Review for manuscript "Technical note: Partial wavelet coherency for improved understanding of scale-specific and localized bivariate relationships in geosciences"

Authors: Wei Hu and Bing Si

Journal: Hydrology and Earth System Sciences Discussions

Summary

In this technical note, the authors propose a method for identifying relationships between two variables for the case where the two variables are correlated to other variables themselves. They apply their 'updated partial wavelet coherency' (PWC) method to a synthetic dataset and two real-world applications and show that this updated PWC model shows similar performance as existing PWC models. They conclude that their model outperforms existing models because it provides phase information and allows for excluding several correlated variables from the PWC.

General remarks

I think that the study addresses a question of interest to the hydrological community, i.e. 'how can we identify the most important driving variables of a certain phenomena at different time scales'. The technical note is generally well structured. However, I think that it lacks a didactical and detailed introduction to the topic, problem, and wavelet analysis. The introduction would significantly benefit from providing examples of when the identification of bivariate relationships are important (i.e. providing a motivation for the study), an in-depth introduction to wavelet analysis (for the readers who are not yet too familiar with the topic), and an introduction to the terminology used. Extending the introduction will increase the length of the note and I suggest removing the practical example number 2 instead. I think it does not provide additional insights regarding the performance of the method proposed compared to the statements that were already made based on the synthetic data and the first practical example. Since the new method does not seem to clearly outperform existing methods, I would better explain why adding phase information and excluding several confounding variables is beneficial for the analysis. I would also add a more detailed discussion of model weaknesses, especially the implications of detecting spurious correlations. In addition, the note would profit from careful language editing.

Major points

1. **Abstract:** The abstract is not very accessible to non-wavelet-specialists. I would provide a short example for when such an analysis would be necessary/beneficial and shortly summarize what wavelet coherence analysis is all about. Please also shortly explain why PWC has been introduced in the first place (l. 12). I would also mention the datasets used for model evaluations (l. 14). I think the statement 'producing more accurate results' (l. 18) needs justification, otherwise it is not very credible. I would exclude lines 21-24 because this is a technical note and specific results regarding the example applications going beyond model performance are in my opinion not of interest here.

2. **Introduction:** The introduction should in my opinion provide a motivation for the use of PWC methods, also for non-specialists on the topic e.g. by providing examples of important bivariate relationships in the geosciences and why we may be interested in them. In addition, an introduction to wavelet analysis in general and wavelet coherence analysis in particular should be provided. The reader should also be made familiar with the terminology used, e.g. what kind of scales are you talking about and what is an 'excluding variable'. A clear motivation for why excluding variables and including phases matters is required to underline the benefits of the methods later on in the results and conclusions sections (l. 57-58). Currently, the introduction does not very well prepare readers for what they are going to read in the methods and results sections.
3. **Theory:** I think that you should start even simpler here and provide a short introduction to wavelet analysis (difference between discrete and complex, terminology) and wavelet coherence analysis. In addition, it is unclear to me what exactly the difference between classical PWC and your proposed method is (l. 74-76). Currently, it is not entirely clear to me how the Monte Carlo experiment was performed (l. 108-110). Could you please slightly expand this section?
4. **Data and analysis:** I would recommend removing the 'soil water content' example (section 4.2.2) because as I can see it does not show anything that has not yet been shown by the 'free evaporation example' in terms of the validity of the model. I would rather invest the space in extending the introduction as outlined in more details above. In the figure captions, I would add a reference to the dataset used to generate it. In addition, I am not sure what you would like to show with the cases where the variable of interest is excluded. I would therefore exclude the results referring to this exercise (e.g. Figure 1 last row and see l. 236-237). I also think that the figures would profit a lot from using labels for subfigures, which would facilitate orientation. To me, the difference between the Mihanovic et al. (2009) model and the proposed model are not evident by looking at the Figures presented (a difference of 0.03 does not seem to be a lot, l.293). Therefore, I think the actual advantages of using this new method should better be worked out and explained before a statement such as 'the new method outperforms the Mihanovic et al. method' (l. 293-294) is made. Please also explain why the inclusion of 'phase information' is an advantage of the new method (l. 312-313).
5. **A proper discussion section is missing:** I would add an in-depth discussion of the weaknesses and benefits of the approach and put the new method into perspective by comparing it to existing methods.
6. **Conclusions:** Given the evidence provided in the results section, statements such as 'the new method produces slightly more accurate coherence' do not seem to be justified. As mentioned earlier the benefits of including phase information and excluding several variables need to be better explained. Some of the material presented in this section could be moved to the new discussion section.
7. **Code availability:** I would provide the Matlab code via a data/file repository such as HydroShare or Zenodo instead of the supplement (l.27). This would be very helpful for the community and potential users.

Minor points

- L. 31: please explain what you mean by 'time and space localization'.

- L.34: ‘among these methods’
- Transition from l. 42 to l. 43: very sharp transition from bivariate relationships to prediction. I would try to establish a clear link between the two things.
- L. 48: what do you mean by ‘this issue’?
- L. 50: what kind of scales? Temporal or spatial?
- L. 53-54: would combine greenhouse gas emissions and climate in one category.
- L. 61: information ‘which will allow to….’
- L. 61: what do you mean by ‘analogy’ in this context. I think that rephrasing may be required.
- L. 62: Be specific with what you mean by ‘it’: ‘the proposed method’.
- L. 76: Please explain to the reader what you mean by ‘scale’ and ‘location’.
- L. 99: same for ‘phase angle’.
- L. 184-185: can in my opinion be removed.
- L. 191: what does data refer to? Soil water content?
- L. 214: ‘significance band’.
- L. 215-216: is this statement underlined by any analysis performed?
- L. 247: what is the purpose of replacing half of the time series by 0?
- L. 261-263: Which feature in the plots actually indicates these ‘abrupt changes’?
- L. 266: I can only see one wavelet band of high significance in Figure 3. Where is the second one you mention here?
- L. 298: introduce term ‘octave’.
- L. 363-366: would move this sentence to discussion section.