Review of Manuscript

'A Wavelet-Based Approach to Streamflow Event Identification and Modeled Timing Error Evaluation' (hess-2020-323)

by E. Towler and J. L. McCreight

Dear Editor, dear Authors,

I have reviewed the aforementioned work. My conclusions and comments are as follows:

1. Scope

The article is within the scope of HESS.

2. Summary

The authors explain a method based on wavelet transform and cross wavelet transform to i) detect relevant events in streamflow time series and ii) to calculate timing errors between a reference time series for which the events were determined (typically an observed time series) and a test time series (typically a model simulation). Relevant regions in the full 2-d space (time and timescale) of wavelet transforms are identified by significance testing as suggested by Torrence and Compo (1998). The timing errors are calculated based on the cross wavelet transform as suggested by Liu et al. (2011), but restricted to the areas of significant events in the reference time series, which imposes a direction on the comparison. The authors illustrate their method with several application examples, which vary by their event characteristics (single event, multiple events, events caused by different processes). The authors conclude that the proposed method offers a systematic, objective, and data-driven method for event identification and timing error calculation, which can be applied to large data sets, and they stress that beyond the particular application (including particular user choices) presented in the paper, other uses of the core method are possible.

3. Overall ranking

Overall, the authors provide an elegant and general solution to the fuzzy problem of event detection and timing error calculation in streamflow time series, which I am sure provides more generality, better reproducibility and better insight than most existing methods, including the ones I was involved with (Series Distance). There are, however, some flaws in the study, in terms of presentation clarity and in terms of demonstrating the generality of the method beyond the particular chosen use case, which should be eliminated. The relevance of the method deserves this extra effort.

4. Evaluation

General points

- In the introduction, please provide a more comprehensive literature review on methods for event detection and timing error calculation. E.g. Mei and Anagnostou (2015), Merz et al. (2006), Koskelo et al. (2012).
- The use case presented in the paper takes observed events as the reference, and calculates timing errors for these (see e.g. P4 L90-92). This neglects other important aspects of event-specific (dis-)agreement of observations and simulations: False alarms and missed events. This is mentioned several times by the authors (e.g. P7 L148-153, P12 L266, P12 L271-272, P17 L388-391), and they also mention that the method could be set up differently if these aspects are of interest, but they do not explain how. False alarms and missed events are important and often-

used features of categorical model evaluation (and the idea of 'event' is categorical). So I suggest that the authors add a short discussion about if and how their method can be used to measure them (I am not asking to actually perform these analyses, but to provide guidance for future uses).

- I found it hard to follow the description of the method, as it extends over several sections of the paper:
  - In section 2, a conceptual overview is given but it misses an at least brief description of how WT and XWT function, which will be helpful for anyone not familiar with the concept. Also, section 2 refers to supplement Table 1 and supplement Fig. 1, which are in fact important to understand the method. I suggest moving these to the main paper. Section 2 refers to Fig. 1, but the concept of event clusters is not explained. This concept only becomes clear in Fig. 3, which is referred to in section 5 for the first time.
  - In section 3, the steps of the method are repeated in more detail, which creates some redundancy with section 2, but still the concept of event clusters only becomes clear in section 5 (Fig. 3). Also, I did not fully understand from the text how the observed WT and the XWT of the observed and simulated time series are related, such that evaluating the XWT in significance regions of the WT is justified (see P 12 L262-263). The authors only mention that significance areas in the WT and the XWT do not necessarily coincide (P10 L237). Please explain and justify in more detail.
  - Section 4.2 provides an application, which is somewhere in the middle between a demonstration case to explain the method (then it would be better to include it into section 3) and a demonstration cases like those in section 5. In section 4 it appears rather orphaned.
  - Overall, I suggest merging sections 2, 3 and 4.2 into one section explaining the method, which includes supplement Table 1, supplement Figure 1, and an illustrative example as shown in Fig. 3 where the concept of event clusters becomes clear.
- The meaning of 'event cluster' is not completely clear to me: From the supplement Table 1, it refers – for a particular choice of timescale – to a time-contiguous set of events (i.e. each horizontal line in Fig. 3d is one event cluster). From P9 L198-200, it seems that it refers to contiguous regions of statistical significance (i.e. the entire colored area in Fig. 3c is one event cluster). Please clarify.
- The concept of identifying relevant timescales by calculating – for every timescale - average power across all relevant events over time and then selecting local and global maxima (see e.g. Fig. 3c, right panel) is not clear to me. What is the meaning/interpretation of such a local maximum of averaged power, and how does it qualify as a selection criterion for relevant timescales? It will work when the relevant timescales are the same for all the rainfall-runoff events in the time series, but it will not if they come from different processes (such as the time series you show in supplemental Fig. 4) and have different characteristic timescales. Would it be better to assign a single characteristic timescale to each 'island of significance' (contiguous region of significant events surrounded by non-significant areas) by finding the maximum power peak in each island, and then calculating the timing error only for this representative (time,timescale)?
- I found it difficult to follow the presentation of the results in section 5, as not the same set of Figures was provided for each case in the paper. I suggest reducing the number of cases, but providing the same set of figures for all of them.

Specific points

- P3 L66: Seibert et al. (2016)
- P5 L 101: selected
- At the beginning of section 4, please add a short justification of your choice of test data
- Fig. 1d: The position and length of the horizontal green line is not clear at this point. Please explain in the text for easier comprehension.

- P14 L318: Despite what the authors state in the text, Fig. 1a and Fig. 2a do not show the same observed time series (however the significance areas in Fig. 1c and Fig. 2c are the same).
- In Fig. 2a, the 'obs' time series is light blue, in all other Figures it is dark blue. Please harmonize.
- P14 L329: I could not find Table 1
- P16 L368: I could not find Table 2
- P18 L423: I could not find Table 3
- All Figures with time series: x-axis (time) is usually given in calendar date, and y-axis (timescale) in hours. Displaying both in unit hours would facilitate the comparison of relevant timescales with the features in the time series.

Yours sincerely,

Uwe Ehret

References

Koskelo, A. I., Fisher, T. R., Utz, R. M., and Jordan, T. E.: A new precipitation-based method of baseflow separation and event identification for small watersheds (< 50 km2), J. Hydrol., 450–451, 267–278, https://doi.org/10.1016/j.jhydrol.2012.04.055, 2012.

Liu, Y., Brown, J., Demargne, J., Seo, D. J., 2011. A wavelet-based approach to assessing timing errors in hydrologic predictions. J. Hydrol. 397(3–4), 210–224. http://doi.org/10.1016/j.jhydrol.2010.11.040

Mei, Y. and Anagnostou, E. N.: A hydrograph separation method based on information from rainfall and runoff records, J. Hydrol., 523, 636–649, https://doi.org/10.1016/j.jhydrol.2015.01.083, 2015.

Merz, R., Blöschl, G., and Parajka, J.: Spatio-temporal variability of event runoff coefficients, J. ydrol., 331, 591–604, https://doi.org/10.1016/j.jhydrol.2006.06.008, 2006.

Torrence, C., Compo, G.P., 1998. A Practical Guide to Wavelet Analysis. Bull. Am. Meteorol. Soc. 79(1), 61-78.