

Supplementary Information for „Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: A Theory-Guided Causal Effect Network approach“

5

Giorgia Di Capua^{1,2}, Marlene Kretschmer¹, Reik V. Donner^{1,3}, Bart van den Hurk^{2,4}, Ramesh Vellore⁵,
Raghavan Krishnan⁵, and Dim Coumou^{1,2}

10

- 15 ¹Potsdam Institute for Climate Impact Research, Potsdam, Germany
²VU University of Amsterdam, Institute for Environmental Studies, Amsterdam, Netherlands
³Magdeburg-Stendal University of Applied Sciences, Magdeburg, Germany
⁴Deltares, Delft, Netherlands
⁵Indian Institute for Tropical Meteorology, Pune, India

20

Correspondence to: Giorgia Di Capua (dicapua@pik-potsdam.com)

Text S1. PCMC algorithm and numerical example

The PC-step has one main free parameter α , which is the significance threshold chosen to retain or discard a certain parent after calculating the partial correlation. In the latest version of the algorithm (TIGRAMITE 3.0, <https://github.com/jakobrunge/tigramite>), this value can either be set as a single value, e.g. $\alpha=0.05$, or as a set of values (the default set is $\alpha=\{0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$). In the second case, the best α value is chosen by applying the Akaike information criterion (AIC), which compares the parents obtained with each element of α .

The PC-step starts by first identifying the set of actors that we want to analyse. As an example we assume that our set of actors is composed by 7 univariate time series which we hypothesize to share causal relationships. This set is referred to as the initial parents and is defined as follows:

$$P = \{A, B, C, D, E, F, G\}$$

Each letter in P represents a univariate time series, which in our case must be given in the form of detrended anomalies. The PC algorithm first calculates plain correlations between the first element (at lag 0) and each of the remaining elements in P at a certain lag τ (here $\tau = 1, 2$). Let us assume that $A_{\tau=0}$ is found to be significantly correlated with 3 other actors, which will form the set of initial parents for $A_{\tau=0}$ (note that actors in \widehat{P}_0 are ordered by the strength of the correlation):

$$\widehat{P}_0(A_{\tau=0}) = \{C_{\tau=-1}, D_{\tau=-2}, E_{\tau=-1}, G_{\tau=-1}\}$$

For each element in $\widehat{P}(A_{\tau=0})$, partial correlations are calculated conditioning on the first strongest correlation. The partial correlation between the variables x and y conditioned on variable z is calculated by first performing linear regressions of x on z and of y on z and then calculating the correlation between the residuals:

$$\rho = \rho(x, y|z) = \rho(\text{Res}(x), \text{Res}(y))$$

If the partial correlation between x and y is still significant at a certain significance level α , x and y are said to be conditionally dependent given variable z , i.e. the correlation between x and y cannot be (exclusively) explained by the influence of variable z . When the opposite happens, the link is thus spurious and therefore filtered out and x and y are conditionally independent.

Following our example, the algorithm would proceed as follows:

$$\rho = \rho(A_{\tau=0}, C_{\tau=-1}|D_{\tau=-2}) = 0.3, p < 0.05$$

$$\rho = \rho(A_{\tau=0}, D_{\tau=-2}|C_{\tau=-1}) = -0.23, p < 0.05$$

$$\rho = \rho(A_{\tau=0}, E_{\tau=-1}|C_{\tau=-1}) = 0.35, p < 0.05$$

$$\rho = \rho(A_{\tau=0}, G_{\tau=-1}|C_{\tau=-1}) = -0.18, p > 0.05$$

In each step, the algorithm calculates the partial correlation between $A_{\tau=0}$ and each of the remaining elements of $\widehat{P}(A_{\tau=0})$ conditional on the first strongest parents ($C_{\tau=-1}$ for all parents except for $C_{\tau=-1}$ itself where the second strongest, $D_{\tau=-2}$, is used). After this step, the set of parents is reduced to

$$\widehat{P}_1(A_{\tau=0}) = \{C_{\tau=-1}, D_{\tau=-2}, E_{\tau=-1}\}$$

Now, the algorithm tests the remaining parents conditioning on a subset of two variables, again starting from the strongest:

$$\rho = \rho(A_{\tau=0}, C_{\tau=-1}|D_{\tau=-2}, E_{\tau=-1}) = 0.29, p < 0.05$$

$$\rho = \rho(A_{\tau=0}, D_{\tau=-2} | C_{\tau=-1}, E_{\tau=-1}) = -0.17, p > 0.05$$

$$\rho = \rho(A_{\tau=0}, E_{\tau=-1} | C_{\tau=-1}, D_{\tau=-2}) = 0.27, p < 0.05$$

The resulting set of parents will now be:

$$\widehat{P}_2(A_{\tau=0}) = \{C_{\tau=-1}, E_{\tau=-1}\}$$

60 In the next step, the algorithm would test a combination of three variables. In our example, this is no longer possible because the set of parents now contains only two variables. When the number of parents becomes smaller or equal to the number of conditions that should be tested, the algorithm converges for A and starts to test the parents of B , following exactly the same process. Let us assume that after testing all the actors in P , we find three sets of parents:

$$\widehat{P}_2(A_{\tau=0}) = \{C_{\tau=-1}, E_{\tau=-1}\}$$

$$65 \quad \widehat{P}_3(C_{\tau=0}) = \{A_{\tau=-1}, D_{\tau=-1}\}$$

$$\widehat{P}_2(E_{\tau=0}) = \{B_{\tau=-1}, F_{\tau=-1}\}$$

The selected sets of parents then enter the second step of PCMC. The PC-step has one main free parameter α , which is the significance threshold chosen to retain or discard a certain parent after calculating the partial correlation.

70 In the MCI-step, the partial correlation between an actor and its set of parents is calculated again but conditioning also on the sets of parents of the parents of the actor we are interested in. Following our example:

$$\rho = \rho(A_{\tau=0}, C_{\tau=-1} | A_{\tau=-2}, D_{\tau=-2}) = 0.26, p < 0.05$$

$$\rho = \rho(A_{\tau=0}, E_{\tau=-1} | B_{\tau=-2}, F_{\tau=-2}) = 0.22, p < 0.05$$

After this final test, both parents pass the MCI test and will then form the final set of parents for $A_{\tau=0}$. When a set of different α is used, the parents with the highest AIC score are used as conditions when calculating partial correlations in the MCI-step
75 (Runge et al., 2017).

Text S2. CEN for a forward (along the z axis) propagating wave

We study the following model for a forward propagating wave shown in Fig. S8

$$80 \quad f(z, t) = A * \cos(k * z - \omega * t) + \epsilon_t, (1)$$

with the following variables and parameters:

A = amplitude, here $A = 1$

$k = \frac{2\pi}{\lambda}$ wave number, here $\lambda = 60^\circ$

$\omega = \frac{2\pi}{T}$ frequency, here $T = 4$ time steps

85 ϵ_t = random samples from a normal (Gaussian) distribution with mean value 0 and s.d.=0.6

t = time dimension

z = space dimension

We estimate the period T and the wave length λ from Figure S7. In the bottom panel (at lag -2), we can identify the upstream high (H1) and the downstream low (L1), which are by definition located at a relative distance of $\frac{1}{2} \lambda$. In the top panel (at lag -1), we can see that both H1 and L1 have moved downstream (eastward) of about $\frac{1}{4} \lambda$ ($\sim 15^\circ$). We can thus estimate the wave length as $\lambda \sim 60^\circ$ and the period (i.e., the time the wave needs to be displaced by 1λ) as $T \sim 4$ time steps. We use these parameters in equation (1) and then calculate the time series for the wave taken at two different spatial locations: one upstream ($z1 \sim 0$) and one downstream ($z3 \sim \frac{1}{2} \lambda$).

95 We then test what a CEN for these two wave variables Wave_z1 and Wave_z3 would look like when a lag = -1 is considered. Our results reveal that if the two points in space are taken at a distance of $\frac{1}{2} \lambda$, the CEN does not show any causal link (Figure not shown). This happens because for half of the time, one wave is increasing while the other is decreasing, but for the remaining time the two wave act in phase. However, as soon as the exact spatial location is displaced by a few degrees, and the time period is not perfectly $T=4$ time steps, the same behaviour as the one shown in Fig. 5 in the

100 main text is detected: the causal link from z1 to z3 has a negative sign, meaning that when the amplitude of the wave is increasing upstream, going forward in time, it will decrease downstream, while the opposite happens for the link from z3 back to z1. In Fig. S8, the following values for the wave parameters are used:

$z1 = 0.5^\circ$ (upstream location)

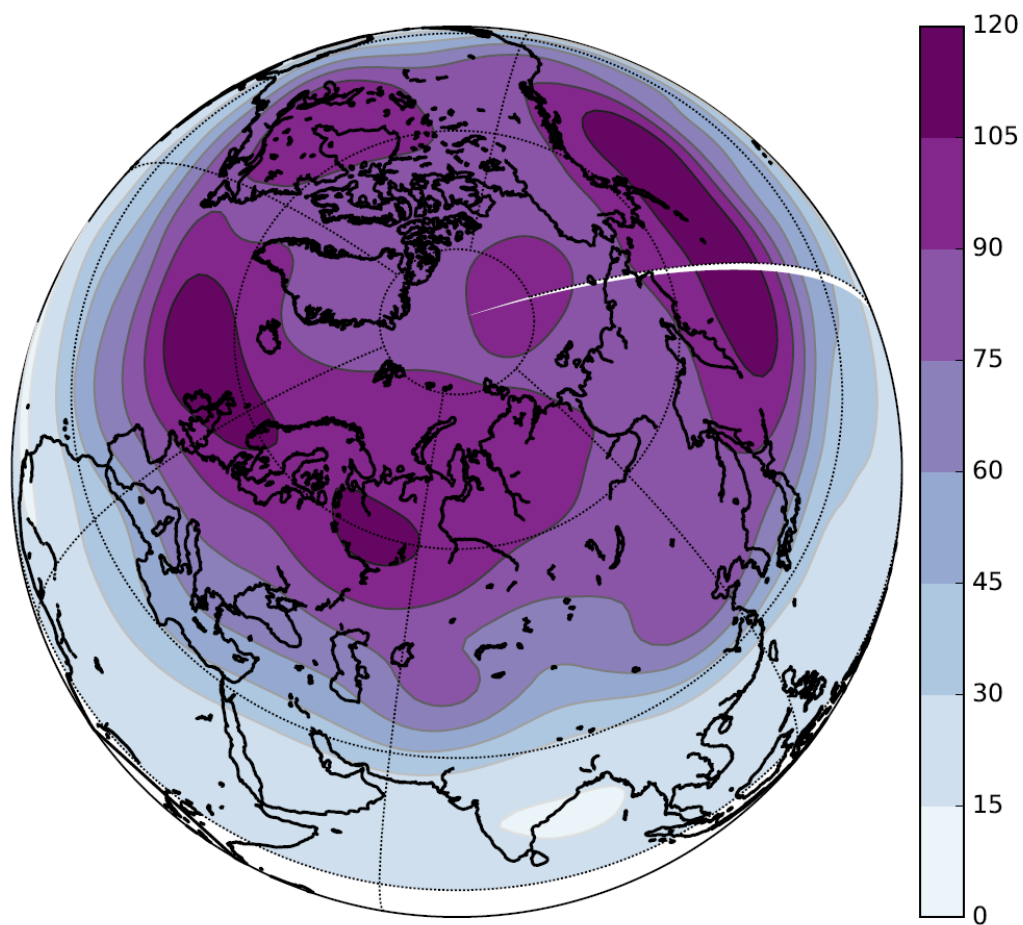
$z2 = 34.5^\circ$ (downstream location centred $\sim \frac{1}{2} \lambda$)

105 $T = 3.8$

Figure S7 shows evidence that the depicted wave train propagating eastward can be seen in the correlation maps between the CGTI and Z200 at lag -1 and -2. From these plots, it is possible to estimate the period T and wave length λ of the depicted wave. Using these values to build an idealized wave function, we provide an example to show how a wave that propagates downstream measured at two different spatial locations would behave in the PC-MCI algorithm (Fig. S8, panel a,b). Figure

110 S8, panel c shows that the CEN built with the time series of a forward propagating wave observed at two different locations in space, one upstream ($z1 \sim 0$) and one downstream ($z3 \sim \frac{1}{2} \lambda$), show the same behaviour as H1 and L1: the causal link from z1 to z3 has a negative sign, meaning that when the amplitude of the wave is increasing upstream, going forward in time, it will decrease downstream, while the opposite happens for the link from z3 back to z1.

Z200 ERA-I (1979-2017) JJAS weekly (AD) - std (m)



115

Figure S1. Standard deviation of weekly detrended anomalies of JJAS Z200 in the Northern Hemisphere for the period 1979-2017.

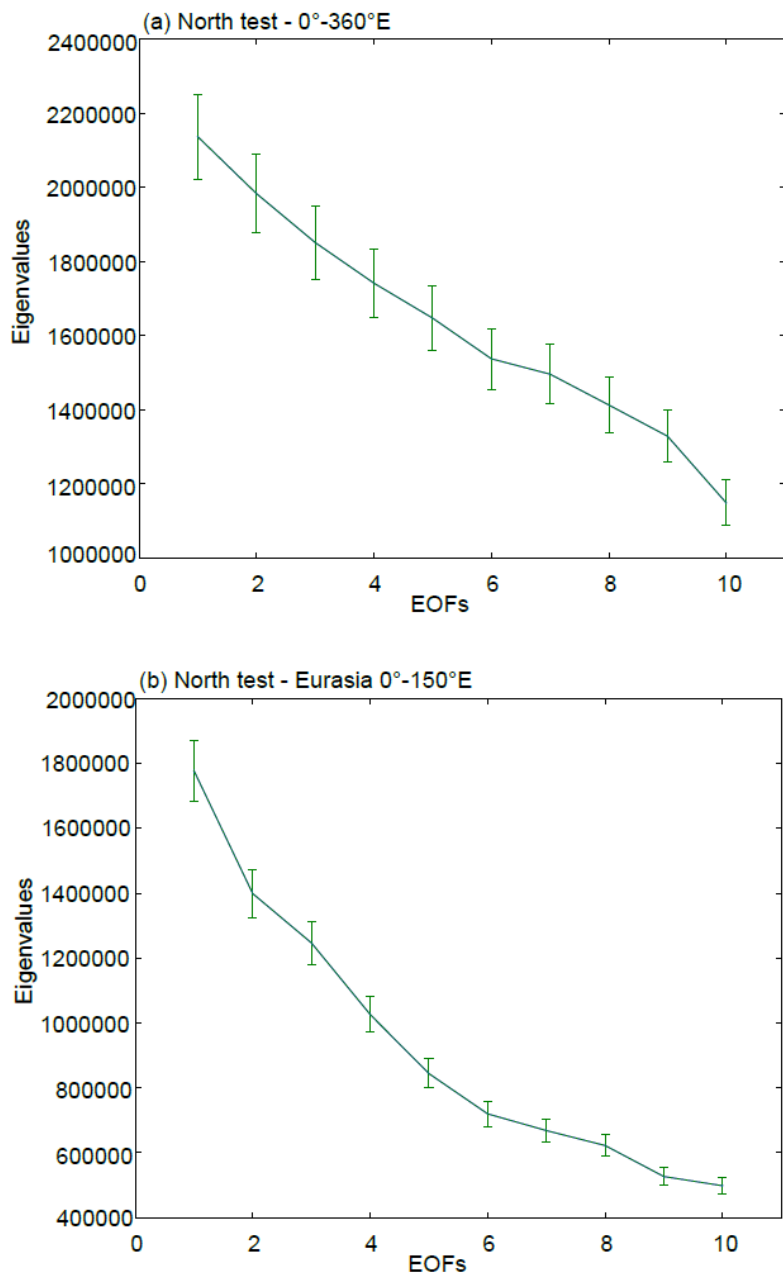
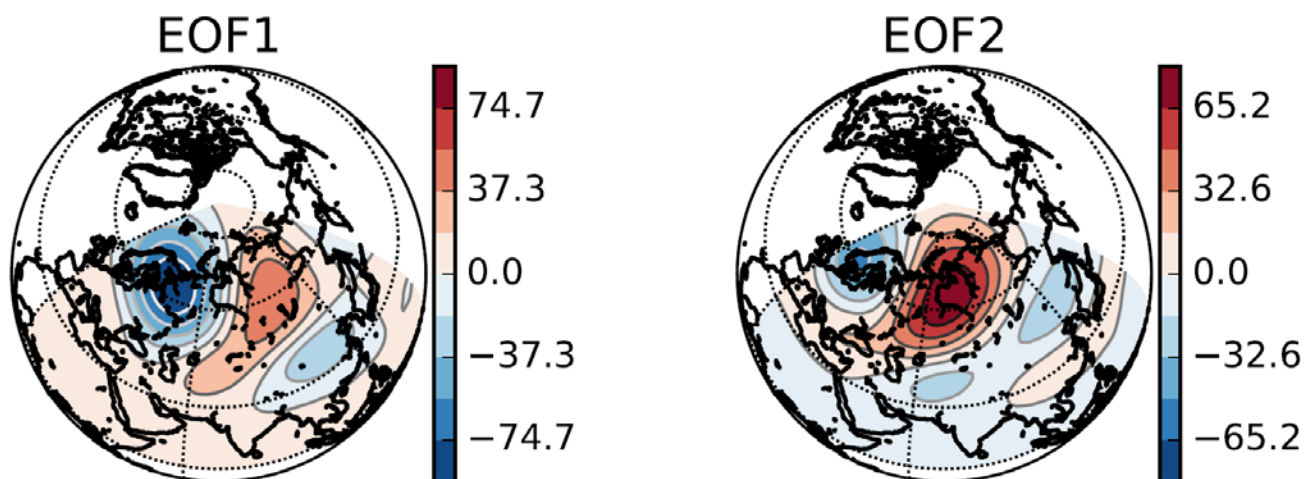
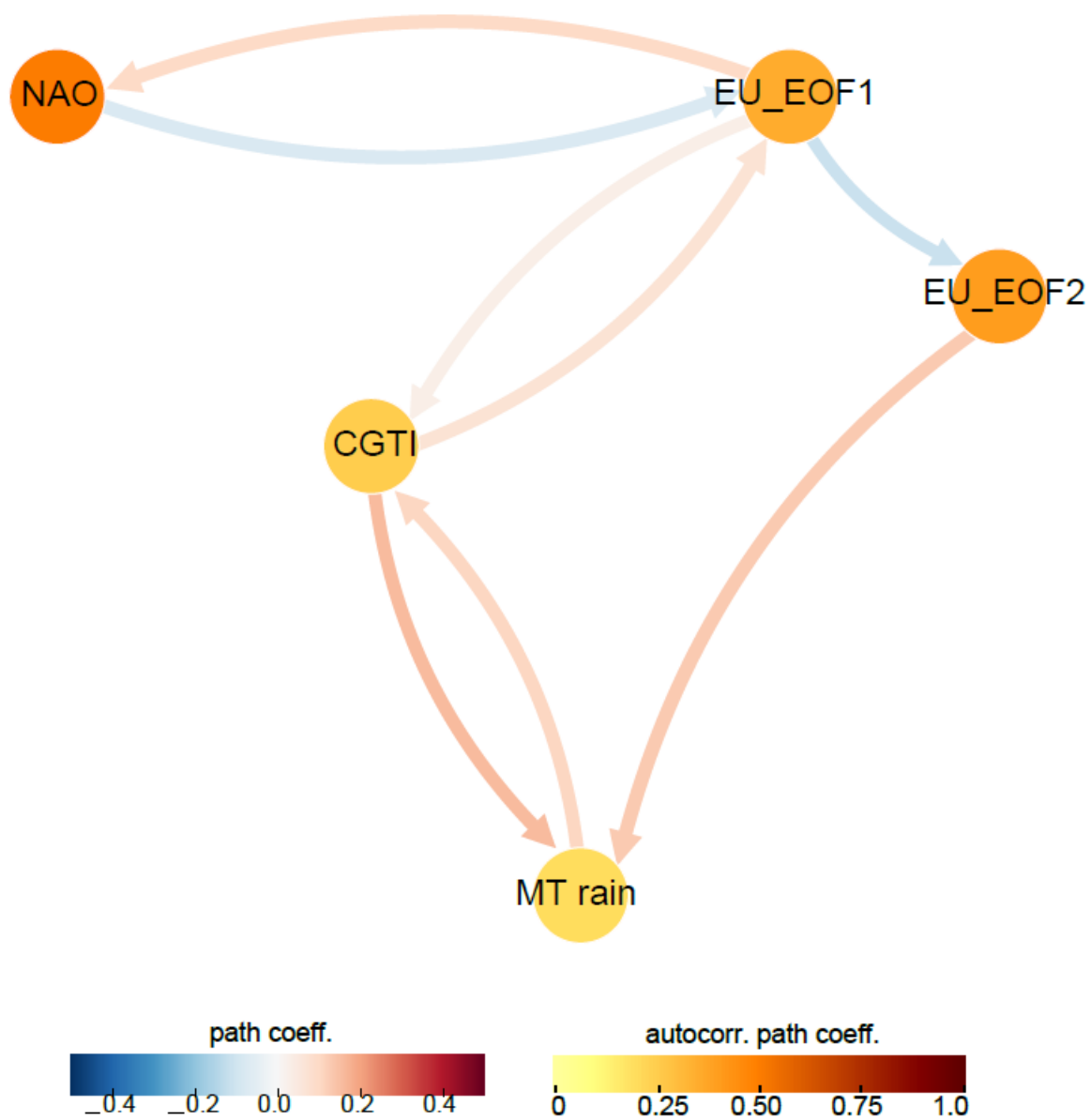


Figure S2. North test for the first ten EOFs for the full Northern Hemisphere (0°-90°N, 0°-360°E, left panel) and for the Eurasian sector (0°-90°N, 0°-150°E, right panel). The North test function assesses the uniqueness of EOF modes through assumptions of error on singular values (λ) as described by North et al (1982) and Hannachi et al (2007). Overlapping error bars between neighbouring λ values indicates a possible mixture of signals.

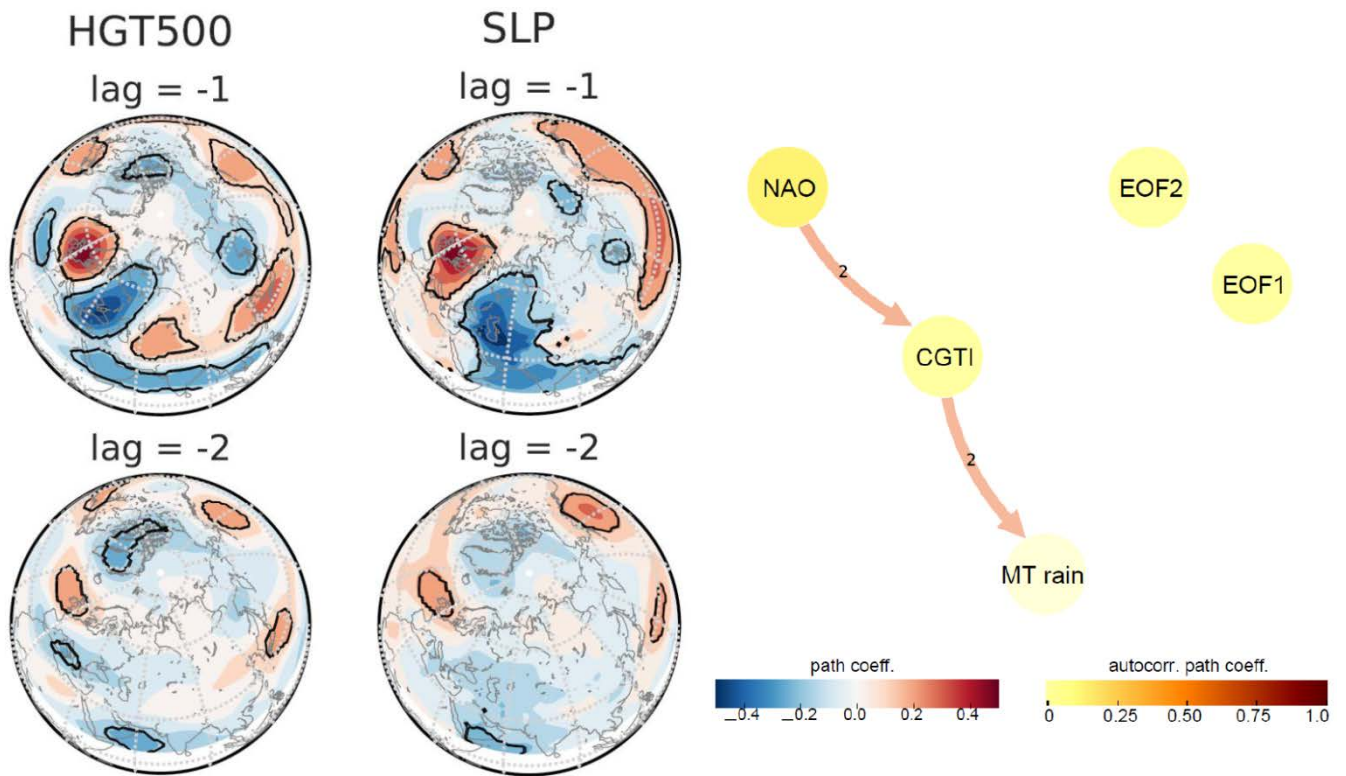


125 **Figure S3.** EOF1 (right panel) and EOF2 (left panel) for the JJAS weekly Z200 field in the Eurasian mid-latitudes (0°-150°E) for the period 1979-2017.



130

Figure S4. Causal Effect Network (CEN) built with CGTI, NAO, the PC of EOF1 and EOF2 defined on the Eurasian sector (shown in Figure S3) and MT rainfall for the period 1979-2017.



135 **Figure S5.** Correlation maps for the CGTI index with Z500 (left panels) and SLP (central panels) for lags -1 and -2 weeks. A CEN similar to that shown in Fig. 3b in the main manuscript is reproduced at lag -2.

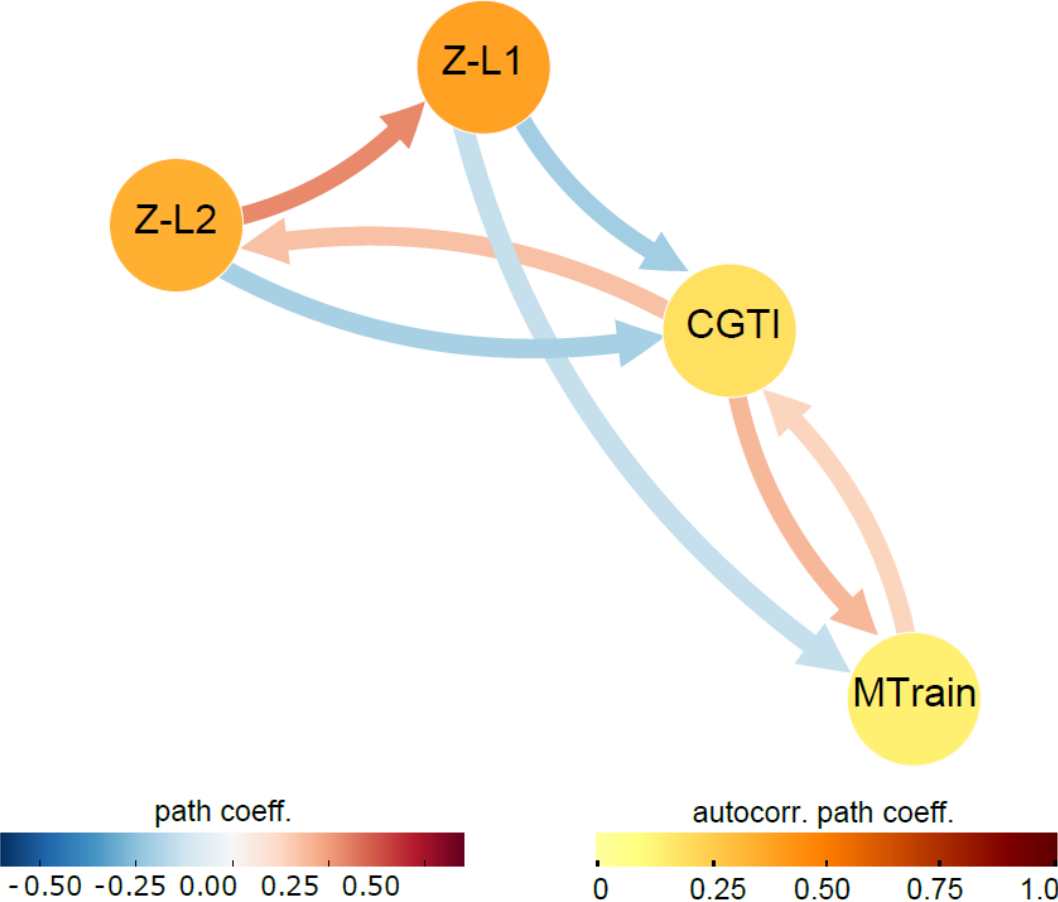
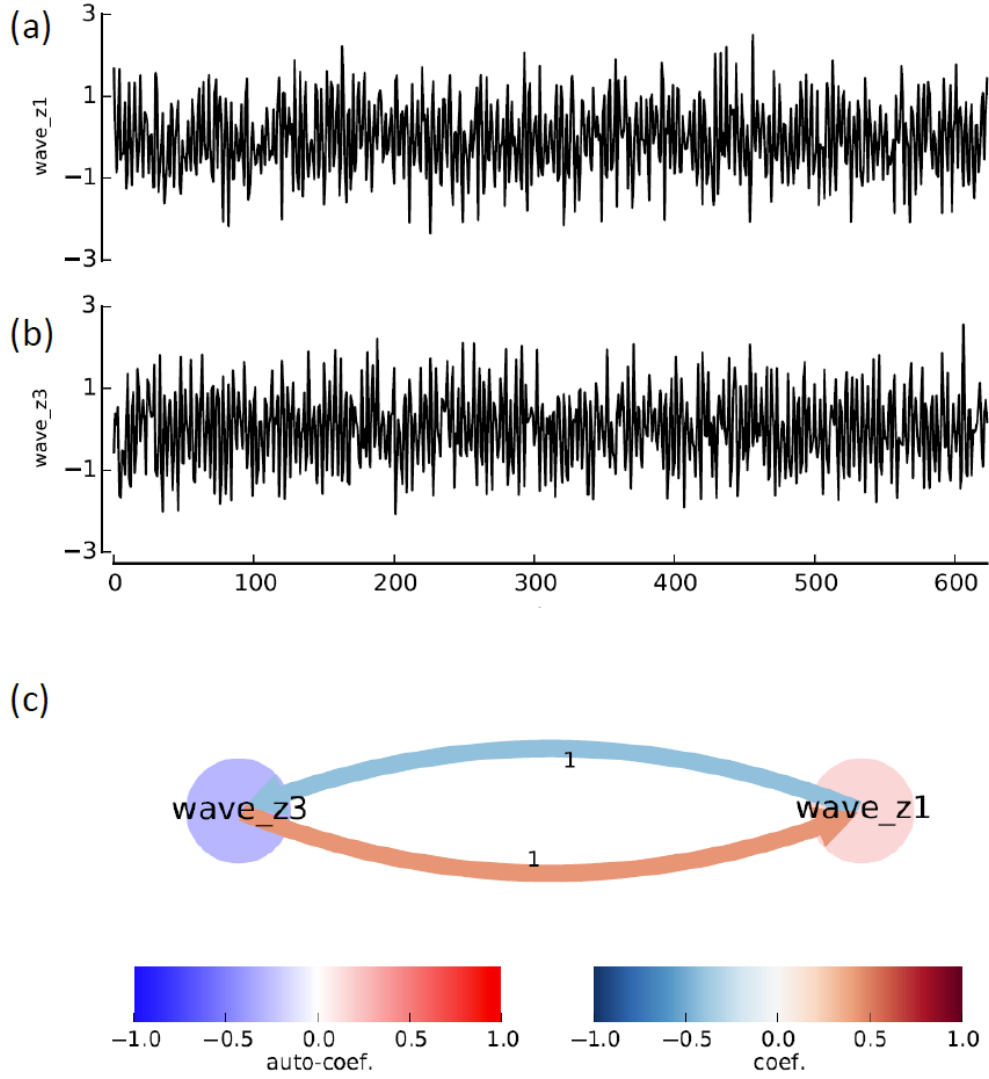
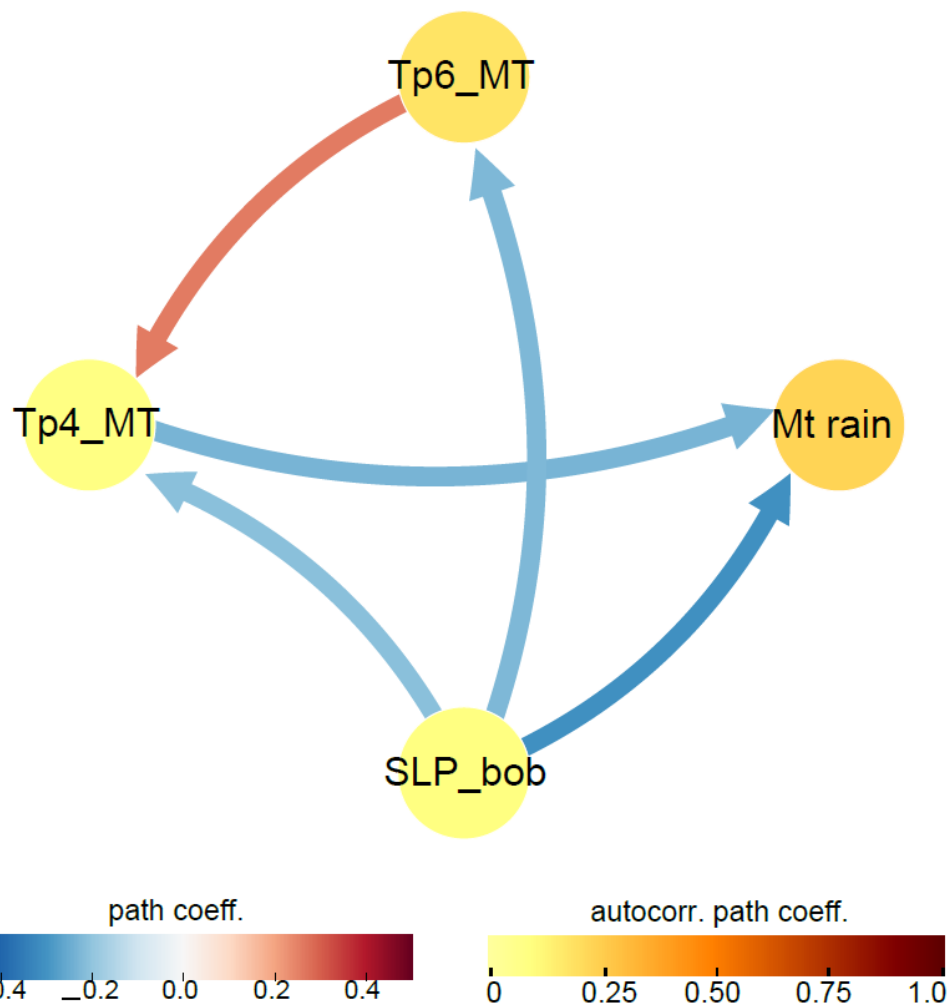


Figure S6. CEN built with weekly CGTI, MT rainfall L1 and L2 (as defined in Figure 4a in the main manuscript). MT rainfall is taken from Pai et al. dataset for the period 1979-2017.



155 **Figure S8.** Panel (a) and (b) show the time series for a forward propagating wave with $T=3.8$ and $\lambda=60^\circ$, taken in $z1=0.5^\circ$ and $z3=34.6^\circ$ respectively. Panel (c) shows the CEN built with the time-series of the propagating wave taken in two different points in space $z1$ and $z3$, where $z1$ is found upstream ($z1 = 0.5^\circ$) and $z3$ is located downstream ($z3 = 34.6^\circ$). See also Text S2 and the main manuscript for further explanation.



160 **Figure S9.** CEN built with MT rainfall from the CPC-NCEP dataset, SLP_bob, Tp4_MT, Tp6_MT for the 1979-2016 period.

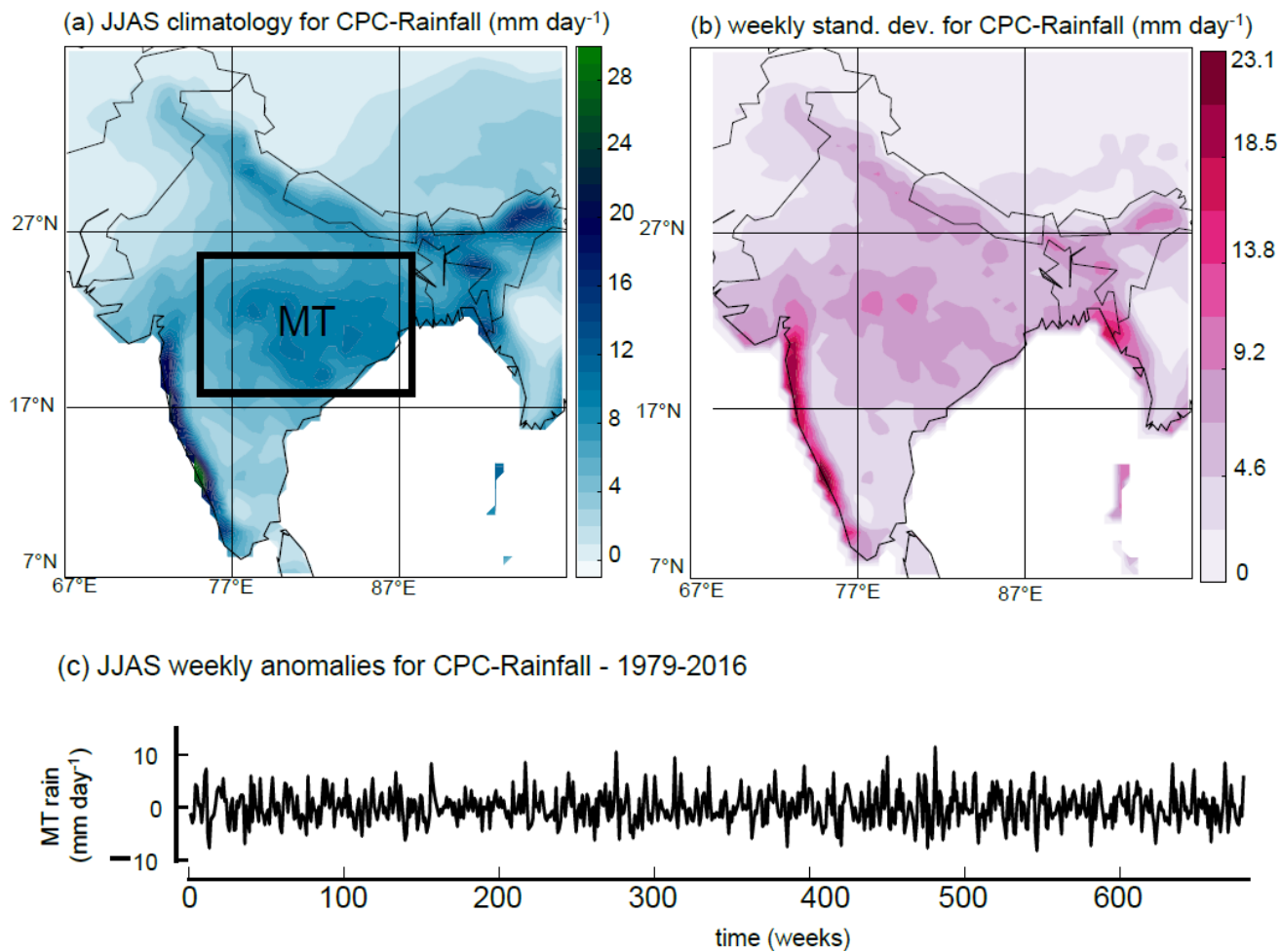


Figure S10. Rainfall climatology over India. Panel (a): JJAS rainfall climatology over the 1979-2016 period from the CPC-NCEP dataset. The black box identifies the MT region. Panel (b): standard deviation for weekly JJAS rainfall over the 1979-2016 period from the CPC-NCEP dataset. Panel (c): time series of weekly MT rainfall over the period 1979-2016; each year contains 18 weeks, with the first week starting on the 27th of May.

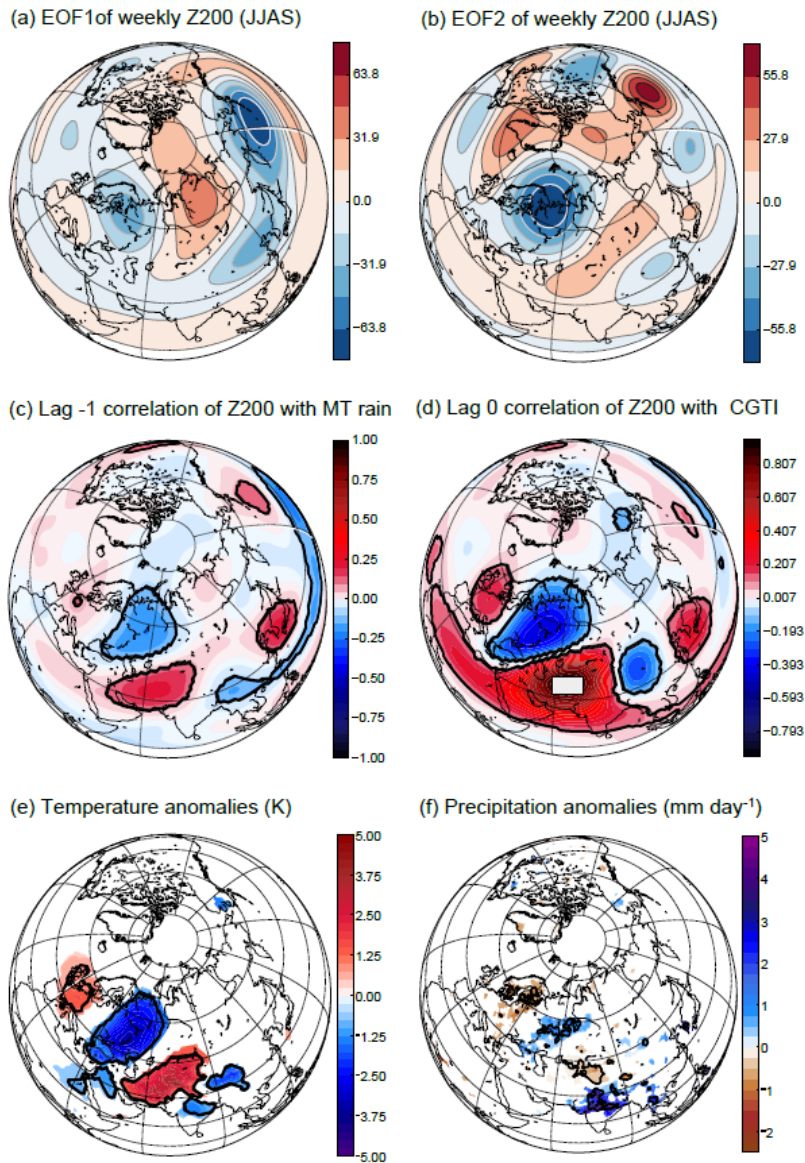


Figure S11. Mid-latitude variability associated with ISM. Panel (a) and (b): EOF1 and EOF2 for the JJAS weekly Z200 field in the northern mid-latitudes for the period 1979-2016. Panel (c): correlation between weekly MT rainfall and Z200 (lag = -1 week) for the period 1979-2016. Panel (d): the CGTI region (white box) and the correlation between CGTI and Z200 (lag = 0), which forms the circumglobal teleconnection pattern for the period 1979-2016. In panels c,d correlation coefficients and anomalies with a p-value of $p < 0.05$ (accounting for the effect of serial correlations) are shown by black contours. Panel (e): Temperature anomalies over the Northern Hemisphere during weeks with CGTI > 1 std of CGTI minus weeks with CGTI < -1 std of CGTI for the period 1979-2016. Panel (f): as panel (e) but for rainfall anomalies. In panels e,f anomalies with a p-value of $p < 0.05$ (accounting for the effect of serial correlations) are shown by black contours, while grid points significant with non-corrected p values are shaded.

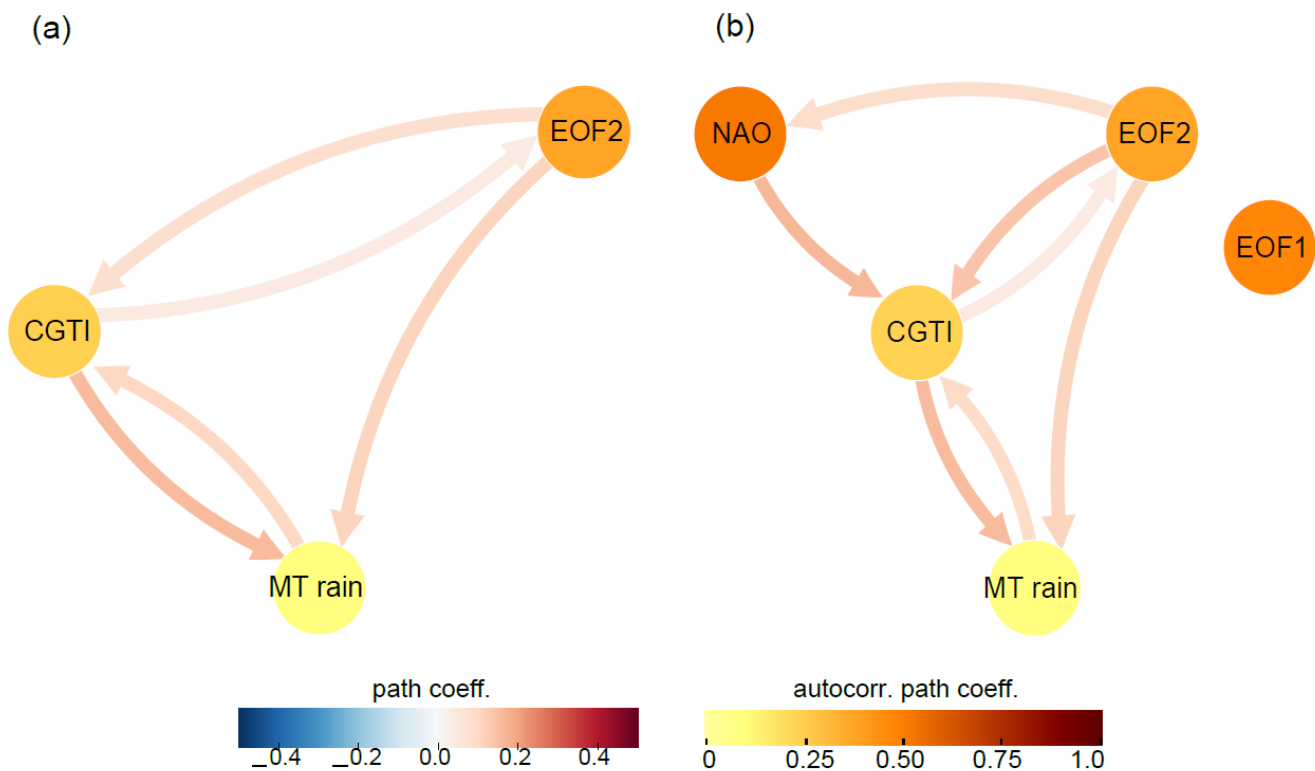


Figure S12. Causal mid-latitude interactions of the ISM. Panel (a): Causal Effect Network (CEN) built with CGTI, the PC of EOF2 and MT rainfall from CPC-NCEP dataset for the period 1979-2016. Panel (b): as panel (a) but with the addition of the EOF1 and NAO.

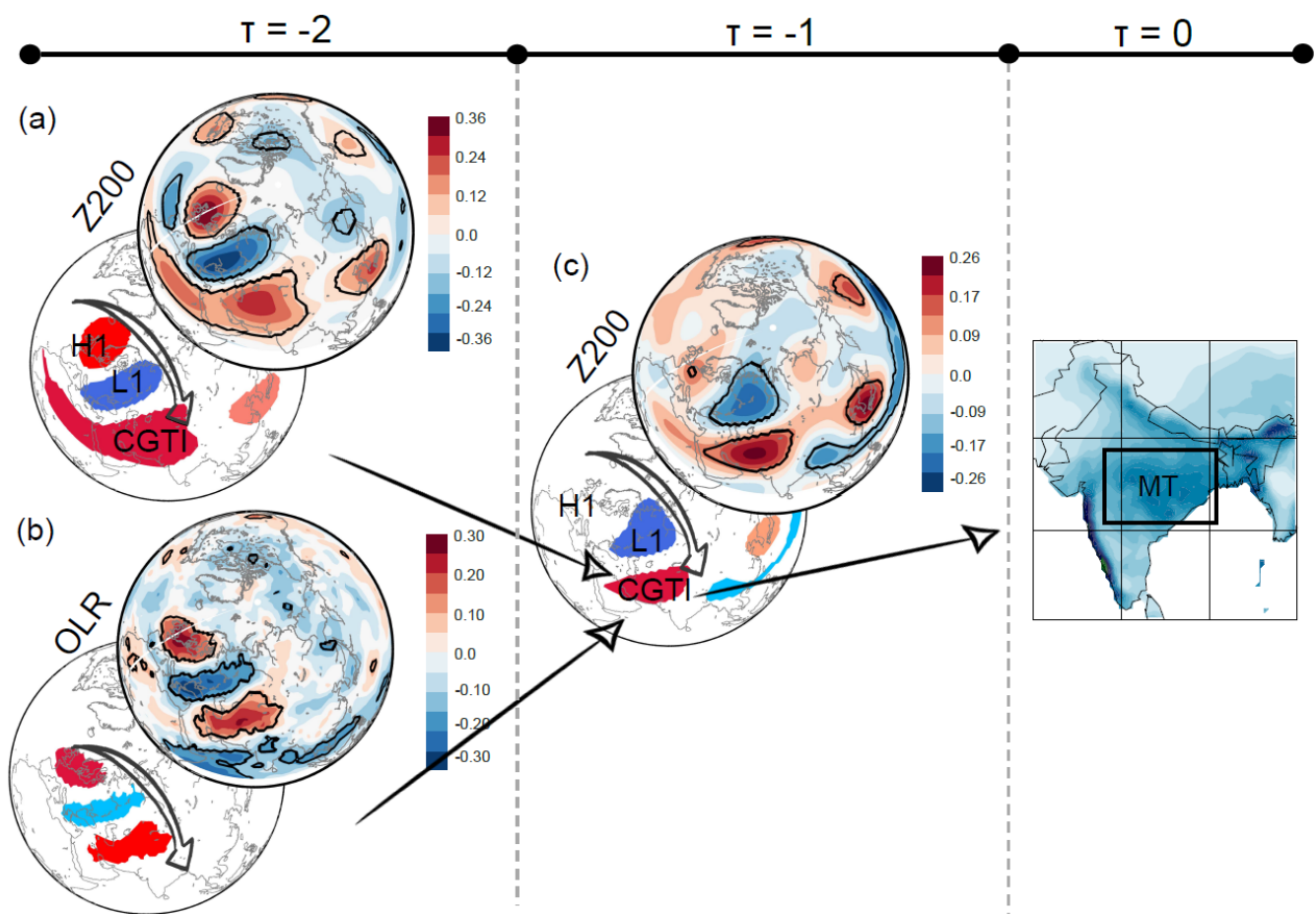


Figure S13. Mid-latitude causal precursors of ISM. Panel (a): correlation of CGTI with Z200 at 1-week lead time (top panel), and the causal precursors of CGTI identified via RG-CPD (bottom panel) for the period 1979-2016. Panel (b): as for panel (a) but for OLR fields. Panel (c): correlation map for weekly MT rainfall and Z200 field at 1-week lead-time (top panel) and the causal precursors identified via RG-CPD (bottom panel) for the period 1979-2016. Panel (d): ISM rainfall over the MT region from the CPC-NCEP dataset for the period 1979-2016.

185

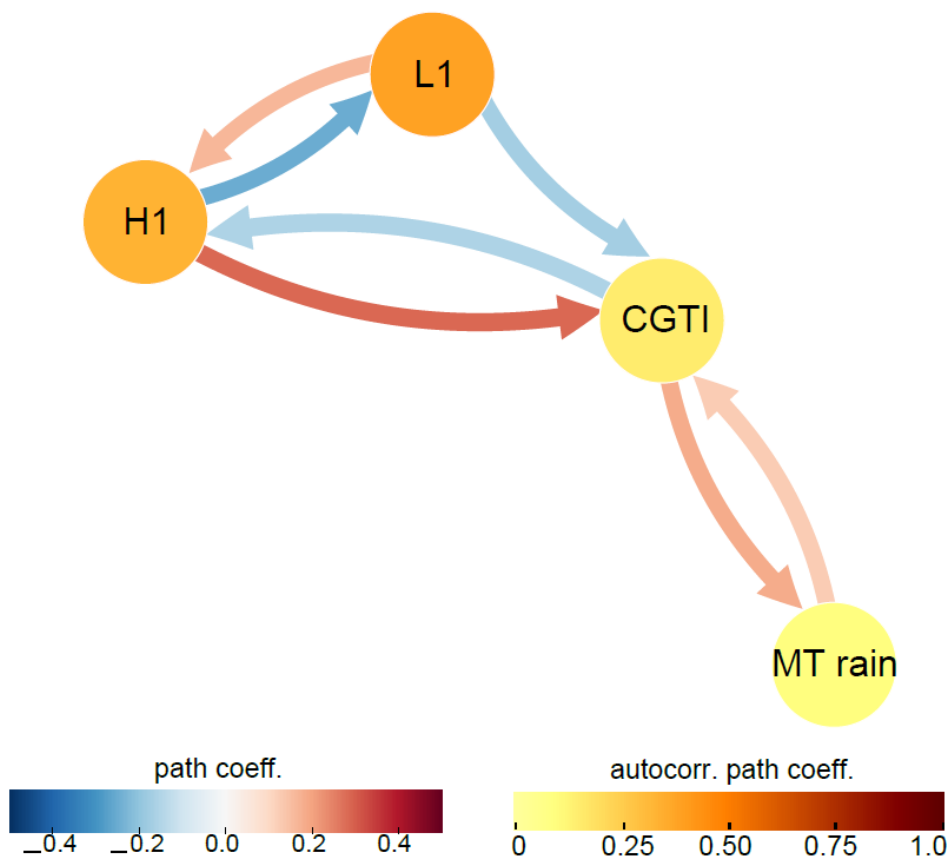


Figure S14. Mid-latitude wave train. CEN built with the MT from the CPC-NCEP dataset rainfall for the period 1979-2016, CGTI, L1 and H1 (as identified in Figure 5a).

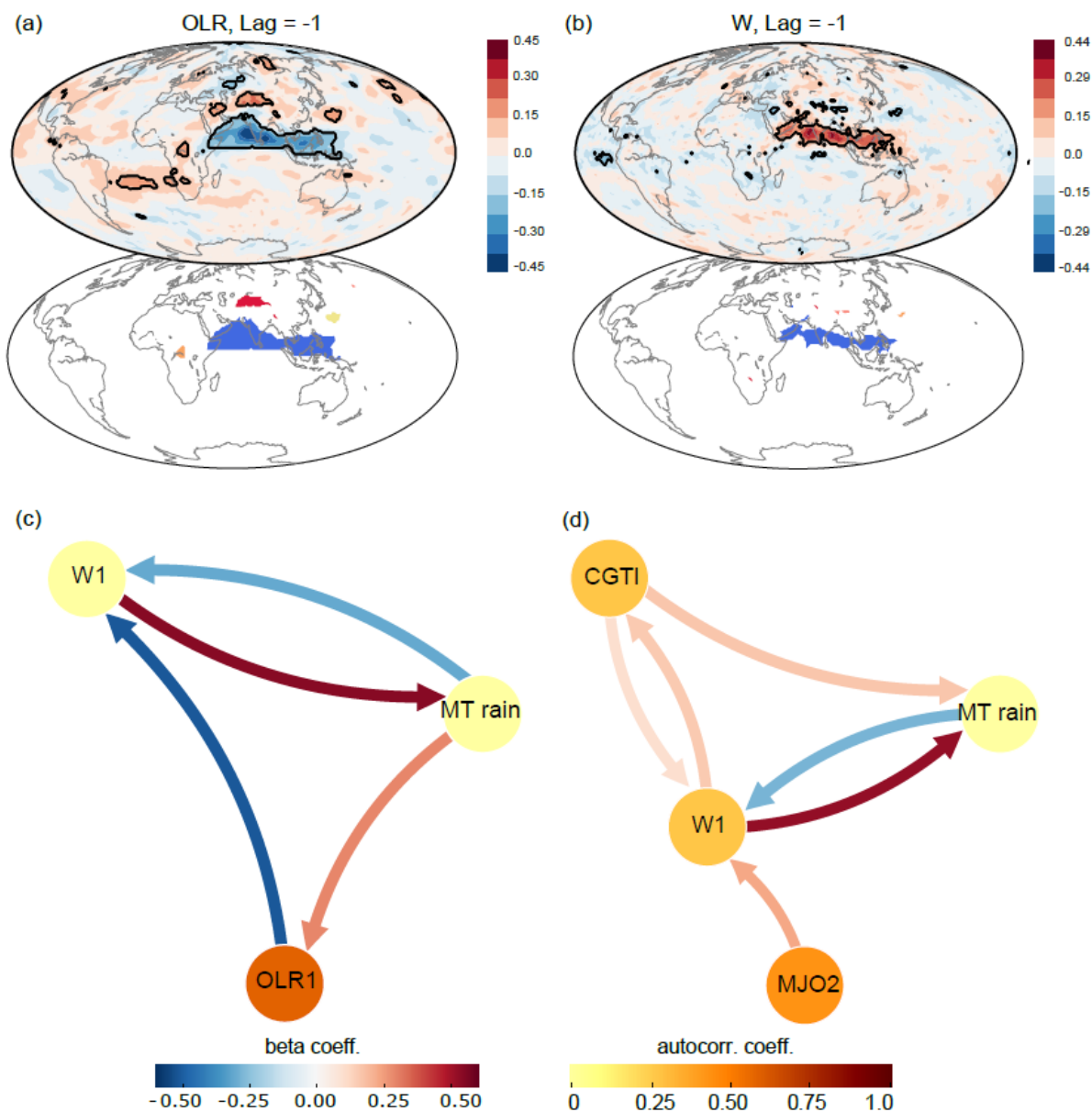


Figure S15. Tropical causal interactions of ISM. Panel (a) shows the correlation map for weekly MT rainfall from the CPC-NCEP dataset for the period 1979-2016 and the global OLR field at 1-week lead-time (top panel) and the causal precursors identified via RG-CPD (bottom panel). Panels (b): as for panel (a) but for W fields. Panel (c) and (d) show the CEN build with W1, OLR1 and MT rainfall and MT rainfall, W1, CGTI and MJO respectively.



200 **Figure S16. Combined mid-latitude and tropical causal interactions of ISM.** CEN built with W1, MJO2, MT rainfall from the CPC-NCEP dataset for the period 1979-2016, NAO, CGTI and EOF2.

Link	Link strength (CE)
$\beta_{W1 \rightarrow MT}$	0.52
$\beta_{MJO2 \rightarrow W1}$	0.35
$\beta_{CGTI \rightarrow MT}$	0.19
$\beta_{W1 \rightarrow MJO2}$	-0.39
$\beta_{MT \rightarrow W1}$	-0.25
$\beta_{EOF2 \rightarrow NAO}$	0.09
$\beta_{NAO \rightarrow CGTI}$	0.16
$\beta_{W1 \rightarrow CGTI}$	0.16
$\beta_{EOF2 \rightarrow CGTI}$	0.12
$\beta_{CGTI \rightarrow EOF2}$	0.06
$\beta_{CGTI \rightarrow W1}$	0.10

Table S1. Causal effect (CE) values. CE values for links presented in Fig. 7 in the main text.

	MT rainfall	NAO	CGTI	EOF2	W1	MJO2
ACE	0.050	0.032	0.070	0.043	0.214	0.070
ACS	0.141	0.019	0.088	0.011	0.141	0.079

Table S2. Average causal effect (ACE) and average casual susceptibility (ACS). ACE and ACS for actors presented in Fig. 7 in the main text.