table A.1 in Appendix **A**.

## 3. Case studies

We used two case studies to illustrate the performances of objective functions during model calibrations on monthly streamflow and Gross Primary Production (GPP)—the two primary variables in hydrology and ecology, respectively.

### 3.1. Study area and model setup for hydrological modeling

To drive SWAT for the hydrological modeling, we used the headwater area of the East River Basin (i.e., the Xunwu River) in South China as the case study, focusing on streamflow calibration, a common concern in hydrology. The Lizhangfeng flow gaging station has a drainage area of 1400 $km^2$, and average annual precipitation is about 1648 mm in this area. The sources of input data (e.g., climate, topography, soil, and land use) are the same with what we used in previous studies where details can be found (Chen and Wu, 2012; Wu and Chen, 2013). In the current study, the SWAT setup with discretization resulted in the delineation of 11 subbasins and 67 Hydrological Response Units (HRUs) for the specific area. The calibration procedure was conducted using R-SWAT-FME with 5 years (1977–1981) of observed monthly streamflow at Lizhangfeng, and six streamflow-related parameters were selected in this study (see Table 1).

### 3.2. Study site and model setup for biogeochemical modeling

For biogeochemical modeling with EDCM, we used a forest flux tower site—the Harvard Forest Environmental Monitoring Site, near Petersham, Massachusetts, in the United States (Curtis et al., 2002; Goulden et al., 1996)—with a focus on calibrating GPP, an elementary term in the carbon cycle. Soil texture data from the Ameriflux website indicate a soil composition of 66% sand, 29% silt, and 5% clay, with a bulk density of 0.9 $g/cm^3$. Monthly precipitation and air temperature data were from the Parameter-elevation Regressions on Independent Slopes Model (PRISM) Climate Group (2012). Other model input data (e.g., soil organic carbon) were from the national data layers for the conterminous United States of the Land Carbon project (Zhu and Reed, 2012, in press). The derived GPP data obtained from the Ameriflux website were used as the observations during the 5-year (2001–2005) model calibration using R-EDCM-FME, and four parameters were involved in this procedure as listed in Table 1.

## 4. Results

Using the R-SWAT-FME framework, we derived a set of optimal parameter values for each objective function in streamflow calibration (with the first case study of hydrological modeling). As listed in Table 1, the four sets of parameter values are quite different under the