Environmental Modelling & Software 90 (2017) 157–166

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Modelling errors calculation adapted to rainfall – Runoff model user expectations and discharge data uncertainties

Audrey Douinot*, Hélène Roux, Denis Dartus

Institut de Mécanique des Fluides de Toulouse (IMFT), Université de Toulouse, CNRS-INPT-UPS, Toulouse France

ARTICLE INFO

Article history:
Received 25 July 2016
Accepted 12 January 2017

Keywords:
Uncertainty propagation
Hydrological modelling
Calibration
Evaluation
Discharge uncertainty
Error isolines

ABSTRACT

A novel objective function for rainfall-runoff model calibration, named Discharge Envelop Catching (DEC), is proposed. DEC meets the objectives of: i) taking into account uncertainty of discharge observations, ii) enabling the end-user to define an acceptable uncertainty, that best fits his needs, for each part of the simulated hydrograph. A calibration methodology based on DEC is demonstrated on MARINE, an existing hydrological model dedicated to flash floods. Calibration results of state-of-the-art objective functions are benchmarked against the proposed objective function. The demonstration highlights the usefulness of the DEC objective function in identifying the strengths and weaknesses of a model in reproducing hydrological processes. These results emphasize the added value of considering uncertainty of discharge observations during calibration and of refining the measure of model error according to the objectives of the hydrological model.

1. Introduction

An objective function converts the outputs of a rainfall-runoff model into a single likelihood measure, according to discharge measurements. This likelihood measure plays a key role, as it controls the model assessment and calibration. As such it provides a comparison basis for models or scenarios. An objective function must provide a meaningful criterion, representative of the errors occurring in the prediction time series. Ideally the objective function must make a distinction between the observed errors coming from data uncertainties and the modelling errors coming from model limitations and parameter uncertainties. Defining such a metric is hard, as model outputs obviously depend on the input data and the observed discharge quality.

The uncertainty of the forcing data (rainfall/snowfall, soil moisture, etc.) is in general not measurable (Villarini and Krajewski (2010); Kirstetter et al. (2010)) whereas discharge uncertainties can be accurately quantified (McMillan and Westerberg, 2015; Coxon et al. (2015); Le Coz et al. (2014)). This makes it possible to integrate uncertainty of the discharge observationst into an objective function. However the classical functions, such as the Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970), or the Kling-Gupta-Efficiency (KGE, Gupta et al., 2009), are based on the difference between the model outputs and the observed discharge, without considering the discharge uncertainty. This can result in the overfitting of a model prediction to uncertain discharge observations.

Some modifications in different calibration approaches are found in the literature in order to integrate uncertainty of the discharge observations. Croke (2007) modified the NSE by weighting the residual vector according to the accuracy of observed discharge measurement. The metric thus emphasizes the prediction of a well known observed discharge at the expense of the observed discharge with high uncertainty. This is especially problematic in the context of flood modelling, where extreme flood discharges are generally marred with high uncertainty. Calibration methods based on Bayesian approach (Kuczera (1983); Engeland and Gottschalk, 2002, Kavetski et al., 2006), formalize an error model, considering among others discharge uncertainty. Formalizations of different type of errors, such as input uncertainty or model uncertainty are based on strong assumptions that require validation, which is not always possible. In the end, the calibration results depend on the definition of the error model. Liu et al. (2009) proposed a calibration method using a "limits-of-acceptability" approach. A parametrization is either accepted or rejected. The limit of acceptability is fixed according to discharge uncertainty. The method is convenient to assess the likelihood of a parameter set for a model, but it does not provide information on the

* Corresponding author.
E-mail address: audrey.douinot@imft.fr (A. Douinot).

http://dx.doi.org/10.1016/j.envsoft.2017.01.007
1364-8152/

relevance of the model.

The aim of the paper is to provide an objective function: i) taking into account uncertainty of the discharge observations; ii) adapting the calibration to user expectations and model assumptions; iii) providing a meaningful score which can be interpreted to assess the relevance of the model.

Section 2 presents the rationale of the paper. It discusses the state of the art of objective functions in the field of hydrologic models, with a focus on the model calibration issue. The proposed objective function, called Discharge Envelop Catching efficiency, is defined in Section 3 and evaluated against three other objectives functions in Section 4. Finally, calibration results are presented and discussed in Section 5.

2. Background and motivation

We begin the section introducing the mathematical concepts used throughout the paper.

2.1. Mathematical notation and symbols

We adopt the fomulation of Vrugt and Sadegh (2013) of model calibration and evaluation issues: "Consider a discrete vector of measurements $\hat{Y} = \{\hat{y}_1, \ldots, \hat{y}_n\}$, observed at times $t = \{1, \ldots, n\}$ that summarizes the response of an environmental system F to forcing variables $\hat{U} = \{\hat{u}_1, \ldots, \hat{u}_n\}$. Let $Y = \{y_1, \ldots, y_n\}$ the corresponding predictions from a dynamic (non linear) model f, with parameter values θ,

$$Y(\theta) = f\left(x_0, \theta, \widehat{U}\right) \tag{1}$$

where $x0$ is the initial state of the system at $t = 0$." The residual vector defines the difference between actual and model-simulated system behaviours:

$$E(\theta) = \widehat{Y} - Y(\theta) = \{e_1(\theta), \ldots, e_n(\theta)\} \tag{2}$$

The error model F that allows for residual vector transformation defines the modelling error vector:

$$\varepsilon(\theta) = F\left[\widehat{Y} - Y(\theta)\right] = \{\varepsilon_1(\theta), \ldots, \varepsilon_n(\theta)\} \tag{3}$$

A function G is used to map the modelling error vector into a metric called likelihood measure. The combination of F and G is the objective function.

Calibration aims to find the values of [$\theta \in \Theta \in \Re^d$] that provide the best likelihood measure. As the optimal parameter set may not be unique and several candidates may minimized equally the objective function, the calibration process faces model equifinality (Beven and Binley, 1992; Beven, 2006). Choosing a way of selecting or weighting behavioural parameter sets according to likelihood measure corresponds to the last step of a calibration methodology.

We now consider the fact that forcing variables $\hat{U}$, initial state x_0 and observed discharges $\hat{Y}$ are uncertain measurements and denote $\sigma_{\hat{U}}$, σ_{X0}, $\sigma_{\hat{Y}}$ the vectors quantifying those uncertainties. Forcing variables and initial state uncertainties affect model predictions and modify equation (1):

$$Y'(\theta) = f\left(x_0 \middle| \sigma_{x_0}, \theta, \widehat{U} \middle| \sigma_{\widehat{U}}\right) \tag{4}$$

where $Y'(\theta)$ is the model prediction with respect to input uncertainties. Similarly, the observed discharge uncertainties modify equation (3):

$$\varepsilon(\theta) = F\left[\widehat{Y} \middle| \sigma_{\widehat{Y}} - Y(\theta)\right] = \{\varepsilon_1(\theta), \ldots, \varepsilon_n(\theta)\} \tag{5}$$

This paper focuses on equation (5) and proposes an error model F that allows for benchmarking a model prediction vector $Y(\theta)$ against uncertain observations ($\hat{Y}$, $\sigma_{\hat{Y}}$). The choice of the optimal function G which maps the modelling error vector into a metric is also discussed.

2.2. Adapting the likelihood measure to the model

As said before, the primary goal of calibration is finding parameter sets that best mimic the observed discharge. The role of the objective function is to define the most appropriate likelihood measure to accurately assess the success of the model to reproduce the hydrological behavior of a catchment system.

In the literature, performance models are usually assessed using statistic scores such as linear correlation, mean, variance or indexes widespread in the hydrology community such as NSE, RMSE or Kling-Gupta-Efficiency (KGE, Gupta et al. (2009)). The use of those scores as conventional likelihood measures is supposed to facilitate model comparison. However, as pointed out by Seibert (2001) or Schaefli and Gupta (2007), a score may reflect poorly the goodness-of-fit of a model, even when established by hydrologists. As an example, a NSE score of 0,6 could equally mean good or poor fit depending on data quality and on the studied catchment. Moussa (2010) and Schaefli et al. (2005) also highlighted the limitations of the NSE for flood event modelling assessment, showing that considering the high value of standard deviation of discharge time series, the residuals might be high and still lead to a good score, due to the NSE definition.

Schaefli and Gupta (2007) suggested to take into account model assumptions and user expectations into the objective function. They defined the benchmark efficiency (BE):

$$BE = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(\hat{y}_i - y_{bi}\right)^2} \tag{6}$$

where y_{bi} is called the benchmark discharge model at time i. The model reference is no more the observed discharge mean as in NSE, but a benchmark model defined as admissible by the hydrologist. The BE definition implies a meaningful score according to what is expected from the model.

All the objective functions seen so far choose to minimize the sum of squared residuals as the calibration objective. As noticed by Beven and Binley (2014), this is not without implication. The combination of all residuals within a single value actually hides the underlying assumption that this score represents at best all the residuals. Assuming that the sum of squared residual is the best representation has two important implications:

- the same importance is attached to all residual values, whatever their position along the hydrograph. Yet, absolute errors during high flows or low flows may not be interpreted the same by hydrologist. This issue could be avoided by weighting residual vector as in mNSE (Croke, 2007) or calculating the sum of squared **relative** errors;
- among the residual distribution, the mean represents the best index to minimize. As residuals are most commonly correlated, heteroscedastic and have non-Gaussian distributions (Schoups and Vrugt (2010)), the relevance of this choice is not certain. Moreover, the mean of the residual distribution is mainly