Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-323-RC2, 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "A Wavelet-Based Approach to Streamflow Event Identification and Modeled Timing Error Evaluation" *by* Erin Towler and James L. McCreight

Cedric David (Referee)

cedric.david@jpl.nasa.gov

Received and published: 24 October 2020

General comments

The following is a review of manuscript hess-2020-323 entitled "A Wavelet-Based Approach to Streamflow Event Identification and Modeled Timing Error Evaluation" by E. Towler and J. L. McCreight being considered by Hydrology and Earth System Sciences.

This manuscripts describes a methodology for evaluating the timing of simulated river discharge hydrographs when compared with in situ observations. The approach first uses wavelet transforms (WTs) to expand observed one-dimensional hydrographs (discharge vs time) into two-dimensional WTs (power vs timescale and power vs time) as a means for event detection. From those events detected in the observations, the methodology then uses cross wavelet transforms (XWTs) to evaluate the difference in timing and duration of events (at multiple power levels) between observations and simulations. The new approach is specifically designed to compare subsequent versions of a given river model and can be used both for evaluation and for diagnosis. The paper uses simulations from subsequent versions of NOAA's National Water Model and observations from the USGS at four selected locations.

I really enjoyed reading this paper and I learnt a lot from it. I agree that the evaluation of hydrograph timing is an important aspect of river model calibration/validation, and one that is often overlooked. I am guilty of that myself! This subject matter is timely given the ongoing explosion of continental scale river models such as the NWM or similar global applications. In my opinion, the authors make a strong case for the value of their work given the complexity of hydrograph shapes and describe clear guidelines for the implementation of their methodology, while also acknowledging the multiple ways in which it could be adapted. This research ought to be of interest to the readership of HESS. I elected to write this review without reading any of the other community comments so that my opinion could be relatively unbiased. My recommendation is to return the manuscript to the authors for minor revisions. My comments are outlined below, in decreasing order of importance. I hope that the authors will find some value in my suggestions. Thank you for the opportunity to review this work.

Specific comments

First, I really want to highlight that I think the authors did a commendable effort in the clarity of their explanations. I specifically enjoyed the inclusion of a "Conceptual Overview" (Section 2) and the use of the simple prescribed "+5 hours" example (Section 4.2). I think the general descriptive approach used really serves the manuscript very well and makes it accessible to many readers, including those (such as myself) who are not familiar with wavelet transform-based event detection methods. I'd like to make two suggestions that may further help in this manner. First, I understand from Supplemental Table 1 that the units of "power" are mˆ6/sˆ2, hence they are squared units compared to discharge. It may be valuable to some readers to specify these units in the manuscript and also to perhaps suggest a hydrologic meaning to this quantity. For example, if one was to take the square root of the power value, would this in any way represent the amplitude of the peaks in the figures? If so, this may be a valuable explanation to add, which could also be graphically illustrated in Figure 1a. A rough estimate from Figure 1 suggests that the maximum power is 60,000 mˆ6/sˆ2 leading to a square root of approximately 245 mˆ3/s which is strikingly close to the amplitude of the hydrograph. Second, I could have used some further hand-holding in understanding Figure 2c. I guess I expected the entirety of the color scale to correspond to "+5 hours" (i.e. all red). I don't really understand why there appears to be a 10-hour minimum to accurately catching errors and why this makes sense because "the [time]scale must be at least double the error". It would be valuable to expand on this concept around Lines 322-325.

I wonder if there could be a good graphical way to explain how the timing error is computed in Figure 2c. The equations provided in Section 3.2.2 are not exactly straightforward, and I assume that making a graphic from those would be challenging, but it seems to be a key component of the study and some readers might benefit from such addition.

The manuscript tends to rely a bit heavily on supplemental figures throughout the text, which makes the reader jump from document to document. I suggest that the authors go through their figures carefully and evaluate whether some of supplemental figures could be combined with main manuscript figures.

The authors make a clear argument that some of the traditional error metrics (e.g. RMSE, NSE) implicitly include errors in timing but don't shed much light on them. Yet, the strength of these metrics is also the simplicity of their computation. In an effort to increase the broad acceptance and use of timing metrics that are less subjective than peak-over-threshold, it might be helpful to for the authors to attempt a recommendation for the simplest possible form of their methodology. I understand that different powers are related to independent peaks and that there is value in looking at them all. I wonder if using powers computed as the square of discharge values from traditional occurrence probability thresholds could help reconcile (connect?) the WT approach and the threshold approach. I am not suggesting more analysis here, more so a paragraph in the discussion (Section 6, around Lines 478-487) where the authors might further expand on the simplest way for others to apply their approach.

I'm not sure I fully understand the True/False (dark grey/light grey) legend in figures 5, 9, and 10. Figure 10 seems to suggest that lighter colors are not statistically significant but it's not clear from the legend. Likewise, the "Avg XWT Signif" color bar in Figures 11 and 12 is a bit mysterious to me as I see no associated colors on the graph. Could the authors rework their legends in these figures?

Technical corrections

Line 296-297: missing "Land" in the acronym for NLDAS.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-323, 2020.