Hydrol. Earth Syst. Sci. Discuss.,
doi:10.5194/hess-2016-192-RC2, 2016

Hydrology and Earth System Sciences
Discussions

Open Access

Interactive comment on "Physically-based distributed hydrological model calibration based on a short period of streamflow data: case studies in two Chinese basins" *by* W. Sun et al.

Anonymous Referee #2

Received and published: 2 September 2016

Two watersheds is not enough to conclude your study provides general conclusions! There are groups that use thousands of watersheds, look up large sample hydrology, for example: http://meetingorganizer.copernicus.org/EGU2015/session/18271 http://www.hydrol-earth-syst-sci.net/18/463/2014/hess-18-463-2014.html

The contribution of the paper is not clear. Such analysis on the quality of calibration data dates back to 1996 (http://www.sciencedirect.com/science/article/pii/0022169495029184) and republication in HESS is not justified!

There is no figure provided of how calibration of SWAT with limited data translates into model simulation! How do I know 1 month of data is enough for calibration if I don't see

Printer-friendly version

Discussion paper

how the model works graphically? NSE is certainly not enough!

Page 2, lines 1-5: I don't agree with your statement that models like SWAT are able to predict droughts and floods! Droughts and floods respond to climatic forcings and climatic models are used to forecast them, certainly not SWAT!

Page 2, line 8-9: "Most parameters of hydrological models are conceptual without explicit physical meaning, which makes it necessary to identify parameter values through model calibration based on streamflow data". This refers to conceptual models mostly. Physically based distributed are supposed to have parameters with clear physical meaning, that can ideally be measured in the field.

Page 2, Lines 18-20: "Many recent works have focused on using in situ or remote sensing observations of hydrological processes other than streamflow for model calibration, e.g., soil moisture (e.g., Silvestro et al., 2015; Vrugt 20 et al., 2002), evapotranspiration (Vervoort et al., 2014; Winsemius et al., 2008), groundwater level(e.g., Khu et al., 2008)." My understanding is that since streamflow measurements are not available, one can alternatively use other variables such as soil moisture, ground water table and evapotranspiration as calibration data. This is certainly not the case, since measuring these variables is much more difficult and costly than streamflow. I suggest you phrase your sentences more carefully to avoid such confusions.

Page 3, line 4: What do you mean by "changing environment"?

Page 3, lines 4-6: You argue "For hydrological simulations or predictions in changing environments, physically-based distributed hydrological models are usually preferred, because of their better description of the spatial heterogeneity and details of the water cycle at the basin scale (Finger et al, 2012; Wu and Liu, 2012)." I understand that some physical modelers would make such arguments, but it is certainly a debated issue, so I wouldn't make such strong claims. This being said, in a changing climate, even physically based models are not proven to be working properly. The argument made by developers of physically based models is that since they use specific description of the watersheds, their models can handle land-use change (change of physical characteristics of watersheds). This also needs a lot of research still.

Equation 2 is all WRONG! You want to use an objective function of NSE, do, but you can't call it a likelihood function and use it as in the Bayes theorem! There is no scaling in Bayes law! You may call this weight, but not posterior likelihood.

I have a hard time with equation 3 also! Weights (or as you call them posterior likelihoods) are calculate based on overall performance of the model (t=1:N), but are used at each time step to estimate the cumulative probability of streaamflow. This is not right! You want to estimate the 95% uncertainty range, take the 2.5 and 97.5th percentiles of your streamflow simulation ensemble.

Page 6, lines 4-6: I don't understand this sentence, and it is critical to evaluate the methodology proposed in this paper: "As a first trial for the distributed model, we sought to explore the highest possible performance using certain short lengths of records, not the general performance of specific lengths."

The entire section 2.4 is poorly written and methodology is poorly described, making it really difficult to assess the paper.

Page 7, line 2: You admit that it is very time consuming to calibrate SWAT using GLUE. Why not using more intelligent calibration approaches like Markov Chain Monte Carlo? It has been shown in the literature that MCMC is orders of magnitude more efficient than GLUE.

Page 7 line 11: "Kim and Kaluarachchi (2009) and Yapo et al. (1996) showed that data from high-flow periods are more informative than data from low-flow periods for model calibration, because most model modules are activated in high-flow periods." I tend to disagree! It depends on your performance metrics, if you use NSE which is sensitive to peak flows, then yes you are right! But if you use metrics such as baseflow index this is not going to hold. Different processes of a model are activated under different forcings, so you can't simply ignore several processes and focus on the one (few) process(es) that are activated at the wet condition.

Page 7 line 24: Uncertainty bound not band! Correct throughout the manuscript.

Page 8, lines 2-3: "For the 1-year period, all three calibrations performed similarly to the benchmark calibration, and the dataset for 2006 even outperformed the benchmark" It is interesting and concerning that a shorter calibration period provides a higher performance. It requires explanation as to how it happened! You can't just leave it like that which might spuriously suggest smaller calibration period is sometimes even better! Here are my thoughts: 1. You are not using a consistent period to evaluate your model! 2. Your calibration approach did not converge to the right posterior distribution (as might happen with GLUE) 3. Your data includes some misinformation, meaning not only it doesn't provide any good information to constrain model parameters, but also it misguides the model! In the cases of 1 & 2, extra data can only be redundant and cannot deteriorate the performance of the model!

Page 8, lines 14-16: "The calibration using the 1-month dataset still achieved similar performance to benchmark calibration. Thus, it is indicated that in the Jinjiang Basin, it is possible to calibrate the SWAT model effectively using only 1-month's continuous daily observations of streamflow." This claim is rather strange to me! One month is enough to capture all the processes? Some processes might not even be activated in one month! Again, this is because you focused all your attention on NSE, and what is most important in NSE is the high peaks. So if you activate the processes that reproduce the high peaks, you get a good performance. This doesn't mean one month is enough to calibrate a model!

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., doi:10.5194/hess-2016-192, 2016.