Journal of Hydrology 528 (2015) 29–37

Contents lists available at ScienceDirect

Journal of Hydrology

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

Technical Note

Modeling residual hydrologic errors with Bayesian inference

Tyler Smith [a,*], Lucy Marshall [b], Ashish Sharma [b]

[a] Department of Civil and Environmental Engineering, Clarkson University, Potsdam, NY 13699, USA
[b] School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia

ARTICLE INFO

Article history:
Received 1 July 2014
Accepted 25 May 2015
Available online 7 June 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Ezio Todini, Associate Editor

Keywords:
Bayesian inference
Residual error
Formal likelihood function
Rainfall–runoff modeling

SUMMARY

Hydrologic modelers are confronted with the challenge of producing estimates of the uncertainty associated with model predictions across an array of catchments and hydrologic flow regimes. Formal Bayesian approaches are commonly employed for parameter calibration and uncertainty analysis, but are often criticized for making strong assumptions about the nature of model residuals via the likelihood function that may not be well satisfied (or even checked). This technical note outlines a residual error model (likelihood function) specification framework that aims to provide guidance for the application of more appropriate residual error models through a nested approach that is both flexible and extendible. The framework synthesizes many previously employed residual error models and has been applied to four synthetic datasets (of differing error structure) and a real dataset from the Black River catchment in Queensland, Australia. Each residual error model was investigated and assessed under a top-down approach focused on its ability to properly characterize the errors. The results of these test applications indicate that a multifaceted assessment strategy is necessary to determine the adequacy of an individual likelihood function.

1. Introduction

The use of Bayesian inferential approaches is typically predicated on the formal likelihood function adequately characterizing the form of the true model errors (Mantovan and Todini, 2006). In hydrology this can be a difficult task, and has led to criticism of the appropriateness of formal Bayesian methods in "real data" cases (Beven et al., 2007). Regardless of this, Bayesian methods have become a common part of the hydrologic modeling literature (e.g., Kuczera and Parent, 1998; Marshall et al., 2004; Smith and Marshall, 2008). Box and Tiao (1973) provide some insight into the development and application of formal likelihood functions within a Bayesian inferential approach, highlighting the importance of iterative improvement, stating "to be on firm ground we must do more than merely postulate a model; we must build and test a tentative model at each stage of investigation" (p. 7).

However, general guidance on how to appropriately construct a residual error model to be used in hydrologic modeling studies has received limited attention (e.g., Kuczera, 1983) beyond application-specific problems. Even application-specific examples of appropriate likelihood function selection are relatively rare (e.g., Bates and Campbell, 2001; Kuczera, 1983; Schaefli et al., 2007; Schoups and Vrugt, 2010; Smith et al., 2010; Sorooshian and Dracup, 1980); especially when compared to the widespread adoption of simple normality assumptions (e.g., Ajami et al., 2007; Campbell et al., 1999; Duan et al., 2007; Hsu et al., 2009; Samanta et al., 2008; Vrugt et al., 2006). Xu (2001, p. 77) points out that "in the field of hydrological modeling, few writers examine and describe any properties of residuals given by their models when fitted to the data". Recently there have been a few notable exceptions to this practice (e.g., Engeland et al., 2010; Evin et al., 2013, 2014; Reichert and Mieleitner, 2009; Schoups and Vrugt, 2010; Smith et al., 2010; Thyer et al., 2009), suggesting it is an opportune time to review and synthesize likelihood selection practices.

In this technical note, we aim to introduce an instructive (rather than prescriptive) framework to direct the selection of a formal likelihood function to be used within a Bayesian context for modeling residual errors in hydrologic modeling applications. The significance and contribution of this work is in the formalization and illustration of the process for development appropriate residual error models in conceptual hydrological modeling, rather than in the idea of adding complexity to likelihood functions (when warranted). The residual error models employed here are meant to represent neither the full range of nor necessarily the best candidates; rather they comprise many of the most often used forms.

* Corresponding author at: Box 5710, 8 Clarkson Avenue, Potsdam, NY 13699, USA. Tel.: +1 (315) 268 2243.
E-mail address: tsmith@clarkson.edu (T. Smith).

http://dx.doi.org/10.1016/j.jhydrol.2015.05.051
0022-1694/

2. Selection of a residual error model

Conceptual hydrologic models can be thought of, in a generic sense, as being of the form

$$Q_t = f(x_t; \theta) + \varepsilon_t \quad (1)$$

where t indexes time, Q_t is the observed discharge, $f(x_t;\theta)$ is the model simulated discharge, x_t is the model forcing data (typically, rainfall and evapotranspiration), θ is the set of unknown model parameters, and ε_t is the error. Proper understanding of the form of the errors (ε_t) is important for using the model for predictions and vital to the success of the inference. Under the Bayesian framework, the errors are modeled via the formal likelihood function.

Debate over the selection and appropriate use of formal likelihood functions has dominated recent hydrologic modeling literature (e.g., Beven et al., 2007, 2008; Mantovan and Todini, 2006; Mantovan et al., 2007; Stedinger et al., 2008). Given the debate over formal likelihood functions, a set of instructive guidelines for examining likelihood function appropriateness are needed to improve the application of Bayesian methods to hydrologic modeling problems and address on of the concerns often raised by opponents of the formal Bayesian approach. Mirroring a top-down hydrologic model development approach (Sivapalan et al., 2003), the residual error model framework presented here advocates adding complexity only when merited by analysis of the residuals (refer to Box and Tiao, 1973).

The residual error model selection framework, introduced in Fig. 1, synthesizes many previously employed approaches aimed at producing more appropriate formal likelihood functions; innovations that account for error characteristics such as autocorrelation (e.g., Sorooshian and Dracup, 1980), non-constant variance (e.g., Bates and Campbell, 2001), and zero-inflation caused by extended periods of no observed streamflow (Smith et al., 2010). Heteroscedasticity is addressed via data transformation using the Box–Cox family of transformations (Box and Cox, 1964)

$$y^T = \begin{cases} \frac{(y+\lambda_2)^{\lambda_1}-1}{\lambda_1} & (\lambda_1 \neq 0; y > -\lambda_2) \quad \text{(a)} \\ \log(y+\lambda_2) & (\lambda_1 = 0; y > -\lambda_2) \quad \text{(b)} \end{cases} \quad (2)$$

where y^T is the transformed data, y is the untransformed data, and λ_1 and λ_2 are the Box–Cox parameters with the constraints as shown. Autocorrelation is accounted for with an autoregressive model

$$y_t = \sum_{j=1}^{p} \phi_j y_{t-j} \quad (3)$$

where y is the data with time index t, p is the order of the autoregressive process, and ϕ are the autoregressive coefficients. Zero-inflation is considered by using the mixture likelihood approach (see Table 1) of Smith et al. (2010), where the likelihood function is represented as the product of three components: (1) zero observations modeled with zero error (n_1), (2) zero observations modeled with nonzero error (n_2), and (3) nonzero

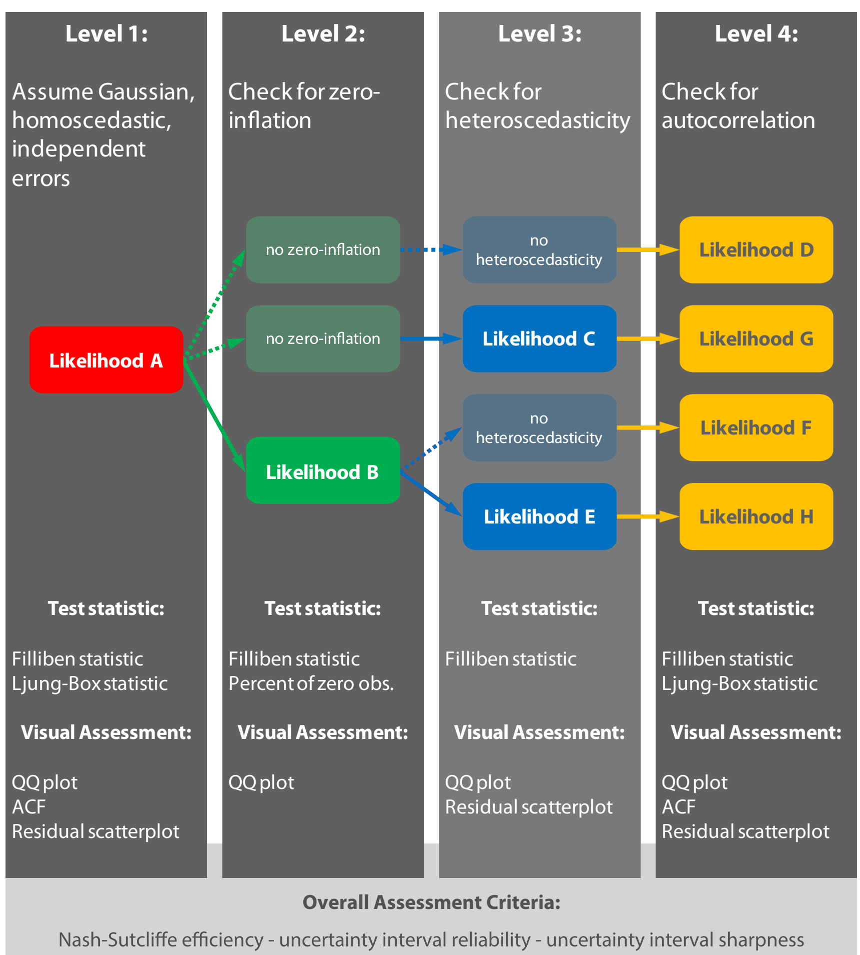

Fig. 1. The likelihood function selection logic. A basic assumption of the form of the residuals is made and followed by checks on typical areas of deficiency at each level of the logic.