Response to Anonymous Referee #2

Comment #1:
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.*

Response #1:

Many thanks for your comment. We think the new method outperforms the existing one from the three aspects: (1) more accurate results because of the theoretical differences (will be explained below); (2) inclusion of phase information; and (3) any number of excluding variables can be considered.
Below we will respond to each of your comments.

Comment #2:

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.*
Response #2:
More detailed information on the general wavelet analysis, PWC, and problem of existing methods will be added in the Introduction and Theory sections (see more details below). The importance of bivariate relationships will be explained by adding "The BWC partitions correlation between two variables into different locations and scales, which are different from the overall relationships at the sampling scale as shown from the traditional correlation coefficient. For example, BWC indicated that soil water content (SWC) of a hummocky landscape in the Canadian Prairies was negatively correlated to soil organic carbon (SOC) content at a slope scale (50 m), but was positively correlated to SOC at a watershed scale (120 m) in summer due to the different processes involved at different scales (Hu et al., 2017). Because the positive correlation cancels with the negative at different scales and/ or locations, the traditional correlation coefficient between SWC and SOC is absent, which is misleading." In terms of wavelet analysis, we will add "The wavelet analyses are based on wavelet transform using mother wavelet function which expands spatial (or

time) series into location-scale (or time-frequency) space for identification of localized intermittent scales (or periodicities)."
The original motivation to have both real datasets is to demonstrate that the proposed method can be used for both spatial and temporal data. We agree that more detailed introduction will increase the length of the paper, so we will remove the results related to SWC. Its suitability to both spatial and temporal data will be mentioned in the revision.

The new method and the existing method have theoretical differences. We will derive the equation for calculating PWC in case of one excluding variable from our Eq. (9). After we carefully check the paper of Mihanovic et al. (2009), it looks like their equation is the same to our equation in case of one excluding variables. Unfortunately, Ng and Chan (2012a) might have misinterpreted the equation of Mihanović et al. (2009) and developed Matlab code for calculating PWC by replacing complex coherence with the real coherence between two variables (please see the detail below). Unfortunately, the code of Ng and Chan (2012a) has been widely used. Not surprisingly, the equation of Ng and Chan (2012a) usually underestimates PWC value relative to the new method. Analysis from the real data has indicated that the differences between two methods can be large. The related discussion will be added either in the main body of the paper or in the supplementary. Meanwhile, more discussion on why adding phase information and excluding several confounding variables is beneficial for the analysis will be added.

A separate discussion section will be added by including a more detailed discussion of model advantages (e.g., the three aspects mentioned in the Response #1) and weaknesses (including spurious correlations and multiple-testing).

Language has been carefully checked by editors from our publication office.

Please see the details below on how we will address the comments you have made.

Major points

Comment #3:
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.*
Response #3:

We will add "It is a measure of correlation between two spatial (or time) series in the location-scale (time-frequency) domain. This method is particularly suited to geoscience where relationships between multiple variables commonly differ with locations or/and scales due to varying processes involved across different scales and locations." to explain what is wavelet coherence and when it would benefit.
The PWC was introduced "to detect the scale-specific and localized bivariate relationships by excluding the effects of other variables".

The description of dataset used for model evaluations will be "Both stationary and non-stationary artificial datasets with response variable being the sum of five cosine waves at 256 locations are used for method tests.".
Why the new method produces more accurate results will be explained by adding "Compared with the previous PWC calculation, the new method produces more accurate results in case of one excluding variable because bivariate real coherence rather than the bivariate complex coherence was mistakenly used in the previous PWC calculation" .
Lines 21-24 from previous submission have been removed.

Comment #4:
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.*
Response #4:

The important bivariate relationships in geosciences will be explained by adding "The BWC gives correlation between two variables at different locations and scales rather than the overall relationships at the sampling scale as obtained from the traditional correlation coefficient. For example, BWC indicated that soil water content (SWC) of a hummocky landscape in the Canadian Prairies was negatively correlated to soil organic carbon (SOC) content at a slope scale (50 m), but was positively correlated to SOC at a watershed scale (120 m) in summer due to the different processes involved at different scales (Hu et al., 2017). Because of the different correlations at different scales, the traditional correlation coefficient indicated that SWC was not correlated to SOC, which is misleading..".

The motivation for the use of PWC method is further explained by adding "Partial correlation analysis is one such method to avoid the misleading relationships resulting from the interdependence between other variables and both predictor and response variables (Kenney and Keeping, 1939)" and "For example, PWC analysis indicated that Southern Oscillation Index and Pacific Decadal Oscillation did not affect precipitation across India, while this was misinterpreted by the BWC analysis because of their interdependence on Niño 3.4 that affects precipitation (Rathinasamy et al., 2017)."

An introduction to wavelet analysis in general will be added as "The wavelet analyses are based on wavelet transform using mother wavelet function which expands spatial (or time) series into location-scale (or time-frequency) space for identification of localized intermittent scales (or periodicities)."

When we talk about scale, it can mean spatial or temporal scale depending on if the dataset are spatial series or time series. To avoid repeatedly addressing if this is related to spatial or time scale, we will define it at the first time by adding "For convenience, we will mainly refer to location and scale irrespective of spatial or temporal series unless otherwise mentioned".

Excluding variable refers to "variable whose influence on the response variable is excluded".

The explanation on the motivation for why excluding variables and including phases matter will be added as "The coherence between response and predictor variables can still be misleading if more than one variable is interdependent with the predictor variable. This is especially true if these variables are correlated with the predictor variable at different locations or/and scales. In addition, the types of correlation (i.e., positive or negative) especially at different locations and scales remains unclear without phase information.".in the introduction.

Comment #5:

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?*

Response #5:

We will add the introduction to wavelet analysis, wavelet coherence analysis and associated equations at start of the Theory section. Here we assume you mean difference between discrete wavelet transform and continuous wavelet transform. The addition will be like:

"Wavelet analysis is based on the calculations of wavelet coefficients using wavelet transform at different scales and locations for each variable involved. Two types of wavelet transform exist including continuous wavelet transform (CWT) and discrete wavelet transform (DWT). While the DWT is mainly used for data compression and noise reduction, the CWT is widely used for extracting scale-specific and localized features as is the case of this study (Grinsted et al., 2004). For the CWT, the Morlet wavelet is used as a mother wavelet function to transform a spatial (or time) series into location-scale (or time-frequency) domain which allows us to identify both location-specific amplitude and phase information of wavelet coefficients at different scales (Torrence and Compo, 1998). From wavelet coefficients, auto- and cross-wavelet power spectra for two variables can be calculated as the product of wavelet coefficient and complex conjugate of itself (auto-wavelet power spectra) or another variable (cross-wavelet power spectra). The BWC is calculated as the ratio of smoothed cross-wavelet power spectra to the product of two auto-wavelet power spectra (Grinsted et al., 2004). Hu and Si (2016) extended wavelet coherence from two to multiple (≥3) variables and developed MWC. Detailed information on the calculations of wavelet coefficients, auto- and cross-wavelet power spectra, BWC, and MWC based on the CWT can be found elsewhere (Torrence and Compo, 1998; Grinsted et al., 2004; Si and Farrell, 2004; Si, 2008; Hu and Si, 2016; Hu et al., 2017). Here, we will only introduce the theory and calculation very relevant to PWC. "

In addition, the derivation of Eq.(1) in the original submission from equations of complex partial spectrum in frequency domain and bivariate complex coherence from time-frequency domain will be demonstrated in the supplement as below:

"S1 Derivation of the complex PWC Eq.(1)

Complex partial spectrum from frequency (scale)domain (Makhtar et al., 2014) can be used to define that of time-frequency (location-scale) domain, $\underset{W}{\leftrightarrow}^{y,x\cdot \mathrm{Z}}(s,\tau)$, which is expressed as

$$\underset{W}{\leftrightarrow}^{y,x\cdot Z}(s,\tau) = \underset{W}{\leftrightarrow}^{y,x}(s,\tau) - \frac{\underset{W}{\leftrightarrow}^{y,Z}(s,\tau)\overline{\underset{W}{\leftrightarrow}^{x,Z}(s,\tau)}}{\underset{W}{\leftrightarrow}^{Z,Z}(s,\tau)} \quad \text{(S1)}$$

where $\underset{W}{\leftrightarrow}$ is the smoothed cross spectrum, $\overline{(\cdot)}$ is the complex conjugate operator, y, x, and Z ($Z = \{Z_1, Z_2, \cdots, Z_q\}$) refer to the response variable, predictor variable, and excluding variables, respectively. s and τ refer to scale (frequency) and location (time), respectively.

Given the definition of coherence between two variables y and x, their complex coherence $\gamma_{y,x}(s,\tau)$ (Eq.(5)) can be re-written as

$$\gamma_{y,x}(s,\tau) = \frac{\underset{W}{\leftrightarrow}^{y,x}(s,\tau)}{\sqrt{\underset{W}{\leftrightarrow}^{y,y}(s,\tau)\,\underset{W}{\leftrightarrow}^{x,x}(s,\tau)}} \quad \text{(S2)}$$

Then we can define complex partial coherence as

$$\gamma_{y,x\cdot Z}(s,\tau) = \frac{\underset{W}{\leftrightarrow}^{y,x\cdot Z}(s,\tau)}{\sqrt{\underset{W}{\leftrightarrow}^{y,y\cdot Z}(s,\tau)\,\underset{W}{\leftrightarrow}^{x,x\cdot Z}(s,\tau)}} \quad \text{(S3)}$$

Based on Eq. (S1) and Eqs 2, 3, and 4 ($R^2_{y,x\cdot Z}(s,\tau) = \frac{\underset{W}{\leftrightarrow}^{y,Z}(s,\tau)\underset{W}{\leftrightarrow}^{Z,Z}(s,\tau)^{-1}\overline{\underset{W}{\leftrightarrow}^{x,Z}(s,\tau)}}{\underset{W}{\leftrightarrow}^{y,x}(s,\tau)}$,

$R^2_{y,Z}(s,\tau) = \frac{\underset{W}{\leftrightarrow}^{y,Z}(s,\tau)\underset{W}{\leftrightarrow}^{Z,Z}(s,\tau)^{-1}\overline{\underset{W}{\leftrightarrow}^{y,Z}(s,\tau)}}{\underset{W}{\leftrightarrow}^{y,y}(s,\tau)}$, and $R^2_{x,Z}(s,\tau) = \frac{\underset{W}{\leftrightarrow}^{x,Z}(s,\tau)\underset{W}{\leftrightarrow}^{Z,Z}(s,\tau)^{-1}\overline{\underset{W}{\leftrightarrow}^{x,Z}(s,\tau)}}{\underset{W}{\leftrightarrow}^{x,x}(s,\tau)}$)

we obtain

$$\underset{W}{\leftrightarrow}^{y,x\cdot Z}(s,\tau) = \underset{W}{\leftrightarrow}^{y,x}(s,\tau)\left(1 - \frac{\underset{W}{\leftrightarrow}^{y,Z}(s,\tau)\,\overline{\underset{W}{\leftrightarrow}^{x,Z}(s,\tau)}}{\underset{W}{\leftrightarrow}^{Z,Z}(s,\tau)\,\underset{W}{\leftrightarrow}^{y,x}(s,\tau)}\right) = \underset{W}{\leftrightarrow}^{y,x}(s,\tau)\left(1 - R^2_{y,x\cdot Z}(s,\tau)\right) \quad \text{(S4)}$$

$$\underset{W}{\leftrightarrow}^{y,y\cdot Z}(s,\tau) = \underset{W}{\leftrightarrow}^{y,y}(s,\tau)\left(1 - \frac{\underset{W}{\leftrightarrow}^{y,Z}(s,\tau)\,\overline{\underset{W}{\leftrightarrow}^{y,Z}(s,\tau)}}{\underset{W}{\leftrightarrow}^{Z,Z}(s,\tau)\,\underset{W}{\leftrightarrow}^{y,y}(s,\tau)}\right) = \underset{W}{\leftrightarrow}^{y,y}(s,\tau)\left(1 - R^2_{y,Z}(s,\tau)\right) \quad \text{(S5)}$$

$$\underset{W}{\leftrightarrow}^{x,x\cdot Z}(s,\tau) = \underset{W}{\leftrightarrow}^{x,x}(s,\tau)\left(1 - \frac{\underset{W}{\leftrightarrow}^{x,Z}(s,\tau)\,\overline{\underset{W}{\leftrightarrow}^{x,Z}(s,\tau)}}{\underset{W}{\leftrightarrow}^{Z,Z}(s,\tau)\,\underset{W}{\leftrightarrow}^{x,x}(s,\tau)}\right) = \underset{W}{\leftrightarrow}^{x,x}(s,\tau)\left(1 - R^2_{x\cdot Z}(s,\tau)\right) \quad \text{(S6)}$$

Inserting Eqs S4, S5, and S6 into Eq. (S3), we have

$$\gamma_{y,x\cdot Z}(s,\tau) = \frac{\underset{W}{\leftrightarrow}^{y,x}(s,\tau)\left(1-R^2_{y,x\cdot Z}(s,\tau)\right)}{\sqrt{\underset{W}{\leftrightarrow}^{y,y}(s,\tau)\left(1-R^2_{y,Z}(s,\tau)\right)\underset{W}{\leftrightarrow}^{x,x}(s,\tau)\left(1-R^2_{x\cdot Z}(s,\tau)\right)}} = \frac{\underset{W}{\leftrightarrow}^{y,x}(s,\tau)\left(1-R^2_{y,x\cdot Z}(s,\tau)\right)}{\sqrt{\underset{W}{\leftrightarrow}^{y,y}(s,\tau)\underset{W}{\leftrightarrow}^{x,x}(s,\tau)}\sqrt{\left(1-R^2_{y,Z}(s,\tau)\right)\left(1-R^2_{x\cdot Z}(s,\tau)\right)}} =$$
$$\frac{\left(1-R^2_{y,x\cdot Z}(s,\tau)\right)\gamma_{y,x}(s,\tau)}{\sqrt{\left(1-R^2_{y,Z}(s,\tau)\right)\left(1-R^2_{x\cdot Z}(s,\tau)\right)}} \quad \text{(S7)}$$

Obviously, Eq. (S7) and Eq. (1) are identical."

The difference between Eq. (9) and Eq. (14) will be explained by derivation of PWC in the case of one excluding variable from Eq. (1) (This is also addressed in the Response #2 to the referee #1).

So, when only one variable (e.g., $Z1$) is excluded, Eq.(9) ($\rho^2_{y,x\cdot Z} = \frac{\left|1-R^2_{y,x\cdot Z}(s,\tau)\right|^2 R^2_{y,x}(s,\tau)}{\left(1-R^2_{y,Z}(s,\tau)\right)\left(1-R^2_{x,Z}(s,\tau)\right)}$)

can be written as

$$\rho^2_{y,x\cdot Z1} = \frac{\left|1-R^2_{y,x\cdot Z1}(s,\tau)\right|^2 R^2_{y,x}(s,\tau)}{\left(1-R^2_{y,Z1}(s,\tau)\right)\left(1-R^2_{x,Z1}(s,\tau)\right)} \tag{RC1}$$

Based on equations (2) in our paper,

$$\rho^2_{y,x\cdot Z1} = \frac{\left|1-\frac{\overleftrightarrow{W}^{y,Z1}(s,\tau)\,\overleftrightarrow{W}^{Z1,Z1}(s,\tau)^{-1}\,\overline{\overleftrightarrow{W}^{x,Z1}(s,\tau)}}{\overleftrightarrow{W}^{y,x}(s,\tau)}\right|^2 \frac{\left|\overleftrightarrow{W}^{y,x}(s,\tau)\right|^2}{\overleftrightarrow{W}^{y,y}(s,\tau)\,\overleftrightarrow{W}^{x,x}(s,\tau)}}{\left(1-R^2_{y,Z1}(s,\tau)\right)\left(1-R^2_{x,Z1}(s,\tau)\right)}$$

$$= \frac{\left|\overleftrightarrow{W}^{y,x}(s,\tau)-\frac{\overleftrightarrow{W}^{y,Z1}(s,\tau)\overline{\overleftrightarrow{W}^{x,Z1}(s,\tau)}}{\overleftrightarrow{W}^{Z1,Z1}(s,\tau)}\right|^2}{\overleftrightarrow{W}^{y,y}(s,\tau)\overleftrightarrow{W}^{x,x}(s,\tau)\left(1-R^2_{y,Z1}(s,\tau)\right)\left(1-R^2_{x,Z1}(s,\tau)\right)}$$

$$= \frac{\frac{1}{\sqrt{\left(\overleftrightarrow{W}^{y,y}(s,\tau)\overleftrightarrow{W}^{x,x}(s,\tau)\right)^2}^{\,2}}\left|\overleftrightarrow{W}^{y,x}(s,\tau)-\frac{\overleftrightarrow{W}^{y,Z1}(s,\tau)\overline{\overleftrightarrow{W}^{x,Z1}(s,\tau)}}{\left(\sqrt{\overleftrightarrow{W}^{Z1,Z1}(s,\tau)}\right)^2}\right|^2}{\left(1-R^2_{y,Z1}(s,\tau)\right)\left(1-R^2_{x,Z1}(s,\tau)\right)}$$

$$= \frac{\left|\frac{\overleftrightarrow{W}^{y,x}(s,\tau)}{\sqrt{\overleftrightarrow{W}^{y,y}(s,\tau)}\sqrt{\overleftrightarrow{W}^{x,x}(s,\tau)}}-\frac{\overleftrightarrow{W}^{y,Z1}(s,\tau)\;\overline{\overleftrightarrow{W}^{x,Z1}(s,\tau)}}{\sqrt{\overleftrightarrow{W}^{y,y}(s,\tau)}\sqrt{\overleftrightarrow{W}^{x,x}(s,\tau)}\sqrt{\overleftrightarrow{W}^{Z1,Z1}(s,\tau)}\;\sqrt{\overleftrightarrow{W}^{Z1,Z1}(s,\tau)}}\right|^2}{\left(1-R^2_{y,Z1}(s,\tau)\right)\left(1-R^2_{x,Z1}(s,\tau)\right)}$$

$$= \frac{\left|\frac{\overleftrightarrow{W}^{y,x}(s,\tau)}{\sqrt{\overleftrightarrow{W}^{y,y}(s,\tau)}\sqrt{\overleftrightarrow{W}^{x,x}(s,\tau)}}-\frac{\overleftrightarrow{W}^{y,Z1}(s,\tau)}{\sqrt{\overleftrightarrow{W}^{y,y}(s,\tau)}\sqrt{\overleftrightarrow{W}^{Z1,Z1}(s,\tau)}}\cdot\frac{\overline{\overleftrightarrow{W}^{x,Z1}(s,\tau)}}{\sqrt{\overleftrightarrow{W}^{x,x}(s,\tau)}\sqrt{\overleftrightarrow{W}^{Z1,Z1}(s,\tau)}}\right|^2}{\left(1-R^2_{y,Z1}(s,\tau)\right)\left(1-R^2_{x,Z1}(s,\tau)\right)}$$

$$= \frac{\left|\gamma_{y,x}(s,\tau)-\gamma_{y,Z1}(s,\tau)\overline{\gamma_{x,Z1}(s,\tau)}\right|^2}{\left(1-R^2_{y,Z1}(s,\tau)\right)\left(1-R^2_{x,Z1}(s,\tau)\right)} \tag{RC2}$$

Namely, when only one variable (e.g., $Z1$) is excluded, Eq.(9) can be written as

$$\rho^2_{y,x\cdot Z1} = \frac{\left|\gamma_{y,x}(s,\tau)-\gamma_{y,Z1}(s,\tau)\overline{\gamma_{x,Z1}(s,\tau)}\right|^2}{\left(1-R^2_{y,Z1}(s,\tau)\right)\left(1-R^2_{x,Z1}(s,\tau)\right)} \tag{RC3}$$

Eq. (RC3) will be added to revision as Eq. (14), and the derivation process for be added to the supplementary.

In the case of one excluding variable ($Z = \{Z_1\}$), Mihanović et al. (2009) suggested that the PWC can be calculated by an equation analogous to the traditional partial correlation squared (Kenney and Keeping, 1939). Their equation is the same to Eq. (14 or RC3). Unfortunately, Ng and Chan (2012a) might have misinterpreted the equation of Mihanović et al. (2009) and developed Matlab code for calculating PWC using the equation expressed as

$$\rho^2_{y,x\cdot Z1} = \frac{\left|R_{y,x}(s,\tau) - R_{y,Z1}(s,\tau)\, R_{x,Z1}(s,\tau)\right|^2}{\left(1 - R^2_{y,Z1}(s,\tau)\right)\left(1 - R^2_{x,Z1}(s,\tau)\right)} \quad \text{(RC4) (or 15 in the revision)}$$

where $R_{y,x}(s,\tau)$, $R_{y,Z1}(s,\tau)$, and $R_{x,Z1}(s,\tau)$ are the square root of $R^2_{y,x}(s,\tau)$, $R^2_{y,Z1}(s,\tau)$, $R^2_{x,Z1}(s,\tau)$, respectively. $R^2_{y,Z1}(s,\tau)$ and $R^2_{x,Z1}(s,\tau)$ can be calculated from Eq. (10) by replacing y and x with their corresponding variables. Eq. (15) has been widely used to calculate PWC in case of one excluding variable (Aloui et al., 2018; Altarturi et al., 2018; Jia et al., 2018; Li et al., 2018; Mutascu and Sokic, 2020; Ng and Chan, 2012b; Rathinasamy et al., 2017; Wu et al., 2020). Note that complex coherence and real coherence are involved in the numerators of Eqs. (14) and (15), respectively, while the denominators are exactly the same. Further comparison indicates that Eq. (RC4) underestimates PWC value relative to Eq. (14) unless unless $\gamma_{y,x}(s,\tau)$ and $\gamma_{y,Z1}(s,\tau)\,\overline{\gamma_{x,Z1}(s,\tau)}$ in Eq. (14) are collinear (i.e., their arguments are identical) under which the two equations produce the same PWC values. Differences between Eqs. (14) and (15) will be discussed further using both artificial data and real dataset. For the comparison purpose, we refer to Eqs. (14) and (15) as new method and classical method, respectively.

Therefore, the differences in PWC values calculated from the two methods are context-specific. As the Referee #2 mentioned, although the difference between the Mihanovic et al. (2009) model (Eq.15) and the proposed model (Eq.14) are small, i.e., the difference of PWC values is only 0.03 for the artificial data, but Eq.14 produces PWC closer to 1.

In addition, the comparison of these two methods using real data indicated that the difference between the two methods can be large. As an example, Figure RC2 and Figure RC3a shows big differences of PWC values between these two methods at scales of around 12 month (1 year). Mean PWC values by the new method were consistently higher than the classical method, and the differences ranged from 0.4 to 0.6 around the scale of 1 year (Figure RC3b).This highlights that the new method produces more accurate results than the classical method.

In the revision copy, we will incorporate these discussions either in the main body of the paper or in the supplementary.

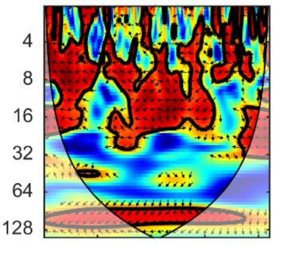

Figure RC2. Partial wavelet coherency (PWC) between evaporation (E) and relative humidity (RH) after excluding the effect of mean temperature (T) calculated by the new method. (subplot of Figure 3d in the revision)

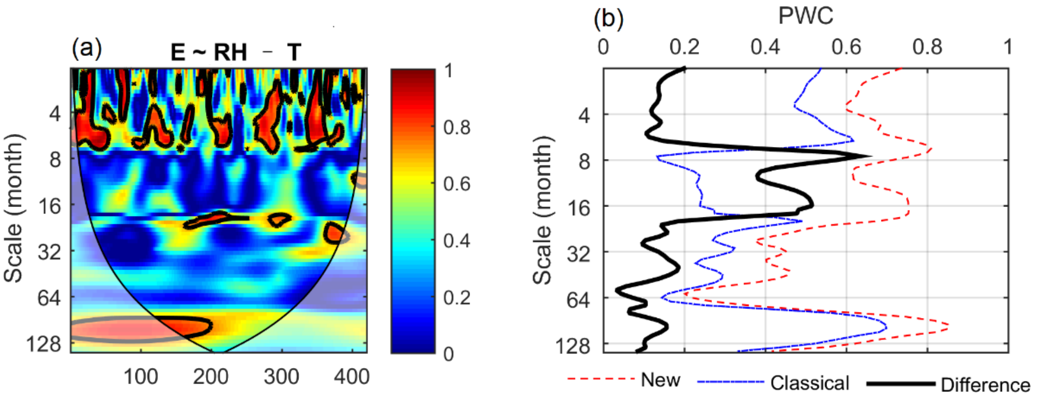

Figure RC3. Partial wavelet coherency (PWC) between evaporation (E) and relative humidity (RH) after excluding the effect of mean temperature (T) using the classical method (a) and differences in PWC between the new method and classical method as a function of scale (b).

Monte Carlo method is explained in more details by adding why we chose AR1 model and how many repeats are needed as "The first-order autoregressive model (AR(1)) is chosen because it can be used to simulate most geoscience data very well (Wendroth et al., 1992; Grinsted et al., 2004; Si and Farrell, 2004). Different combinations of r1 values (i.e., 0.0, 0.5, and 0.9) were used to generate 10 to 10 000 AR(1) series with three, four and five variables. Our results indicate that the noise combination has little impact on the PWC values at the 95% confidence level as also found by Grinsted et al. (2004) for the BWC case (data not shown). The relative difference of PWC at the 95% confidence level compared to that calculated from 10 000 AR(1) series decreases with increase in number of AR(1) series. When the number of AR(1) is above 300, very low maximum relative difference (e.g., <2%) is observed (Fig. S1 of Sect. S3 in the Supplement). Therefore, repeating number of 300 seems to be efficient for significance test. If calculation time is not a barrier, however, bigger repeating number such as ≥1000 is recommended.".

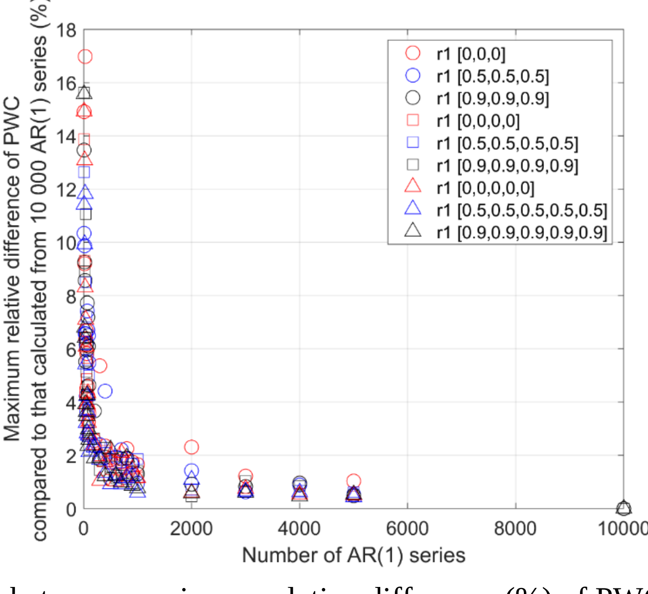

Figure S1. Relationship between maximum relative difference (%) of PWC compared to that calculated from 10 000 AR(1) series (surrogate dataset) versus the number of AR(1) series during the significance test using the Monte Carlo test. Number of scales per octave is 12.

Comment #6:

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).*

Response #6:

Thanks for this advice. We removed the soil water content example.

Reference to the dataset used to generate the figures will be added in the figure caption as "All variables were generated by following Yan and Gao (2007) and Hu and Si (2016) and are explained in Section 3.1 and are shown in Fig. S2 of Sect. Ss in the Supplement."

The purpose of showing the cases of variable of interest being excluded is to basically show that the PWC values should be theoretically zero in that case. As we have the similar results in the case of two excluding variables (Figure 3 in the original submission), we will remove this in Figure 1.

We will add a label for each subfigure in the revision.

As we explained below, some theoretical differences exist between these two methods. We will add this into the Theory section as shown above.

In the new discussion section, we will highlight the advantages and weakness of the new method (Please see the details in the Response #7 below).

Comment #7:

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.*

Response #7:

Advantages and weaknesses of the method will be added as:

“5. Discussion on the advantages and weaknesses of the new method

5.1 Advantages

We extend the partial coherence method from the frequency (scale) domain (Koopmans, 1995) to the time-frequency (location-scale) domain. The new method is an extension of previous work on PWC and MWC (Mihanović et al., 2009; Hu and Si, 2016). Method test and application has verified that it has the advantage of dealing with more than one excluding variable and providing the phase information associated with the PWC. In case of one excluding variable, Mihanović et al. (2009) has suggested to calculate PWC by an equation analogous to the traditional partial correlation squared (Eq. 14), which can be derived from our Eq. (9). However, their equation was widely used by replacing the complex coherence in Eq. (14) with real coherence as expressed in Eq. (15).

The differences between the new method (Eq.14) and the classical method (Eq. 15) are compared using both the artificial and real datasets. Except for the phase information, the two methods generally produce comparable coherence for the artificial dataset for the case of one excluding variable (Fig. S5 of Sect. S3 in the Supplement). However, the new PWC method produces consistently slightly higher coherence than the classical method. For example, their mean PWCs between y and y_2 at the scale of 8 after excluding the effect of y_4 are 1.00 and 0.97, respectively. This indicates that the new method produces coherence between y and y_2 at the scale (8) of y_2 closer to 1 as we expect. While the classical method produces similar PWC between E and other meteorological factors in most cases especially for the coherence between E and T after excluding the effects of others (Fig. S6 of Sect. S3 in the Supplement), large differences between these two methods can also be observed. For example, while the new method recognizes the strong coherence between E and RH after excluding the effect of T at scales of around 1 year (Fig. 3d), this coherence was negligible by the classical method (Fig. 5a). Mean PWC values by the new method were consistently higher than the classical method, and the differences ranged from 0.4 to 0.6 around the scale of 1 year (Fig. 5b). Considering the real coherence (Eq.15) rather than complex coherence (Eq.14) between every two variables in the numerators can potentially result in large underestimation of the partial wavelet coherence. Therefore, the new method produces more accurate results than the classical method is one of the advantages.

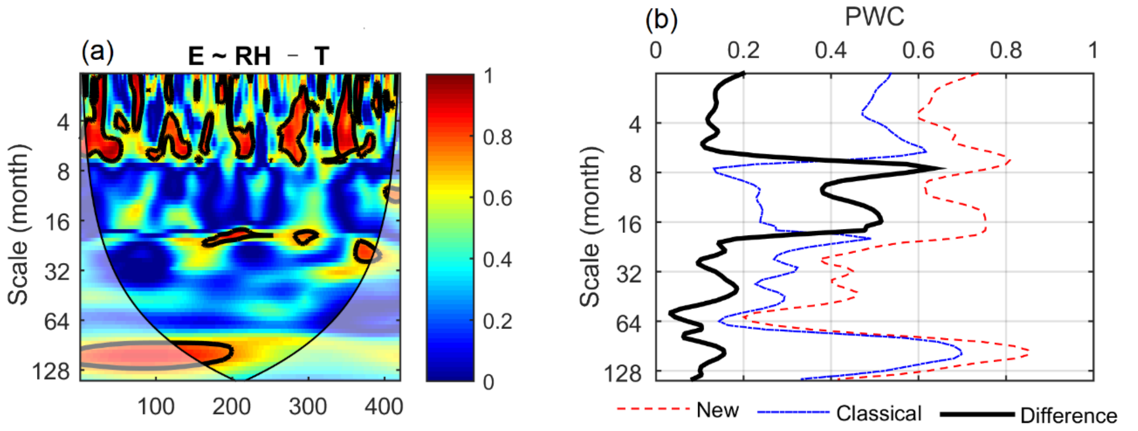

Figure 5.

Partial wavelet coherency (PWC) between evaporation (E) and relative humidity (RH) after excluding the effect of mean temperature (T) using the classical method (a) and differences in PWC between the new method and classical method as a function of scale (b).

Compared with the Mihanović et al. (2009) method, inclusion of phase information in the new PWC is another advantage of this method. This is because phase information is directly related to the type of correlation, i.e., in-phase and out-of-phase indicating positive and negative correlation, respectively. Different types of correlations were usually found at different locations and scales (Hu et al., 2017), the inclusion of phase information will be useful to understand the differences in associated mechanisms or processes at different locations and scales. In addition, the inclusion of phase information will allow us to detect the changes in not only degree of correlation (i.e., coherence) but also the type of correlation after excluding the effect of other variables. For example, E and RH were positively correlated at the 1-year cycle (8–16 months) from year 1979 to 1995 because higher evaporation usually occurs in summer when high T coincides with high RH as influenced by the monsoon climate in the area where data were collected (Fig. S4 of Sect. S3 in the Supplement). Interestingly, after excluding the effect of T, E was negatively correlated with RH at the scale of 1-year as we expect (Fig. 3d).

Moreover, our new PWC method can be used to deal with situations with more than one excluding variable, which is a knowledge gap. When multiple variables are correlated with both the predictor and responsible variables, the correlations between predictor and responsible variables may be misleading if the effects of all these multiple variable were not removed. For example, at the dominant scale (i.e., 1-year) of E variation, the effects of RH on E existed after excluding the effects of T or SH. However, their contrasting correlations (Fig. 3d-e) resulted in negligible effects of RH on E at this scale after the effects of all other variables were excluded (Fig. 4b). In this case, the dominant role of mean temperature in driving free water evaporation was proved at the 1-year cycle (Fig. 4a). This also further verifies the suitability of the Hargreaves model (only air temperature and incident solar radiation required) (Hargreaves, 1989) for estimating potential evapotranspiration on the Chinese Loess Plateau (Li, 2012).

5.2 Weaknesses

Similar to the Mihanović et al. (2009) method, the new method has the risk to produce spurious high correlations after excluding the effect from other variables. Take the artificial dataset for example, at a scale of 32, PWC values between y and y_2 after excluding y_4 are not significant, but relatively high, partly because of small octaves per scale (octave refers to the scaled distance between two scales with one scale being twice or half of the other, default of 1/12). This spurious unexpected high PWC is caused by low values in both the numerator (partly associated with the low coherence between response y and predictor variables y_2 at scale of 32) and denominator (partly associated with the high coherence between response y and excluding variable y_4 at a scale of 32) in Eq. (9). The same problem also exists in the Mihanović et al. (2009) method (Fig. S5 of Sect. S3 in the Supplement). So, caution should be taken to interpret those results. However, it seems that the domain with spurious correlation calculated by the new method is very limited and it is located mainly outside of the cones of influence. Anyway, the unexpected results can be easily ruled out with knowledge of BWC between response and predictor variables. We would expect that the correlation between two variables should not be increased after the effects of excluding variables are removed. Therefore, BWC analysis is suggested for better interpretation of the PWC results.

Similar to BWC and MWC, the confidence level of PWC calculated from the Monte Carlo is based on a single hypothesis which is tested one by one. But in reality, confidence level of PWC values at all locations and scales needs to be tested simultaneously. Therefore, the significance test suffers from the multiple-testing problem (Schaefli et al., 2007; Schulte et al., 2015). The new method may benefit from a better statistical significance testing method. Options for multiple-testing can be the Bonferroni adjusted *p* test (Westfall and Young, 1993) or false discovery rate (Abramovich and Benjamini, 1996; Shen et al., 2002) which is less stringent than the former. "

Comment #8:

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.*

Response #8:
As we replied above, we think 'the new method produces more accurate coherence' is justified by considering both the theoretical differences and the example of real data (Figure RC3) above. The benefits of including phase information and excluding several variables will be discussed in the new discussion section as we explained in Response #7.
Yes, a large part from the conclusions part will be moved to the Discussion section as shown in Response #7.

Comment #9:

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.*

Response #9:
We have provided the Matlab code to the figshare (https://figshare.com/s/bc97956f43fe5734c784). Meanwhile, we have also put the updated codes for multiple wavelet coherence (MWC) which is necessary for calculating PWC in the same repository. We have improved the calculation time for MWC.

Minor points

Comment #10:
L. 31: please explain what you mean by 'time and space localization'.
Response #10:
We will add an example to show the localization "For example, time series of air temperature usually fluctuates periodically at different scales (e.g., daily and yearly), but abrupt changes (e.g., extremely high or low) in air temperature may occur at a certain instant of time as a result of extreme weather and climate events (e.g., heat and rain).".

Comment #11:

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.

Response #11:
We will change "Among which" to "Among these wavelet methods".

We're sorry that we are not sure we understood this comment. But we end up with the wide application of multiple wavelet coherence (MWC) method in the previous graph, and the next paragraph we start with what the MWC application has told us. Namely more predictor variables does not necessarily explain more variations in the response variable because predictor variables are usually cross-correlated. Because of the same reason, bivariate relationships can be misleading. Then we call the need to develop partial wavelet coherence (PWC). Now in the revision, we will put them in the same paragraph.

Comment #12:

L. 48: what do you mean by 'this issue'?

Response #12:
We mean "the misleading relationships resulting from the interdependence between other variables and both predictor and response variables".

Comment #13:
L. 50: what kind of scales? Temporal or spatial?

Response #13:
We mean either temporal or spatial scales depending on if the dataset are time series or spatial series. For avoiding repeatedly saying this, we will clarify this at the first time by adding "For convenience, we will mainly refer to location and scale irrespective of spatial or temporal series unless otherwise mentioned.".

Comment #14:
L. 53-54: would combine greenhouse gas emissions and climate in one category.

Response #14:
Actually we mean different things. We mean precipitation by climate, so we will change climate to meteorological science for avoiding confusing.

Comment #15:
L. 61: information 'which will allow to….'

Response #15:
We will add "which will allow to better understand the magnitude and type of bivariate relationships after removing effects from all other interdependent variables"

Comment #16:
L. 61: what do you mean by 'analogy' in this context. I think that rephrasing may be required.

Response #16:
We will change "in analogy with" simply to "from".

Comment #17:

L. 62: Be specific with what you mean by 'it': 'the proposed method'.

Response #17:
We will change it to "The proposed method".

Comment #18:
L. 76: Please explain to the reader what you mean by 'scale' and 'location'.

Response #18:
Scale and location for spatial series correspond to frequency (periodicity) and time, respectively. "For convenience, we will mainly refer to location and scale irrespective of spatial or temporal series unless otherwise mentioned."

Comment #19:
L. 99: same for 'phase angle'.

Response #19:
We will add its explanation in the bracket as "(i.e., angle between two complex numbers)".

Comment #20:
L. 184-185: can in my opinion be removed.

Response #20:
We will remove this sentence.

Comment #21:
L. 191: what does data refer to? Soil water content?

Response #21:
It refers to soil water datasets. Now removed as you suggested.

Comment #22:
L. 214: 'significance band'.

Response #22:
We will change it to significance band.

Comment #23:
L. 215-216: is this statement underlined by any analysis performed?

Response #23:
Yes. The number is obtained from calculation.

Comment #24:
L. 247: what is the purpose of replacing half of the time series by 0?

Response #24:
As we highlighted in the methodology section, "second half of the original series of y2 (or z2) are replaced by 0 to simulate abrupt changes (i.e., transient and localized feature) of the spatial series".

Comment #25:
L. 261-263: Which feature in the plots actually indicates these 'abrupt changes'?

Response #25:
The abrupt changes were captured by the abrupt transition from coherence of 0 to coherence of 1 as shown in figure 2a and 2e (top 2 at the left hand side of figure 2 in the original submission).

Comment #26:
L. 266: I can only see one wavelet band of high significance in Figure 3. Where is the second one you mention here?

Response #26:
We did not show the results here, but it was shown in Fig. 2 of our previous paper (Hu and Si, 2016). For this reason, the citation of "(Hu and Si, 2016)" will be added here.
Comment #27:
L. 298: introduce term 'octave'.

Response #27:
We will add "octave refers to the scaled distance between two scales with one scale being twice or half of the other."

Comment #28:
L. 363-366: would move this sentence to discussion section.
Response #28:
Yes, we will move this sentence to the discussion section.

Thanks again for your constructive comment.