Hydrol. Earth Syst. Sci. Discuss., 12, C5964–C5970, 2016
www.hydrol-earth-syst-sci-discuss.net/12/C5964/2016/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "Parameter regionalization of a monthly water balance model for the conterminous United States" *by* A. R. Bock et al.

A. R. Bock et al.

abock@usgs.gov

Received and published: 5 January 2016

Author's Responses to referee #1:

Reviewer Main Points:

-[1] The parameter regionalisation procedure could be explained more effectively. In the first place, it would be good to have a schematic of the procedure to clearly see what is the role, inputs and outputs of each step (sensitivity analysis, classification of regions, individual calibration, grouped calibration, etc.).

AB: A schematic is a good idea. We will create an example for inclusion into the paper,

as well as locations within the text where references to the schematic would be helpful.

-The structure of Sec 2-3-4 could be revised to better separate out the methodology from the illustration of results. For instance, I find a bit odd that sensitivity analysis results are presented in Sec. 3.1, before describing how they will be used in the proposed methodology. Another example is the first paragraph of Sec. 4.1, which explains why the individual streamgage calibration is needed, it would fit better in a "methods" section rather than the "results"section.

AB: We built the main methodology presented in the paper based on the results from the sensitivity analysis. Because we wanted to stress the independence of the sensitivity analysis from the calibration and regionalization procedure, we pushed the sensitivity analysis results to 3.1. The way sensitivity estimates are used for regionalization (described on page 10034,line 7 onward) needs to be explained more clearly, especially since this is the most novel aspect of the proposed methodology. Specifically:

-What is the connection between the first and second classification? They are independent from each other and then intersected to obtain the actual classification? Please clarify

-Description of the second classification (lines 10-11) is also unclear. What are the "unique combinations of parameter sensitivities"? How are they defined? What is their meaning?

AB: Correct, the two classifications are independent. The first classification (p. 10034, lines 6-9) derives regions based on hydrologic response units (HRUs) with unique combinations of magnitudes of the five parameter sensitivities (highest to lowest). The intent is to identify geographic regions of similar model response or behavior based on the numerical orders of the sensitivities.

AB: The second classification (p.10034, lines 10-11) identifies regions based on HRUs with unique combinations of parameters with FAST-based parameter sensitivities that

exceed 5% (of the cumulative 100%). The intent with the second classification is to identify geographic regions with similar important parameters identified by the sensitivity analysis). The resulting polygons of these two classifications are merged to create the final region classification.

-Be more specific on how the two classification approaches work. Sentence on lines 7-8 of page 10034 is too generic, does it mean that the parameter ranking is the same in each region?

AB: Yes, all hydrologic response units (HRUs) identified within each region in classification 1 (p. 10034, lines 6-9) have the same ranking of the 5 model parameters from highest sensitivity to lowest sensitivity. Additionally, all HRUs within each region identified in classification 2 (p. 10034, lines 10-11) had the identical subset of parameters which exceeded 5% from FAST.

-From lines 17-20, I understand that the sensitivity-based classification is further refined using a more 'conventional' approach that looks at proximity and topographic divides. How does this refinement step works? Does it introduce significant changes in the classification? This is important to know in order to understand the value of the proposed sensitivity-based classification versus proximity or topography-based classification.

AB: This more conventional approach was necessary because of the lack of streamgages available for calibration in some of the calibration regions. The density of the stream gage network can be very sparse for some geographic regions of the U.S., especially in arid/semi-arid areas (see Kiang et al., 2013 in the discussion paper reference list). Following the "unsupervised" merging of the two classifications, the authors conducted a "supervised" classification of regions with less than 3 gages, where these regions where merged with geographically proximate regions with adequate streamgage representation that also shared the most similar parameter sensitivity results.

AB: The primary topographic divides utilized were topographically-derived boundaries

from the NHDPlus or Hydrologic Unit Code boundaries (black lines in Figure 2). These are sub-boundaries of both the model discretization and many USGS water resource management efforts. We felt it was important to maintain these boundaries, especially for the western United States where orographic climate effects are very important to the hydrologic cycle.

-[2] Some of the numerical results are a bit surprising and should be double-checked. In particular, in Fig. 5.a the fact that one parameter has sensitivity of exactly 100 and all others of exactly 0 seems odd.

AB: For regions with homogeneous, sub-tropical type climatic conditions, such as the Southeast, results such as Figures 5a and 5c were consistent across many of the objective functions we had used to measure parameter sensitivity with FAST (including parameter sensitivity measured for NSE at select reference streamgages). In areas such as the Lower Mississippi (Region 8, Figure 2), the amount of snowfall is negligible, so the three parameters that control snowfall and snowpack accumulation have negligible effect on total runoff. If there is minimal occurrence of snow in a region, then snow parameters won't be important, even in a complex model. For further discussion, a colleague has submitted a paper to HESS that strictly examines results of FAST applied to a 35-parameter daily streamflow model across the U.S.

-Also, the result of Figure 13 is very puzzling. As the authors note on page 10041 (line 27 onward), the groupNSE values are expected to be lower than the gageNSE values. Figure 13 instead shows many cases were groupNSE is much larger than gageNSE. I really struggle to believe that NSE can be increased so much and so often when using a model calibrated with a different objective function. The only explanation I can think of is that either the calibration algorithm in the gageNSE failed (for instance by getting stuck in a local minimum or being terminated too early) or that the comparison is not fair (for example that NSE refer to different time periods?). This needs clarification.

AB: What we wanted to emphasize with this plot is not a comparison of two calibration

methods, but that the grouped calibration strategy that focuses on the sensitive parameters can provide just as much information as traditional NSE-based individual streamgage calibration. There is quite a bit of difference between the two calibration methods compared: the individual gage calibration used the entire period of record for each streamgage, while the grouped calibration uses an odd/even calibration/validation calibration strategy, different objective functions are used (NSE versus multi-term weighted objective function for the grouped calibration), and climate adjustments are derived for each streamgage in the individual calibration, and for the entire region in the grouped calibration.

Author's response to remaining minor remarks

-p. 10030, lines 14-15: please justify why you do not incorporate the adjustment factors in the FAST analysis

AB: We viewed the adjustment factors as more related to the forcing data itself and independent from the model structure.

-p. 10031, lines 6-7 "parameter ranges were based..." Are these the ranges in Table 1 and already commented on p.10028, line 24? If so, just refer to the Table here.

AB: The parameters listed in Table 1 were the bounding ranges. We will make the adjustment to the text.

-p. 10031, lines 7-9: What do you mean by "standard application"? Also, I suppose the R package uses the equation N = 2Nharm max(!) + 1 to determine the minimal number of runs. If so, better cite Cuckier et al (1973), which is where the formula comes from. Also, please mention what is this number in your case, it would help readers to get an idea of how computationally demanding is the proposed approach.

AB: "In standard application" is ambiguous and should be removed. We accept the referee's recommendations and the sentence on lines 8-10 should be re-worded to: "The fast R package can determine the minimal number of runs necessary to estimate

the output variance all parameters (Cukier et al., 1973). For our application we generated an ensemble of 1000 sets (above the minimal number of 71 suggested by fast) to have the potential capability for further research into comparison of different sensitivity analysis methods."

-p. 10038, lines 1-4: The definition of the multi-term objective function is unclear. Are the four terms summed up? Why considering both mean monthly runoff and annual runoff (I would imagine that they convey the same information, the former being equal to the latter divided by 12)? I think inserting an equation with the mathematical definition of the objective function would help here.

AB: The objective function minimizes the sum of difference between the Z-scores of measured and simulated variables for four terms: Mean Monthly Streamflow (As shown in figures 8 and 10a), Monthly Streamflow (Raw monthly time series), Annual Streamflow (Time Series aggregated to annual time steps), and Mean Monthly SWE with a 25% error bound. The first three terms of the objective function were chosen because they conveyed information that can be used to easily inform other models (such as daily time-step models). Just to note, annual and mean monthly objective functions as we defined them are very different; the former is not equal to the latter/12.

We can include an equation with the appropriate text to help elucidate the different terms of the objective function.

-p. 10038, lines 14-16: Please clarify how the error bounds were taken into account. Did you modify the definition of the Z score for the SWE?

AB: Z-scores were calculated from basin mean monthly measured (SNODAS) and simulated SWE values. Upper and lower bounds of 25% were calculated for the mean monthly SWE Z-score (Zobs *1.25 for the upper bound; Zobs *.75 for the lower bound). If the simulated MWBM SWE value was contained within the upper and lower 25% bounds, the absolute difference was designated as 0. If the simulated MWBM SWE Z-score value was above the upper 25% error bound, the absolute Z-score difference

was calculated between the simulated Z-score value and the value of the upper 25% Z-score bound. If the simulated Z-score was below than the lower 25% error bound, the absolute Z-score difference was calculated between the simulated Z-score value and the value of the lower 25% Z-score bound.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 10023, 2015.