Hydrol. Earth Syst. Sci. Discuss., 3, S1804–S1809, 2007
www.hydrol-earth-syst-sci-discuss.net/3/S1804/2007/

Hydrology and Earth System Sciences Discussions

HESSD

3, S1804–S1809, 2007

Interactive Comment

Interactive comment on "Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation" *by* Y. Tang et al.

P. Reed

preed@engr.psu.edu

Received and published: 9 January 2007

For the sake of documenting our full review response, we have included our response to a detailed editorial review of our manuscript. Our final submitted letter to the editor will provide a detailed list of changes made in the revised manuscript beyond those detailed in our online author comments.

RESPONSE TO EDITOR REVIEW:

We appreciate your thorough review and management of our manuscript. Our responses to your editorial comments are given below.

Overall Editor Comment:

This is a well-written paper which reports a great deal of computational work and carefully-considered analysis, although its conclusions (that model parameter sensitivities vary with the time-interval used in modeling, and with the methods used to assess model sensitivity) are not really surprising.

Authors' Response: We also feel that this work contributes a comprehensive assessment of the 4 tested sensitivity methods. To clarify these conclusions and this contribution the following text was added to the abstract.

This study also contributes a comprehensive assessment of the repeatability, robustness, efficiency, and ease-of-implementation of the four sensitivity methods. Overall ANOVA and Sobol's method were shown to be superior to RSA and PEST. Relative to one another, ANOVA has reduced computational requirements and Sobol's method yielded more robust sensitivity rankings.

Specific Editor Comments

1. Regarding PEST, the authors say that "The optimization problem is solved by linearizing the relationship between a model's output and its parameters. The linearization is conducted using a Taylor series expansion Ě.". Where the structure of a hydrological model contains thresholds, however, as almost all such models do, the derivatives with respect to the parameter in any Taylor expansion will not change smoothly at any threshold. This seems (to me) to cast doubt on any attempt to derive a Taylor expansion and, therefore, to make PEST an unsatisfactory approach. It may also explain the authors' conclusion that results given by PEST were often different and "contradictory" (their word). They say, in their Section 7, that "PEST is more prone to misclassify sensitivitiesĚ since the method's derivatives are computed at a single pointĚ", but the lack of smoothness in derivatives at a threshold is the important issue, in my opinion. The strengths claimed for PEST by the authors ("computational efficiency, ease-of-implementation, ease-of-use") may not mean very much if the foun-

HESSD

3, S1804–S1809, 2007

Interactive Comment

dation of the method is questionable.

Authors' Response: We agree with the editor that PEST is limited in this study since the hydrological model lacks smoothness in its derivatives at a threshold. However, we did not intend to endorse PEST but to test it since it has become a popular tool in hydrologic applications. We give references in our manuscript which show that PEST is being applied both in groundwater and surface hydrology (See line 4 of the last paragraph in Section 2.1 on page 2). To emphasize the potential limits of PEST indicated by the editor, we have added a sentence at line 14 of the first paragraph in Section 7 on page 11 shown below.

Readers should be aware that the linearization of the relationship between model's output and its parameters will adversely impact PEST applications for hydrologic models with thresholds because of their impacts on the derivatives in the Taylor's series expansion.

2. A feature of several of the tables presented - in particular, Tables 4, 5, 6, 7 (particularly), 8 and 12 - show the sensitivities of the SCA-SMA model to its parameter ADIMP. It is not clear why this should be so, and I could find no discussion in the paper that casts light on it. Indeed, the paper contains almost no information about what ADIMP actually is; Table 2 defines it as "additional impervious area" but to what is it additional, and why does it have such influence? The authors' Figure 2, showing model structure diagrammatically, is no help, nor is their section 3.2.

Authors' Response: To clarify the major forms of runoff and the meanings of the parameters PCTIM and ADIMP in SAC-SMA, we have added text before the sentence " Prior work Ě" located on in the in section 3.2 on page 6. We have also added text at the end of section 3.2. The added text is shown below.

Sentences added before the sentence "Prior work": It is indicated in the figure that there are four principal forms of runoff generated by SAC-SMA: 1) direct runoff on the impervious area, 2) surface runoff when the upper zone free water storage is filled and

HESSD

3, S1804–S1809, 2007

Interactive Comment

the precipitation intensity is greater than percolation and interflow rate, 3) the lateral interflow from upper zone free water storage, and 4) primary baseflow. The direct runoff is composed of the impervious runoff over the permanent impervious area and the direct runoff on the temporal impervious area. The permanent impervious area, represented by parameter PCTIM (percent of impervious area), represents constant impervious areas such as pavements. The temporal impervious area, represented by parameter ADIMP (additional impervious area), includes the filling of small reservoirs, marshes, and temporal seepage outflow areas which become impervious when the upper zone tension water is filled.

Sentences added at the end of the paragraph: As shown in Table 2, the maximum allowable value of ADIMP specified by the author is 0.4 indicating that 40% of the watershed area is the temporal impervious area, which can lead to large direct runoff under wet conditions.

3. The use of techniques from the statistical methodology of experimental design (Latin hypercubes, fractional factorials, ANOVA) is interesting and attractive. One wonders whether other procedures from this area (such as response-surface designs) will also find application. The authors quote Box, Hunter and Hunter (1973) as evidence for restricting their ANOVA models to include only two-factor interactions, but the statement by Box et al. was really given for the kinds of industrial experimentation considered in their book; it does not necessarily follow (in my opinion) that it is also true for computer experiments to determine model sensitivities. It would be straightforward to check this by extending the authors' model in equation (2) to include higher-order interactions, and to test whether these are indeed small.

Authors' Response: We did not include higher-order interactions in our analysis because of two reasons: 1) computational constraints, and 2) The results of Sobol's method confirm our hypothesis that the first order and second order effects account for more than 90% of the total variance as shown in Table 10 making it reasonable to neglect higher order interactions.

4. There are a number of small (possible) errors in the paper. I will not list all of them; one of the more important appears in the legends to Tables 8, 11 and 12, which says that the values in brackets provide the 95% confidence interval for F-values, then quote: "(i.e., the mean the bracketed value yields the confidence interval)". Some words seem to be missing between "mean" and "the". And if this text implies that the confidence interval is symmetric about the mean value, how can this be if, as I understand it, the F-distributions were derived by bootstrap methods? Wouldn't you want to use the quantiles of the distribution of bootstrap F-values? The fact that the single-parameter sensitivities given in these tables show no trend with increasing time-interval also deserves some comment. Why, for example, in Table 8, does the sensitivity measure of the parameter LZTWM go from 1011 to 1724 and then down to 1149, as the time-step increases from 1 to 24 hours?

Authors' Response: The suggested edits related to the small errors in the legends in Tables 7, 8, 11, and 12 have been made. In Tables 7, 8, 11, and 12, the values outside the brackets are the calculated F-values or sensitivity indices. The value inside the bracket is half of the confidence interval, i.e., the value outside the bracket ś the value inside the bracket = the confidence interval. The confidence interval represents the uncertainty of the estimated F-value or the sensitivity index. We bootstrapped the empirical distribution of the statistical metric (F-value or sensitivity index) with respect to the resamples by using the moment method, which is uses large sample theory and assumes a symmetric 95% confidence interval. The moment method has the advantage in obtaining reliable estimates of standard error with smaller resample size relative to the percentile method. Bootstrapping was not used to derive the F-distribution. It provides the sampling distribution of a parameter's F-value's estimation. The F-values for different parameters are calculated using the ANOVA methodology. The resample size, 2000, we used in this study is based on the discussions of a prior study and our own computational experiments confirming that the distribution is reasonably symmetric with this sample size. To further clarify our bootstrapping results, we have added and modified sentences on text in Section 5.3. The modifications are shown below.

HESSD

3, S1804–S1809, 2007

Interactive Comment

Because of the uncertainties of data and the estimation errors, it makes more sense to look at the relative rankings of different parameters instead of the absolute values. It is difficult to distinguish the two values like 1011 and 1149. Additionally, please note that the Sobol indices show a similar trend for the specific parameters discussed by the editor, which is likely the result of a short-time scale threshold in the SNOW-17/SAC-SMA model.

Modified sentences in Section 5.3: The moment method (Archer et al., 1997) was adopted for acquiring the bootstrap confidence intervals (BCIs) for this paper. The moment method is based on large sample theory and requires a sufficiently large resampling dimension to yield symmetric 95% confidence intervals. In this study, the resample dimension N was set to 2,000 based on prior literature discussions as well as computational experiments that confirmed a symmetric distribution for standard errors. Readers interested in detailed descriptions of the bootstrapping method used in this paper can reference the following sources (Archer et el., 1997, Efron and Tibshirani, 1993).

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 3, 3333, 2006.