Unraveling the spatial diversity of Indian precipitation teleconnections via nonlinear multi-scale approach

Jürgen Kurths¹,²,³⁺, Ankit Agarwal¹,²,⁴⁺, Norbert Marwan¹, Maheswaran Rathinasamy¹, Levke Caesar¹,⁵, Raghvan Krishnan⁶, and Bruno Merz²,⁴⁺

¹Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Telegrafenberg, Potsdam, Germany
²Institute of Earth and Environmental Science, University of Potsdam, Potsdam, Germany
³Institute of Physics, Humboldt Universität zu Berlin, Germany
⁴GFZ German Research Centre for Geosciences, Section 5.4: Hydrology, Telegrafenberg, Potsdam, Germany
⁵Institute of Physics and Astronomy, University of Potsdam, Potsdam, Germany
⁶Indian Institute of Tropical Meteorology, Pune, India

*these authors equally share first authorship.

Correspondence: Ankit Agarwal* (agarwal@pik-potsdam.de/aagarwal@uni-potsdam.de)

S1. Event synchronization (ES)

Event synchronization is a promising measure for determining the time synchronization and delay pattern between two event-like signals (Quian Quiroga et al., 2002). We define an event when a value in the signal \( x(t) \) (or \( y(t) \)) exceeds a threshold (selected by a \( \alpha \) percentile). Events in \( x(t) \) and \( y(t) \) occurring at time \( t_l^x \) and \( t_m^y \), where \( l = 1,2,3,4 \ldots S_x \) and \( m = 1,2,3,4 \ldots S_y \), are considered to be synchronized when they occur within a time lag \( \pm \tau_{lm}^{xy} \) which is defined as (Agarwal et al., 2017a)

\[
\tau_{lm}^{xy} = \min \{ t_{l+1}^x - t_l^x, t_l^x - t_{l-1}^x, t_{m+1}^y - t_m^y, t_m^y - t_{m-1}^y \} / 2
\] (1)

\( S_x \) and \( S_y \) are the total number of events (greater than threshold \( \alpha \)) that occurred in the signal \( x(t) \) and \( y(t) \), respectively. This definition of the time lag helps to separate independent events. Then we count the number of times an event occurs in the signal \( x(t) \) after the maximum time lag \( \tau_{lm}^{xy} \) of an event that appears in the signal \( y(t) \) and vice versa, resulting in the quantities \( C(x|y) \) and \( C(y|x) \):

\[
C(x|y) = \sum_{l=1}^{S_x} \sum_{m=1}^{S_y} J_{xy} \quad \text{and} \quad C(y|x) = \sum_{l=1}^{S_x} \sum_{m=1}^{S_y} J_{yx}
\] (2)

with

\[
J_{xy} = \begin{cases} 
1 & \text{if} \quad 0 < t_l^x - t_m^y < \tau_{lm}^{xy}, \\
\frac{1}{2} & \text{if} \quad t_l^x = t_m^y, \\
0 & \text{else}, 
\end{cases}
\] (3)

From these quantities, we define a measure of the strength of event synchronization (\( Q_{xy} \)) between \( x(t) \) and \( y(t) \) by
\[ Q_{xy} = \frac{C(x|y) + C(y|x)}{\sqrt{(S_x - 2)(S_y - 2)}}. \]  

(4)

\( Q_{xy} \) is normalized to \( 0 \leq Q_{xy} \leq 1 \). \( Q_{xy} = 1 \) refers to perfect synchronization between the signals \( x(t) \) and \( y(t) \).

Event synchronization (ES) has been specifically designed to identify nonlinear associations among event time series with varying lags between them.

**S2. Maximum overlap discrete wavelet transformation**

Time series of continuous geophysical variables can be interpreted as the superposition of variations occurring at different scales. Different physical processes drive these patterns, and a partitioning of the variability at different scales can help to isolate and characterize the underlying processes (Sturtevant et al., 2016). Wavelets have been successfully used to characterize the time scales of interactions between hydrometeorological variables (Molini et al., 2010).

The wavelet transform of a signal decomposes it into a set of components with predefined central frequencies and spectral bandwidths. Here we use the maximal overlap discrete wavelet transform (Percival and Walden, 2000) (MODWT), because the orthogonal discrete wavelet transform (DWT) results in a pyramid of wavelet coefficients which does not contain the time synchronization of the events. Further, our experience with DWT suggests that it suffers from ‘shift sensitivity’ also known as ‘shift variance’ which is undesirable because DWT coefficients fail to distinguish between input-signal shifts (Rathinasamy and Khosa, 2012). Although MODWT has considerable redundancy but it is shift invariant, and this property renders the MODWT more suited for time series analysis.

MODWT decomposes the time series into different time scales or frequency components. The wavelet decomposition is realized using the two basis functions known as father wavelets and mother wavelet. Any function \( f(t) \) can be expressed in these basis functions and their scaled and translated versions as given in Eq.(5)

\[ f(t) = \sum_k s_{j,k} \varphi_{j,k}(t) + \sum_k d_{j,k} \psi_{j,k}(t) + \sum_k d_{j-1,k} \psi_{j-1,k}(t) + \cdots + \sum_k d_{1,k} \psi_{1,k}(t) \]  

(5)

where \( J \) is the number of multiresolution components (scales), and \( k \) is in the range of 1 to the number of coefficient in the specified component. The coefficients \( s_{j,k} \) are the approximation coefficients and \( d_{j,k}, \ldots, d_{1,k} \) are the wavelet transform coefficients, while the functions \( \varphi_{j,k}(t) \) and \( \{\psi_{j,k}(t)\} | \ j = 1, \ldots, J - 1, J \} \) are the approximating wavelet function and detailed wavelet functions respectively.

These basis functions are defined in terms of father and mother wavelets as follows:

\[ \varphi_{j,k}(t) = 2^{-j/2} \varphi(2^{-j}t - k) \]  

(6)

\[ \psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \]  

(7)

Further,

\[ s_{j,k} \approx \int \varphi_{j,k}(t) f(t) dt , \]  

(8)
\[ d_{j,k} = \int \psi_{j,k}(t)f(t)dt, \quad j = 1, \ldots, J - 1, \]

where the scaling coefficients \( s_{j,k} \) capture the smooth trend of the time series at the coarse scale \( 2^j \), which are also called smooth coefficients; and the wavelet coefficients \( d_{j,k} \), also known as detail coefficients can detect deviations from the coarsest scale to the finest scale.

The original series \( f(t) \) can be reconstructed by the summing the detailed components and the smooth components.

\[ f(t) = S_{J,k} + D_{J,k} + D_{J-1,k} + \cdots + D_{1,k} \]

where

\[ S_{J,k} = \sum_k s_{j,k} \varphi_{j,k}(t), \quad D_{J,k} = \sum_k d_{j,k} \psi_{j,k}(t), \quad \ldots, \quad D_{1,k} = \sum_k d_{1,k} \psi_{1,k}(t) \]

Eq (10) defines a multiresolution analysis (MRA) of \( f(t) \); i.e., we express the series \( f(t) \) as the sum of a constant vector \( S_j \) and \( J \) other vectors \( D_j, j = 1, \ldots, J \), each of which contain a timeseries related to variations in \( f(t) \) at a certain scale. We refer to \( D_j \) as the \( j^{th} \) level wavelet detail. Fig.S1 shows the MODWT decomposition of a sample signal up to 7 scales resulting in 7 detailed components (\( D_1 - D_7 \)) and one approximate Component (\( S_7 \)).

Let \( Y_t \) represents a time series history of a geophysical process. In order to partition the variability of the process at different scales \( j = 1 \ldots J \), the signal \( Y_t \) is transformed into the wavelet space which provides the required information at different scales. This is obtained by convolving \( Y_t \) with a set of low pass \( (g) \) and high pass \( (h) \) filters. For instance, at each scale \( j \), the MODWT applies a high pass wavelet filter \( h_{j,1} \) and a lower pass filter \( g_{j,1} \) of length \( (l) \) to the time

![Fig.S1 Scheme of multi-scale decomposition of signals using maximum overlap discrete wavelet transformation (MODWT). The relationship between signal \( Y_t \) (blue), detailed component \( D_j \) (black), and approximate component \( S_j \) (red), is shown.](image-url)

Let \( Y_t \) represents a time series history of a geophysical process. In order to partition the variability of the process at different scales \( j = 1 \ldots J \), the signal \( Y_t \) is transformed into the wavelet space which provides the required information at different scales. This is obtained by convolving \( Y_t \) with a set of low pass \( (g) \) and high pass \( (h) \) filters. For instance, at each scale \( j \), the MODWT applies a high pass wavelet filter \( h_{j,1} \) and a lower pass filter \( g_{j,1} \) of length \( (l) \) to the time
series $Y$ to respectively yield the wavelet coefficients $W_{j,t}$ and $V_{j,t}$ for every point $t$ in the time series (Percival and Walden 2000).

\[
W_{j,t} = \sum_{l=0}^{L_j - 1} h_{j,l} Y_{t-l \mod 2N} \\
V_{j,t} = \sum_{l=0}^{L_j - 1} g_{j,l} Y_{t-l \mod 2N}
\] (12)

The $W_{j,t}$ wavelet coefficients distinguish fluctuations in the time series of scale $2^{j-1}$, while the $V_{j,t}$ coefficients provide information about the variations at scale $2^j$ and higher. Let the maximum level of decomposition be $j = J$. This would result in a total $J'$ series of wavelet coefficients with $W_{j,t}, j = 1, 2, 3 \ldots J'$, and one series of $V_{J,t}$.

Let us now define $D_j$ which represents the time domain reconstruction of $W_j$. It represents the portion of $Y$ attributable to scale $j$. Let $S_j$ represent the time domain reconstruction of $V_j$. For the maximum level of decomposition, $V_j$ has all of its elements equal to the sample mean of $Y$.

Therefore, we can write

\[
Y = \sum_{j=1}^{J} D_j + S_j
\] (13)

S3. Significance test for similarity measure

To evaluate the statistical significance of the ES values, a surrogate test is used as proposed by Agarwal et al. (Agarwal et al., 2017c). We randomly reshuffle each time series 100 times (arbitrary number) but keeping the distribution same. The reshuffling will ensure that any potential synchronization between the even series will be destroyed and that they will be equivalent to independent random series. Then, for each pair of time series (precipitation and climate time series), we calculate the MSES values for the different scales. At each scale, the empirical test distribution of the 100 MSES values for the reshuffled time series is compared to the MSES values of the original time series. Using a 1% significance level, we assume that synchronization cannot be explained by chance if the MSES value at a certain scale of the original time series is larger than the $99^{th}$ percentile of the test distribution.

S4. Wavelet Coherence

To compare the results obtained from the MSES, we use wavelet coherence (WC). Wavelet coherence analysis has been used as a robust tool in identifying the relationship between two variables at multiple scales. The wavelet coherence between time series $\{X_t\}$ and $\{Y_t\}$ was defined by (Torrence and Compo, 1998) as
\[ R^2(j,t) = \frac{\zeta\left(v^{-1}W_{xy}(j,t)\right)^2}{\zeta\left(v^{-1}\left|W_x(j,t)\right|^2\right)\zeta\left(v^{-1}\left|W_y(j,t)\right|^2\right)} \]  (14)

Where \( R^2(j,t) \) takes a value between 0 and 1; \( \zeta \) is a smoothing operator and can be written as

\[ \zeta(W) = \zeta_{scale}(\zeta_{time}(W(j,t))) \]

\( W_{xy} \) represents the cross-wavelet coefficient between X and Y. \( W_x(j,t) \) and \( W_y(j,t) \) denote the wavelet coefficients obtained from wavelet transform of X and Y respectively at scale \( j \) and time \( t \).

The global wavelet coherence at a certain scale \( j \) is defined as the time-averaged value of the wavelet coefficients at the scale with the COI. It is estimated by

\[ R^2(j) = \frac{1}{n_j} \sum_{t=t_1}^{t_2} R^2(j,t) \]  (15)

Where \( n_j \) is the number of points with COI and \( n_j = t_2 - t_1 + 1 \).

Global wavelet coherence is a useful measure to examine the common characteristic periodicities between \( x \) and \( y \). Grinsted et al. showed the applicability of WC analysis of the association of precipitation with climate variables (Grinsted et al., 2004). More detailed description of wavelet coherence analysis is can be found in (Grinsted et al., 2004).

**Reference**


