Technical Note: Multiple wavelet coherence for untangling scale-specific and localized multivariate relationships in geosciences

Wei Hu[2,3] and Bing Cheng Si[1,3]

[1]*College of Hydraulic and Architectural Engineering, Northwest A&F University, Yangling 712100, China*

[2]*New Zealand Institute for Plant & Food Research Limited, Private Bag 4704, Christchurch 8140, New Zealand*

[3]*University of Saskatchewan, Department of Soil Science, Saskatoon, SK S7N 5A8, Canada*

Correspondence to: Wei Hu (wei.hu@plantandfood.co.nz) and Bing Cheng Si (bing.si@usask.ca)

Abstract

The scale-specific and localized bivariate relationships in geosciences can be revealed using bivariate wavelet coherence. The objective of this study was to develop a multiple wavelet coherence method for examining scale-specific and localized multivariate relationships. Stationary and non-stationary artificial datasets, generated with the response variable as the summation of five predictor variables (cosine waves) with different scales, were used to test the new method. Comparisons were also conducted using existing multivariate methods, including multiple spectral coherence and multivariate empirical mode decomposition (MEMD). Results show that multiple spectral coherence is unable to identify localized multivariate relationships, and underestimates the scale-specific multivariate relationships for non-stationary processes. The MEMD method was able to separate all variables into components at

the same set of scales, revealing scale-specific relationships when combined with multiple correlation coefficients, but has the same weakness as multiple spectral coherence. However, multiple wavelet coherences are able to identify scale-specific and localized multivariate relationships, as they are close to 1 at multiple scales and locations corresponding to those of predictor variables. Therefore, multiple wavelet coherence outperforms other common multivariate methods. Multiple wavelet coherence was applied to a real dataset and revealed the optimal combination of factors for explaining temporal variation of free water evaporation at Changwu site in China at multiple scale-location domains. Matlab codes for multiple wavelet coherence were developed and are provided in the supplement.

1. Introduction

Geoscience data such as topography, climate, and ocean waves usually present cyclic patterns, with high-frequency (small-scale) processes being superimposed on low-frequency (large-scale) processes (Si, 2008). More often than not, geoscience data are transient, consisting of a variety of frequency regimes that may be localized in space or time (Torrence and Compo, 1998; Si and Zeleke, 2005; Graf et al., 2014). The transient characteristics exist widely in non-stationary processes, but also sometimes occur in stationary processes (Feldstein, 2000). The wavelet method is a common tool for detecting multi-scale and localized features of transient processes in geosciences. Bivariate wavelet coherency has been widely used for untangling scale-specific and localized relationships for transient processes in areas including

43 geophysics (Lakshmi et al., 2004; Müller et al., 2008), hydrology (Labat et al., 2005;
44 Das and Mohanty, 2008; Tang and Piechota, 2009; Carey et al., 2013; Graf et al.,
45 2014), soil science (Si and Zeleke, 2005; Biswas and Si, 2011), meteorology
46 (Torrence and Compo, 1998), and ecology (Polansky et al., 2010). This method,
47 however, is limited to two variables. Processes in geosciences are usually complex
48 and may be affected by more than two environmental factors. A method is needed for
49 analyzing multivariate (>2 variables) and localized relationships at multiple scales.

50 Several methods have been used for characterizing multivariate relationships. For
51 example, multiple spectral coherence (MSC) has been used to explore the
52 scale-specific relationships between soil saturated hydraulic conductivity (K_s) and
53 multiple soil physical properties (Koopmans, 1974; Si, 2008), but requires a stationary
54 data series, which is rare in geosciences. Multivariate empirical mode decomposition
55 (MEMD), a data-driven method, decomposes each variable into different components
56 (intrinsic mode functions (IMFs)) with each IMF corresponding to a “common scale”
57 inherent in multiple variables (Rehman and Mandic, 2010). The MEMD method is
58 meritorious due to its ability to deal with both transient and nonlinear systems. The
59 combination of the squared multiple correlation coefficient and MEMD (MCC_{memd})
60 has been used to explore the multivariate control of soil water content and K_s at
61 multiple scales (Hu and Si, 2013; She et al., 2013, 2015; Hu et al., 2014). However,
62 the sum of variances from different components did not typically equal the total
63 variance of the original series, which may produce misleading MCC_{memd} results. In
64 addition, multivariate relationships in geosciences are most likely to change with time

or space due to the transient nature of the processes involved. However, localized multivariate relationships are not available using any of the existing multivariate methods. Therefore, extending the wavelet coherence from two variables to multiple variables is required .

An attempt to extend wavelet coherence from two to three variables has been made by Mihanović et al. (2009). Their method was also applied later in the marine sciences (Ng and Chan, 2012a, b). Limitations arise when using the trivariate wavelet coherence: first, only two predictor variables are considered; second, the two predictor variables must be orthogonal. Otherwise, extremely high or low (spurious) coherence (>>1 or <0) may be produced. This spuriousness is inconsistent with the definition of coherence, which may limit the application of this method in geosciences where environmental variables are usually cross-correlated. Therefore, a robust method for calculating MWC, which produces coherence in the closed interval of [0, 1], is needed.

The objective of this paper is to develop an MWC that applies to cases where there are multiple environmental variables, of which may be cross-correlated. This method is first tested with artificial datasets to demonstrate its advantages over existing multivariate methods. The superiority of the new method over others can be assessed by determining whether the known major features of the artificial data are demonstrated by these methods. The new method is then applied to a temporal series of evaporation (E) from free water surface and meteorological factors at Changwu site in Shaanxi, China.

2. Theory

Bivariate wavelet coherence can be understood as the traditional correlation coefficient localized in the scale-location domain (Grinsted et al., 2004). Just as correlation coefficients can be extended from two variables to multiple (>2) variables, wavelet coherence between two variables may also be extended to multiple variables. Similar to bivariate wavelet coherence, MWC is based on a series of auto- and cross-wavelet power spectra, at different scales and spatial (or temporal) locations, for the response variable and all predictor variables.

Following Koopman (1974), a matrix representation of the smoothed auto- and cross-wavelet power spectra for multiple predictor variables X ($X = \{X1, X2, \ldots, Xq\}$) can be written as

$$\overleftrightarrow{W}^{X,X}(s,\tau) = \begin{bmatrix} \overleftrightarrow{W}^{X1,X1}(s,\tau) & \overleftrightarrow{W}^{X1,X2}(s,\tau) & \cdots & \overleftrightarrow{W}^{X1,Xq}(s,\tau) \\ \overleftrightarrow{W}^{X2,X1}(s,\tau) & \overleftrightarrow{W}^{X2,X2}(s,\tau) & \cdots & \overleftrightarrow{W}^{X2,Xq}(s,\tau) \\ \vdots & \vdots & & \vdots \\ \overleftrightarrow{W}^{Xq,X1}(s,\tau) & \overleftrightarrow{W}^{Xq,X2}(s,\tau) & \cdots & \overleftrightarrow{W}^{Xq,Xq}(s,\tau) \end{bmatrix}, \quad (1)$$

where $\overleftrightarrow{W}^{Xi,Xj}(s,\tau)$ is the smoothed auto-wavelet power spectra (when $i=j$) or cross-wavelet power spectra (when $i \neq j$) at scale s and spatial (or temporal) location τ, respectively. For the detailed calculation of smoothed auto- and cross-wavelet power spectra, see Supplement, Sect. S1.

The matrix of smoothed cross wavelet power spectra between response variable Y and predictor variables Xi can be defined as

$$\overleftrightarrow{W}^{Y,X}(s,\tau) = \begin{bmatrix} \overleftrightarrow{W}^{Y,X1}(s,\tau) & \overleftrightarrow{W}^{Y,X2}(s,\tau) & \cdots & \overleftrightarrow{W}^{Y,Xq}(s,\tau) \end{bmatrix}, \quad (2)$$

where $\overleftrightarrow{W}^{Y,Xi}(s,\tau)$ is the smoothed cross-wavelet power spectra between *Y* and *Xi* at scale *s* and spatial (or temporal) location τ.

The smoothed wavelet power spectrum of response variable *Y* is $\overleftrightarrow{W}^{Y,Y}(s,\tau)$.

Following Koopmans (1974), the MWC at scale *s* and location τ, ${\rho_m}^2(s,\tau)$, can be written as

$$ {\rho_m}^2(s,\tau) = \frac{\overleftrightarrow{W}^{Y,X}(s,\tau)\overleftrightarrow{W}^{X,X}(s,\tau)^{-1}\overline{\overleftrightarrow{W}^{Y,X}(s,\tau)}}{\overleftrightarrow{W}^{Y,Y}(s,\tau)}. \quad (3) $$

When only one predictor variable (e.g., *X1*) is included in *X*, Eq. (3) is the equation for bivariate wavelet coherence, $\rho_b^2(s,\tau)$, which can be expressed as (Torrence and Webster, 1999; Grinsted et al., 2004):

$$ \rho_b^2(s,\tau) = \frac{\overleftrightarrow{W}^{Y,X1}(s,\tau)\overline{\overleftrightarrow{W}^{Y,X1}(s,\tau)}}{\overleftrightarrow{W}^{X1,X1}(s,\tau)\overleftrightarrow{W}^{Y,Y}(s,\tau)}. \quad (4) $$

Therefore, bivariate wavelet coherence is consistent with multiple wavelet coherence if only one predictor variable is included. In addition, the wavelet phase between a response variable (*Y*) and a predictor variable (*X1*) is

$$ \phi(s,\tau) = \tan^{-1}\left(\mathrm{Im}\left(W^{Y,X1}(s,\tau)\right)/\mathrm{Re}\left(W^{Y,X1}(s,\tau)\right)\right), \quad (5) $$

where Im and Re denote the imaginary and real part of $W^{Y,X1}(s,\tau)$, respectively. Note that the phase information between a response variable *Y* and multiple predictor variables *X* cannot be obtained.

Multiple wavelet coherence at the 95% confidence level is calculated using the Monte Carlo method (Grinsted et al., 2004). Surrogate spatial series (i.e., red noise) of all variables are generated with a Monte Carlo simulation based on their first-order autocorrelation coefficient (AR1). The MWC at each scale and location is calculated

using the simulated spatial series. This is repeated an adequate number of times (e.g., 1000) (Grinsted et al., 2004). At each scale, MWCs at all locations outside the cones of influence, from all simulations are ranked in ascending order. The value at the 95th percentile represents the 95% confidence level for the MWC at that scale. The Matlab codes and user manual document for calculating MWC and significance level are provided in the Supplement (Sect. S2–S4).

3. Data and analysis

3.1 Artificial data for method test

The method is tested using a stationary and non-stationary artificial dataset, generated following Yan and Gao (2007). The response variable (y for the stationary case and z for the non-stationary case) encompasses five cosine waves (y1 to y5 for the stationary case and z1 to z5 for the non-stationary case), with different dimensionless scales (Fig. 1). For the stationary case, $y1=\cos(2\pi x/4)$, $y2=\cos(2\pi x/8)$, $y3=\cos(2\pi x/16)$, $y4=\cos(2\pi x/32)$, and $y5=\cos(2\pi x/64)$, where x=0, 1, 2, …, 255. There is one regular cycle every 4, 8, 16, 32, and 64 locations, representing dimensionless scales of 4, 8, 16, 32, and 64 for y1, y2, y3, y4, and y5, respectively (Fig. 1a). The regular cycles make each predictor and response series stationary. For the non-stationary case, $z1=\cos(500\pi(x/1000)^{0.5})$, $z2=\cos(250\pi(x/1000)^{0.5})$, $z3=\cos(125\pi(x/1000)^{0.5})$, $z4=\cos(62.5\pi(x/1000)^{0.5})$, and $z5=\cos(31.25\pi(x/1000)^{0.5})$, where x=0, 1, 2, …, 255. The equation containing the square root of the location term results in the gradual change in frequency (scale), with the greatest dimensionless

148 scales of 4, 8, 16, 32, and 64 at the right hand side for z1, z2, z3, z4, and z5,
149 respectively (Fig. 1b). The average scales for these predictor variables are 3, 5, 9, 17,
150 and 32, respectively. The location-varying scales make each predictor and response
151 variable non-stationary.

152 For both the stationary and non-stationary series, the variance of the response
153 variable is 2.5. The predictor variables, each with a variance of 0.5, are orthogonal to
154 each other, and contribute equally to the total variance of the response variable. The
155 cosine-like artificial datasets mimic many time series such as seismic signals,
156 turbulence, air temperature, precipitation, hydrologic fluxes, and the El
157 Niño-Southern Oscillation. They also mimic geoscientific spatial series such as ocean
158 waves, seafloor bathymetry, land surface topography, and soil water content along a
159 hummocky landscape. Therefore, they are representative of a geoscience data series
160 and are suitable for testing the new method.

161 Multiple wavelet coherence between the response variable y (or z) and two (y2 and
162 y4, or z2 and z4) or three (y2, y3, and y4, or z2, z3, and z4) predictor variables were
163 calculated. The advantage of the artificial data is that the known scale- and localized
164 features for all variables, and the known relationships between the response and each
165 predictor variable, are exact. By definition, the coherence is 1 at scales corresponding
166 to those of the included predictor variables, and 0 at other scales.

167 To demonstrate the advantages of MWC in dealing with abrupt changes (a type of
168 transient and localized feature), the second half of the original series of y2 (or z2) or
169 y4 (or z4) are replaced by 0, and MWC between the response variable and new set of

predictor variables is calculated. We anticipate that the coherence changes from 1 to 0 at the location where the new predictor variable becomes 0.

Predictor variables may not be as regular as that shown in Fig. 1, and may also be cross-correlated to one another. For these reasons, zero-mean white noises with standard deviations of 0.3, 1, and 4 are added to the predictor variables of y2 (or z2) and y4 (or z4). The resulting noised series have correlation coefficients of 0.9, 0.5, and 0.1, respectively, with their original predictor variable. Therefore, we will refer to them as weakly, moderately, and highly noised series, respectively. Multiple wavelet coherences between the response variable and different predictor variables (original and noised series) are calculated to demonstrate the performance of MWC when noised or correlated predictor variables are involved. Only the non-stationary case will be demonstrated, because the performances of MWC for stationary and non-stationary cases are similar.

The MWC is compared to the MSC (Koopmans, 1974; Si, 2008) and MCC_{memd} (Hu and Si, 2013), which are widely used for spatial or temporal series analysis in different disciplines. The advantages of the new method over these two methods will be demonstrated mainly in terms of relationships between response and predictor variables at various scales of the response variable. The MSC is calculated based on the calculated auto- and cross- power spectra, using an equation similar to Eq. (3). The detailed introduction of this method can be found in Si (2008). For the calculation of MCC_{memd}, a set of response and predictor variables form a multivariate data series for MEMD. The MEMD is a data driven method and has the ability to align “common

192 scales" present within multivariate data. Please refer to Rehman and Mandic (2010)
193 and Hu and Si (2013) for the MEMD analysis, and the website
194 (http://www.commsp.ee.ic.ac.uk/~mandic/research/emd.htm) for the related Matlab
195 codes. The original series of response and predictor variables can be decomposed by
196 the MEMD, into different components (IMFs) with varying scales. For IMFs at the
197 same scale, multiple stepwise regressions are conducted between response and
198 predictor variables, and the multiple correlation coefficients for each scale-specific
199 IMF are calculated.

200 **3.2 Real data for application**

201 Daily evaporation (*E)* from free water surfaces in an E601 evaporation pan (pan
202 diameter of 61.8 cm), and other meteorological factors (i.e., relative humidity, mean
203 temperature, sun hours, and wind speed) were collected from January 1, 1979 to
204 December 31, 2013, at Changwu site in Shaanxi, China. The Changwu site is a
205 transition area between semi-arid and subhumid climates, where agricultural
206 productivity is mainly limited by water. Monthly averages of all variables were used
207 in this study, because we are mainly interested in seasonal and inter-annual variability.

208 # 4. Results and discussion

209 **4.1 MWC with orthogonal predictor variables**

210 For the stationary data, there are two narrow, horizontal bands (red color)
211 representing an MWC value of around 1, at the respective scales of 8 and 32 for all

212 locations (Fig. 2a). These two bands also correspond to the scales of 8 and 32,
213 respectively, for the two predictor variables. When an additional predictor variable
214 with the scale of 16 is introduced, a wide band appears from 6 to 40, signifying that
215 the MWC equals approximately 1 at all locations, at the scales of 8, 16, and 32. As
216 anticipated, when all five predictor variables with scales ranging from 4 to 64 are
217 included, coherence values of close to 1 are found in the whole scale-location domain
218 (data not shown).

219 The application of MWC to the non-stationary datasets shows that the scales with
220 significant MWC values gradually increase as distance increases. This increase in the
221 scales is due to the non-stationarity of the variables (Fig. 2b). For example, when
222 predictor variables of z2 and z4 are included, scales of the two bands corresponding to
223 MWC around 1 increase from 4 to 8 and from 8 to 32, respectively. Furthermore, as
224 expected, for only one predictor variable (stationary and non-stationary), MWC
225 reduces to bivariate wavelet coherence; there is only one band of coherence around 1,
226 which corresponds to the scale of that predictor variable (data not shown). Note that
227 the significant MWC values for both stationary and non-stationary cases are not
228 exactly 1 at all scales or locations, due to the smoothing effect along both scales and
229 locations. However, the mean MWC values of the significant bands are very high (i.e.,
230 0.94–1.00), and the MWC values at the centre of the significant band are 1, which
231 corresponds to the exact scale of a predictor variable.

232 When the point values in the second half of the data series of a predictor variable
233 are replaced by 0, the MWC values in that half of the data series are almost 0 at scales

corresponding to that predictor variable (Fig. 3). For the stationary case, when the point values in the second half of the data series of predictor variable y2 (or y4) are replaced by 0, the MWC values are around 1 at the scale of 8 (or 32) in the first half of the transect, and 0 in the second half (Fig. 3a). Similar results are also found for the non-stationary case (Fig. 3b). This is expected because the constant series of 0 is not correlated to the response variables at any scale. Much like bivariate wavelet coherence, the MWC method is able to detect abrupt changes in the data series, and has the advantages of dealing with localized multivariate relationships.

4.2 MWC with noised and correlated predictor variables

When z2 and a noised series derived from z2 are included as predictor variables, there is only one band of coherence close to 1 at scales corresponding to z2, irrespective of the correlation between z2 and a noised series of z2 (Fig. 4a). When z2 and a noised series of z4 are included as predictor variables, the coherence depends on the degree of the noise (Fig. 4b). For weakly noised series, there are two bands of coherence of around 1, corresponding to the scales of z2 and z4, respectively. The percentage area of significant coherence (PASC) is 23%, which equals that of when z2 and z4 are included. With the increasing magnitude of noise, the coherence and corresponding PASC at the scales corresponding to z4 decrease. When z2 and a strongly noised series of z4 are considered, the band of coherence around 1, at scales corresponding to z4, disappears.

The inclusion of a third noised z4 variable substantially increases the area with high

coherence (in red) as compared to the case when only z2 and z4 are included (Fig. 4c). This indicates that MWC will increase as the number of predictor variables increases, with the highest coherence less or equal to 1, irrespective of the number of predictor variables. However, the area of significant coherence may not necessarily increase because of the simultaneously increased statistical significance threshold (Ng and Chan, 2012a). In fact, the PASC values for three predictor variables (19–20%) are lower than those of the two predictor variables (23%). This indicates that, in this case, two predictor variables are better than three in terms of explaining the variations of the response variable. This occurs because the variance of the response variable that is explained by the noised variable is already accounted for by other variables. Therefore, only an additional variable that can independently explain a fair amount of variance could contribute significantly to explaining variations of a response variable (Fig. 4b). This may also explain why there is only one band of coherence around 1 at scales corresponding to z2, when z2 and a noised series of z2 are included (Fig. 4a). This information is helpful in choosing predictor variables for developing scale-specific predictions, especially when predictor variables are correlated.

4.3 Comparison with other multivariate methods

4.3.1 MSC

The MSC as a function of scale is shown in Fig. 5a. For the stationary case, when y2 and y4 are included as predictor variables, there are two plateaus centered at the scales of 8 and 28, representing a coherence of 1. As expected, when an additional

predictor variable y3 is added, the corresponding scale of 16 also shows coherence of 1. The MSC produces similar scale-specific relationships, as MWC does for a stationary dataset, with exception given to the centered scale (i.e., 28) with a coherence of 1. Here, the scale with a unity MSC deviates from the expected value (i.e., 32) for predictor variable y4. For the non-stationary case, however, the MSC is much lower than 1 for the predictor variables of z2 and z4; an MSC of 1 is present only at the scale of 8 when an additional predictor variable z3 is added. Obviously, the MSC underestimates the multivariate relationships, and is not suitable for non-stationary processes (Si, 2008) due to its inability to deal with localized features. The MSC at a specific scale provides the average of multivariate relationships, across all locations. Due to the change in scale of a predictor variable with location for the non-stationary case, the MSC deviates greatly from 1.

The MSC decreases at scales when the second half of the included predictor variable series are replaced by 0 for both the stationary and non-stationary series (Fig. 5b). For example, when the second half of the y4 series in the stationary case are replaced by 0, the MSC at scales of around 32 decreases from 1 to 0.52. Although the MSC, throughout the second half of the series, can detect the decrease of coherence at the scales corresponding to the 0 values, the exact locations for the decrease cannot be identified. In fact, the coherence decreases only in the second half of the series, and does not change in the first half of the series. The location for the decrease can be easily identified by the MWC, but not by MSC. This further demonstrates the inability of the MSC to deal with localized features.

298 4.3.2 MCC_{memd}

299 Five intrinsic mode functions (IMFs) with non-negligible variance, are obtained for
300 multivariate data series. While the obtained scales for the response variable y are in
301 agreement with the true scales for the stationary case, the obtained scales (i.e., 3, 6, 11,
302 21, and 43) for the response variable z deviate slightly from the average scales for the
303 non-stationary case. For the response variable, the contribution of IMFs to the total
304 variance generally decreases (20% to 13% for stationary, and 27% to 11% for
305 non-stationary) from IMF1 to IMF5. This disagrees with the fact that each scale
306 contributes equally (i.e., 20%) to the total variance. In addition, the sum of variances
307 over all IMFs for each variable is less than 100% (ranging from 84% to 93%),
308 indicating that MEMD cannot capture all the variances. For the detailed results of
309 MEMD, see Supplement, Sect. S5.

310 The MCC_{memd} as a function of scale, is shown in Fig. 6a. For the stationary case,
311 when predictor variables of y2 and y4 are included, the MCC_{memd} values are 0.98 and
312 0.93, respectively, at scales corresponding to those of y2 and y4. When a predictor
313 variable of y3 is included, the MCC_{memd} values are 1.00, 1.00, and 0.96, respectively,
314 at scales corresponding to those of y2, y3, and y4. For the non-stationary, two
315 predictor variable case, the corresponding MCC_{memd} values are 0.80 and 0.85. For the
316 non-stationary, three predictor variable case, the corresponding MCC_{memd} values are
317 0.95, 0.99, and 0.91, respectively. Therefore, the MCC_{memd} can be used to determine
318 the scale-specific multivariate relationships. Similar to MSC, however, the MCC_{memd}
319 underestimates the multivariate relationships, especially for the non-stationary case

320 with less predictor variables. On the contrary, the MCC_{memd} also overestimates the
321 multivariate relationships. For example, when considering only predictor variables
322 corresponding to scales of 8, 16, and 32, the MCC_{memd} value for the stationary case is
323 0.47 at the scale of 64. This deviates much from the expected MCC_{memd} value of 0
324 (Fig. 6a). The possible underestimation and overestimation by the MCC_{memd} may
325 come from the decomposition errors inherent in the MEMD algorithm (Rehman and
326 Mandic, 2010).

327 Similar to MSC, the localized multivariate relationships cannot be obtained from
328 MCC_{memd}. This can be better explained by the decrease of MCC_{memd} when half of the
329 series of the predictor variables are replaced by 0 (Fig. 6b). Take the stationary case
330 for example, the MCC_{memd} values at the scales corresponding to y2 and y4 decrease
331 from 0.98 to 0.49, and from 0.93 to 0.62, respectively, when the second half of the y2
332 and y4 series are replaced by 0.

333 As explained above, the MWC has advantages in untangling localized multivariate
334 relationships as compared to the common multivariate methods. It is important to
335 reveal the multivariate relationships which vary with time or space, that are associated
336 with different processes. For example, discharge usually occurs on knolls, while
337 recharge usually occurs in neighboring depressions (Gates et al., 2011). Therefore, the
338 controlling factors of soil water storage may vary with the land element characteristics
339 of a location. Local controls may be more important on knolls, while non-local
340 controls may be more important in depressions (Grayson et al., 1997). In a temporal
341 domain, vegetation transpiration contributes more to the evapotranspiration in the

342 growing seasons, which may result in the changes of environmental factors explaining
343 temporal variations of evapotranspiration in different seasons.

344 **4.4 Application of the MWC**

345 Each meteorological factor was significantly correlated to *E*, but the dominant
346 factors explaining variations in *E* differed with scale. For example, the relative
347 humidity was the dominating factor at small (2–8 months) and large (>32 months)
348 scales, while temperature was the dominating factor at the medium (8–32 months)
349 scales. Overall, the relative humidity corresponded to the greatest mean MWC (0.62)
350 and PASC value (40%) at multiple scale-location domains. For the detailed
351 relationships between *E* and each factor, see Supplement, Sect. S6.

352 The MWC analysis shows that the combination of relative humidity and mean
353 temperature produced the greatest mean MWC (0.82) and PASC (49%) among all
354 two-factor cases. This suggested that relative humidity and mean temperature were
355 the most appropriate factors for explaining variations in *E* at multiple scale-location
356 domains (Fig. 7a). However, adding an additional factor such as sun hours, which was
357 the best among all three-factor cases, increased the average coherence (0.91), but
358 slightly decreased the PASC to 48% (Fig. 7b). This indicated that sun hours was not
359 significantly different from red noise in explaining additional variation in *E*. Similar
360 results were found when the wind speed was added. This occurs because most areas
361 with significant coherence between *E* and sun hours or wind speed were a subset of
362 areas with significant coherence between *E* and relative humidity or mean

363 temperature (see Supplement, Sect. S3). Therefore, relative humidity and mean
364 temperature were adequate for explaining the temporal variation of *E* at various scales
365 at this site. This was consistent with Li et al. (2012), who indicated that relative
366 humidity and mean temperature were the two main contributors to the temporal
367 change of potential evapotranspiration on the Chinese Loess Plateau.

368 **5. Conclusions**

369 Multiple wavelet coherence was developed to determine scale-specific and
370 localized multivariate relationships in geosciences. The new method was tested and
371 compared with existing multivariate methods, using an artificial dataset. The new
372 method can be used to determine the proportion of the variance of a response variable
373 that is explained by predictor variables, at a specific scale and location (spatially or
374 temporally). As compared with bivariate wavelet coherence, more variation may be
375 explained at multiple scale-location domains by the MWC. Including more variables
376 is only beneficial if the variables are not strongly cross-correlated, and can
377 independently explain a fair amount of variability in a response variable. Therefore,
378 the best combinations of variables that explain multivariate, spatial or temporal
379 variability at multiple scales can be determined. This is important for optimizing
380 variables to develop scale-specific prediction.

381 The MSC and MCC_{memd} can determine multivariate relationships at multiple scales,
382 but localized multivariate relationships are not available. Furthermore, both MSC and
383 MCC_{memd} are likely to underestimate the degree of multivariate relationships for

384 non-stationary processes. In addition, the performance of MCC_{memd} relies on the
385 performance of MEMD, which needs further development. Application of the MWC
386 into the real dataset indicates that the combination of relative humidity and mean
387 temperature are the optimal factors that can be used to explain temporal variations of
388 *E* at the Changwu site in China.

389 Limitations of the new method also exist. Theoretically, any number of predictor
390 variables can be included in the multiple wavelet analysis. However, the statistical
391 significance threshold usually increases with the number of predictor variables
392 (Grinsted et al., 2004; Ng and Chan, 2012a). In addition, the inclusion of too many
393 predictor variables may result in the statistical significance threshold at particular
394 wavelet scales (e.g., the lowest and largest scales) to approach unity. This would
395 restrict the availability of statistical information. Furthermore, similar to bivariate
396 wavelet analysis, the new method also suffers from the multiple-testing problem
397 (Maraun and Kurths, 2004; Maraun et al., 2007; Schulte et al., 2015; Schulte, 2016).
398 Therefore, a more robust statistical significance testing method may be beneficial to
399 the new method.

400 In summary, multiple wavelet coherence has advantages over existing multivariate
401 methods, and provides an effective vehicle for untangling complex spatial or temporal
402 variability for multiple controlling factors at multiple scales and locations. It may also
403 be used as a data-driven tool for modeling and predicting various processes in the area
404 of geosciences, such as precipitation, drought, soil water dynamics, stream flow, and
405 atmospheric circulation.

406 **Acknowledgements**

407 The Matlab codes for calculating multiple wavelet coherence are developed based on
408 the codes provided by A. Grinsted
409 (http://noc.ac.uk/using-science/crosswavelet-wavelet-coherence) and, together with
410 the user manual, are available in the Supplement (Sect. S2-S4). The project was
411 partially funded by the Natural Sciences and Engineering Research Council of Canada
412 (NSERC) and Agriculture Development Fund of Saskatchewan. We thank the two
413 anonymous reviewers for their constructive comments.

414 **References**

415 Biswas, A. and Si, B. C.: Identifying scale specific controls of soil water storage in a
416 hummocky landscape using wavelet coherency, Geoderma, 165, 50–59, doi:
417 10.1016/j.geoderma.2011.07.002, 2011.

418 Carey, S. K., Tetzlaff, D., Buttle, J., Laudon, H., McDonnell, J., McGuire, K., Seibert,
419 J., Soulsby, C., and Shanley, J.: Use of color maps and wavelet coherence to discern
420 seasonal and interannual climate influences on streamflow variability in northern
421 catchments, Water Resour. Res., 49, 6194–6207, doi: 10.1002/wrcr.20469, 2013.

422 Das, N.N. and Mohanty, B. P.: Temporal dynamics of PSR-based soil moisture across
423 spatial scales in an agricultural landscape during SMEX02: A wavelet approach,
424 Remote Sens. Environ., 112, 522–534, doi:10.1016/j.rse.2007.05.007, 2008.

425 Feldstein, S. B.: The timescale, power spectra, and climate noise properties of
426 teleconnection patterns, J. Clim., 13, 4430–4440,

427 doi:10.1175/15200442(2000)0132.0.CO, 2000.
428 Gates, J. B., Scanlon, B. R., Mu, X. M., and Zhang, L.: Impacts of soil conservation
429 on groundwater recharge in the semi-arid Loess Plateau, China, Hydrogeol. J., 19,
430 865–875, 2011.
431 Graf, A., Bogena, H. R., Drüe, C., Hardelauf, H., Pütz, T., Heinemann, G., and
432 Vereecken, H.: Spatiotemporal relations between water budget components and soil
433 water content in a forested tributary catchment, Water Resour. Res., 50, 4837–4857,
434 doi:10.1002/2013WR014516, 2014.
435 Grayson, R. B., Western, A. W., Chiew, F. H. S., and Blöschl, G.: Preferred states in
436 spatial soil moisture patterns: local and nonlocal controls, Water Resour. Res., 33,
437 2897–2908, doi: 10.1029/97WR02174, 1997.
438 Grinsted, A., Moore, J. C., and Jevrejeva, S.: Application of the cross wavelet
439 transform and wavelet coherence to geophysical time series, Nonlinear Proc. Geoph.,
440 11, 561–566, 2004.
441 Hu, W., Biswas, A., and Si, B. C.: Application of multivariate empirical mode
442 decomposition for revealing scale- and season-specific time stability of soil water
443 storage, Catena, 113, 377–385, doi:10.1016/j.catena.2013.08.024, 2014.
444 Hu, W. and Si, B. C.: Soil water prediction based on its scale-specific control using
445 multivariate empirical mode decomposition, Geoderma, 193–194,180–188, doi:
446 10.1016/j.geoderma.2012.10.021, 2013.
447 Koopmans, L. H.: The spectral analysis of time series, Academic Press, New York,
448 1974.

449 Labat, D.: Recent advances in wavelet analyses: Part I. A review of concepts, J.
450 Hydrol., 314, 275–288, doi: 10.1016/j.jhydrol.2005.04.003, 2005.
451 Lakshmi, V., Piechota, T., Narayan, U., and Tang, C. L.: Soil moisture as an indicator
452 of weather extremes, Geophys. Res. Lett., 31, L11401, doi:10.1029/2004GL019930,
453 2004.
454 Li, Z., Zheng, F. L., and Liu, W. Z.: Spatiotemporal characteristics of reference
455 evapotranspiration during 1961–2009 and its projected changes during 2011–2099 on
456 the Loess Plateau of China, Agric. For. Meteorol., 154-155, 147–155,
457 doi:10.1016/j.agrformet.2011.10.019, 2012.
458 Maraun, D. and Kurths, J.: Cross wavelet analysis: significance testing and pitfalls,
459 Nonlin. Processes Geophys., 11, 505–514, doi: 10.5194/npg-11-505-2004, 2004.
460 Maraun, D., Kurths, J., and Holschneider, M.: Nonstationary Gaussian processes in
461 wavelet domain: synthesis, estimation, and significance testing, Phys. Rev. E., 75,
462 016707, doi:10.1103/PhysRevE.75.016707, 2007.
463 Mihanović, H., Qrlić, M., and Pasrić, Z.: Diurnal thermocline oscillations driven by
464 tidal flow around an island in the Middle Adriatic, J. Marine Syst., 78, S157–S168,
465 doi: 10.1016/j.jmarsys.2009.01.021, 2009.
466 Müller, W. A., Frankignoul, C., and Chouaib, N.: Observed decadal tropical
467 Pacific–North Atlantic teleconnections, Geophys. Res. Lett., 35, L24810,
468 doi:10.1029/2008GL035901, 2008.
469 Ng, E. K. W. and Chan, J. C. L.: Geophysical applications of partial wavelet
470 coherence and multiple wavelet coherence, J. Atmos. Ocean. Tech., 29, 1845–1853,

471 doi: 10.1175/JTECH-D-12-00056.1, 2012a.
472 Ng, E. K. W. and Chan, J. C. L.: Interannual variations of tropical cyclone activity
473 over the north Indian Ocean, Int. J. Climatol., 32, 819–830, doi: 10.1002/joc.2304,
474 2012b.
475 Polansky, L., Wittemyer, G., Cross, P. C., Tambling, C. J., and Getz, W. M.: From
476 moonlight to movement and synchronized randomness: Fourier and wavelet analyses
477 of animal location time series data, Ecology, 91, 1506–1518, doi: 10.1890/08-2159.1,
478 2010.
479 Rehman, N. and Mandic, D. P.: Multivariate empirical mode decomposition, Proc. R.
480 Soc. A., 466, 1291–1302, doi:10.1098/rspa.2009.0502, 2010.
481 Schulte, J. A.: Cumulative areawise testing in wavelet analysis and its application to
482 geophysical time series, Nonlin. Processes Geophys., 23, 45–57,
483 doi:10.5194/npg-2345-2016, 2016.
484 Schulte, J. A., Duffy, C., and Najjar, R. G.: Geometric and topological approaches to
485 significance testing in wavelet analysis, Nonlin. Processes Geophys., 22, 139–156,
486 doi:10.5194/npg-22-139-2015, 2015.
487 She, D. L., Liu, D. D., Peng, S. Z., and Shao ,M. A.: Multiscale influences of soil
488 properties on soil water content distribution in a watershed on the Chinese Loess
489 Plateau, Soil Sci., 178, 530–539, doi: 10.1097/SS.0000000000000021, 2013.
490 She, D. L., Zheng, J. X., Shao, M. A., Timm, L. C., and Xia, Y. Q.: Multivariate
491 empirical mode decomposition derived multi-scale spatial relationships between
492 saturated hydraulic conductivity and basic soil properties, Clean-Soil Air Water, doi:

493 10.1002/clen.201400143, 2015.
494 Si, B. C.: Spatial scaling analyses of soil physical properties: A review of spectral and
495 wavelet methods, Vadose Zone J., 7, 547–562, doi: 10.2136/vzj2007.0040, 2008.
496 Si, B. C. and Zeleke, T. B.: Wavelet coherency analysis to relate saturated hydraulic
497 properties to soil physical properties, Water Resour. Res., 41, W11424,
498 doi:10.1029/2005WR004118, 2005.
499 Tang, C. L. and Piechota, T. C.: Spatial and temporal soil moisture and drought
500 variability in the Upper Colorado River Basin, J. Hydrol., 379, 122–135, doi:
501 10.1016/j.jhydrol.2009.09.052, 2009.
502 Torrence, C. and Compo, G. P.: A practical guide to wavelet analysis, Bull. Am.
503 Meteorol. Soc., 79, 61–78, doi: 10.1175/1520-0477(1998)079<0061:apgtwa>2.0.co;2,
504 1998.
505 Torrence, C. and Webster, P. J.: Interdecadal changes in the ENSO-monsoon system, J.
506 Clim., 12, 2679–2690, doi: 10.1175/1520-0442(1999)012<2679:ICITEM>2.0.CO;2,
507 1999.
508 Yan, R. and Gao, R. X.: A tour of the Hilbert–Huang transform: an empirical tool for
509 signal analysis, IEEE Instrum. Meas. Mag., 10, 40–45, doi:
510 10.1109/MIM.2007.4343566, 2007.

511 **Figure captions**

512 **Figure 1**. (a) Stationary and (b) non-stationary series of response variables (y for
513 stationary and z for non-stationary case) encompassing five cosine waves (y1 to y5
514 for stationary and z1 to z5 for non-stationary case) with different dimensionless
515 scales.

516 **Figure 2.** Multiple wavelet coherence (a) between response variable y and predictor
517 variables y2 and y4; (b) between response y and predictors y2, y3, and y4; (c)
518 between response z and predictors z2 and z4; and (d) between response z and
519 predictors z2, z3, and z4. The artificial data series (y) encompasses five cosine waves
520 (y1, y2, y3, y4, and y5) with different scales for the stationary case, and the artificial
521 data series (z) encompasses five cosine waves (z1, z2, z3, z4, and z5) with different
522 scales for the non-stationary case. The predictor variables, connected by a hyphen, are
523 shown in the top right corner of each subplot. Thin solid lines demarcate the cones of
524 influence, and thick solid lines show the 95% confidence levels.

525 **Figure 3.** Multiple wavelet coherence (a) between y and y2h0 and y4; (b) between y
526 and y2 and y4h0; (c) between z and z2h0 and z4; and (d) between z and z2 and z4h0.
527 The artificial data series (y) encompasses five cosine waves (y1, y2, y3, y4, and y5)
528 with different scales for the stationary case and the artificial data series (z)
529 encompasses five cosine waves (z1, z2, z3, z4, and z5) with different scales for the
530 non-stationary case. The variables y2h0 (or z2h0) and y4h0 (or z4h0) refer to the new
531 series of y2 (or z2) and y4 (or z4), in which the second half are replaced by 0. The
532 predictor variables, connected by a hyphen, are shown in the top right corner of each

subplot. Thin solid lines demarcate the cones of influence and thick solid lines show the 95% confidence levels.

Figure 4. Multiple wavelet coherence of an artificial data series (z) encompassing five cosine waves (z1, z2, z3, z4, and z5) with different scales and (a) z2 and noised z2, (b) z2 and noised z4, and (c) z2, z4, and noised z4 for the non-stationary case. The predictor variables are connected by a hyphen and shown in the top right corner of each subplot. z2wn (z4wn), z2mn (z4mn), and z2sn (z4sn) indicate weakly, moderately, and strongly noised z2 (z4) series, respectively. Weakly, moderately, and strongly noised series are correlated with original series, having correlation coefficients of 0.9, 0.5, and 0.1, respectively. Thin solid lines demarcate the cones of influence and thick solid lines show the 95% confidence levels.

Figure 5. Multiple spectral coherence (MSC) of an artificial data series (y or z) encompassing five cosine waves (y1 to y5; or z1 to z5) with different scales and (a) two (y2 and y4; or z2 and z4) or three (y2, y3, and y4; or z2, z3, and z4) data series, and (b) two (y2 and y4; or z2 and z4) data series when the second half of one data series are replaced by 0. The variables y2h0 (or z2h0) and y4h0 (or z4h0) refer to the new series of y2 (or z2) and y4 (or z4) in which the second half are replaced by 0.

Figure 6. Multiple correlation coefficient between multivariate empirical mode decomposition (MCC_{memd}) of an artificial series (y or z) and (a) two (y2 and y4; or z2 and z4) or three (y2, y3, and y4; or z2, z3, and z4) data series, and (b) two (y2 and y4; or z2 and z4) data series when the second half of one data series are replaced by 0. The variables y2h0 (or z2h0) and y4h0 (or z4h0) refer to the new series of y2 (or z2)

555 and y4 (or z4) in which the second half are replaced by 0.

556 **Figure 7.** Multiple wavelet coherence between evaporation (*E*) from water surfaces
557 and meteorological factors ((a) relative humidity and mean temperature and (b)
558 relative humidity, mean temperature, and sun hours) at Changwu site in Shaanxi,
559 China. Thin solid lines demarcate the cones of influence, and thick solid lines show
560 the 95% confidence level.

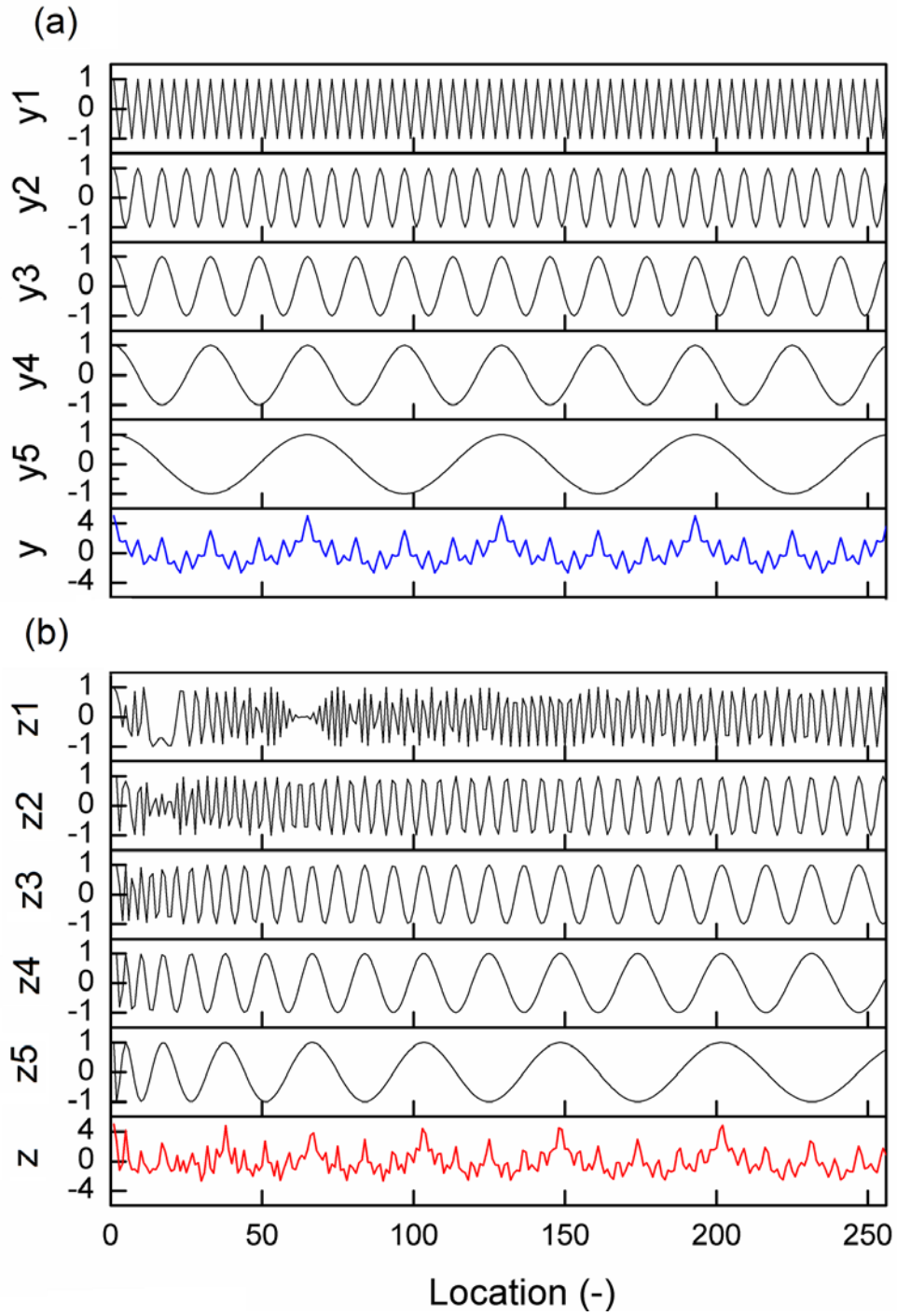

Figure 1. (a) Stationary and (b) non-stationary series of response variables (y for stationary and z for non-stationary case) encompassing five cosine waves (y1 to y5 for stationary and z1 to z5 for non-stationary case) with different dimensionless scales.

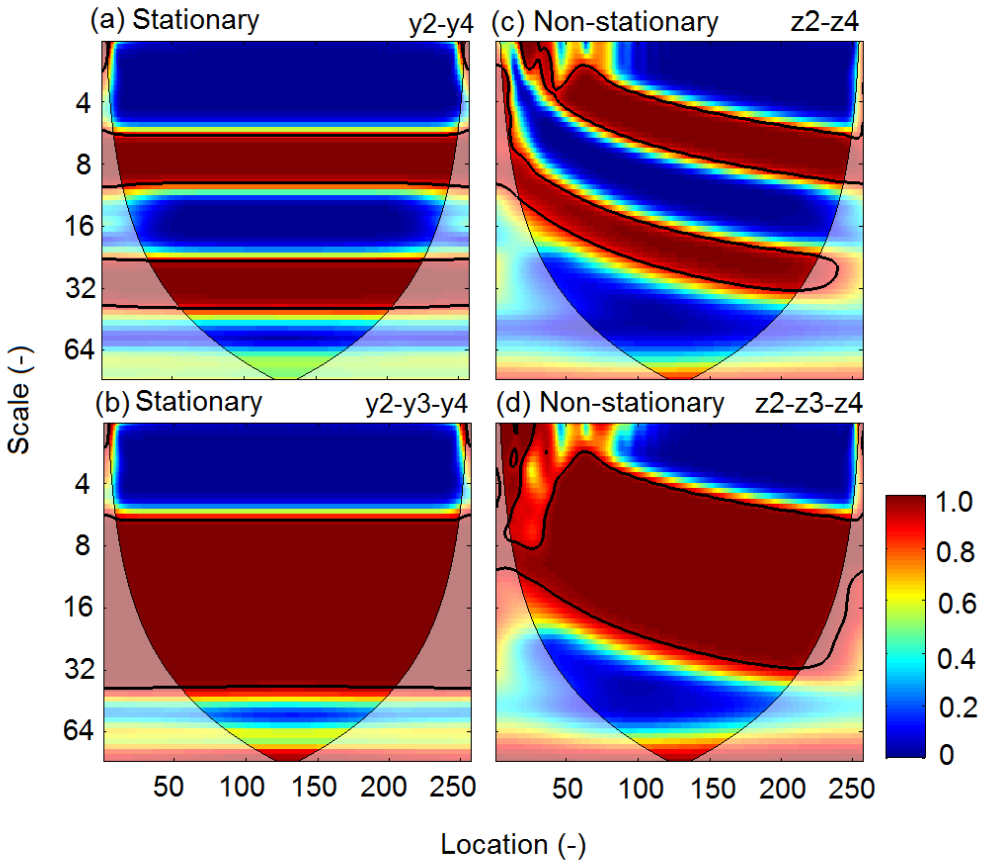

Figure 2. Multiple wavelet coherence (a) between response variable y and predictor variables y2 and y4; (b) between response y and predictors y2, y3, and y4; (c) between response z and predictors z2 and z4; and (d) between response z and predictors z2, z3, and z4. The artificial data series (y) encompasses five cosine waves (y1, y2, y3, y4, and y5) with different scales for the stationary case, and the artificial data series (z) encompasses five cosine waves (z1, z2, z3, z4, and z5) with different scales for the non-stationary case. The predictor variables, connected by a hyphen, are shown in the top right corner of each subplot. Thin solid lines demarcate the cones of influence, and thick solid lines show the 95% confidence levels.

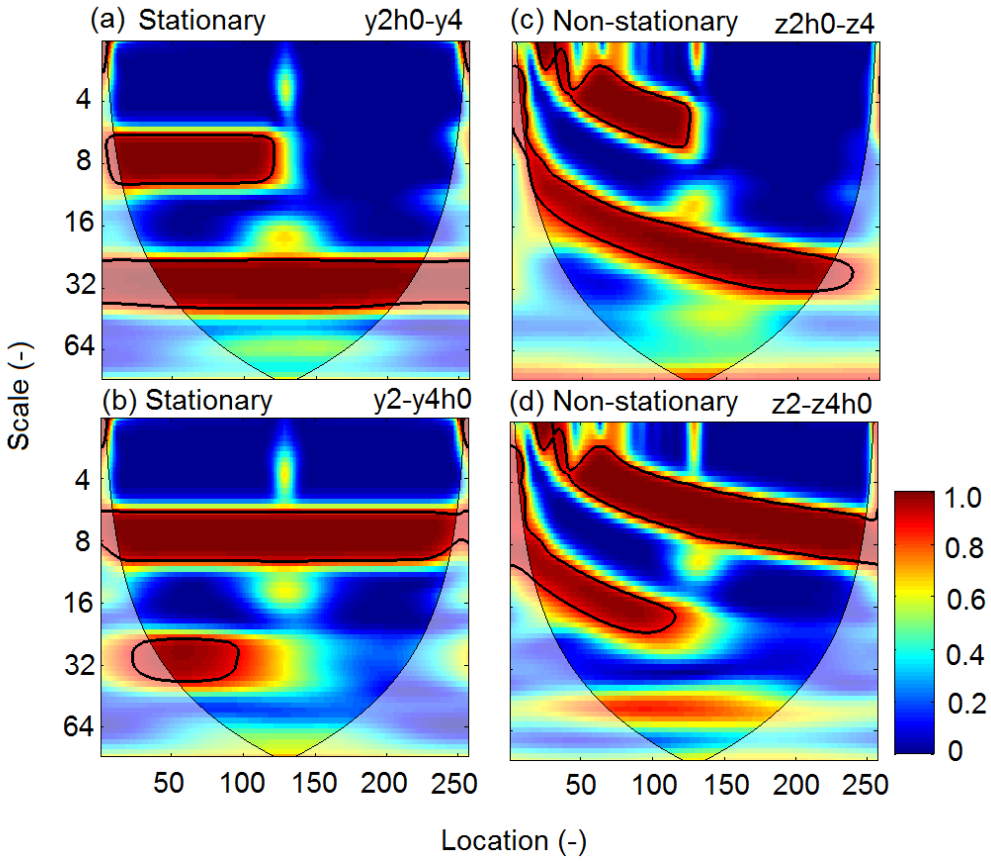

Figure 3. Multiple wavelet coherence (a) between y and y2h0 and y4; (b) between y and y2 and y4h0; (c) between z and z2h0 and z4; and (d) between z and z2 and z4h0. The artificial data series (y) encompasses five cosine waves (y1, y2, y3, y4, and y5) with different scales for the stationary case and the artificial data series (z) encompasses five cosine waves (z1, z2, z3, z4, and z5) with different scales for the non-stationary case. The variables y2h0 (or z2h0) and y4h0 (or z4h0) refer to the new series of y2 (or z2) and y4 (or z4), in which the second half are replaced by 0. The predictor variables, connected by a hyphen, are shown in the top right corner of each subplot. Thin solid lines demarcate the cones of influence and thick solid lines show the 95% confidence levels.

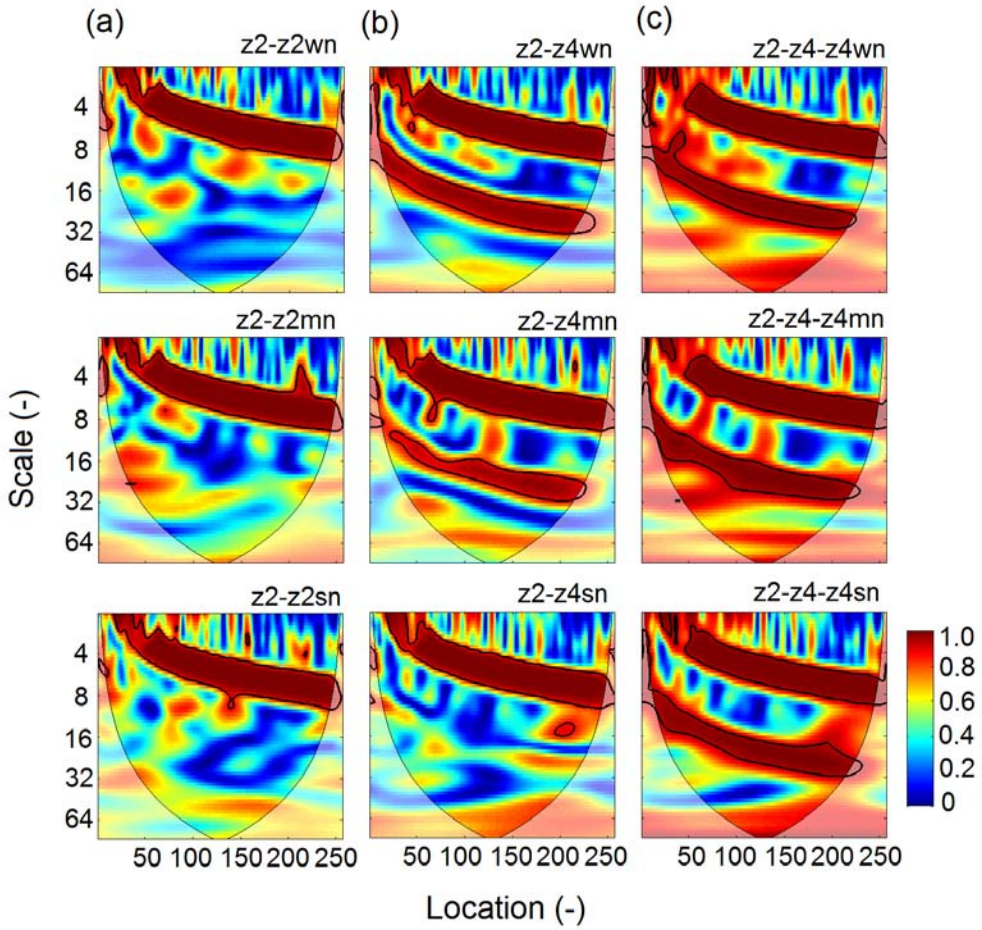

Figure 4. Multiple wavelet coherence of an artificial data series (z) encompassing five cosine waves (z1, z2, z3, z4, and z5) with different scales and (a) z2 and noised z2, (b) z2 and noised z4, and (c) z2, z4, and noised z4 for the non-stationary case. The predictor variables are connected by a hyphen and shown in the top right corner of each subplot. z2wn (z4wn), z2mn (z4mn), and z2sn (z4sn) indicate weakly, moderately, and strongly noised z2 (z4) series, respectively. Weakly, moderately, and strongly noised series are correlated with original series, having with correlation coefficients of 0.9, 0.5, and 0.1, respectively. Thin solid lines demarcate the cones of influence and thick solid lines show the 95% confidence levels.

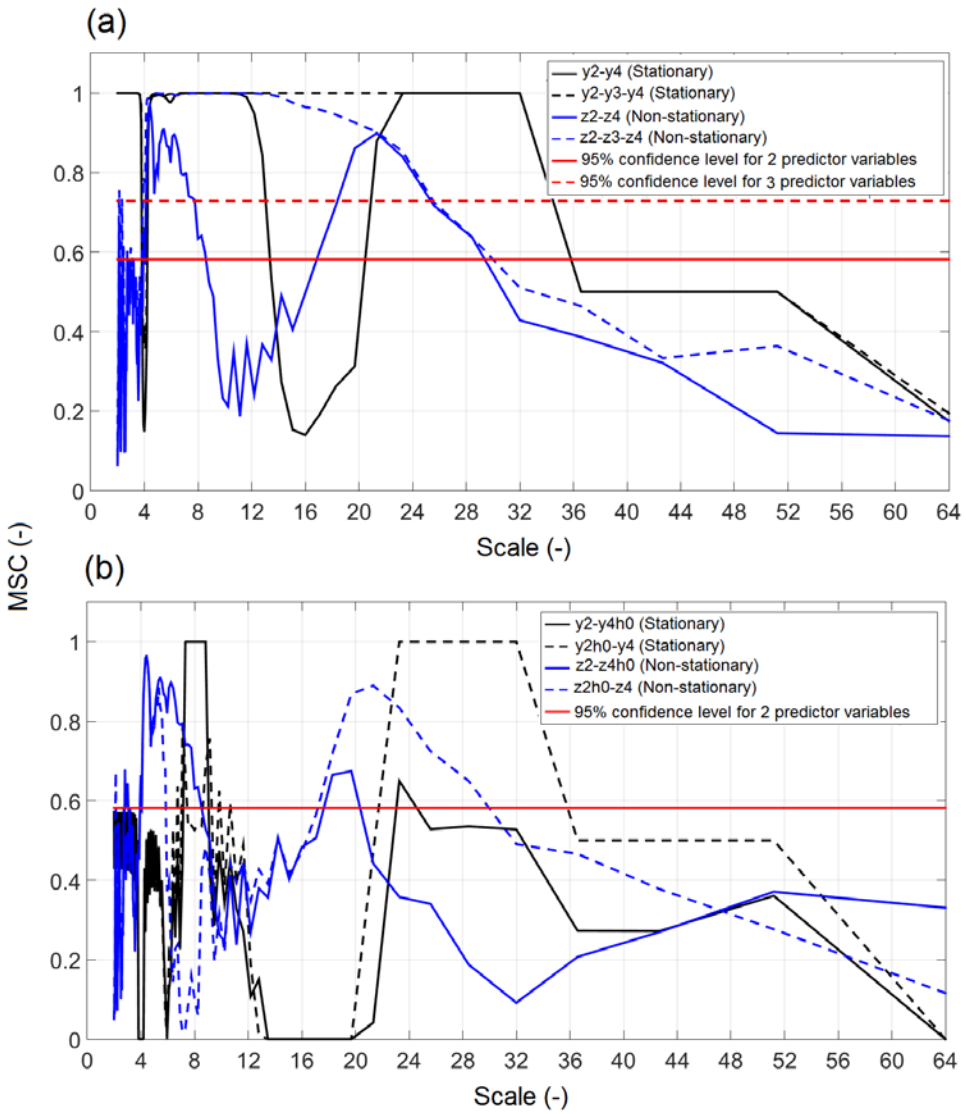

Figure 5. Multiple spectral coherence (MSC) of an artificial data series (y or z) encompassing five cosine waves (y1 to y5; or z1 to z5) with different scales and (a) two (y2 and y4; or z2 and z4) or three (y2, y3, and y4; or z2, z3, and z4) data series, and (b) two (y2 and y4; or z2 and z4) data series when the second half of one data series are replaced by 0. The variables y2h0 (or z2h0) and y4h0 (or z4h0) refer to the new series of y2 (or z2) and y4 (or z4) in which the second half are replaced by 0.

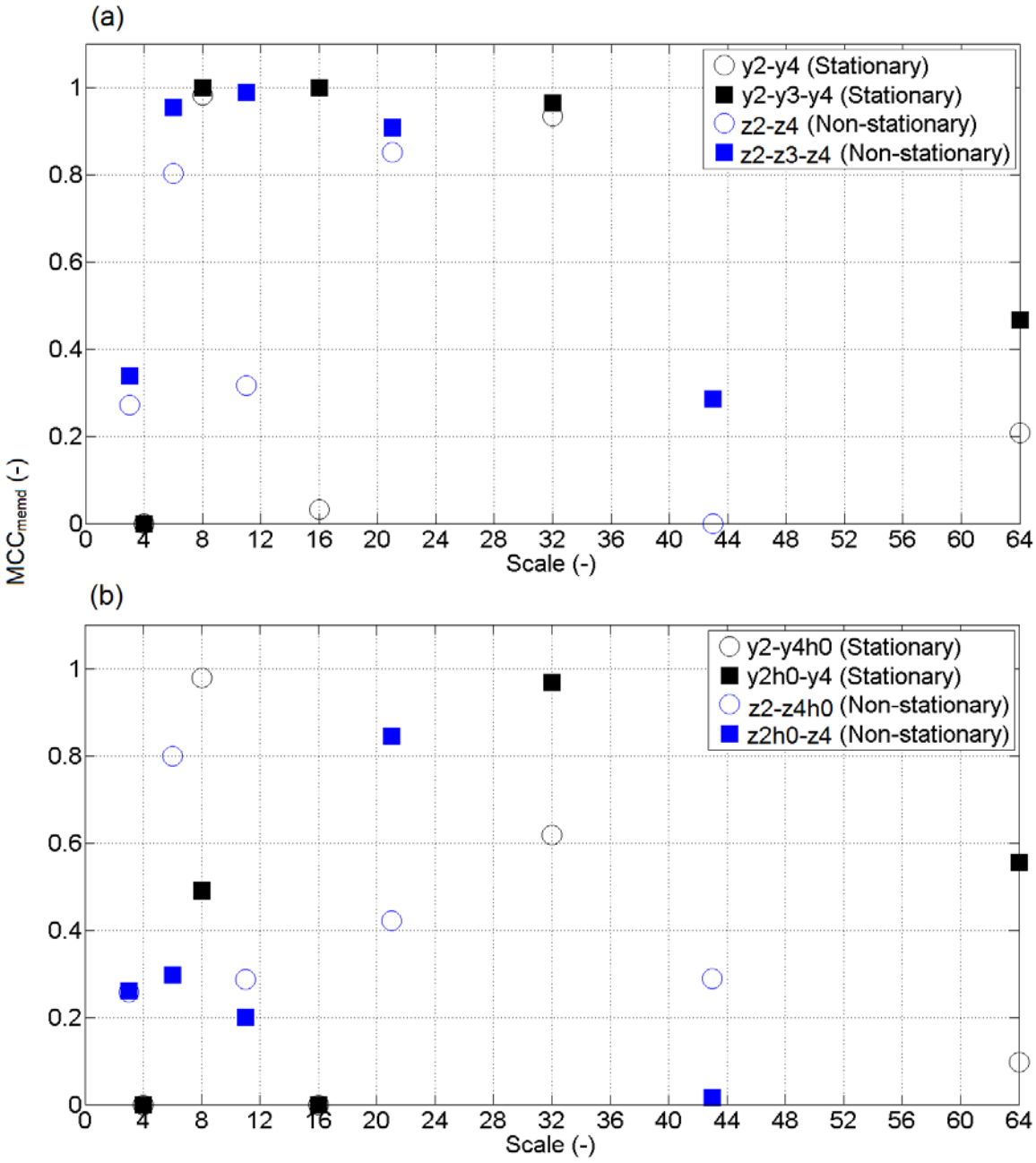

Figure 6. Multiple correlation coefficient between multivariate empirical mode decomposition (MCC_{memd}) of an artificial series (y or z) and (a) two (y2 and y4; or z2 and z4) or three (y2, y3, and y4; or z2, z3, and z4) data series, and (b) two (y2 and y4; or z2 and z4) data series when the second half of one data series are replaced by 0. The variables y2h0 (or z2h0) and y4h0 (or z4h0) refer to the new series of y2 (or z2) and y4 (or z4) in which the second half are replaced by 0.

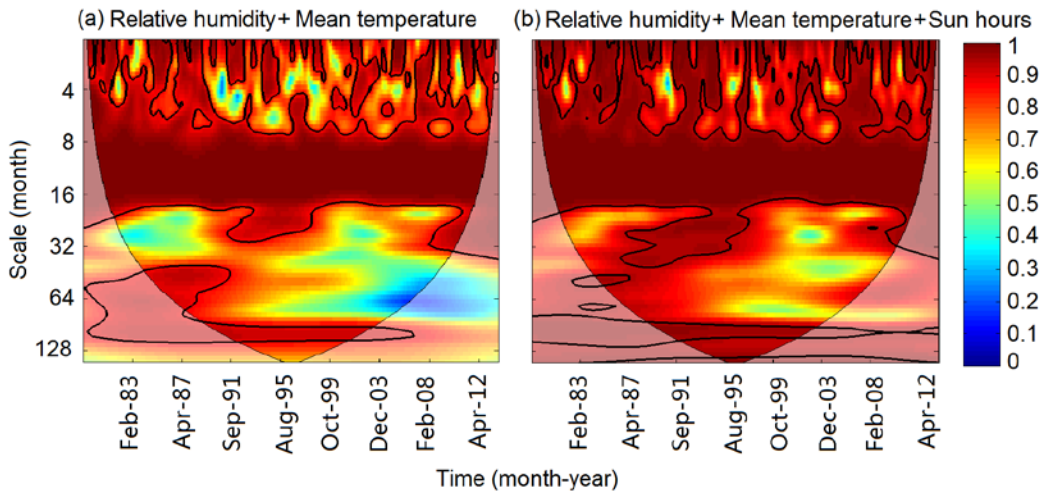

Figure 7. Multiple wavelet coherence between evaporation (E) from water surfaces and meteorological factors ((a) relative humidity and mean temperature and, (b) relative humidity, mean temperature, and sun hours) at Changwu site in Shaanxi, China. Thin solid lines demarcate the cones of influence, and thick solid lines show the 95% confidence level.