Environmental Modelling & Software 51 (2014) 190–194

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

Short communication

A new approach to visualizing time-varying sensitivity indices for environmental model diagnostics across evaluation time-scales

Carolina Massmann [a,*], Thorsten Wagener [b], Hubert Holzmann [a]

[a] Institute of Water Management, Hydrology and Hydraulic Engineering, University of Natural Resources and Life Sciences, Vienna, Austria
[b] Department of Civil Engineering, Queen's School of Engineering, University of Bristol, Bristol, UK

ARTICLE INFO

Article history:
Received 3 July 2013
Received in revised form
27 September 2013
Accepted 30 September 2013
Available online 29 October 2013

Keywords:
Global sensitivity analysis
FAST
Sensitivity indices visualization

ABSTRACT

Assessing the time-varying sensitivity of environmental models has become a common approach to understand both the value of different data periods for estimating specific parameters, and as part of a diagnostic analysis of the model structure itself (i.e. whether dominant processes are emerging in the model at the right times and over the appropriate time periods). It is not straightforward to visualize these results though, given that the window size over which the time-varying sensitivity is best integrated generally varies for different parameters. In this short communication we present a new approach to visualizing such time-varying sensitivity across time scales of integration. As a case study, we estimate first order sensitivity indices with the FAST (Fourier Amplitude Sensitivity Test) method for a typical conceptual rainfall–runoff model. The resulting plots can guide data selection for model calibration, support diagnostic model evaluation and help to define the timing and length of spot gauging campaigns in places where long-term calibration data are not yet available.

1. Introduction

A wide range of studies has estimated the time-varying sensitivity of hydrologic and other environmental models over the last decade (Cloke et al., 2008; Guse et al., 2013; Herman et al., 2013; Reusser and Zehe, 2011; Reusser et al., 2011; Sieber and Uhlenbrook, 2005; Wagener et al., 2001, 2002, 2003; Wagner and Harvey, 1997; Wlostowski et al., 2013). These studies vary in their approaches to sensitivity calculation (e.g. local versus global strategies), and in their study goals (e.g. support model calibration or diagnose model structures). If the goal is model calibration, then understanding which periods of a time series are most helpful in identifying a particular parameter or a specific group of parameters is an important part of the calibration process. Often, some of the model parameters will represent processes that only matter during specific time periods, i.e. specific modes of the system, for example recession constants or parameters controlling the extent of saturated areas in a catchment during a flood event. Such parameters are only likely to be identifiable if these periods can be isolated, or if they sufficiently impact a global objective function (i.e. one that aggregates residuals over the full time series). It is therefore often observed that parameters which are important during low flow periods, when errors are generally small, or parameters which are only important for a very short time, are not easily identifiable. A second goal when applying time-varying sensitivity analysis can be the diagnostic evaluation of the realism of a model structure (Wagener, 2003). Each model parameter (or a set of parameters) represents a specific process assumed to reflect the real world system under study. A modeller will have an expectation under which conditions a specific process should matter, and hence when the relevant parameter should impact or even control the model output. This expectation can be tested by checking whether the parameters are sensitive at the right times.

Sensitivity analysis methods used for time-varying analysis include local approaches and global approaches, and those explicitly considering interactions between parameters and those that do not. Regardless of the method applied, they are generally used to estimate sensitivity at each time step or for a running window, i.e. by calculating the value of the variable of interest for a fixed number of time steps, assigning this value to the central (or another) point of this subset and then doing the same for the consecutive time steps. The advantage of using a running window is that the estimates are less affected by outliers in the data (e.g. noisy measurements). It is further recommended to calculate the sensitivities for the middle of the window, since not doing this can make the sensitivity appear before or after the actual periods in

* Corresponding author. Fax: +43 1 47654 5549.
E-mail address: carolina.massmann@boku.ac.at (C. Massmann).

1364-8152/$ – see front matter
http://dx.doi.org/10.1016/j.envsoft.2013.09.033

which the model parameters are sensitive. Also, since different processes act over different time-scales an analysis across time scales is needed to ensure that no information is lost (Wagener et al., 2003). We therefore generally have to produce multiple time-series of sensitivity (for different window sizes) for each parameter. The visualization of these results is however not straightforward, since simple line plots would produce complex figures that are hard to interpret or we would have to produce a large number of plots if shown separately (e.g. as pattern plots). Because visualizing research data and results in an accessible manner is an important step for any analysis and can obscure or strengthen our ability to communicate study outcomes (Kelleher and Wagener, 2011), we propose an improvement in the visualization of sensitivity indices across window sizes by adapting wavelet analysis plots for our purpose.

This brief communication shows how sensitivity indices for different temporal evaluation periods can be plotted together. This is exemplified with the FAST (Fourier Amplitude Sensitivity Test) method, which has been used for some hydrological applications focussing however only on one time scale (Guse et al., 2013; Reusser and Zehe, 2011; Reusser et al., 2011). The study is carried out for a small Austrian catchment modelled with a conceptual rainfall–runoff model considering two objective functions. We regularly refer to other papers for further discussion so that we can comply with the required brevity of this short communication.

2. Methodology

2.1. Catchment description

Rosalia is a small catchment of 2.35 km^2 located in lower Austria. It is covered mainly by forests and ranges between 400 and 725 m.a.s.l. in elevation. The estimated average annual precipitation equals 780 mm and the average monthly depth of runoff is 14 mm. The model was set up for the time between January 1989 and July 1993, but only the period after July 1990 was considered in the analysis while the first year served as warm-up period.

2.2. Hydrological model

The conceptual rainfall–runoff model used (Fig. 1) is a modified version of the model introduced by Holzmann and Nachtnebel (2002). Its main processes are briefly discussed here. Snow accumulation and modelling depend on the hypsometric lapse rate (*hypgrad*), which is used for adjusting the temperature according to elevation. Snow melt is modelled with a day-degree factor (*fak*). Hortonian flow is activated when the infiltration capacity (*inf_cap*) of the soil is smaller than the rainfall intensity. Water in the Hortonian flow storage is then routed to the stream as a function of the recession constant *akval*. All water infiltrating into the soil is first held by the root storage, which has a height specified by parameter *rstdepth*. From this storage the water is released to the soil storage via a spill mechanism. The size of the soil storage is defined by parameter *adepms* and its three outlets represent saturation flow, interflow and percolation, which are described by the recession constants *k1*, *k2*, and *k3* respectively. Percolating water is retained by the groundwater storage from which baseflow is released as a function of the recession constant (*k4*). The potential evapotranspiration is estimated with the Thornthwaite method and the actual evapotranspiration (*etp*) is a function of the potential evapotranspiration and the water content in the root and soil storages.

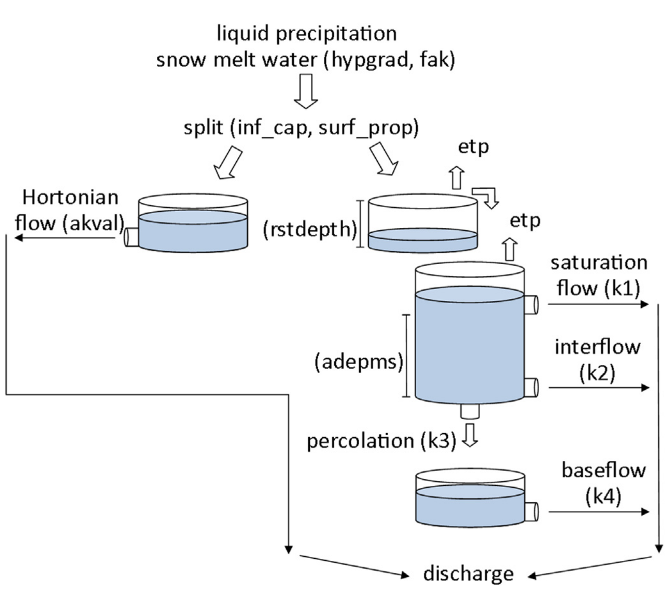

Fig. 1. Scheme of the hydrological model used, with the considered parameters shown in parentheses to aid interpretation of the sensitivity analysis plots.

2.3. FAST sensitivity analysis method

FAST is a global sensitivity method belonging to the category of variance decomposition approaches. These methods allocate the total variance of the model output to the different parameters (or other model input factors) and interactions between them:

$$Var = \sum_{i}^{n} Var_i + \sum_{i}^{n}\sum_{j=i+1}^{n} Var_{ij} + \sum_{i}^{n}\sum_{j=i+1}^{n}\sum_{k=j+1}^{n} Var_{ijk} + \ldots + Var_{12\ldots n} \quad (1)$$

where *Var* describes the total variance of the model output, the indices *i*, *j*, *k* stand for the considered parameters and *n* equals the total number of parameters. The terms Var_i and Var_{ij} represent the variance explained by the parameter *i* on its own and in combination with the parameter *j*, respectively. By adding all terms of Equation (1) in which the index *i* does not appear, we obtain $Var \sim i$ which represents the variance that is not explained by parameter *i*. The first order sensitivity (Eq. (2)) explains the proportion of the total variance due to the uncertainty in parameter *i* on its own:

$$S_i = \frac{Var_i}{Var} \quad (2)$$

The main characteristic of the FAST method is that the parameter space is sampled by oscillating all parameters so that they cover the whole input domain. The search function for obtaining the parameter sets with which the hydrological model is run was taken from Saltelli et al. (1999):

$$x_i(s) = \frac{1}{2} + \frac{1}{\pi}\arcsin(\sin(\omega_i s)) \quad (3)$$

where x_i refers to a vector with values for the parameter *i* and ω_i stands for the angular frequency with which this parameter oscillates. Each parameter has a different angular frequency selected according to specific rules for making sure that no frequency can be obtained from a linear combination of other considered frequencies using integer coefficients (Saltelli et al., 1999). The variable *s* is a vector with N_s elements forming an arithmetic progression and defining a “search curve”.

When the hydrological model is run with the oscillating parameter values obtained with Equation (3), there will be an oscillation of the model output which can be disaggregated into the impact of different frequencies (and their corresponding parameters) using a Fourier transformation (Cukier et al., 1978).

2.4. Implementation of the sensitivity analysis

The FAST method was implemented based on the descriptions provided by Saltelli et al. (1999) and the code published by Ekström (2005). The result of such a sensitivity analysis depend on the probability function assumed for the parameters and on the ranges between which these parameters are allowed to vary (e.g. a parameter which could have an important effect on the variance will only show a limited effect if the range of values over which it is allowed to vary is very small). Due to the lack of additional information, it was assumed that the parameters are uniformly distributed. With respect to the parameter ranges, the minimum and maximum parameter values for the sensitivity analysis were defined by adding and subtracting 15% of the optimum calibrated value, respectively. After confirming that a sample size of 20.000 yielded stable first order indices, the parameter sets were sampled according to Equation (3). The hydrological model was subsequently run with each of these parameter sets. Based on the simulated and modelled discharges, the values for two error metrics (objective functions) were calculated:

$$\text{Mean squared error } MSE_w\left(t_{(a+w/2)}\right) = \frac{1}{w}\sum_{i=a}^{a+w}\left[Q_o(t_i) - Q_s(t_i)\right]^2 \quad (4)$$

$$\text{Correlation coefficient } CC_w\left(t_{(a+w/2)}\right) = \frac{\frac{1}{w}\sum_{i=a}^{a+w}\left[\left(Q_o(t_i) - \overline{Q}_{o,(a+w/2)}\right) - \left(Q_s(t_i) - \overline{Q}_{s,(a+w/2)}\right)\right]^2}{\sigma_{o,(a+w/2)}\sigma_{s,(a+w/2)}} \quad (5)$$

with *w* standing for the window size, which took the values of 2, 3, 4, 8, 16, 30, 60, 120, 240, 360, 500, 700, 1000 and 1080 days. The subscripts *o* and *s* identify the observed and measured values of the discharge (*Q*) and the standard deviation (*σ*).