Dear Editor,

Herewith we submit the revised version of our manuscript " Dominant patterns of interaction between the tropics and mid-latitudes in boreal summer: Causal relationships and the role of timescales" by Di Capua et al. to the Copernicus Journal *Weather and Climate Dynamics*.

We have addressed all comments and suggestions made by the two anonymous reviewers. The revised manuscript has improved both in clarity of the content and robustness of the analysis.

The revised version of the manuscript contains all changes as indicated in the two responses to the reviewers' comments published in the public discussion on WCDD. A point-by-point response to all reviewers' comments are found below, together with a "track changes" version of the manuscript.

We very much hope that *Weather and Climate Dynamics* will considers this manuscript for publication.

Yours sincerely,

Potsdam, 20.07.2020

Giorgia Di Capua on behalf of all authors

Response to review #1

We thank anonymous referee #1 for the constructive review and helpful comments that have greatly helped us to improve our work in the revised manuscript. The main improvement are summarized as follows:

- We have performed a series of sensitivity tests which show that our findings are robust, and which have reduced noise in some plots and thereby their visual appearance
- We have improved the description of the methodology section and the visualisation of the results
- We have added further background literature on the use of causality methods in atmospheric science

We have taken into account all suggestions made and a point-by-point response to each comment is reported below. Please note that in the following text the referee's comments are highlighted in bold font, while our answers are in regular font.

Specific comments

1. **There is no literature review of other approaches to identifying causal relationships in climate data. One example is I. Horenko, S. Gerber, T.J. O'Kane, J.S. Risbey and D. Monselesan (2017) On inference and validation of causality relations in climate teleconnections, (In Nonlinear and Stochastic Climate Dynamics. Cambridge University Press, Eds. C. Franzke and T.J. O'Kane)**
We thank the anonymous reviewer for this suggestion. We included in our revised version of the manuscript a paragraph briefly describing other causal approaches applied to atmospheric sciences (lines 114-121): "*In recent years, several approaches have been applied to identify causal relationships in climate and atmospheric sciences (Runge et al., 2019b), ranging from Granger causality (McGraw and Barnes, 2018, 2020; Samarasinghe et al., 2019) to causal (Bayesian) graphical models (Pearl, 2000, Ebert-Uphoff and Deng, 2012a, 2012b; Horenko et al., 2017) and conditional independence-based network discovery methods for time series (Runge et al., 2019a). These studies have shown the ability of causal discovery tools to improve our understanding of several atmospheric circulation interactions such as Arctic – mid-latitudes connections (McGraw and Barnes, 2020; Samarasinghe et al., 2019), synoptic-scale disturbances between boreal summer and boreal winter (Ebert-Uphoff and Deng, 2012a) and the relationship between ENSO and surface temperature in the American continent (McGraw and Barnes, 2018).* "

2. **The initial application of MCA appears to perform a basic dimension reduction. The authors assert that "expert knowledge" is required in choosing the particular variables to calculate the cross covariances however there is no indication that any other combinations were examined. For example, OLR could be replaced with velocity potential – as in indices for the MJO – with similar results.**
We thank the anonymous reviewer for suggesting this interesting test and we have expanded our analysis by considering the results obtained when other variables are used. We applied MCA on mid-latitude Z200 and tropical OLR because we are interested in studying the relationship between mid-latitude circulation patterns and tropical convection. Thus, we are focussing on variables representing tropical convection when attempting to provide a comparable analysis. We originally selected OLR because it captures strong convective clouds (which is a smoother signal than direct rainfall estimates), and because OLR is also used, for example, to define the BSISO index that describes the essential evolution of convective activity over the Indian Ocean region. In the revised version of the manuscript, we will provide a series of sensitivity tests for the identified MCA patterns by substituting OLR with velocity potential or vertical velocity (a proxy of convection). Figure S5 (in the revised version of the Supplementary Material) shows the first two MCA patterns for mid-latitude Z200 paired with tropical vertical velocity (note that upward motion has a negative sign since vertical velocity is expressed in Pa/s), while Fig. S6 shows the same for Z200 paired with velocity potential. The MCA patterns obtained when pairing vertical velocity with Z200 (figure S5) show highly consistent results with respect to those found for Z200 and OLR (Fig. 2 in the main text), demonstrating the robustness of the

original MCA results obtained with OLR. When we use velocity potential (Fig. S6), the MCA 1 pattern strongly resembles that originally obtained using OLR (with a wave-5 pattern in Z200 and low velocity potential over the Indian summer monsoon region). The MCA 2 pattern however shows less agreement: It correctly captures the OLR pattern in the western Indian Ocean but does not represent the WNPSM convective activity patterns. A reason for this discrepancy is that velocity potential provides a much smoother proxy for upper-level divergence than OLR, which is very strong in the Indian monsoon region, and apparently less pronounced in relation to the WNPSM. We briefly comment on this in the revised main text (lines 350-355): "*We also investigate whether the obtained MCA patterns are sensitive to the choice of OLR in representing tropical convective activity. Using vertical velocity, another proxy for tropical convection where strong convective activity is represented by enhanced upward motions, shows qualitatively the same patterns as those in Figs. 2b,d (see Fig. S5 in the Supplementary Material). When velocity potential is used instead of OLR, the first MCA mode still closely resembles the OLR/Z200 MCA mode 1, while the second MCA mode only partly captures features in the western Indian Ocean (see Fig. S6 in the Supplementary Material).*"

The methodology applied here seems to be unable to answer if a sufficient set of covariates has been chosen apart. How, for example, do you test if the combination of actors is sufficient or even parsimonius? Can some form of information theoretic approach be applied for example Akaike or Bayesian?

We consider causal discovery here and not a prediction task of any of the actors, for which criteria such as those mentioned are indeed important. Hence, the choice of included actors is subject to the hypothesis underlying the analysis setup. One could, however, phrase causal discovery, as in Granger's work, as a prediction problem. On the other hand, a causal interpretation rests on a number of assumptions and we discuss limitations related to causal sufficiency and other assumptions made in the discussion in the revised manuscript (lines 644-659): "*Finally, it should not be forgotten that in the context of the present work, causal interpretation rests upon several assumptions, such as the causal Markov condition, faithfulness, causal sufficiency, stationarity of the causal links and assumptions about the dependence-type (Runge, 2018). These assumptions can be violated in a real system and it is important to be aware of the associated typical challenges for causal discovery in Earth system sciences (Runge et al., 2019). Causal sufficiency requires that all relevant actors in a specific system are accounted for. Here, given the limited set of actors analysed, we cannot rule out that other excluded actors may act as important (common) drivers. Therefore, the obtained links can be considered causal only with respect to the specific set of actors used here. However, the absence of a link can still be interpreted as a likely indication that no direct physical connection among the respective variables exists. Moreover, we assume linear dependencies and stationarity for the detection of the causal links. While linearity has been shown to be a useful assumption in previous work (Di Capua et al., 2020), monsoon dynamics behaves partly nonlinearly and therefore, our causal networks only capture some part of the underlying mechanisms by construction. Also, the SAM teleconnections might well behave in an nonstationary manner on decadal time-scales (Di Capua et al., 2019; Robock et al., 2003). We therefore cannot rule out that (multi-)decadal oscillations such as the Pacific Decadal Oscillation may influence our results. However, the amount of reliable data is limited and this prohibits the application of nonlinear measures or study of effects of nonstationarity.*"

3. **Given the leading two modes of MCA appear to be in quadrature, how does MCA compare to EOF/PCA or even k-means?**

We thank the anonymous reviewer for raising this point. We have now performed a comparison between MCA patterns and EOF patterns. In Fig. S4 (in the revised version of the Supplementary Material), which will be included in the revised Supplementary Material, we show the first 5 EOF patterns for both Z200 and OLR. We calculate the spatial correlation between all EOF and MCA patterns. For Z200, MCA 1 shows the strongest correlation with EOF 2 (r ~ 0.8). This is consistent with previous literature showing that the circumglobal teleconnection pattern (as captured by Z200 of

MCA1), is linked to the second EOF of Z200 (Ding and Wang 2005, Di Capua et al. 2020). MCA 2 has a strong spatial correlation (r ~ 0.6) with EOF 1. For OLR, MCA 1 shows the strongest correlation with EOF 2 (r ~ 0.5), while MCA 2 has the strongest correlation with EOF 5 (r ~0.4). Thus, with only the exception of OLR MCA 2, all MCA patterns are closely related to the first two EOFs for both Z200 and OLR. This comparison shows that the identified MCA patterns are also on a regional level important in explaining the variability. Note that the fraction of variance explained is relatively low (for all EOFs), but this relates to the prior removal of interannual variability, thus leaving only the disturbances from the year-specific mean state. In our present work, we are interested in identifying those patterns that evolve simultaneously (due to the dynamical coupling between the two fields), and therefore we applied MCA to identify those patterns that can explain *shared* covariance, which is not captured by separate EOF analyses. We briefly comment on this in the main text (lines 342-349): "*We compare the patterns obtained with MCA with those obtained with EOF analysis of Z200 and OLR fields (see Fig. S4 in the Supplementary Material). We find that the closest match of the Z200 MCA mode 1 pattern is with Z200 EOF 2 (spatial correlation ~ 0.8), while the closest match of Z200 MCA mode 2 is with EOF 1 (spatial correlation ~ 0.6). OLR MCA mode 1 has the closest match with EOF 2 (spatial correlation ~ 0.5), while OLR MCA mode 2 has the closest match with EOF 5 (spatial correlation ~ 0.4). Thus, in general our MCA patterns also reflect the first two EOFs of Z200 and OLR indicating that they explain an important fraction of the regional variability. Nevertheless, here we are interested in those patterns that can explain shared covariance, which cannot be achieved by using EOF analysis alone. Therefore, we use the MCA-defined patterns for the following part of the analysis.*" .

Apparently, many of the underlying assumptions are the same i.e stationarity etc It would help greatly if the authors could indicate if their approach is causal in the sense of Grainger given there appears to be no underlying stochastic model?
Our definition of causal graphs follows Pearl's causal Bayesian networks (Pearl 2000) and our approach to estimate these graphs from data comes from the constraint-based causal discovery framework (Spirtes 2000), here adapted to time series (Runge et al. 2019). In the constraint-based causal discovery framework, the existence (or absence) of causal relations is based on conditional independencies among subsets of the lagged variables together with a number of assumptions (as listed in our Discussion section). If Granger causality is only applied to pairs of variables, Granger causality does not account for common drivers or indirect links as is the case in our framework. Further, the constraint-based causal discovery framework *in general* goes beyond Granger causality since it can also account for contemporaneous causal links. Here we only focus on lagged links. If Granger causality is meant in a full multivariate setting, our approach is asymptotically equivalent to Granger causality, but for finite samples Granger causality has much lower detection power since it does not deal well with the curse of dimensionality as investigated in detail in Runge et al. (2019).

4. **The analysis and attribution of the causal relationships is ultimately largely empirical, at times overly complicated and in some parts exceedingly verbose in description. The "causal maps" are very noisy and the reported relationships are very poorly represented from the patterns in the causal maps presented.**
We have taken the issue of noisiness raised by the anonymous reviewer very seriously, and combining this suggestion with the corresponding comment by anonymous reviewer #2, we have designed a robustness test that has removed much of the noise in the causal maps, greatly improving their visual appearance and interpretation. As a result, some of the more scattered regions that were described in the first version of the paper are now removed, and we can purely focus our description on the main, robust patterns. We describe this robustness in the revised manuscript (lines 295-303): "*Finally, to test the robustness of our causal maps to the choice of time period, we calculate causal maps for a range of sub-periods. In 10 trials we removed 10% of the record (4 years). For ENSO-phase dependent causal maps, we have shorter time series and we thus remove one year in each trial, leaving a set of 14 causal maps for La Niña events and 13 causal maps for El Niño events. As a result, we obtain an*

ensemble of causal maps and apply the false discovery rate correction to p-values of each single map. Then, both for the full period (1979-2018) and for El Niño and La Niña years separately, we masked out areas where less than 70% of the trials indicated a significant causal link, giving an indication of the robustness of our findings and at the same time suppressing noise."
This results in reduced noise in the new causal maps (see Fig. 3-5 in the revised version of the manuscript).

5. **It would greatly help the reader if the methodology was described in sufficient detail and better placed in context with other approaches, both in terms of dimension reduction and causal inference. This, in combination with a more concise discussion of the physical properties of the modes would allow the reader to better judge the merits of the approach.**
In the revised manuscript, we have improved the methodology section by adding a concrete example showing how the PCMCI algorithm works (also following the comments by the second reviewer, see point 1 in our response to reviewer #2), lines 213-257:

"*In this analysis, A and B represent the two MCA scores obtained for a selected MCA mode, while C(lat,lon) represents the grid point time series of a 2D field, e.g. T2m or Z200. In its first step, PCMCI iterates through partial correlations with increasing cardinality of conditions to remove the influence of common drivers and indirect links and estimate a preliminary set of parents. The first iteration of PC (cardinality 0) calculates the correlation between a selected time series, e.g.* $A_{\tau=0}$*, and the past of any other available time series,* $\{ A_{\tau=-1}, B_{\tau=-1}, C(lat,lon)_{\tau=-1}, \ldots, A_{\tau=-\tau max}, B_{\tau=-\tau max}, C(lat,lon)_{\tau=-\tau max}\}$*, including its own past* $A_{\tau=-1,\ldots,-\tau max}$*. For illustration purposes, we here provide an example for C(lat,lon), where ρ denotes the correlation and τ is the lag that is being used in the network (in this example,* $\tau_{max} = -2$*):*

$$\rho(C(lon,lat)_{\tau=o}, A_{\tau=-1}) = 0.32, p = 0.01 \quad (5)$$

$$\rho(C(lon,lat)_{\tau=o}, A_{\tau=-2}) = 0.13, p = 0.1$$

$$\rho(C(lon,lat)_{\tau=o}, B_{\tau=-1}) = 0.35, p = 0.005$$

$$\rho(C(lon,lat)_{\tau=o}, B_{\tau=-2}) = 0.23, p = 0.058$$

$$\rho(C(lon,lat)_{\tau=o}, C(lon,lat)_{\tau=-1}) = 0.41, p = 0.01$$

$$\rho(C(lon,lat)_{\tau=o}, C(lon,lat)_{\tau=-2}) = -0.16, p = 0.06$$

Applying a significance level α = 0.05, only three actors are significantly correlated with C(lat,lon) at the chosen time lag. These form the initial preliminary set of parents for C(lat,lon) and are ordered by the strength of their correlation:

$$P^0_{C(lon,lat)} = \{C(lat,lon)_{\tau=-1}, B_{\tau=-1}, A_{\tau=-1}\} \quad (6)$$

Next, partial correlations between C(lat,lon) and each actor in $P^0_{C(lon,lat)}$ *are calculated by conditioning on the strongest preliminary parent:*

$$\rho(C(lat,lon)_{\tau=o}, C(lat,lon)_{\tau=-1} | B_{\tau=-1}) = 0.35, p = 0.02 \quad (7)$$

$$\rho(C(lat,lon)_{\tau=o}, B_{\tau=-1} | C(lat,lon)_{\tau=-1}) = 0.28, p = 0.03$$

$$\rho(C(lat,lon)_{\tau=o}, A_{\tau=-1} | C(lat,lon)_{\tau=-1}) = 0.25, p = 0.04$$

Parents with significant partial correlations will enter the second set of preliminary parents:

$$P^1_{C(lat,lon)} = \{C(lat,lon)_{\tau=-1}, B_{\tau=-1}, A_{\tau=-1}\} \quad (8)$$

Next, the partial correlation is calculated conditioning on the two strongest parents:

$$\rho(C(lat,lon)_{\tau=o}, C(lat,lon)_{\tau=-1}|B_{\tau=-1}, A_{\tau=-1}) = 0.31, p = 0.03 \quad (9)$$

$$\rho(C(lat,lon)_{\tau=o}, B_{\tau=-1}|C(lat,lon)_{\tau=-1}, A_{\tau=-1}) = 0.23, p = 0.04$$

$$\rho(C(lat,lon)_{\tau=o}, A_{\tau=-1}|C(lat,lon)_{\tau=-1}, B_{\tau=-1}) = 0.12, p = 0.08$$

Since it is not possible to further increase the dimension of the condition set, from the PC step, the preliminary parents converge to:

$$P^2_{C(lon,lat)} = \{C(lat,lon)_{\tau=-1}, B_{\tau=-1}\} \quad (10)$$

By repeating this step for each variable, preliminary sets of parents are estimated. Let's assume that in our example we also obtain:

$$P^3_A = \{\, B_{\tau=-1}\,, A_{\tau=-2,}\} \quad (11)$$

$$P^2_B = \{\, B_{\tau=-1}\}$$

In the MCI step, partial correlation is calculated again between each pair of actors (at different time lags) conditional on the above estimated sets of preliminary parents, whereby both sets of parents are conditioned upon. To give one example, this would lead to:

$$\rho\big(C(lat,lon)_{\tau=o}, A_{\tau=-1}\big|P^2_{C(lat,lon)}, P^3_A\big) =$$

$$= \rho(C(lat,lon)_{\tau=o}, A_{\tau=-1}|C(lat,lon)_{\tau=-1}\, B_{\tau=-1}, B_{\tau=-2}, B_{\tau=-3}) = 0.1, p = 0.3 \quad (12)$$

Note that the parents of $A_{\tau=-1}$ are shifted in time by τ = -1. After repeating (12) for each pair of actors shown in (5) and for time lags from 0 to τ_{max}, those parents that are significant in the MCI test will then form the final set of causal parents for each actor. We refer to Runge et al. (2019a) for a more detailed discussion and explanation of the algorithm design and extensive numerical experiments."

Moreover, we have improved the physical interpretation of each mode in the results section (lines 314-341):

"*The first MCA mode explains 18% of the squared covariance (squared covariance fraction, SCF) and shows a CGT-like wave-5 pattern in mid-latitude Z200. The Pearson correlation between the two time series of MCA scores for the first mode is r ~ 0.5. The spatial correlation with the weekly CGT pattern, as defined by Ding and Wang 2005, is 0.52 (Fig. 2a). The CGT pattern also represents the second most important pattern in boreal summer mid-latitude circulation (Di Capua et al., 2020; Ding and Wang, 2005). This wave-5 pattern is linked to the South Asian monsoon (SAM) activity via its positive centre of action east of the Caspian Sea (see Fig. 2a). Applying MCA, we find that the CGT pattern co-varies with a band of enhanced tropical convective activity that extends from the Arabian Sea towards Southeast Asia, with a peak of convective activity over the Bay of Bengal (Fig. 2b) (Kang et al., 1999).. Using OLR composites and the Kikuchi Boreal Summer Intraseasonal Oscillation (BSISO) index, we explicitly show that the temporal evolution of the SAM convective activity as defined in Fig. 2b at weekly time-scales resembles the evolution of the BSISO (Goswami and Ajaya Mohan, 2001; Saha et al., 2012) (see Figs. S1-S2 and further discussion in the Supplementary Material). Therefore, we explicitly link the region of low OLR identified in Fig. 2b over the northern Indian Ocean and the Indian subcontinent to the SAM activity as described in the literature. Note that we name each MCA pattern after a characteristic regional feature, but the analysis is applied to the larger geographical domains as shown in Figure 2.*

The second mode of co-variability explains a SCF of 14% and is characterized by a region of strong positive Z200 anomalies located at ~ 45° N, over the western North Pacific, directly to the west of the dateline (i.e. the most prominent centre of action of the mid-latitude wave). The Pearson correlation between the two time series of MCA scores for the first mode is r ~ 0.6. We will refer to this pattern as the North Pacific High (NPH) (Fig. 2c). The NPH is the summer counterpart of the North Pacific subtropical high, which characterizes boreal winter. During summer, this high pressure region is displaced northward by the start of the monsoon season in the western Pacific Ocean and replaces the Aleutian Low (Lu, 2001; Riyu, 2002). The NPH is associated with a region of enhanced convection over the sub-tropical western North Pacific, related to the western North Pacific summer monsoon (WNPSM) convective activity (Fig. 2d) (Li and Wang, 2005; Nitta, 1987; Wang et al., 2001). The WNPSM core domain extends from 110°-160°E and 10°-20°N, while the boundary with the SAM is located over the South China Sea (Murakami and Matsumoto, 1994). The WNPSM is characterized by a late sudden onset (end of July) and a peak in rainfall activity during August and September, which is different from the SAM that features an earlier onset (in June) and peak rainfall activity during July-August.".

Response to review #2

We thank anonymous referee #2 for the constructive review and helpful comments that have greatly helped us to improve our work in the revised manuscript. The main improvements in the response to reviewer #2 are summarized as follows:

- We have added a robustness test to check the sensitivity of the detected causal links when the time period is changed
- We have improved the visualization of the causal maps by reducing the noise and adding labels to better identify each region when described in the text
- We have expanded the explanation in the Methods section
- We have checked the description of each region in the Results section

We have taken into account all suggestions made by the reviewer and a point-by-point response to each comment is reported below. Please note that in the following text the referee's comments are highlighted in bold font, while our answers are in regular font.

Specific comments

1) Clarification on methodology: - Section 2.2: The choice of MCA is not clear as compared to other methods of dimension reduction.
We choose MCA over other methods of dimension reduction because we are interested to identify those patterns that evolve simultaneously and may be causally related (via e.g. dynamical coupling between multiple variables). Thus, we applied MCA to identify those patterns that can explain shared covariance, which is an objective that cannot be addressed by using EOF analysis alone. We explain this point explicitly in the revised manuscript (lines 169-174): "A*mong the available correlation based methods to highlight strong co-variability and reduce the dimensionality of a spatiotemporal dataset, MCA allows identification of patterns in pair of variables that evolve simultaneously and may be causally related (via e.g. dynamical coupling between multiple climatological fields). MCA detects patterns that can explain shared covariance, which cannot be achieved using other dimensionality reduction methods that consider individual variables separately, such as empirical orthogonal function (EOF) analysis. However, for providing a complete picture we will also discuss the corresponding EOF patterns and the fraction of variance explained for comparison with our MCA results.*"

It would be helpful to describe what will happen with MCA modes after section 2.2.
We now explain in more detail what happens to times series identified by using MCA in section 2.2 (lines 188-190): "*Here, we select the first two MCA modes representing the dominant patterns of co-variability between tropical convection and mid-latitude circulation, and calculate time series for each MCA mode. These time series will be used as input for the causal discovery algorithm (see sections 2.3 and 2.4).*"

In Figure 2, the legend suggests four time series but one can only recognize two time series.
We agree that in the first version of Fig. 2 it was difficult to recognise two time series. We have changed the colours to represent the two pairs of time series and adopted a different aspect ratio for the axes to better show the four time series. See Fig. 2 in the revised version of the document document.

Section 2.4 is very generic; it would also be useful to know at some point what "A, B, C" are in the current analysis.
We will include this suggestion in the revised manuscript. In the Results section, we will make explicit how the variables used compare to the examples given in the method section. For example lines 379-381: "*Referring to the schematic illustrated in Fig. 1 and following the PCMCI algorithm explanation (section 2.3), here A and B time series are represented by the SAM and CGT time series respectively,*

while C(lon, lat) is represented by Z200, OLR and T2m fields." Moreover, we have also added a more detailed explanation on how these time series are used in the causal discovery algorithm (see response to reviewer #1, point 4).

Adding a table describing indexes and abbreviations separated in cause and response actors used for the causal effect analysis would be helpful.
Following the reviewer's suggestion we have added a table (Table 1 in this in the revised version of the manuscript) to better identify each time series/field used (see also point 5 in this response).

A discussion on the sensitivity of results to data-length would also be useful.
We address this comment by providing a robustness test by repeated calculation of the causal maps and screening for robust regions in the final results. This step also makes the causal maps less noisy, such that robust patterns emerge better, improving the visual appearance and interpretability of Figures 3, 4 and 5 in the revised version of the manuscript. We describe in detail how this test is performed (lines 295-303): "*Finally, to test the robustness of our causal maps to the choice of time period and to reduce non-robust small-scale features, we repeatedly calculate causal maps for reduced time series length. In 10 trials we removed a consecutive time record of ~10% (4 years) of the entire period. For ENSO dependent causal maps, we have shorter time series and we thus remove only one year in each trial, leaving a set of 14 causal maps for La Niña events and 13 causal maps for El Niño events. As a result, we obtain an ensemble of causal maps and apply the false discovery rate correction to their p-values. Then, both for the 1979-2018 period and for El Niño and La Niña years separately, we masked out areas where less than 70% of the trials indicated a significant causal link. This gives an indication of robustness of our findings and suppresses noise.*" The masks obtained in this way and used to produce new Fig. 3,4 and 5 are shown in the revised Supplementary Material (Figs. S8,S9 and S12).

L219-223 and L310-315 should be in Methods because this text describes methodology and not the results.
We have moved those lines "*To extract the dominant co-variability patterns reflecting interactions between mid-latitude circulation in the Northern Hemisphere and tropical convection at intraseasonal time-scales, we follow Ding et al. (2011) and apply maximum covariance analysis (MCA) to OLR fields (used as a proxy for convective activity) in the tropical belt (15°S-30°N, 0°-360°E) paired with Z200 fields in the northern mid-latitudes (25°N-75°N, 0°-360°E).*" (now lines 162-166) and "*Here, we will derive causal maps using the time series obtained with MCA for modes 1 and 2 and Z200, OLR and T2m fields both for the entire time period (1979-2018) and for two subsets depicting different ENSO phases, to assess how the ENSO background state influences the causal relationships. El Niño (La Niña) summers are defined as summers preceding the El Niño (La Niña) peak in boreal winter. We thus obtain 14 La Niña years and 13 El Niño years (see Table 1 in the Supplementary material for a list of corresponding years and Fig. S1 for the associated SST anomaly composites). Although the strongest SST anomalies related to the ENSO phase are found in winter, warm (cold) SST patterns related to El Niño (La Niña) phases are already clearly developed during the preceding summers.*" (now lines 288-295) to the Methods section as suggested.

2) **Clarification on results and discussion: L250-259: It is not clear what the purpose of this paragraph is.**
We agree with anonymous reviewer #2 that a detailed description of BSISO can distract the reader from the main story line. We have moved this explanation into the SI and now refer to it only briefly in the main text (lines 322-325): "*Using OLR composites, we explicitly show that the temporal evolution of the SAM convective activity at weekly time-scales resembles the evolution of the Boreal Summer Intraseasonal Oscillation (BSISO) (Goswami and Ajaya Mohan, 2001; Saha et al., 2012) (see Fig. S5-S6 and further discussion in the Supplementary Material).*"

L266-268: Mentioned patterns do not look "similar" at all to me. I would suggest to specify

regions where similarities are seen by authors.
We thank the anonymous reviewer for pointing out that it was difficult to recognize in the figures the regions that we are interpreting in the text. We have addressed this comment by adding labels that are referred to in the main text, including the Results section. See new Figs. 3,4 and 5 in this document.

Explaining some of the results, authors interpret patches of beta-values on causal maps that look like noise. E.g., L280: "Although the CGT influence is mostly concentrated in the mid-latitude regions, one can see a negative causal effect of the CGT pattern on OLR values over the Bay of Bengal (Fig. 3f).“ It looks like the effect that authors describe is a small dash over the Bay of Bengal, I cannot even see the color of the region, just the black contour color. Does the method behind causal maps take care of spatial noise?
We have taken the issue of robustness and potential noise in our causal maps seriously (see also our earlier reply and Figs. S8,S9 and S12 in the revised version of the Supplementary Material). The new figures 3, 4 and 5 (in the revised version of the manuscript) are now all produced using the robustness test described above (see our response to comment #1). As a result, the specific region described on line 280 (original manuscript) is now indeed masked out and we have updated the text correspondingly (lines 400-403): "*The CGT influence is mostly concentrated in the mid-latitude regions, and a significant and consistent negative causal effect of the CGT pattern on OLR values in the tropical regions can only be seen in a small area in the western Indian Ocean (Fig. 3f).*" In general, using the robustness test described above noisy patterns have been removed, enabling us to only discuss the main, large-scale patterns of interest.

L282: "Asia and North America are strongly affected by the CGT.“ It would be useful to support the qualitative judgment of the link-strength by providing beta-coefficient values in parentheses for this particular example and throughout the text, where link's strength from causal maps is described.
We thank the anonymous reviewer for this useful suggestion. We will add the values of the beta coefficients throughout the revised text to help the reader in the interpretation of the results.

L455: "apparent paradox": I am not sure there is any paradox. Studies cited by the authors describe a trend in current observations and future climate change projections, which cover two different time periods, thus such comparison is not consistent.
We have removed the sentence referring to the apparent paradox and rephrased the paragraph to make our point more carefully. The revised paragraph now reads (lines 620-626): "*Future projections describe an increase in monsoon precipitation associated with increasing global mean temperature and thermodynamic arguments (Menon et al., 2013; Turner and Annamalai, 2012). Quantifying teleconnections between the tropics and mid-latitudes is important in order to better understand and constrain future changes in boreal summer circulation, as uncertainty may arise due to changing connections to remote regions. While simulations show great uncertainty in the ENSO response to global warming (Cai et al., 2015; Chen et al., 2017a, 2015, 2017b), observations show a La Niña-like warming trend in central-western Pacific SST (Kohyama et al., 2017; Mujumdar et al., 2012).* "

L435-440: A comparison of teleconnections acting on subseasonal timescales from this study with those from other studies on interannual and decadal timescales is odd.
By comparing interannual and intraseasonal studies, we do not intend to imply that a similarity in the results obtained at different time scales *should* be expected. Nevertheless, a similarity in the pattern *is found* and this represents an outcome of our analysis that we believe needs to be discussed. In the discussion, we elaborate on what possible explanations for these findings there may be. Moreover, the similarity in these patterns between various time scales strongly suggests that there are interactions between the time scales – see for example the arguments of Sperber et al. (2000) who found a common mode of variability on intraseasonal and interannual time scales. Such commonality of patterns is necessary in order for the large scale forcing to be able to perturb the PDF at shorter time scales. See: Sperber et al. (2000) "Predictability and the relationship between subseasonal and interannual variability during the Asian summer monsoon", Quarterly Journal of the Royal Meteorological Society, 126: 2545-2574.

L56 and L496: A statement about paving the way to better predictions without further explanation is a bit bold. The CEN method has a potential to improve our understanding of climate processes but authors need to explain better how exactly this method can improve climate predictions.
We have added more in depth information on how CEN may help improving seasonal forecast in the revised version of the manuscript (lines 330-339): "*A better understating of these teleconnections in observation can help to improve S2S forecasts. Verifying the existence and strength of causal teleconnections in forecast models, could help diagnose the origin of model biases. E.g. one could disentangle whether lower forecast skill (such as in the mid-latitude regions in summer) is related to local processes or to a misrepresentation of remote drivers. Beverley et al. (2019) showed that the CGT representation in seasonal forecasts is too weak. The CGT is important for predictability of summer extremes and its relationship with the SAM may provide some information to improve predictability. Therefore, these methods could help answering the question "where do model biases come from?" and help developing a physics-based bias correction. At the same time, CEN provide an encoded predictive model, which can be used for actual forecasting (Di Capua et al., 2019; Kretschmer et al., 2017; Lehmann et al., 2020).*"

3) **Inaccurate region description: L295: "Russia/Scandinavia": I would say "northern and eastern Europe" because this where non-zero beta values actually are. On the other hand, what does "non-corrected p values" from the caption mean, I do not find it explained.**
We now show only p-values that are corrected using the false discovery rate correction, to reduce noise and non-robust results. We have also carefully checked the description of each region in the Result section.

L323: "over Kazakhstan" I would say "north of Kazakhstan" if the region enclosed by the contour is meant. Moreover, Kazakhstan is located north-east of the Caspian Sea not north-west of the Caspian Sea.
This region did not pass the new robustness test and was removed.

L319: "a few areas": Indeed these are three regions which can be named.
We have added regional labels in the causal maps and use those references in the text.

L412: "European Russia". I would rather say "northern and eastern Europe".
We will implement this suggestion in the revised manuscript.

4) Figure 5: During El Nino years, there is a link between SAM and Z200 in the tropical Pacific, which is not present during the La Nina years, therefore the concluding statement in the results, conclusions and abstract about strong effect of El-Nino only for the second MCA mode is confusing.
We thank the anonymous reviewer for pointing out this discrepancy. We now mention that both phases of ENSO affect the relationship between SAM and Z200 (lines 490-493): "*Thus, the second MCA mode (the WNPSM-NPH pair) has its strongest effect during El Niño summers, whereas the first MCA mode (SAM-CGT pair) is important during both La Niña and El Niño summers but with different characteristics*" and "*Nevertheless, during La Niña summers, the effect of the SAM-CGT mode is reinforced over Europe, North Africa and the Indian subcontinent and reaches northward towards Canada while during El Niño summers the effect of the SAM is mainly confined to the tropical belt. For the WNPSM-NPH pattern, a clear asymmetry between El Niño and La Niña summers is shown, with a stronger signal during El Niño (Fig. 5e,f) that is absent during La Niña years.*" (lines 596-600).

NPH and mode 2 results are not described in the text.
We now describe the results related to Mode 2 (lines 485-490): "*In the western North Pacific, the most notable feature is the presence of both the WNPSM and NPH on the North Pacific only during El Niño*

summers (Figs. 5e,f). During those summers, the positive causal effect of the WNPSM over the western North Pacific (Region 1 and 2 in Fig. 5e) intensifies in magnitude (absolute beta ~ 0.3-0.4) relative to the 1979-2018 mean pattern (Fig. 4c), although the geographical extent of Region 1 shrinks. Over the western tropical Pacific, in correspondence with the La Niña warm pool, a region of positive causal effect is shown (Region 2 in Fig. 5e). These features disappear during La Niña summers."

L417: "the pattern identified in Fig 5f with a low over central Europe and high over western Russia". I do not see a low-high dipole, the figure shows beta coefficients not geopotential. We have removed this sentence as this statement in not supported by the stricter robustness test applied in the new causal maps.

L419: ": : :wave-trains initiated by La Nina: : :" I do not follow this explanation.
We have removed this sentence as this statement in not supported by the stricter robustness test applied in the new causal maps.

Figure 5f is about El Nino effects. Similarly, L456-458: ": : : if La Nina conditions would become: : :(Fig. 5f)". Figure 5f is about El Nino effects.
This mistake had been corrected by including the correct panel for Fig. 5c.

5) An extensive use of abbreviations makes the paper a bit difficult to follow. – Adding a table describing CEN actors abbreviations would be very helpful. - Abbreviation is introduced but never used in the manuscript such as EASM (L92) and SRP (L439). - BSISO abbreviation in L138 is not introduced.
Following the suggestion of the anonymous reviewer, we have added a table showing the full name of each abbreviation used throughout the manuscript and, when useful, its dimensions. We have removed abbreviations for EASM and SRP since they are not used later in the text. We now introduce the term BSISO both at its first appearance and in the abbreviation table.

Dominant patterns of interaction between the tropics and mid-latitudes in boreal summer: Causal relationships and the role of ~~time-scale~~timescales

Giorgia Di Capua[1,2], Jakob Runge[3], Reik V. Donner[1,4], Bart van den Hurk[2,5], Andrew G. Turner[6,7], Ramesh Vellore[8], Raghavan Krishnan[8], and Dim Coumou[1,2]

[1]Potsdam Institute for Climate Impact Research, Potsdam, Germany
[2]VU University of Amsterdam, Institute for Environmental Studies, Amsterdam, Netherlands
[3]German Aerospace Centre, Institute of Data Science, Jena, Germany
[4]Magdeburg-Stendal University of Applied Sciences, Magdeburg, Germany
[5]Deltares, Delft, Netherlands
[6]Department of Meteorology, University of Reading, Reading, United Kingdom
[7]National Centre for Atmospheric Science, University of Reading, Reading, United Kingdom
[8]Indian Institute for Tropical Meteorology, Pune, India

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

Abstract. Tropical convective activity represents a source of predictability for mid-latitude weather in the Northern Hemisphere. In winter, the El Niño–Southern Oscillation (ENSO) is the dominant source of predictability in the tropics and extra-tropics, but its role in summer is much less pronounced and the exact teleconnection pathways are not well understood. Here, we assess how tropical convection interacts with mid-latitude summer circulation at different intraseasonal ~~time-scale~~timescales and how ENSO affects these interactions. First, we apply maximum covariance analysis (MCA) between tropical convective activity and mid-latitude geopotential height fields to identify the dominant modes of interaction. The first MCA mode connects the South Asian monsoon with the mid-latitude circumglobal teleconnection pattern. The second MCA mode connects the western North Pacific summer monsoon in the tropics with a wave-5 pattern centred over the North Pacific High in the mid-latitudes. We show that the MCA patterns are fairly insensitive to the selected intraseasonal ~~time-scale~~timescale from weekly to 4-weekly data. To study the potential causal interdependencies between these modes and with other atmospheric fields, we apply the causal ~~effect networks (CEN)~~discovery method PCMCI at different ~~time-scale~~timescales. ~~CENs~~ PCMCI extends standard correlation analysis by removing the confounding effects of autocorrelation, indirect links and common drivers. In general, there is a two-way causal interaction between the tropics and mid-latitudes but the strength and sometimes sign of the causal link are ~~time-scale~~timescale dependent. We introduce causal maps that plot the regionally specific causal effect from each MCA mode. Those maps confirm the dominant patterns of interaction and in addition, highlight specific mid-latitude regions that are most strongly connected to tropical convection. In general, the identified causal teleconnection patterns are only mildly affected by ENSO and the tropical-mid-latitude linkages remain similar. Still, La Niña strengthens the South Asian monsoon generating a stronger response in the mid-latitudes, while during El Niño years, the Pacific pattern is reinforced. This study paves the way for process-based validation of boreal summer teleconnections in (sub-)seasonal forecast models and climate models and therefore works towards improved ~~helps to improve~~ sub-seasonal and climate projections.

1 Introduction

Tropical – mid-latitude teleconnections in boreal summer can have a great impact on surface weather conditions in the northern mid-latitudes (Ding and Wang, 2005; O'Reilly et al., 2018; Wang et al., 2001). Still, the direct influence of the El Niño-Southern Oscillation (ENSO) on the mid-latitude circulation is weaker in summer than in winter (Branstator, 2002; Schubert et al., 2011; Thomson and Vallis, 2018). Instead, in summer, convective activity related to the Northern Hemisphere tropical monsoon systems can profoundly influence surface weather conditions in the mid-latitudes (Branstator, 2014; Ding and Wang, 2005; O'Reilly et al., 2018; Rodwell and Hoskins, 1996). Vice versa, mid-latitude wave trains and cyclonic activity at intraseasonal ~~time-scale~~timescales can modulate the tropical monsoons, and have been linked to extreme rainfall events in the Indian region (Lau and Kim, 2011; Vellore et al., 2014, 2016). Therefore, tropical and mid-latitude regions are likely connected in complex, two-way, teleconnection patterns operating at a range of sub-seasonal ~~time-scale~~timescales (Di Capua et al., 2020; Ding and Wang, 2005, 2007).

During boreal summer, the South Asian monsoon (SAM), represents one of the most ~~important and~~ powerful features of the tropical/subtropical circulation. Characterized by heavy rainfall over central India and the Bay of Bengal, the SAM has strong intraseasonal variability associated with alternating active and break phases, linked to the boreal summer intraseasonal oscillation (BSISO, Choudhury and Krishnan, 2011; Gadgil and Joseph, 2003; Goswami et al., 1998; Krishnamurti and Surgi, 1987; Krishnan et al., 2000; Rao, 1976; Saha et al., 2012; Suhas et al., 2012). The western North Pacific summer monsoon (WNPSM) represents the Pacific counterpart to the SAM and is identified by strong rainfall over the sub-tropical western North Pacific (Li and Wang, 2005). Similar to the SAM, the WNPSM also exhibits strong intraseasonal oscillations (Wang and Xu, 1997).

Latent heat release due to strong monsoonal rainfall can influence subtropical and mid-latitude regions via Rossby wave teleconnections. The SAM has been connected to subtropical arid conditions in the North African region via the so-called monsoon – desert mechanism, creating reinforced descending motions over the Sahara during strong SAM phases (Rodwell and Hoskins, 1996; Stephan et al., 2019). This mechanism is fairly well captured by both climate (Cherchi et al., 2014) and seasonal forecast models (Beverley et al., 2019).The SAM is also connected to mid-latitude circulation via its interaction with the circumglobal teleconnection pattern (CGT), a wave pattern with 5 centres of action encircling the northern mid-latitudes and affecting temperature and precipitation there (Ding and Wang, 2005; Kripalani et al., 1997). This wave-5 like CGT pattern can be identified through interannual to intraseasonal (weekly) ~~time-scale~~timescales and it is likely connected with the SAM via two-way causal links (Di Capua et al., 2020). Seasonal forecast models are biased in their representation of the CGT, with typically a too weak CGT signal (Beverley et al., 2019). Therefore, seasonal forecasts miss an important source of predictability on intraseasonal ~~time-scale~~timescales, primarily in summer (Weisheimer and Palmer, 2014).

The WNPSM has been shown to influence precipitation anomalies over North America via its relation to the Western Pacific – North America (WPNA) pattern (Wang et al., 2001). The WNPSM is shown to be related to surface pressure conditions over East Asia, with high pressure anomalies during years characterized by stronger WNPSM activity (Nitta, 1987). The WNPSM area also represents a genesis region for tropical cyclones in the North Pacific (Briegel and Frank, 1997). The WNPSM is weaker during the decaying phase of El Niño (Wang et al., 2001) and its related circulation anomalies provide a link from ENSO to the East Asian summer monsoon ~~(EASM)~~ (Yim et al., 2008). Thus, in summary, the SAM appears particularly important for sub-seasonal variability over Eurasia, while the WNPSM is important for the Pacific-North American sector.

ENSO, operating at interannual ~~time-scale~~timescales, might primarily influence the mid-latitude circulation via its effect on the SAM strength (Ding et al., 2011). During boreal summers preceding La Niña phases, a strengthening of the Walker circulation can enhance SAM rainfall, while El Niño phases often have an opposite effect (Joseph et al., 2011; Ju and Slingo, 1995; Terray et al., 2003; Wu et al., 2012). However, this relationship depends on the longitudinal position of the strongest El Niño related sea surface temperature (SST) anomalies (Krishna Kumar et al., 2006) and potentially has weakened over recent decades (Chakraborty and Krishnamurti, 2003; Krishna Kumar et al., 1999; Srivastava et al., 2017; Xavier et al., 2007). At interannual ~~time-scale~~timescales, anomalous tropical convection in the central-eastern Pacific related to ENSO has also been shown to affect mid-latitude circulation over the Euro-Atlantic sector as well as temperature and precipitation anomalies over Europe during boreal summer (O'Reilly et al., 2018). Trends in tropical SSTs play a crucial role in the interdecadal changes of this tropical-extratropical teleconnection (O'Reilly et al., 2019).

In general, a major challenge faced by teleconnection research is to understand the underlying physical processes and associated cause-effect relationships. Past observational studies have typically employed correlation analysis or linear regression techniques. Such analyses can however be dominated by spurious correlations and therefore can give only limited insight into cause-effect relationships. On the other hand, model-based studies can be affected by biases in the representation of circulation and precipitation characteristics (Beverley et al., 2019; Schubert et al., 2011; Weisheimer and Palmer, 2014), which can feed back on each other. Also, although perturbation and sensitivity experiments can point towards potential causal relationships, they do not necessarily reveal the causal links between tropical and mid-latitudinal features, since the uncovered relationship may not be the dominant one.

In recent years, several approaches have been applied to identify causal relationships in climate and atmospheric sciences (Runge et al., 2019b), ranging from Granger causality (McGraw and Barnes, 2018, 2020; Samarasinghe et al., 2019) to causal (Bayesian) graphical models (Ebert-Uphoff and Deng, 2012a, 2012b; Horenko et al., 2017; Pearl, 2000) and conditional independence-based network discovery methods for time series (Runge et al., 2019a). These studies have shown the ability of causal discovery tools to improve our understanding of several atmospheric circulation interactions such as Arctic – mid-latitudes connections (McGraw and Barnes, 2020; Samarasinghe et al., 2019), synoptic-scale disturbances between boreal summer and boreal winter (Ebert-Uphoff and Deng, 2012a) and the relationship between ENSO and surface temperature in the American continent (McGraw and Barnes, 2018).

Here, we use a causal inference approach to study the relationships between the Northern Hemisphere mid-latitudes and the tropical belt during boreal summer at different intraseasonal timescales. We apply a causal discovery approach making use of the so-called PCMCI (Peter & Clark algorithm combined with the Momentary Conditional Independence approach, see Section 2.3) method (Runge et al., 2019a) ~~.~~and then estimate physically interpretable causal links weights by (standardized) multivariate regression. The resulting weighted network representation of causal interdependencies is referred to as a ~~This method detects causal links and~~ Causal Effect Network~~s~~ (CEN) (Kretschmer et al., 2016). ~~to quantify their effect strength, in order to study the relationships between the Northern Hemisphere mid-latitudes and the tropical belt during boreal summer at different intraseasonal time-scale~~timescales~~.~~ Expanding our understanding of the corresponding physical~~these~~ mechanisms has the potential to improve seasonal and subseasonal forecasts in boreal summer. The main advantage of ~~such~~ causal discovery tools is that they can identify and remove spurious correlations (Runge et al., 2015b; Runge, 2018; Runge et al., 2019a) and thus provide insight into the potential causal relationships (McGraw and Barnes, 2018; Runge et al., 2014). Building upon this advanced methodology, we introduce a new concept called causal maps, ~~identifying~~ visually highlighting causally related spatial structures. Finally, we assess the role of the ENSO background state on the identified causal relationships between the tropical belt and the mid-latitude circulation. The remainder of this paper is organized as follows: Section 2 presents the data and methods used in this analysis. Section 3 describes the results obtained by applying CEN and causal maps to the identified research questions. Section 4 ~~illustrates~~ provides a~~the~~ discussion of the obtained results in the context of the existing literature and finally, Section 5 presents a short summary and conclusions.

2 Data and Methods

2.1 Data

In our analysis, we diagnose monsoon characteristics and Northern Hemisphere circulation features using outgoing longwave radiation (OLR) at the top of the atmosphere, geopotential height at 200 hPa (Z200) and 2m surface temperature (T2m) data from the ERA-Interim Reanalysis (Dee et al., 2011) for the period 1979-2018 (1.5°x1.5°). Strong tropical convection is characterized by high cloud tops and thus by low emission temperatures, which in turn correspond to low OLR values (Krishnan et al., 2000). In the tropical belt, OLR can be used as a proxy of convective activity, and therefore, rainfall. To select different ENSO phases, we use the monthly Niño3.4 index from NOAA (data are available at https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Nino34/), representing the central Pacific SST anomalies. El Niño and La Niña events are discerned by periods of December-to-February Niño3.4 index values larger than 0.5°C or smaller than -0.5°C respectively. Then, we identify El Niño summers as those preceding the El Niño peak in winter and La Niña summers as those preceding the La Niña peak in winter. We use the Niño3.4 index since it has been shown to have a relatively strong connection to Indian monsoon rainfall (Krishna Kumar et al., 2006).

We also use the BSISO index as defined by Kikuchi et al. (2012) and Kikuchi and Wang (2010) (data are available at http://iprc.soest.hawaii.edu/users/kazuyosh/ISO_index/data/BSISO_25-90bpfil_pc.txt) in order to describe the phase and amplitude of the BSISO characterising the large-scale driver of active and break events over India. Causal discovery tool techniques require detrended anomalies centred at zero. Therefore, all data are linearly detrended and anomalies are calculated relative to an individual year's mean seasonal state by removing both the mean seasonal cycle and the year's mean seasonal state (i.e. the seasonal average from May to September, MJJAS) (Di Capua et al., 2020; Ding and Wang, 2007). Removing the year's mean seasonal state, and thus excluding the influence of interannual variations of the involved mechanisms, is essential to analyse intraseasonal variability of atmospheric components that present a strong interannual variably, such as the SAM.

2.2 Maximum covariance analysis

To extract the dominant co-variability patterns reflecting interactions between mid-latitude circulation in the Northern Hemisphere and tropical convection at intraseasonal timescales, we follow Ding et al. (2011) and ~~the tropics and mid-latitude, we first~~ apply maximum covariance analysis (MCA) to ~~tropical~~ OLR fields (used as a proxy for convective activity) in the tropical belt (15°S-30°N, 0°-360°E) paired with Z200 fields in the northern mid-latitudes ~~and northern mid-latitude Z200~~ (25°N-75°N, 0°-360°E)~~ fields~~.

MCA identifies the patterns that explain the *greatest squared covariance between two different fields* (Ding et al., 2011; Wiedermann et al., 2017) and ranks them according to their explained squared covariance fraction (SCF) (Wilks, 2011). Among the available correlation based methods to highlight strong co-variability and reduce the dimensionality of a spatiotemporal dataset, MCA allows identification of patterns in pairs of variables that evolve simultaneously and may be causally related (via e.g. dynamical coupling between multiple climatological fields). MCA detects patterns that can explain shared covariance, which cannot be achieved using other dimensionality reduction methods that consider individual variables separately, such as empirical orthogonal function (EOF) analysis. However, for providing a complete picture we will also discuss the corresponding EOF patterns and the fraction of variance explained for comparison with our MCA results.

Each MCA mode ~~thus~~ provides two coupled (2D) spatial patterns (one for tropical OLR and one for mid-latitude Z200) and two associated (1D) time series (the time-dependent MCA scores or pattern amplitudes for both fields), describing the magnitude (prominence) and phase (sign) of those patterns for each time step. These (1D) time series are obtained by calculating the scalar product between each MCA spatial pattern (2D field) and the original spatial field of the associated variable at each time step as

$$A = \boldsymbol{u}^T \boldsymbol{X} \tag{1}$$

$$B = \boldsymbol{v}^T \boldsymbol{Y} \tag{2}$$

where A and B represent the two MCA scores for Z200 and OLR, $\boldsymbol{X}$ and $\boldsymbol{Y}$ are two matrices representing the Z200 and OLR fields, $\boldsymbol{u}$ and $\boldsymbol{v}$ are the coupled patterns that maximize their covariance c, defined as:

$$c = cov[A, B] = cov[\boldsymbol{u}^T\boldsymbol{X}, \boldsymbol{v}^T\boldsymbol{Y}] = \frac{1}{n-1}[\boldsymbol{u}^T\boldsymbol{X}(\boldsymbol{v}^T\boldsymbol{Y})^T] = \boldsymbol{u}^T\boldsymbol{C}_{xy}\boldsymbol{v} \tag{3}$$

and

$$\boldsymbol{C}_{xy} = \frac{1}{n-1}\boldsymbol{X}\boldsymbol{Y}^T \tag{4}$$

with n denoting the number of observation times.

Here, we select the first two MCA modes that represent the dominant patterns of co-variability between tropical convection and mid-latitude circulation, and calculate time series for each MCA mode. These time series will be used as inputs for the causal discovery algorithm (see sections 2.3 and 2.4).

2.3 PCMCI and Causal Effect Networks

PCMCI is a causal discovery method based on the PC algorithm (named after its inventors Peter and Clark, see Spirtes et al., 2000) combined with the Momentary Conditional Independence approach (MCI, Runge et al., 2019). Given a set of univariate time series (called *actors*), PCMCI estimates their time series graph ~~that represents~~representing the conditional independencies among the time-lagged actors. In the context of the present work, actors are user-selected based on theoretical knowledge to represent either a specific component of the atmospheric circulation or surface conditions estimated with MCA (*A*, *B*) or an individual grid point time series *C(~~lat, lon~~lat,lon)*. Assuming linear dependencies, PCMCI uses partial correlations to iteratively test conditional independencies and remove spurious links arising from autocorrelation effects, indirect links, or common drivers. For example, if an actor *Z* drives *X* at lag -1 and *Y* at lag -2, then *X* and *Y* will be correlated, but the partial correlation $\rho(X_{\tau\cancel{t}-1}, Y_t \mid Z_{\tau\cancel{t}-2})$ will be zero. PCMCI efficiently conducts partial correlation tests to identify which links cannot be explained by other time-lagged actors. Compared to the standard PC algorithm, PCMCI better deals with autocorrelation and high-dimensional sets of actors (Runge et al., 2019a). The output of PCMCI is a *p*-value for each time-lagged causal link.

It is important to note that the term causal rests on specific assumptions (Runge, 2018; Spirtes et al., 2000), most importantly that it should be understood as "causal relative to the set of analysed actors". Therefore, adding (or removing) an actor can alter the result of PCMCI, highlighting the importance of having an expert-guided hypothesis underlying the choice of the selected set of actors. In addition, using partial correlation for a conditional independence test implies further assumptions such as the stationarity and linearity of the relationships. To~~ ensure that~~ control for multiple testing~~ does not inflate p-values or the multiple~~ among the multiple grid locations in causal maps, we apply ~~the~~ a false discovery rate (FDR) correction (Benjamini and Hochberg, 1995).

Based on the reconstructed network among the actor variables (at some significance level α), we determine the causal parents as the incoming links to each actor (*C(~~lat, lon~~lat,lon)*, *A*, *B*), which can come from the pasts of *A*, *B*, or *C(~~lat, lon~~lat,lon)* , i.e., $\{A_{\tau\cancel{t}=-1}, B_{\tau\cancel{t}=-1}, C(\cancel{lat, lon}\,lat,lon)_{\tau\cancel{t}=-1}, \ldots, A_{\tau\cancel{t}=-\tau max}, B_{\tau\cancel{t}=-\tau max}, C(\cancel{lat, lon}\,lat,lon)_{\tau\cancel{t}=-\tau max}\}$. In this analysis, *A* and *B* represent the two MCA scores obtained for a selected MCA mode, while *C(lat,lon)* represents the grid point time series of a 2D field, e.g. T2m or Z200. In its first step, PCMCI iterates through partial correlations with increasing cardinality of conditions to remove the influence of common drivers and indirect links and estimate a preliminary set of parents. The first iteration of PC (cardinality 0) calculates the correlation between a selected time series, e.g. $A_{\tau=0}$, and the past of any other available time series, $\{A_{\tau=-1}, B_{\tau=-1}, C(lat,lon)_{\tau=-1}, \ldots, A_{\tau=-\tau max}, B_{\tau=-\tau max}, C(lat,lon)_{\tau=-\tau max}\}$, including its own past $A_{\tau=-1,\ldots,-\tau max}$. For illustration purposes, we here provide an example for *C(lat,lon)*, where ρ denotes the correlation and τ is the lag that is being used in the network (in this example, $\tau_{max} = -2$):

$$\rho(C(lon,lat)_{\tau=o}, A_{\tau=-1}) = 0.32, p = 0.01 \tag{5}$$

$$\rho(C(lon,lat)_{\tau=o}, A_{\tau=-2}) = 0.13, p = 0.1$$

$$\rho(C(lon,lat)_{\tau=o}, B_{\tau=-1}) = 0.35, p = 0.005$$

$$\rho(C(lon,lat)_{\tau=o}, B_{\tau=-2}) = 0.23, p = 0.058$$

$$\rho(C(lon,lat)_{\tau=o}, C(lon,lat)_{\tau=-1}) = 0.41, p = 0.01$$

$$\rho(C(lon,lat)_{\tau=o}, C(lon,lat)_{\tau=-2}) = -0.16, p = 0.06$$

Applying a significance level α = 0.05, only three actors are significantly correlated with *C(lat,lon)* at the chosen time lag. These form the initial preliminary set of parents for *C(lat,lon)* and are ordered by the strength of their correlation:

$$P^0_{C(lon,lat)} = \{C(lat,lon)_{\tau=-1}, B_{\tau=-1}, A_{\tau=-1}\} \tag{6}$$

Next, partial correlations between *C(lat,lon)* and each actor in $P^0_{C(lon,lat)}$ are calculated by conditioning on the strongest preliminary parent:

$$\rho(C(lat,lon)_{\tau=o}, C(lat,lon)_{\tau=-1} | B_{\tau=-1}) = 0.35, p = 0.02 \qquad (7)$$

$$\rho(C(lat,lon)_{\tau=o}, B_{\tau=-1} | C(lat,lon)_{\tau=-1}) = 0.28, p = 0.03$$

$$\rho(C(lat,lon)_{\tau=o}, A_{\tau=-1} | C(lat,lon)_{\tau=-1}) = 0.25, p = 0.04$$

Parents with significant partial correlations will enter the second set of preliminary parents:

$$P^1_{C(lat,lon)} = \{C(lat,lon)_{\tau=-1}, B_{\tau=-1}, A_{\tau=-1}\} \qquad (8)$$

Next, the partial correlation is calculated conditioning on the two strongest parents:

$$\rho(C(lat,lon)_{\tau=o}, C(lat,lon)_{\tau=-1} | B_{\tau=-1}, A_{\tau=-1}) = 0.31, p = 0.03 \qquad (9)$$

$$\rho(C(lat,lon)_{\tau=o}, B_{\tau=-1} | C(lat,lon)_{\tau=-1}, A_{\tau=-1}) = 0.23, p = 0.04$$

$$\rho(C(lat,lon)_{\tau=o}, A_{\tau=-1} | C(lat,lon)_{\tau=-1}, B_{\tau=-1}) = 0.12, p = 0.08$$

Since it is not possible to further increase the dimension of the condition set, from the PC step, the preliminary parents converge to:

$$P^2_{C(lon,lat)} = \{C(lat,lon)_{\tau=-1}, B_{\tau=-1}\} \qquad (10)$$

By repeating this step for each variable, preliminary sets of parents are estimated. Let's assume that in our example we also obtain:

$$P^3_A = \{ B_{\tau=-1}, A_{\tau=-2,} \} \qquad (11)$$

$$P^2_B = \{ B_{\tau=-1} \}$$

In the MCI step, partial correlation is calculated again between each pair of actors (at different time lags) conditional on the above estimated sets of preliminary parents, whereby both sets of parents are conditioned upon. To give one example, this would lead to:

$$\rho\left(C(lat,lon)_{\tau=o}, A_{\tau=-1} \middle| P^2_{C(lat,lon)}, P^3_A\right) =$$

$$= \rho(C(lat,lon)_{\tau=o}, A_{\tau=-1} | C(lat,lon)_{\tau=-1}\ B_{\tau=-1}, B_{\tau=-2}, B_{\tau=-3}) = 0.1, p = 0.3 \qquad (12)$$

Note that the parents of $A_{\tau=-1}$ are shifted in time by τ = -1. After repeating (12) for each pair of actors shown in (5) and for time lags from 0 to τ_{max}, those parents that are significant in the MCI test will then form the final set of *causal parents* for each actor. We refer to Runge et al. (2019a) for a more detailed discussion and explanation of the algorithm design and extensive numerical experiments.

~~Then~~ Finally, we estimate the Causal Effect Network (CEN) (Kretschmer et al., 2016; Runge et al., 2015a) among *A*, *B* and *C(~~lat, lon~~lat,lon)* by applying standardized multiple regression of each actor onto its causal parents identified via PCMCI, i.e., for $Y \in$ in A_t, B_t, $C($~~lat, lon~~*lat,lon*$)_t$ and the parents *P*:

$$Y_t = \sum_i \beta_i X_i + \eta_Y \qquad (~~5~~13)$$

where $X_i \in P\{Y\}$, $i = 1, .., N$, i.e. the set of N parents of Y. Note that there can be different numbers N of parents for each actor. Finally, the strength of a causal link $X_{t-\tau} \rightarrow Y_t$ is expressed in terms of the path coefficient β, which can be interpreted as the change in the expectation of Y_t (in units of its standard deviation (s.d.)) induced by raising $X_{t-\tau}$ by 1 s.d., while keeping all other parents of Y_t constant. Thus, for $\beta = 0.5$, a change in a causal parent of 1 s.d. at lag -1 corresponds to a change 0.5 s.d. in the analysed actor at lag 0 (Runge et al., 2015a). The influence of an actor on itself is referred to as the autocorrelation path coefficient, which must not be confused with the Pearson autocorrelation. A detailed description of the PCMCI algorithm is available in Runge et al. (2019), while recent applications can be found in Kretschmer et al. (2016, 2018) and (Di Capua et al., 2019, 2020)~~Di Capua et al. (2019)~~.

2.4 Causal maps

To explore the causal effects that a specific actor has on a selected 3D (~~lat, lon~~lat,lon, time) atmospheric field, we introduce the concept of *causal maps*. Conceptually, causal maps are similar to correlation maps, as they show the spatial pattern of the relationship between a 3D climate data set (covering two spatial dimensions plus time) and a 1D time series. However, instead of computing correlations between the time variations at each grid point and *one* additional time series, we apply here the PCMCI+CEN approach with actors consisting of the two MCA scores time series (*A*, *B*) and each individual grid point time series (*C*(~~*lat, lon*~~*lat,lon*)). The causal map then plots the path coefficient β from one of the MCA scores (as one actor) to this gridpoint, conditioned on all remaining actors. For a set of two actor timeseries (*A* and *B* in Fig. 1) and one time-varying atmospheric field *C*, we can thus derive two causal maps: one from *A* to *C*(~~*lat, lon*~~*lat,lon*) conditioned on *B* and on the autocorrelation in all actors, and one from *B* to *C*(~~*lat, lon*~~*lat,lon*) conditioned on *A* and on all autocorrelation effects. Figure 1 provides an illustrative example of this type of analysis. Both correlation maps (Fig. 1a) indicate a positive value for a specific geographical location highlighted with the black diamond. The CEN constructed for *A*, *B* and *C*(~~*lat, lon*~~*lat,lon*) at this gridpoint is plotted in Fig. 1b and shows that only *B* is causally connected to *C*. The correlation between *A* and *C* is thus due to an indirect link via *B* (or to a common driver not included in the CEN). This is also seen in the causal maps plotting the path coefficient β which for the $B \rightarrow C$(~~*lat, lon*~~*lat,lon*) link is positive (right panel) but is non-significant for the $A \rightarrow C$(~~*lat, lon*~~*lat,lon*) link (left panel). In causal map visualization we can directly illustrate the effect of a specific actor on a global field (taking into account the influence of autocorrelation), indirect links and common driver effects due to other competing variables.

Here, we will derive causal maps using the time series obtained with MCA for modes 1 and 2 and Z200, OLR and T2m fields both for the entire time period (1979-2018) and for two subsets depicting different ENSO phases, to assess how the ENSO background state influences the causal relationships. El Niño (La Niña) summers are defined as summers preceding the El Niño (La Niña) peak in boreal winter. We thus obtain 14 La Niña years and 13 El Niño years (see Table 1 in the Supplementary material for a list of corresponding years and Fig. S1 for the associated SST anomaly composites). Although the strongest SST anomalies related to the ENSO phase are found in winter, warm (cold) SST patterns related to El Niño (La Niña) phases are already clearly developed during the preceding summers.

Finally, to test the robustness of our causal maps to the choice of time period, we calculate causal maps for a range of sub-periods. In 10 trials we removed 10% of the record (4 years). For ENSO-phase dependent causal maps, we have shorter time series and we thus remove one year in each trial, leaving a set of 14 causal maps for La Niña events and 13 causal maps for El Niño events. As a result, we obtain an ensemble of causal maps and apply the false discovery rate correction to p-values of each single map. Then, both for the full period (1979-2018) and for El Niño and La Niña years separately, we masked out areas where less than 70% of the trials indicated a significant causal link, giving an indication of the robustness of our findings and at the same time suppressing noise.

A summary of the abbreviation and variable used in this analysis can be found in Table 1, while the parameters used for the PCMCI algorithm are reported in the Supplementary Material.

3 Results

3.1 Tropical – mid-latitude interactions: maximum covariance analysis

Figures 2a-d show the coupled patterns for the first two MCA modes between weekly tropical OLR and mid-latitude Z200. Figure 2e shows the associated time series of MCA scores for all four patterns (two for each MCA mode), obtained as explained in Section 2.2.

The first two MCA modes highlight the two key patterns of boreal summer monsoonal activity in the tropics along with the co-varying mid-latitude Z200 patterns. In both modes, the mid-latitudes are characterized by a zonally oriented circumglobal wave pattern with a wavenumber close to 5 (i.e. roughly 5 centres of action). However, the two wave patterns are phase shifted, aligned with the longitudinal position of the strongest monsoonal convection in the tropics.

The first MCA mode explains 18% of the squared covariance (squared covariance fraction, SCF) and shows a CGT-like wave-5 pattern in mid-latitude Z200. The Pearson correlation between the two time series of MCA scores for the first mode is r ~ 0.5. T~~(t~~he spatial correlation with the weekly CGT pattern, as defined by ~~Ding and Wang 2005~~Di Capua et al. 2020, is 0.52 ~~)~~ ~~in mid-latitude Z200~~ (Fig. 2a). The CGT pattern also represents the second most important pattern in boreal summer mid-latitude circulation (Di Capua et al., 2020; Ding and Wang, 2005). This wave-5 pattern is linked to the South Asian monsoon (SAM) activity via its positive centre of action east of the Caspian Sea (see Fig. 2a). Applying MCA, we find that t~~T~~he CGT pattern co-varies with a band of enhanced tropical convective activity that extends from the Arabian Sea towards Southeast Asia, with a peak of convective activity over the Bay of Bengal (Fig. 2b) (Kang et al., 1999). ~~We will refer to this pattern as the South Asian monsoon (SAM).~~Using OLR composites and the Kikuchi Boreal Summer Intraseasonal Oscillation (BSISO) index, we explicitly show that the temporal evolution of SAM convective activity as defined in Fig. 2b at weekly timescales closely resembles the evolution of the BSISO (Goswami and Ajaya Mohan, 2001; Saha et al., 2012) (see Figs. S2-S3 and further discussion in the Supplementary Material). Therefore, we explicitly link the region of low OLR identified in Fig. 2b over the northern Indian Ocean and the Indian subcontinent to SAM activity as described in the literature. Note that we name each MCA pattern after a characteristic regional feature, but the analysis is applied to the larger geographical domains as shown in Figure 2. ~~The Pearson correlation between the two time series of MCA scores for the first mode is r ~ 0.5.~~

The second mode of co-variability explains a SCF of 14% ~~between the two fields~~ and is characterized by a region of strong positive Z200 anomalies located at ~ 45° N, over the western North Pacific, directly to the west of the dateline (i.e. the most prominent centre of action of the mid-latitude wave). The Pearson correlation between the two time series of MCA scores for the first mode is r ~ 0.6. We will refer to this pattern as the North Pacific High (NPH) (Fig. 2c). The NPH is the summer counterpart of the North Pacific subtropical high, which characterizes boreal winter. During summer, this high pressure region is displaced northward by the start of the monsoon season over the western Pacific Ocean and replac~~i~~es~~ng~~ the Aleutian Low (Lu, 2001; Riyu, 2002). ~~As can be seen in MCA mode 1, t~~The NPH is associated with a region of enhanced convection over the sub-tropical western North Pacific, related to the western North Pacific summer monsoon (WNPSM) convective activity (Fig. 2d) (Li and Wang, 2005; Nitta, 1987; Wang et al., 2001). The WNPSM core domain extends from 110°-160°E and 10°-20°N, while the boundary with the SAM is located over the South China Sea (Murakami and Matsumoto, 1994). The WNPSM is characterized by a late sudden onset (end of July) and a peak in rainfall activity during August and September, which is different from the SAM that features an earlier onset (in June) and peak rainfall activity during July-August. ~~The Pearson correlation between the two time series of MCA scores for the first mode is r ~ 0.6.~~

We compare the patterns obtained with MCA with those obtained with EOF analysis of Z200 and OLR fields (see Fig. S4 in the Supplementary Material). We find that the closest match of the Z200 MCA mode 1 pattern is with Z200 EOF 2 (spatial correlation ~ 0.8), while the closest match of Z200 MCA mode 2 is with EOF 1 (spatial correlation ~ 0.6). OLR MCA mode 1 has the closest match with EOF 2 (spatial correlation ~ 0.5), while OLR MCA mode 2 has the closest match with EOF 5 (spatial correlation ~ 0.4). Thus, in general our MCA patterns also reflect the first two EOFs of Z200 and OLR indicating that they explain an important fraction of the regional variability. Nevertheless, here we are interested in those patterns that can explain shared covariance, which cannot be achieved by using EOF analysis alone. Therefore, we use the MCA-defined patterns for the following part of the analysis.

We also investigate whether the obtained MCA patterns are sensitive to the choice of OLR in representing tropical convective activity. Using vertical velocity, another proxy for tropical convection where strong convective activity is represented by enhanced upward motion, shows qualitatively the same patterns as those in Figs. 2a-d (see Fig. S5 in the Supplementary Material). When velocity potential is used instead of OLR, the first MCA mode still closely resembles the OLR/Z200 MCA mode 1, while the second MCA mode only partly captures features in the western Indian Ocean (see Fig. S6 in the Supplementary Material).

Application of ~~the~~ MCA to 4-weekly data gives nearly identical MCA patterns but with somewhat lower magnitude of the Z200 and O~~R~~LR anomalies (see Fig. ~~S1~~ S7 in the Supplementary Material). In this case, we define both 4-weekly and weekly MCA scores by projecting 4-weekly MCA patterns onto 4-weekly and weekly data respectively (see Fig. ~~S1e~~S7e-f in the Supplementary Material). In this way, we check whether the analysis is robust given different definitions of the MCA patterns.

~~Using OLR composites, we explicitly show that the temporal evolution of the SAM convective activity at weekly time-scale~~timescale~~s resembles the evolution of the Boreal Summer Intraseasonal Oscillation (BSISO) (Goswami and Ajaya Mohan, 2001; Saha et al., 2012) (see Fig. S2 in the Supplementary Material). The OLR pattern depicted by the first MCA mode represents phase 4-5 of the BSISO evolution (Fig. S2). The BSISO is characterized by a rainfall band tilted from northeast to southwest propagating from the tropical Indian Ocean toward Southeast Asia with a period of about one to two months. To further explore this hypothesis, we present a Wheeler-Hendon diagram using the BSISO index as defined by Kikuchi (2010) and plot (using different colours) BSISO phases that correspond to different lags (as defined considering the MCA mode 1 pattern for OLR as lag 0, see Fig. S2). The results show that each lag tends to cluster consistently around the corresponding BSISO phase (see Fig. S3a in the Supplementary Material). This suggests that the BSISO may exert a large-scale tropical control on mid-latitude anomalies, using variations in SAM rainfall as a pathway. When the same approach is applied to the WNPSM pattern, no consistent behaviour can be identified (Fig. S3b).~~

3.2 Influence of tropical – mid-latitude MCA modes on Northern Hemisphere circulation

To show how each MCA mode affects the circulation and surface conditions in the Northern Hemisphere, we calculate causal maps for the influence of SAM, CGT, WNPSM and NPH time series (as defined in Fig. 2e) on selected atmospheric fields in the Northern Hemisphere (15°S-75°N, 0°-360°E). Although we use τ_{max} = -~~4~~ 2 and τ_{min} = 0, we plot only β values for lag -1 (week), as β values for longer time lags are mostly nonsignificant. This way ~~also~~ the past behaviour of each actor, with potential confounding effects, is also accounted for in the corresponding grid-point CEN. Note that we only show robust links as defined in Section 2.4 and the masks used to plot the results are shown in Figs. S8-S9 in the Supplementary Material.

Figure 3 shows the causal maps for weekly Z200, OLR and T2m fields with SAM and CGT time series, including correlation maps for weekly Z200 fields with SAM and CGT time series. Referring to the schematic illustrated in Fig. 1 and following the PCMCI algorithm explanation (section 2.3), here the *A* and *B* time series are represented by the SAM and CGT time series respectively, while *C(lat,lon)* is represented by Z200, OLR and T2m fields. In the mid-latitudes, the correlation map between Z200 and SAM (Fig. 3a) shows ~~a similar wave pattern as that shown~~some similarities in the correlation map between Z200 and CGT (Fig. 3b), with negative correlation regions over central Europe and Scandinavia (Region 2 and Region 4) and over the eastern North Pacific and eastern Canada visible in both plots (regions 3 and 6). Both correlation maps also display a positive correlation over northern Africa (Region 1), though with smaller values in the CGT plot. The causal map for the link SAM $_{\tau=-1}$ → Z200 $_{\tau=0}$ (after removing the effects of the CGT and of the past of both SAM and Z200) shows that the path coefficient β remains pronounced over northern India and northern Africa (Region 1 in Fig. 3c), with values β ~ 0.3-0.4. Interestingly, those regions disappear completely in the causal map for the link CGT $_{\tau=-1}$ → Z200 $_{\tau=0}$ (after removing the effects from SAM and of the past of both CGT and Z200). Thus, in this way, we can separate the signal coming from SAM convective activity from signals originating from the CGT pattern. Also, the causal maps in Figs. 3c and 3d indicate that the influence of SAM on Europe (negative path coefficients shown by Region 2 in~~,~~ Fig. 3c) and the North Pacific (seesaw of positive and negative path coefficients shown by Region 3 in Fig. 3c) is *not* mediated via the CGT. This influence is weaker than that over the North Africa, with values of β ~ 0.1-0.3. In turn, the influence of SAM on other mid-latitude regions (over the North Atlantic, Region 4, over some parts of East Asia, Region 5 and Canada/USA, Region 6 in Fig. 3d) is mediated via the CGT~~-~~, with values of β ~ 0.1-0.3~~(Fig. 3d)~~.

In the mid-latitudes, the causal maps for OLR and T2m (Figs. 3e-h) are largely consistent with those obtained for Z200, with negative β~~w~~ values for OLR representing wet anomalies ~~(negative OLR)~~ overlapping with negative β values for T2m representing colder ~~T2m~~ anomalies and dry anomalies (positive β values for OLR) overlapping with warm T2m (positive β values for T2m). ~~Although t~~The CGT influence is mostly concentrated in the mid-latitude regions, where the same Regions 4 to 6 identified in Fig. 3d, can also be found in Figs. 3f and 3h.~~, one can see~~ A~~a~~ significant and consistent negative causal effect of the CGT pattern on OLR values ~~over the Bay of Bengal~~ in the tropical regions can only be seen in a small area in the western Indian Ocean (Region 1 in Fig. 3f). Again, the OLR and T2m causal maps indicate that the SAM has a direct influence on northern Africa and Europe as well as tropical Africa (Region 1 in Figs. 3e and 3g). ~~Asia and North America are strongly affected by the CGT.~~ Over the Indian peninsula and Indochina, strong convective motions (negative β values for OLR in Fig. 3e) are accompanied by colder temperature (negative β values for T2m in Fig. 3g), related to increased precipitation and consequently, decreased surface temperatures during active SAM activity. The influence of SAM on the western North Pacific identified by Region 3 in Fig. 3c is also detected in OLR (Region 3 in Fig. 3e). Negative β values found over Region 2 in Fig. 3c are only slightly visible in OLR and T2m. However, we should also remind the reader that our causal maps show only the most robust links (see Section 2.4).

Figure 4 shows the same set of results but now for the second MCA mode consisting of WNPSM and NPH pattern related time series. Here, our *A* and *B* time series are represented by the WNPSM and NPH time series while *C(lat,lon)* is again represented by either Z200, OLR or T2m fields. As expected, both correlation maps resemble the Z200 field of MCA 2 (Fig. 2c,d) with two characteristic features: an arch-shaped wave pattern in the North Pacific (Regions 7,8 and 9 in Figs. 4a and 4b) ~~A wave-5 pattern~~ with a prominent positive correlation over the NPH region and over western North America and two weaker centres of action over the Eurasian continent (Region 10 and 11 in Fig. 4b and Region 11 in Fig. 4a). The corresponding causal maps based on CENs are given in Figs. 4c (path coefficient β for the link $\text{WNPSM}_{\tau = -1} \rightarrow \text{Z200}_{\tau = 0}$) and 4d (for the link $\text{NPH}_{\tau = -1} \rightarrow \text{Z200}_{\tau = 0}$). If we compare the correlation maps (Figs. 4a,b) with the causal maps (Figs. 4c,d), we find great similarity in the spatial structures of the Z200 patterns over the North Pacific in both figures (Region 7 in Figs. 4c,d) with β ~~,~~~ 0.1-0.3 for the influence of NPH on Z200 and β ~ 0.1-0.2 for the influence of WNPSM on Z200~~although the magnitudes have reduced in Fig. 4c~~. These causal maps show that the influence of the WNPSM on Z200 (after removing the effect of the NPH) is confined to the North Pacific alone (Regions 7 and 8 in Fig. 4c). The causal effect of the NPH pattern on Z200 (after removing effects of WNPSM) shows ~~The~~ two most prominent regions displaying a significant positive path coefficient β ~ 0.2-0.4 ~~are found~~ over the NPH region (Region 7 in Fig. 4d) and over the US west coast (Region 9 in Fig. 4d)~~. In contrast, the causal effect of the NPH pattern on Z200 (after removing effects of WNPSM) shows a wave train that encircles the mid-latitudes~~

~~(Fig. 4d)~~. ~~Like the MCA mode 2 pattern itself, this wave train shows positive centres over Russia/Scandinavia, the western North Pacific and western US coast. The two regions~~Region 7 in the Pacific sector coincides with ~~those~~ that found for the WNPSM causal map (Fig. 4c). This suggests that the NPH is reinforced both by convective activity of the WNPSM and by the mid-latitude wave pattern localized in the North Pacific. Regions 10 and 11 found in Figs. 4a and 4b disappear in both Figs. 4c and 4d, showing that the correlation in these regions is mostly explained by Z200 activity in the mid-latitude itself (note that in the CEN we also condition on the past of Z200) or by other factors not considered in this analysis.

Next, we compute the causal maps for the influence of WNPSM and NPH on weekly OLR and T2m fields (Figs. 4e,f and 4g,h respectively). ~~For T2m causal maps, the results are largely consistent with those obtained for Z200, with positive Z200 anomalies hinting to warm and dry weather in the mid-latitudes and strong convective motions being accompanied by colder temperatures in the tropical belt. Thus, these results highlight the importance of the NPH in shaping surface temperatures across the northern mid-latitudes.~~ The impact of the WNPSM on OLR and T2m fields is very weak, though it is possible to recognise some negative $\beta \sim$ 0.1-0.2 over the Arabian Sea and over the WNPSM area (Region 8 in Fig. 4e)~~pattern is confined to the western North Pacific~~. ~~Further, convective activity in the WNPSM region is reinforced by the NPH pattern as indicated by a negative OLR anomaly over the WNPSM region (Fig. 4f). These results thus indicate that there is a two-way causal relationship between the WNPSM and the NPH.~~ ~~F~~The impact of NPH on OLR and ~~or~~ T2m causal maps shows some similarities with the correlation map shown in Fig. 4b. T2m and OLR show the strongest effect over North America (Region 9 in Figs. 4f and 4h) with $\beta \sim$ 0.2-0.4, and this~~the~~ result~~s~~ ~~are~~is largely consistent with ~~those~~that obtained for Z200, with positive Z200 anomalies hinting ~~t~~at~~o~~ warm and dry weather in the mid-latitudes related to an active WNPSM. Over Eurasia, it is possible to recognize Regions 10 and 11 in both ORL and T2m causal maps but with smaller regions and lower $\beta \sim$ 0.1-0.2 (Figs. 4f and 4h). ~~and strong convective motions being accompanied by colder temperatures in the tropical belt.~~ Thus, these results highlight the importance of the NPH pattern in shaping surface temperatures across the northern mid-latitudes.

Using weekly MCA scores obtained from 4-weekly MCA patterns (Fig. ~~S1~~S7) gives consistent results, showing that the analysis is robust when a different definition of the MCA pattern is chosen (see Figs. ~~S4~~S10-~~S5~~ S11 in the Supplementary Material). Causal maps calculated for 4-weekly Z200 for both MCA 1 and MCA 2 show less significant results, likely due to the limited time series length (not shown).

3.3 The influence of ENSO on tropical – mid-latitude causal interactions

Next, we assess how the ENSO background state influences the causal relationships between mid-latitude and tropical patterns in boreal summer. We recalculate the causal maps for both MCA scores and Z200 fields, where *A* and *B* time series are represented by the SAM and CGT time series for the first MCA mode and by the WNPSM and NPH time series for the second MCA mode, while *C(lat,lon)* is represented by Z200. As for Fig. 3 and 4, the robustness mask used to plot the results shown in Fig. 5 is shown in Fig. S12 in the Supplementary Material.

~~To do so, we compute causal maps, similar to those in Figs. 3 and 4 but for El Niño and La Niña summers separately (Fig. 5), where El Niño (La Niña) summers are defined as summers preceding the El Niño (La Niña) peak in boreal winter. Figure S6 in the Supplementary Material shows composites of SST anomalies for summers associated with the different types of ENSO phases as defined in this way. The warm (cold) SST patterns related to El Niño (La Niña) phases are already clearly developed during preceding summers, though the strongest anomalies are found in winter.~~

In general, the strength and the sign of the patterns seen in the causal maps (Fig. 3 and 4) are not markedly affected by ENSO, though we can see higher absolute values of β ~ 0.1 in Fig. 5 with respect to Figs. 3 and 4~~ with some notable exceptions~~. During La Niña years, the effect of SAM on the Sahara Desert intensifies and also its effects on the Tibetan Plateau and in the mid-latitudes are more pronounced (Region 1, 2 and 6 in Fig. 5c). This is likely related to stronger SAM convective activity during La Niña summers. ~~During La Niña, we also see a few affected areas over Eurasia and North America that are not present during El Niño.~~ The region of negative causal effect of SAM on central Europe, is ~~also~~ present only during La Niña summers (with a β ~ 0.2-0.3 and larger in area when compared to~~in~~ the 1979-2018 causal map in ~~(~~Fig. 3c), ~~however here the signal intensifies (~~and it disappears during El Niño summers~~)~~. This signal is possibly linked to the strong positive causal effect over the Sahara Desert (Region 1). ~~At the north-west of the Caspian Sea, a region of positive causal effect appears over Kazakhstan, linked to the region of positive causal effect in the south, over the Tibetan Plateau.~~ The region of positive causal effect over the Tibetan Plateau ~~further extends to the South China Sea, the Korean peninsula and southern Japan~~ shows the same intensity as for the 1979-2018 period (β ~ 0.2-0.3) though it remains more confined over the Indian subcontinent and the Tibetan Plateau with respect to the 1979-2018 period (Region 1 in Fig. 5c). During La Niña, we also see an area of positive β values over North America (Region 6 in Fig. 5c) that is not present during El Niño.~~The SAM influence continues towards the east and shows two regions of positive causal effect (in central North Pacific and central Canada) and one region of negative causal effect over eastern North Pacific.~~

During El Niño summers, the influence of the SAM is ~~less pronounced~~almost completely absent in the mid-latitude regions, with only one region of low β values over the eastern North Pacific still being present (Region 3 in Fig. 5a). ~~However, it~~The positive β values over the tropical Pacific found in Fig. 3c disappear during La Niña and only some residues of this region are seen during El Niño years ~~extends towards the entire tropical Pacific~~ (Region 8 in Fig. 5a).

In the western North Pacific, the most notable feature is ~~the much stronger~~the ~~influence~~ presence of both the WNPSM and NPH on the North Pacific only during El Niño summers (Figs. 5e,f). During those summers, the positive causal effect of the WNPSM over the western North Pacific (Region 7 and 8 in Fig. 5e) intensifies in magnitude (absolute β ~ 0.3-0.4) with respect to the 1979-2018 ~~causal map~~mean pattern (Fig. 4c), although the geographical extent of Region 7 shrinks. Over the western tropical Pacific, in correspondence with the La Niña warm pool, a region of positive causal effect is shown (Region 8 in Fig. 5e). ~~Both~~ These features disappear during La Niña summers. ~~Thus, during El Niño summers, t~~Thus, the second MCA mode (the WNPSM-NPH pair) ~~shows more intense causal maps~~has its strongest effect during El Niño summers, whereas ~~during la~~

~~Niña summers~~ the first MCA mode (SAM-CGT pair) is ~~more~~ important during both La Niña and El Niño summers but with different spatial characteristics.

Calculating MCA pattern during different ENSO phases does not change the results in a qualitative way, although the order of the patterns is reversed in La Niña summers (see Fig. ~~S7~~S13, in the Supplementary Material). Moreover, we have checked whether the distribution of the spatial correlation between each MCA mode and the respective Z200/OLR fields changes during different ENSO phases and found that ENSO does not affect the frequency of each pattern in a significant way (see Fig. ~~S8~~S14, in the Supplementary Material).

As for the 1979-2018 causal maps, when weekly MCA scores obtained from 4-weekly MCA patterns (Fig. ~~S1~~S7) are used to provide ENSO-dependent causal maps, consistent results are obtained (see Figs. ~~S9~~ S15 in the Supplementary Material). Further analysis of possible physical explanations is provided in Section 4.

3.4 MCA causal interactions

Finally, we study the role of ~~time-scale~~timescales on the causal interaction patterns presented above. We create CENs between the two time series of scores for each MCA mode, as identified in Fig. 2 and Fig. ~~S1~~S7, and do so for weekly and 4-weekly data for the 1979-2018 period. Figure 6 plots the path coefficient β for two separate sets of CENs built for MCA mode 1 (SAM with CGT, Fig. 6a) and MCA mode 2 (WNPSM with NPH, Fig. 6b) for both 4-weekly and weekly ~~time-scale~~timescales. As for the causal maps, we use τ_{max} = -~~4~~ 2 for weekly data and τ_{max} = -1 for 4-weekly data. In both cases, τ_{min} = 0.

At the 4-weekly ~~time-scale~~timescale, the ~~pair~~ WNPSM-NPH pair does not show significant causal links (Figs. 6b). The SAM-CGT pair shows two fairly strong causal links with absolute values β~0.3-0.4, though with different signs (Figs. 6a). The northward link, i.e. SAM$_{\tau=-1}$ → CGT$_{\tau=0}$, shows a positive β ~ 0.4: a 1 s.d. shift in the SAM leads to a ~ 0.4 s.d. positive shift in the CGT 4 weeks later (Fig. 6a). The southward link, i.e. CGT$_{\tau=-1}$ → SAM$_{\tau=0}$, shows β ~ -0.3, meaning that at this ~~time-scale~~timescale a more intense CGT pattern leads to a weakening of the SAM pattern 4 weeks later (Fig. 6a).

At the weekly ~~time-scale~~timescale, both the WNPSM$_{\tau=-1}$ → NPH$_{\tau=0}$ and the NPH$_{\tau=-1}$ → WNPSM$_{\tau=0}$ links show a β ~ 0.1-0.2, indicating that the two-way link has a similar magnitude in both southward and northward directions (Fig. 6b). At this ~~time-scale~~timescale, the path coefficient β for the SAM$_{\tau=-1}$ → CGT$_{\tau=0}$ link is about a factor 4 smaller than that for the 4-weekly ~~time-scale~~timescale (Fig. 6a). The southward link, CGT$_{\tau=-1}$ → SAM$_{\tau=0}$, shows a positive β ~ 0.2 that is about twice as strong as the northward link (Fig. 6a). Thus, the influence of the SAM on the CGT pattern is weak (but present) at shorter (weekly) ~~time-scale~~timescales, but much stronger at longer (4-weekly) ~~time-scale~~timescales.

Finally, we tested how the CENs change when the 4-weekly signal is removed from the weekly time series: Each 4-weekly mean is removed from the four values of the corresponding weekly data (Fig. S16). This way, we attempt to isolate the dominant ~~time-scale~~timescales of physical processes behind the different MCA patterns. This is similar in rationale to removing the effects of interannual variability before quantifying intraseasonal variability. Results for the first MCA (SAM and CGT) shows that the path coefficient β for CGT$_{\tau=-1}$ → SAM$_{\tau=0}$ link remains almost unaffected (see Fig. ~~S8a~~S16a, in the

Supplementary Material). This suggests that this southward link typically operates at weekly ~~time-scale~~timescales (rather than 4-weekly), which is rather intuitive since mid-latitude variability dominates at synoptic ~~time-scale~~timescales. In contrast, the path coefficient β for the northward link (SAM $_{\tau=-1}$ → CGT $_{\tau=0}$) becomes insignificant when the 4-weekly signal is removed from the weekly time series, suggesting that the influence from the tropics via the SAM pattern operates at longer, 4-weekly ~~time-scale~~timescales. Removing the 4-weekly signal from the weekly time series for the second MCA (WNPSM and NPH) roughly halves the path coefficient β for both the northward and southward link (see Fig. S~~8~~16b in the Supplementary Material).

4 Discussion

In our analysis, we have found that the dominant patterns of interaction between the tropics and mid-latitudes remain qualitatively similar across different sub-seasonal ~~time-scale~~timescales (weekly and 4-weekly averages) (Fig. 2 and Fig. S7~~1~~ in the Supplementary Material). Two pairs of co-varying patterns are identified: a) convective activity of the South Asian monsoon (SAM) paired with a mid-latitude wavenumber-5 wave train resembling the circumglobal teleconnection (CGT) pattern and b) convective activity over the western North Pacific, related to the western North Pacific summer monsoon (WNPSM) paired with a second wave-5 circumglobal wave pattern with its strongest action centre represented by the North Pacific High (NPH) and phase shifted with respect to the CGT pattern, to the longitudinal position of WNPSM monsoonal convection in the tropics. These patterns of sub-seasonal co-variability between the mid-latitudes and tropics during boreal summer are remarkably similar to those identified by Ding et al. (2011) for interannual ~~time-scale~~timescales. This consistency across ~~time-scale~~timescales (from weekly over monthly to interannual) suggests that the interannual patterns originate from a summation of the same patterns at sub-seasonal ~~time-scale~~timescales. Still, the strength and sign of the causal interactions are ~~time-scale~~timescale dependent. At longer ~~time-scale~~timescales (from monthly to seasonal) slowly varying components such as tropical SST and associated regions of convective activity dominate. Therefore, on these ~~time-scale~~timescales the causal links from the tropics to mid-latitudes tend to be stronger. At shorter (weekly) ~~time-scale~~timescales, in general a two-way positive feedback between the tropics and mid-latitudes is found, although strong variability in the mid-latitudes dominates over the tropical convection and thus the reversed southward pathways become stronger (Fig. 6). Moreover, we have introduced a novel visualization approach – termed causal maps – that can provide regionally specific information on the causal influence of a specific teleconnection source, and how this signal gets mediated. In t~~T~~his way, we identify mid-latitude regions and surface weather conditions that are influenced by tropical drivers by taking into account the linear influence of both MCA patterns together (for each MCA mode). The strongest causal effect of SAM convection is found over the Saharan region, and depicts the monsoon-desert mechanism (Rodwell and Hoskins, 1996). Also important is the effect of SAM on the central Asian CGT centre of action (see Di Capua et al., 2020). The SAM also appears to directly influence geopotential heights in the North Pacific and central European surface temperatures one week ahead (Fig. 3c,e). The influence of the CGT pattern is stronger over the mid-latitude regions, nevertheless some influence on the Indian Ocean is detected (Fig. 3d,f), further supporting the results shown in Fig. 6. In the North Pacific there is a clear ~~two-way~~ positive influence ~~between~~ from the

WNPSM ~~and~~ towards the NPH patterns reflecting a Hadley cell-like circulation (Fig. 4d). The direct causal effect from the NPH on surface weather conditions ~~This system has a~~is particularly strong ~~influence i~~on central North America~~n~~ ~~and Scandinavian surface weather conditions~~ while its direct effect back on the tropics is weak (Fig. 4d,f,h). Thus, for MCA mode 2, the signal from the WNPSM towards the NPH is consistent both in simple CEN (Fig. 6) and causal maps (Fig. 4), while the direct influence of the NPH on the tropical belt is present but weaker and less robust (see Fig S9 in the Supplementary Material).

In the tropical belt, the processes behind the identified MCA patterns are linked to strong convection related to ~~the~~ monsoon activity. Though tropical convection is characterized by heavy precipitation with a typical duration of less than a day, the latent heat release can act as a Rossby wave source for up to two weeks after the initial forcing is removed (Branstator, 2014). Moreover, while individual convective events are short-lived, the regions of dominant convective activity in the tropics change on much longer ~~time-scale~~timescales, such as in response to the BSISO (30-60 days). Thus, this finding could serve as a possible explanation for why the patterns identified at weekly and 4-weekly ~~time-scale~~timescales show great similarity (see also the discussion surrounding Figs. S2 and S3 in the Supplementary Material). It appears reasonable to assume that the tropics operate at longer ~~time-scale~~timescales providing potential sources of predictability at seasonal-to-subseasonal (S2S) ~~time-scale~~timescales. In contrast, mid-latitude circulation in summer is weaker than in winter and is characterized by circumglobal wave trains with typical ~~time-scale~~timescales of about one or two weeks (Di Capua et al., 2020; Ding and Wang, 2007; Kornhuber et al., 2016).

On the western North Pacific side, our findings linking the WNPSM convective activity to the NPH, and in turn to a wave-5 circumglobal wave train that affects surface weather condition in the mid-latitudes, further support~~s~~ results from previous studies suggesting that convective activity related to this oceanic monsoon system can enhance the high pressure found in the North Pacific mid-latitudes and that this affects weather conditions in western North America (Chou et al., 2003; Wang et al., 2001). ~~Three~~ Four centres of action over northern and eastern Europe, central Asia ~~European Russia~~, central North America and the central North Pacific are identified in the T2m causal map (Fig. 4h). Eastern Europe and central North America were also identified to be teleconnected regions associated with a boreal summer wave-5 ~~strongly correspond to the phase locking regions identified for boreal summer wave-5~~ by Kornhuber et al. (2020), who highlighted the risk for ~~thus linking the NPH and the WNPSM activity to~~ potential bread-basket failures. Recent evidence indicate that land-atmosphere interactions and increased aridity in mid-latitude regions such as North America and Europe may constitute an enhancing mechanism for the amplification of circumglobal quasi-stationary Rossby wave events during boreal summer (Teng and Branstator, 2019). ~~Moreover, the pattern identified in Fig. 5f with a low over central Europe and a high over western Russia, resembles the results shown by previous studies that link positive geopotential height anomalies over western Russia during summer 2010 with low-frequency wave trains initiated by La Niña-related convection in the tropical Pacific (Drouard and Woollings, 2018; Hong et al., 2011; Schneidereit et al., 2012; Trenberth and Fasullo, 2012).~~ Therefore, our results support the importance of the role of Pacific forcing for this wave-5 circumglobal wave pattern. Although other mechanisms could also be relevant in exciting and maintaining this pattern, the link to the WNPSM convection may hold the potential to affect seasonal forecasts and climate risks, such as heat waves and simultaneous crop failures.

We have applied our causal map analyses to specific ENSO phases to assess the role of El Niño and La Niña in modulating the interactions between mid-latitude circulation and tropical convection in boreal summer. These analyses suggest that in general the ENSO phase does not change the qualitative nature of the causal relationships between different MCA patterns: the signs and strengths of the causal links are largely unaffected (see Fig. 5~~a-d~~). Moreover, the same MCA patterns occur both in La Niña and El Niño summers, and their frequency is hardly affected (Figs. ~~S7~~S13-S14~~8~~). Nevertheless, during La Niña summers, the effect of the SAM-CGT mode is reinforced over Europe, North Africa and the Indian subcontinent and reaches northward towards Canada~~,~~ while during El Niño summers the effect of the SAM is mainly confined to the tropical belt. For the WNPSM-NPH pattern, a clear asymmetry between El Niño and La Niña summers is shown, with a stronger signal during El Niño years (Fig. 5e,f) that is absent during La Niña years~~ dominates~~. Although~~,~~ this effect is not very large, it is still important. During La Niña summers, SAM exerts a stronger causal effect on the Tibetan High, along with a reinforced monsoon-desert mechanism and a stronger effect on European circulation. This could be ~~due to the fact that~~because under La Niña conditions, the SAM circulation is supported by a favourable Walker circulation (Ju and Slingo, 1995; Terray et al., 2003). The same applies to the southward link: although ENSO does not alter the ~~normalized~~ standardized causal effect from the CGT to SAM, a stronger CGT pattern in the mid-latitudes would have a stronger absolute effect on SAM activity at the weekly ~~time-scale~~timescale. At interannual ~~time-scale~~timescales, Ding et al. (2011) show that the SAM-CGT pair is strongest during the developing phase of ENSO. Therefore, our results further support the hypothesis that ENSO acts on the CGT pattern via its influence on SAM activity, in agreement with Ding et al. (2011). This finding is also in agreement with previous work showing that, at decadal ~~time-scale~~timescales, the CGT and Silk Road ~~(SRP)~~ patterns intensify under PDO negative (i.e. La Niña-like) forcing (Stephan et al., 2019). During El Niño summers, the SAM shows a more prominent effect on the tropical Pacific. Nevertheless, since we condition on the effect of the CGT, we cannot exclude that this strong signal over the Niño-3.4 region may be due to an element outside our CEN. In the North Pacific, causal maps for different ENSO phases show stronger activity of both WNPSM and NPH links during El Niño summers, consistent with previous literature (Chou et al., 2003; Liu et al., 2016). During El Niño events, tropical convection shifts together with SST anomalies towards the central-eastern Pacific, which may favour WNPSM convective activity. In contrast, during La Niña summers convection is shifted towards the Maritime Continent and the western tropical Indian Ocean, reducing convective activity over the central Pacific and WNPSM region. A weaker WNPSM system may in turn be more prone to the influence of mid-latitude variability on the NPH.

~~Quantifying the teleconnections between tropics and mid-latitudes is important in order to better understand and constrain future changes in boreal summer circulation.~~ Future projections describe an increase in monsoon precipitation associated with increasing global mean temperature and thermodynamic arguments (Menon et al., 2013; Turner and Annamalai, 2012).

Quantifying teleconnections between the tropics and mid-latitudes is important in order to better understand and constrain future changes in boreal summer circulation, as ~~However,~~ uncertainty may arise due to changing connections to remote regions. While simulations show great uncertainty in the ENSO response to global warming ~~, suggesting an enhanced warming trend in eastern Pacific SST and stronger extreme ENSO events~~ (Cai et al., 2015; Chen et al., 2017a, 2015, 2017b), ~~whereas~~ observations show a La Niña-like warming trend in central-western Pacific SST (Kohyama et al., 2017; Mujumdar et al., 2012). ~~Recent work has shown that this apparent paradox between observations and models might be due to systematic biases in the models in their representation of tropical Pacific SSTs . Based on our results, if La Niña conditions would become more frequent and SAM activity increases due to global warming, this should favour the influence of the SAM on the CGT pattern and thus on Z200 and T2m patterns across all northern mid-latitudes (Fig. 5f). Finally, a~~A better understating of these teleconnections in observations can help to ~~paves the way for~~ improve~~d~~ S2S forecasts. Verifying the existence and strength of causal teleconnections in forecast models could help diagnose the origin of model biases. E.g. one could disentangle whether lower forecast skill (such as in the mid-latitude regions in summer) is related to local processes or to a misrepresentation of remote drivers. Beverley et al. (2019) showed that the CGT representation in seasonal forecasts is too weak. The CGT is important for predictability of summer extremes and its relationship with the SAM may provide some information to improve predictability. Therefore, these methods could help answering the question “where do model biases come from?” and help developing a physics-based bias correction. At the same time, CEN provide an encoded predictive model, which can be used for actual forecasting (Di Capua et al., 2019; Kretschmer et al., 2017; Lehmann et al., 2020).~~, in particular for the mid-latitude regions, which currently suffer from low seasonal forecasts skill in summer. O~~Our analysis shows that at 4-weekly ~~time-scale~~timescale, the effect of SAM on the CGT pattern has a path coefficient $\beta \sim 0.4$, thus indicating potential for predictability. Further work should analyse how the causal links between these teleconnection patterns are reproduced in corresponding state-of-the-art S2S forecast and climate models, respectively.

Finally, it should not be forgotten that in the context of the present work, ~~the term~~ *causal interpretation* rests upon several assumptions, such as the causal Markov condition, faithfulness, causal sufficiency, stationarity of the causal links and ~~the~~ assumptions ~~made on the~~about the dependence-type (Runge, 2018). These assumptions can be violated in a real system and it is important to be aware of the associated typical challenges for causal discovery in Earth system sciences (Runge et al., 2019a). Causal sufficiency requires that all ~~the important~~ relevant actors in a specific system are accounted for. Here, given the limited set of actors analysed, ~~it~~we cannot ~~be excluded~~ rule out that other ~~not in~~excluded actors may act as important (common) drivers. Therefore, the obtained links can be considered *causal* only with respect to the specific set of actors used here. However, the *absence* of a link can still be interpreted as a likely indication that no direct physical connection among the respective variables exists (Runge, 2018). Moreover, we assume linear dependencies and stationarity for the detection of ~~the~~ causal links. While linearity has been shown to be a useful assumption in previous work (Di Capua et al., 2020), monsoon dynamics behaves partly nonlinearly and therefore, our causal networks ~~by construction~~ only capture some part of the underlying mechanisms by construction. Also, the SAM teleconnections might well behave in an nonstationary manner on decadal ~~time-scale~~timescales (Di Capua et al., 2019; Robock et al., 2003). We therefore cannot ~~rule out~~exclude the possibility that ~~decadal and interdecadal~~ (multi-)decadal oscillations such as the Pacific Decadal Oscillation may influence our results. However, the amount of reliable data is limited and this prohibits the application of nonlinear measures or ~~to~~ the study of effects due to nonstationarity.

5 Conclusions

We have analysed the interdependencies and spatial effects of the two main MCA modes of co-variability between tropical convection and Northern Hemisphere mid-latitude circulation in boreal summer. The first MCA pair connects the circumglobal teleconnection (CGT) pattern in the mid-latitudes with the South Asian monsoon (SAM) convection, while the second MCA pair connects the western North Pacific summer monsoon (WNPSM) convection with a second circumglobal pattern related to the North Pacific High (NPH). These patterns appear qualitatively independent of the analysed ~~time-scale~~timescales and emerge in weekly, 4-weekly and interannual analyses. The strength of the causal links *is* ~~time-scale~~timescale dependent. In particular, the influence of SAM on CGT is strongest at the 4-weekly ~~time-scale~~timescale, while the reversed link is stronger at weekly ~~time-scale~~timescale. The patterns and sign of the standardized causal effect links are also not strongly affected by ENSO. ~~Still, d~~During La Niña years the effect of the SAM on the mid-latitudes intensifies, ~~–~~while we find statistically significant links ~~during El Niño years~~ for the WNPSM effect on the mid-latitudes only for El Niño years ~~dominates~~. Moreover, the boreal summer intraseasonal oscillation exerts strong control on the SAM convection at various lags.

Furthermore, we have introduced causal maps, a new application of the concept of causal effect networks and have highlighted how this method can overcome limitations of correlation maps by removing spurious links. These causal maps further confirm our findings by showing a general positive two-way causal relationship between the dominant modes. Moreover, they highlight specific regions in the mid-latitudes that are particularly affected by the tropical modes (e.g. Eurasia, North America). These findings provide an improved understanding of the interactions between tropical convective activity and circumglobal wave trains that characterize mid-latitude circulation in boreal summer. This may help ~~paves the way for~~ improving sub-seasonal forecasts as well as constraining future projections of boreal summer circulation. Further work shall assess whether these causal relationships are captured by general circulation models and whether this knowledge can be used to improve seasonal forecasts over the mid-latitudes.

References

Benjamini, Y. and Hochberg, Y.: Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing,

J. R. Stat. Soc. Ser. B, 57(1), 289–300, doi:10.2307/2346101, 1995.

Beverley, J. D., Woolnough, S. J., Baker, L. H., Johnson, S. J. and Weisheimer, A.: The northern hemisphere circumglobal teleconnection in a seasonal forecast model and its relationship to European summer forecast skill, Clim. Dyn., 52(5–6), 3759–3771, doi:10.1007/s00382-018-4371-4, 2019.

Branstator, G.: Long-lived response of the midlatitude circulation and storm tracks to pulses of tropical heating, J. Clim., 27(23), 8809–8826, doi:10.1175/JCLI-D-14-00312.1, 2014.

Branstator, G. W.: Circumglobal teleconnections, the jet stream waveguide, and the North Atlantic Oscillation, J. Clim., 15, 1893–1910, doi:10.1175/1520-0442(2002)015<1893:CTTJSW>2.0.CO;2, 2002.

Briegel, L. M. and Frank, W. M.: Large-scale influences on tropical cyclogenesis in the western North Pacific, Mon. Weather Rev., 125(7), 1397–1413, doi:10.1175/1520-0493(1997)125<1397:LSIOTC>2.0.CO;2, 1997.

Cai, W., Santoso, A., Wang, G., Yeh, S. W., An, S. Il, Cobb, K. M., Collins, M., Guilyardi, E., Jin, F. F., Kug, J. S., Lengaigne, M., Mcphaden, M. J., Takahashi, K., Timmermann, A., Vecchi, G., Watanabe, M. and Wu, L.: ENSO and greenhouse warming, Nat. Clim. Chang., 5(9), 849–859, doi:10.1038/nclimate2743, 2015.

Di Capua, G., Kretschmer, M., Runge, J., Alessandri, A., Donner, R. V., van den Hurk, B., Vellore, R., Krishnan, R. and Coumou, D.: Long-Lead Statistical Forecasts of the Indian Summer Monsoon Rainfall Based on Causal Precursors, Weather Forecast., 34(5), 1377–1394, doi:10.1175/waf-d-19-0002.1, 2019.

Di Capua, G., Kretschmer, M., Donner, R. V., Van Den Hurk, B., Vellore, R., Krishnan, R. and Coumou, D.: Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: A theory-guided causal effect network approach, Earth Syst. Dyn., 11(1), 17–34, doi:10.5194/esd-11-17-2020, 2020.

Chakraborty, A. and Krishnamurti, T. N.: A coupled model study on ENSO, MJO and Indian summer monsoon rainfall relationships, Meteorol. Atmos. Phys., 84(3–4), 243–254, doi:10.1007/s00703-002-0601-7, 2003.

Chen, C., Cane, M. A., Wittenberg, A. T. and Chen, D.: ENSO in the CMIP5 Simulations: Life Cycles, Diversity, and Responses to Climate Change, J. Clim., 30(2), 775–801, doi:10.1175/jcli-d-15-0901.1, 2017a.

Chen, L., Li, T. and Yu, Y.: Causes of strengthening and weakening of ENSO amplitude under global warming in four CMIP5 models, J. Clim., 28(8), 3250–3274, doi:10.1175/JCLI-D-14-00439.1, 2015.

Chen, L., Li, T., Yu, Y. and Behera, S. K.: A possible explanation for the divergent projection of ENSO amplitude change under global warming, Clim. Dyn., 49(11–12), 3799–3811, doi:10.1007/s00382-017-3544-x, 2017b.

Cherchi, A., Annamalai, H., Masina, S. and Navarra, A.: South Asian summer monsoon and the eastern Mediterranean climate: The monsoon-desert mechanism in CMIP5 simulations, J. Clim., 27(18), 6877–6903, doi:10.1175/JCLI-D-13-00530.1, 2014.

Chou, C., Tu, J. Y. and Yu, J. Y.: Interannual variability of the Western North Pacific summer monsoon: Differences between ENSO and non-ENSO years, J. Clim., 16(13), 2275–2287, doi:10.1175/2761.1, 2003.

Choudhury, A. D. and Krishnan, R.: Dynamical Response of the South Asian Monsoon Trough to Latent Heating from Stratiform and Convective Precipitation, J. Atmos. Sci., 68(6), 1347–1363, doi:10.1175/2011JAS3705.1, 2011.

Dee, D. P., Uppala, S. M., Simmons, a. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. a., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, a. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, a. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, a. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N. and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137(656), 553–597, doi:10.1002/qj.828, 2011.

Ding, Q. and Wang, B.: Circumglobal teleconnection in the Northern Hemisphere summer, J. Clim., 18(17), 3483–3505, doi:10.1175/JCLI3473.1, 2005.

Ding, Q. and Wang, B.: Intraseasonal teleconnection between the summer Eurasian wave train and the Indian Monsoon, J. Clim., 20(15), 3751–3767, doi:10.1175/JCLI4221.1, 2007.

Ding, Q., Wang, B., Wallace, J. M. and Branstator, G.: Tropical-extratropical teleconnections in boreal summer: Observed interannual variability, J. Clim., 24(7), 1878–1896, doi:10.1175/2011JCLI3621.1, 2011.

Ebert-Uphoff, I. and Deng, Y.: A new type of climate network based on probabilistic graphical models: Results of boreal winter versus summer, Geophys. Res. Lett., 39(18), 1–7, doi:10.1029/2012GL053269, 2012a.

Ebert-Uphoff, I. and Deng, Y.: Causal discovery for climate research using graphical models, J. Clim., 25(17), 5648–5665, doi:10.1175/JCLI-D-11-00387.1, 2012b.

Gadgil, S. and Joseph, P. V: On breaks of the Indian monsoon, Proc. Indian Acad. Sci. (Earth Planet. Sci.), 112(4), 529–558, doi:https://doi.org/10.1007/BF02709778, 2003.

Goswami, B. N. and Ajaya Mohan, R. S.: Intraseasonal Oscillations and Interannual Variability of the Indian Summer Monsoon, J. Clim., 14, 1180–1198, doi:10.1007/BF02842260, 2001.

Goswami, B. N., Sengupta, D. and Suresh Kumar, G.: Intraseasonal oscillations and interannual variability of surface winds over the Indian monsoon region, Proc. Indian Acad. Sci. Earth Planet. Sci., 107(1), 45–64, doi:10.1007/BF02842260, 1998.

Horenko, I., Gerber, S., O’Kane, T. J., J.S., R. and Monselesan, D.: On inference and validation of causality relations in climate teleconnections, in Nonlinear and Stochastic Climate Dynamics, edited by C. L. E. Franzke and T. J. O’Kane, pp. 184–208, Cambridge University Press., 2017.

Joseph, S., Sahai, A. K., Chattopadhyay, R. and Goswami, B. N.: Can El Niño-Southern Oscillation (ENSO) events modulate intraseasonal oscillations of Indian summer monsoon?, J. Geophys. Res. Atmos., 116(20), 1–12, doi:10.1029/2010JD015510, 2011.

Ju, J. and Slingo, J.: The Asian summer monsoon and ENSO, Q. J. R. Meteorol. Soc., 121, 1133–1168, doi:10.1002/j.1477-8696.1997.tb06267.x, 1995.

Kang, I. S., Ho, C. H., Lim, Y. K. and Lau, K. M.: Principal modes of climatological seasonal and intraseasonal variations of the Asian summer monsoon, Mon. Weather Rev., 127(2–3), 322–340, doi:10.1175/1520-0493(1999)127<0322:pmocsa>2.0.co;2, 1999.

Kikuchi, K. and Wang, B.: Formation of tropical cyclones in the Northern Indian ocean associated with two types of tropical intraseasonal oscillation modes, J. Meteorol. Soc. Japan, 88(3), 475–496, doi:10.2151/jmsj.2010-313, 2010.

Kikuchi, K., Wang, B. and Kajikawa, Y.: Bimodal representation of the tropical intraseasonal oscillation, Clim. Dyn., 38(9–10), 1989–2000, doi:10.1007/s00382-011-1159-1, 2012.

Kohyama, T., Hartmann, D. L. and Battisti, D. S.: La Niña-like mean-state response to global warming and potential oceanic roles, J. Clim., 30(11), 4207–4225, doi:10.1175/JCLI-D-16-0441.1, 2017.

Kornhuber, K., Petoukhov, V., Petri, S., Rahmstorf, S. and Coumou, D.: Evidence for wave resonance as a key mechanism for generating high-amplitude quasi-stationary waves in boreal summer, Clim. Dyn., 49(5–6), 1961–1979, doi:10.1007/s00382-016-3399-6, 2016.

Kornhuber, K., Coumou, D., Vogel, E., Lesk, C., Donges, J. F., Lehmann, J. and Horton, R. M.: Amplified Rossby waves enhance risk of concurrent heatwaves in major breadbasket regions, Nat. Clim. Chang., 20, 48–53, doi:10.1038/s41558-019-0637-z, 2020.

Kretschmer, M., Coumou, D., Donges, J. F. and Runge, J.: Using Causal Effect Networks to analyze different Arctic drivers of mid-latitude winter circulation, J. Clim., 29, 4069–4081, doi:10.1175/JCLI-D-15-0654.1, 2016.

Kretschmer, M., Runge, J. and Coumou, D.: Early prediction of extreme stratospheric polar vortex states based on causal precursors, Geophys. Res. Lett., 44(16), 1–9, doi:10.1002/2017GL074696, 2017.

Kretschmer, M., Cohen, J., Matthias, V., Runge, J. and Coumou, D.: The different stratospheric influence on cold-extremes in Eurasia and North America, npj Clim. Atmos. Sci., 1(1), 1–10, doi:10.1038/s41612-018-0054-4, 2018.

Kripalani, R. H. Ã., Kulkarni, A. and Singh, S. V.: Association of the Indian summer monsoon with the northern hemisphere mid-latitude circulation, Int. J. Climatol., 17(10), 1055–1067, doi:https://doi.org/10.1002/(SICI)1097-0088(199708)17:10<1055::AID-JOC180>3.0.CO;2-3, 1997.

Krishna Kumar, K., Rajagopalan, B. and Cane, M.: On the weakening relationship between the indian monsoon and ENSO, Science, 284(5423), 2156–2159, doi:DOI 10.1126/science.284.5423.2156, 1999.

Krishna Kumar, K., Rajagopalan, B., Hoerling, M., Bates, G. and Cane, M.: Unraveling the Mystery of Indian Monsoon Failure During El Nino, Science (80-.)., 314(October), 115–119, doi:10.1126/science.1131152, 2006.

Krishnamurti, T. N. and Surgi, N.: Observational aspects of the summer monsoon, in Monsoon Meteorology, edited by C.-P. Chang and T. N. Krishnamurti, pp. 3–25, Oxford University Press., 1987.

Krishnan, R., Zhang, C. and Sugi, M.: Dynamics of Breaks in the Indian Summer Monsoon, J. Atmos. Sci., 57(1969), 1354–1372, 2000.

Lau, W. K. M. and Kim, K.-M.: The 2010 Pakistan Flood and Russian Heat Wave : Teleconnection of Hydrometeorological Extremes, J. Hydrometeorol., 13, 392–403, doi:10.1175/JHM-D-11-016.1, 2011.

Lehmann, J., Kretschmer, M., Schauberger, B. and Wechsung, F.: Potential for early forecast of Moroccan wheat yields based on climatic drivers, Geophys. Res. Lett., (2008), 1–10, doi:10.1029/2020gl087516, 2020.

Li, T. and Wang, B.: A review on the western North Pacific monsoon: Synoptic-to-interannual variabilities, Terr. Atmos. Ocean. Sci., 16(2), 285–314, doi:10.3319/TAO.2005.16.2.285(A), 2005.

Liu, F., Li, T., Wang, H., Deng, L. and Zhang, Y.: Modulation of boreal summer intraseasonal oscillations over the Western

North Pacific by ENSO, J. Clim., 29(20), 7189–7201, doi:10.1175/JCLI-D-15-0831.1, 2016.

Lu, R.: Interannual variability of the summertime North Pacific subtropical high and its relation to atmospheric convection over the warm pool, J. Meteorol. Soc. Japan, 79(3), 771–783, doi:10.2151/jmsj.79.771, 2001.

McGraw, M. C. and Barnes, E. A.: Memory matters: A case for granger causality in climate variability studies, J. Clim., 31(8), 3289–3300, doi:10.1175/JCLI-D-17-0334.1, 2018.

McGraw, M. C. and Barnes, E. A.: New insights on subseasonal arctic-midlatitude causal connections from a regularized regression model, J. Clim., 33(1), 213–228, doi:10.1175/JCLI-D-19-0142.1, 2020.

Menon, A., Levermann, A., Schewe, J., Lehmann, J. and Frieler, K.: Consistent increase in Indian monsoon rainfall and its variability across CMIP-5 models, Earth Syst. Dyn., 4(2), 287–300, doi:10.5194/esd-4-287-2013, 2013.

Mujumdar, M., Preethi, B., Sabin, T. P., Ashok, K., Saeed, S., Pai, D. S. and Krishnan, R.: The Asian summer monsoon response to the La Niña event of 2010, Meteorol. Appl., 19(2), 216–225, doi:10.1002/met.1301, 2012.

Murakami, T. and Matsumoto, J.: NII-Electronic Library Service, J. Meteorol. Soc. Japan, 72(5), 719–745, doi:https://doi.org/10.2151/jmsj1965.72.5_719, 1994.

Nitta, T.: Convective Activities in the Tropical Western Pacific and Their Impact on the Northern Hemisphere Summer Circulation, J. Meteorol. Soc. Japan, 65(3), 373–390, 1987.

O'Reilly, C. H., Woollings, T., Zanna, L. and Weisheimer, A.: The impact of tropical precipitation on summertime euro-Atlantic circulation via a circumglobal wave train, J. Clim., 31(16), 6481–6504, doi:10.1175/JCLI-D-17-0451.1, 2018.

O'Reilly, C. H., Woollings, T., Zanna, L. and Weisheimer, A.: An interdecadal shift of the extratropical teleconnection from the tropical Pacific during boreal summer, Geophys. Res. Lett., doi:10.1029/2019gl084079, 2019.

Pearl, J.: Causality: Models, Reasoning, and Inference, Cambridge University Press., 2000.

Rao, Y. P. P.: Southwest monsoon, METEOROLOG., Meteorological Monograph: Synoptic Meteorology., 1976.

Riyu, L.: Indices of the Summertime Western North Pacific Subtropical High, Adv. Atmos. Sci., 19(6), doi:10.1007/s00376-002-0061-5, 2002.

Robock, A., Mu, M., Vinnikov, K. and Robinson, D.: Land Surface Conditions over Eurasia and Indian summer monsoon rainfall, J. Geophys. Res., 108(D4), 1–17, doi:10.1029/2002JD002286, 2003.

Rodwell, M. J. and Hoskins, B.: Monsoons and the dynamics of deserts, Q. J. R. Meteorol. Soc., 122, 1385–1404, 1996.

Runge, J.: Causal network reconstruction from time series : From theoretical assumptions to practical estimation, Chaos, 28, 075310, doi:10.1063/1.5025050, 2018.

Runge, J., Petoukhov, V. and Kurths, J.: Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models, J. Clim., 27(2), 720–739, doi:10.1175/JCLI-D-13-00159.1, 2014.

Runge, J., Petoukhov, V., Donges, J. F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D., Marwan, N., Paluš, M. and Kurths, J.: Identifying causal gateways and mediators in complex spatio-temporal systems, Nat. Commun., 6, 9502, doi:10.1038/ncomms9502, 2015a.

Runge, J., Donner, R. V. and Kurths, J.: Optimal model-free prediction from multivariate time series, Phys. Rev. E - Stat.

Nonlinear, Soft Matter Phys., 91(5), 052909, doi:10.1103/PhysRevE.91.052909, 2015b.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. and Sejdinovic, D.: Detecting causal associations in large nonlinear time series datasets, Sci. Adv., 5, eaau4996, 2019a.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Muñoz-Marí, J., van Nes, E. H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Schölkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K. and Zscheischler, J.: Inferring causation from time series in Earth system sciences, Nat. Commun., 10(1), 1–13, doi:10.1038/s41467-019-10105-3, 2019b.

Saha, S. K., Halder, S., Suryachandra Rao, A. and Goswami, B. N.: Modulation of ISOs by land-atmosphere feedback and contribution to the interannual variability of Indian summer monsoon, J. Geophys. Res. Atmos., 117(13), 1–14, doi:10.1029/2011JD017291, 2012.

Samarasinghe, S. M., McGraw, M. C., Barnes, E. A. and Ebert-Uphoff, I.: A study of links between the Arctic and the midlatitude jet stream using Granger and Pearl causality, Environmetrics, 30(4), 1–14, doi:10.1002/env.2540, 2019.

Schubert, S., Wang, H. and Suarez, M.: Warm season subseasonal variability and climate extremes in the northern hemisphere: The role of stationary Rossby waves, J. Clim., 24(18), 4773–4792, doi:10.1175/JCLI-D-10-05035.1, 2011.

Spirtes, P., Glymour, C. and Scheines, R.: Causation, prediction, and search, Boston: The MIT Press., 2000.

Srivastava, A., Pradhan, M., Goswami, B. N. and Rao, S. A.: Regime shift of Indian summer monsoon rainfall to a persistent arid state: external forcing versus internal variability, Meteorol. Atmos. Phys., (0123456789), 1–14, doi:10.1007/s00703-017-0565-2, 2017.

Stephan, C. C., Klingaman, N. P. and Turner, A. G.: A mechanism for the interdecadal variability of the Silk Road Pattern, J. Clim., 32, 717–736, doi:10.1175/JCLI-D-18-0405.1, 2019.

Suhas, E., Neena, J. M. and Goswami, B. N.: Interannual Variability of Indian Summer Monsoon arising from Interactions between Seasonal Mean and Intraseasonal Oscillations, J. Atmos. Sci., 69(6), 1761–1774, doi:10.1175/JAS-D-11-0211.1, 2012.

Teng, H. and Branstator, G.: Amplification of Waveguide Teleconnections in the Boreal Summer, Curr. Clim. Chang. Reports, 5(4), 421–432, doi:10.1007/s40641-019-00150-x, 2019.

Terray, P., Delecluse, P., Labattu, S. and Terray, L.: Sea surface temperature associations with the late Indian summer monsoon, Clim. Dyn., 21(7–8), 593–618, doi:10.1007/s00382-003-0354-0, 2003.

Thomson, S. I. and Vallis, G. K.: Atmospheric response to SST anomalies. Part II: Background-state dependence, teleconnections, and local effects in summer, J. Atmos. Sci., 75(12), 4125–4138, doi:10.1175/JAS-D-17-0298.1, 2018.

Turner, A. G. and Annamalai, H.: Climate change and the south Asian summer monsoon, Nat. Clim. Chang., 2(June), doi:10.1038/NCLIMATE1495, 2012.

Vellore, R. K., Krishnan, R., Pendharkar, J., Choudhury, A. D. and Sabin, T. P.: On the anomalous precipitation enhancement over the Himalayan foothills during monsoon breaks, Clim. Dyn., 43(7–8), 2009–2031, doi:10.1007/s00382-013-2024-1, 2014.

Vellore, R. K., Kaplan, M. L., John, R. K., Sabade, S., Deshpande, N. and Singh, B. B.: Monsoon - extratropical circulation interactions in Himalayan extreme rainfall, Clim. Dyn., 46(11), 3517–3546, doi:10.1007/s00382-015-2784-x, 2016.

Wang, B. and Xu, X.: Northern Hemisphere summer monsoon singularities and climatological intraseasonal oscillation, J. Clim., 10(5), 1071–1085, doi:10.1175/1520-0442(1997)010<1071:NHSMSA>2.0.CO;2, 1997.

Wang, B., Wu, R. and Lau, K. M.: Interannual variability of the asian summer monsoon: Contrasts between the Indian and the Western North Pacific-East Asian monsoons, J. Clim., 14(20), 4073–4090, doi:10.1175/1520-0442(2001)014<4073:IVOTAS>2.0.CO;2, 2001.

Weisheimer, A. and Palmer, T. N.: On the reliability of seasonal climate forecasts, J. R. Soc. Interface, 11(96), 20131162, doi:10.1098/rsif.2013.1162, 2014.

Wiedermann, M., Donges, J. F., Handorf, D., Kurths, J. and Donner, R. V.: Hierarchical structures in Northern Hemispheric extratropical winter ocean–atmosphere interactions, Int. J. Climatol., 37(10), 3821–3836, doi:10.1002/joc.4956, 2017.

Wilks, D. S.: Canonical Correlation Analysis (CCA), Int. Geophys., 100(1992), 563–582, doi:10.1016/B978-0-12-385022-5.00013-0, 2011.

Wu, R., Chen, J. and Chen, W.: Different types of ENSO influences on the Indian summer monsoon variability, J. Clim., 25(3), 903–920, doi:10.1175/JCLI-D-11-00039.1, 2012.

Xavier, P. K., Marzin, C. and Goswami, B. N.: An objective definition of the Indian summer monsoon season and a new perspective on the ENSO–monsoon relationship Prince, R. Meteorol. Soc., 133, 749–764, doi:10.1002/qj, 2007.

Yim, S. Y., Yeh, S. W., Wu, R. and Jhun, J. G.: The influence of ENSO on decadal variations in the relationship between the East Asian and western North Pacific summer monsoons, J. Clim., 21(13), 3165–3179, doi:10.1175/2007JCLI1948.1, 2008.

ABBREVIATION	FULL NAME	DIMENSIONS
BSISO	Boreal summer intraseasonal oscillation	1D time series
SAM	South Asian monsoon - MCA mode 1 OLR	2D spatial pattern + 1D time series
CGT	Circumglobal teleconnection pattern – MCA mode 1 Z200	2D spatial pattern + 1D time series
WNPSM	Western North Pacific summer monsoon – MCA mode 2 OLR	2D spatial pattern + 1D time series
NPH	North Pacific High – MCA mode 2 Z200	2D spatial pattern + 1D time series
Z200	Geopotential height at 200 hPa	2D field + time
OLR	Outgoing longwave radiation	2D field + time
T2M	2m temperature	2D field + time

Table 1. Abbreviations.

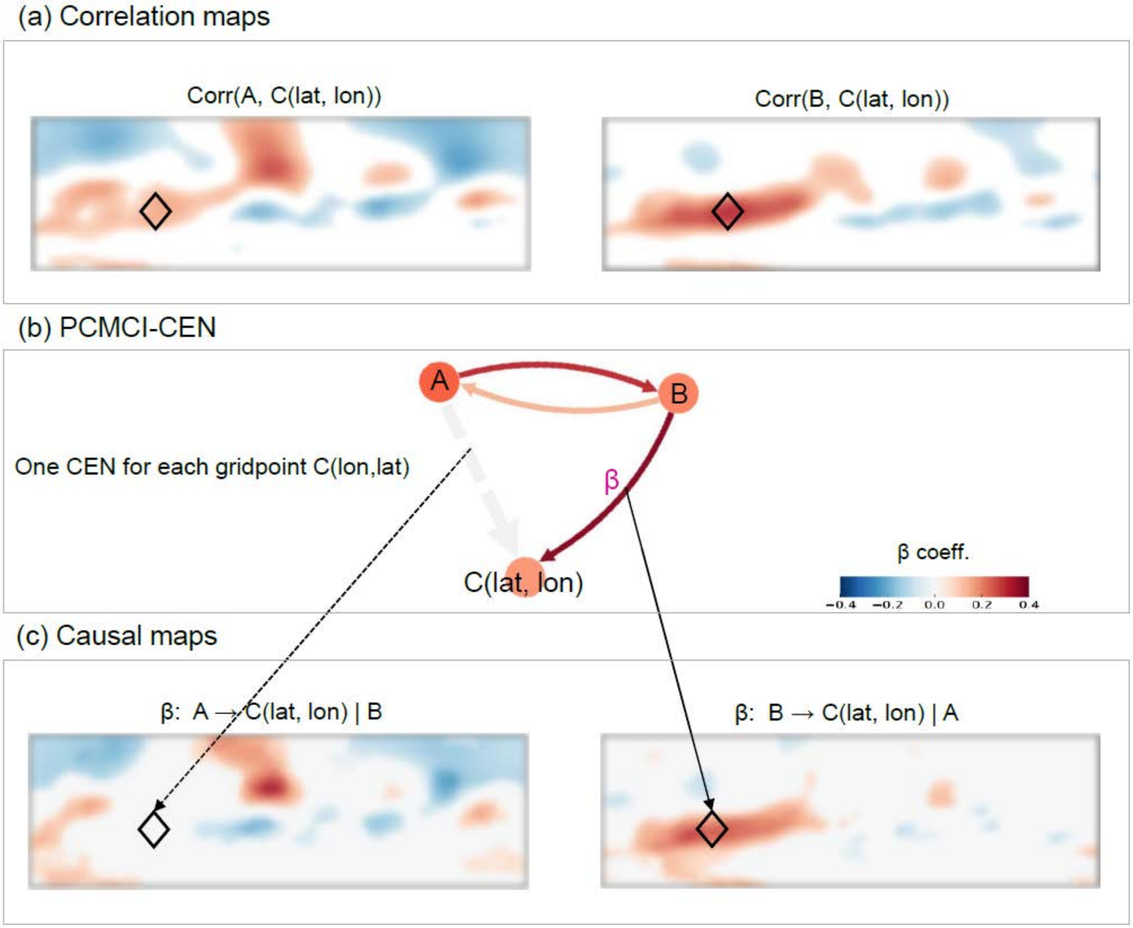

Figure 1: Schematic explanation of causal maps. Panel (a) shows the correlation maps for *A* with *C(*~~*lat, lon*~~*lat,lon)* (left panel) and *B* with *C(*~~*lat, lon*~~*lat,lon)* (right panel). Panel (b) shows an example of a CEN constructed with *A*, *B* and *C(*~~*lat, lon*~~*lat,lon)* for a specific geographical position (identified with a diamond in the 2D maps). Panel (c) shows the corresponding causal maps showing the path coefficients β from *A* to *C*, conditioned on *B* and all autocorrelations (bottom-left panel) and from *B* to *C*, conditioned on *A* and all autocorrelations (bottom-right panel). The "|" denotes the conditioned-out actor: *A* for the right panel and *B* for the left panel. See text for further description.

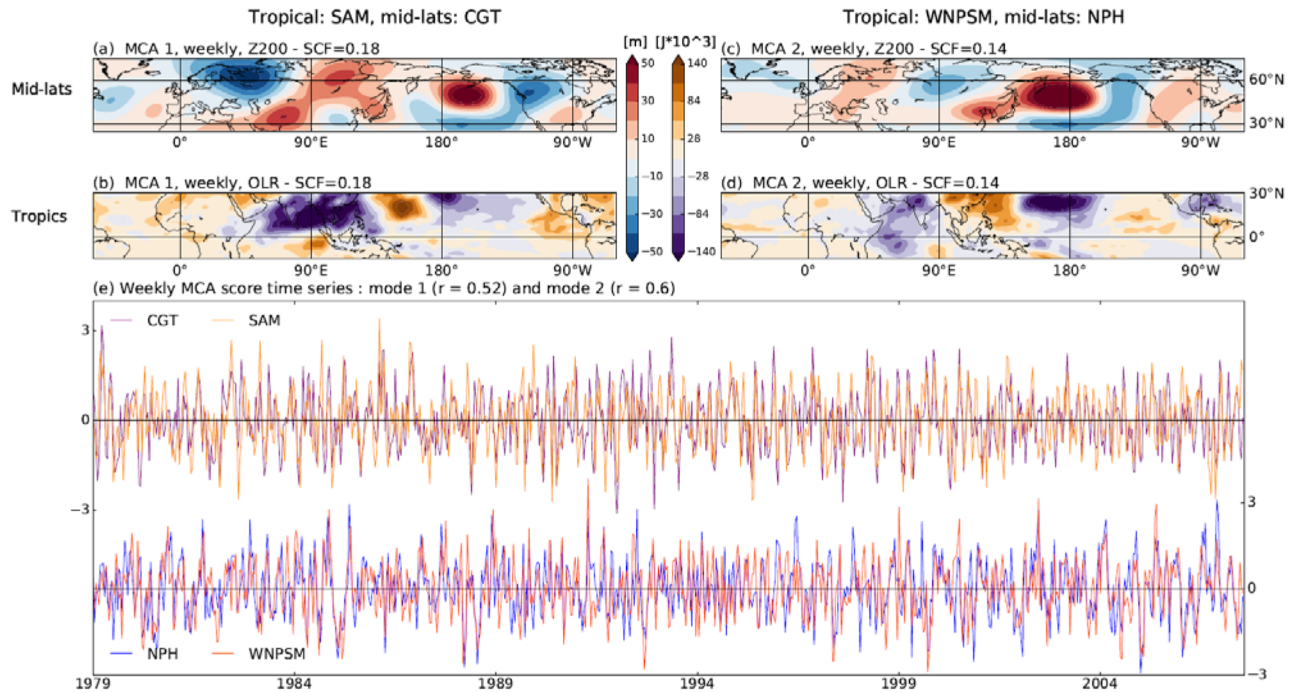

Figure 2: MCA of mid-latitude Z200 and tropical OLR at intraseasonal ~~time-scale~~timescaletimescales. Panels (a) and (b) show the first MCA mode for mid-latitude Z200 (25-75° N) and tropical OLR (15°S-30°N), respectively, at the weekly ~~time-scale~~timescaletimescale. The first MCA highlights the circumglobal teleconnection (CGT) pattern in the mid-latitudes and the South Asian monsoon (SAM) in the tropical belt. Panels (c) and (d): Same as for panel (a) and (b) but for the second MCA mode. This mode depicts the North Pacific High (NPH) in the mid-latitudes and the western North Pacific summer monsoon (WNPSM) in the tropical belt. The squared covariance fraction (SCF) of each MCA mode is given on top of the panels. Panel (e) shows the time series of MCA scores for the two MCA modes at weekly ~~time-scale~~timescaletimescale. Each MCA pattern has its own time series, i.e. one for tropical OLR and one for mid-latitude Z200 (note that different *y*-axes are used).

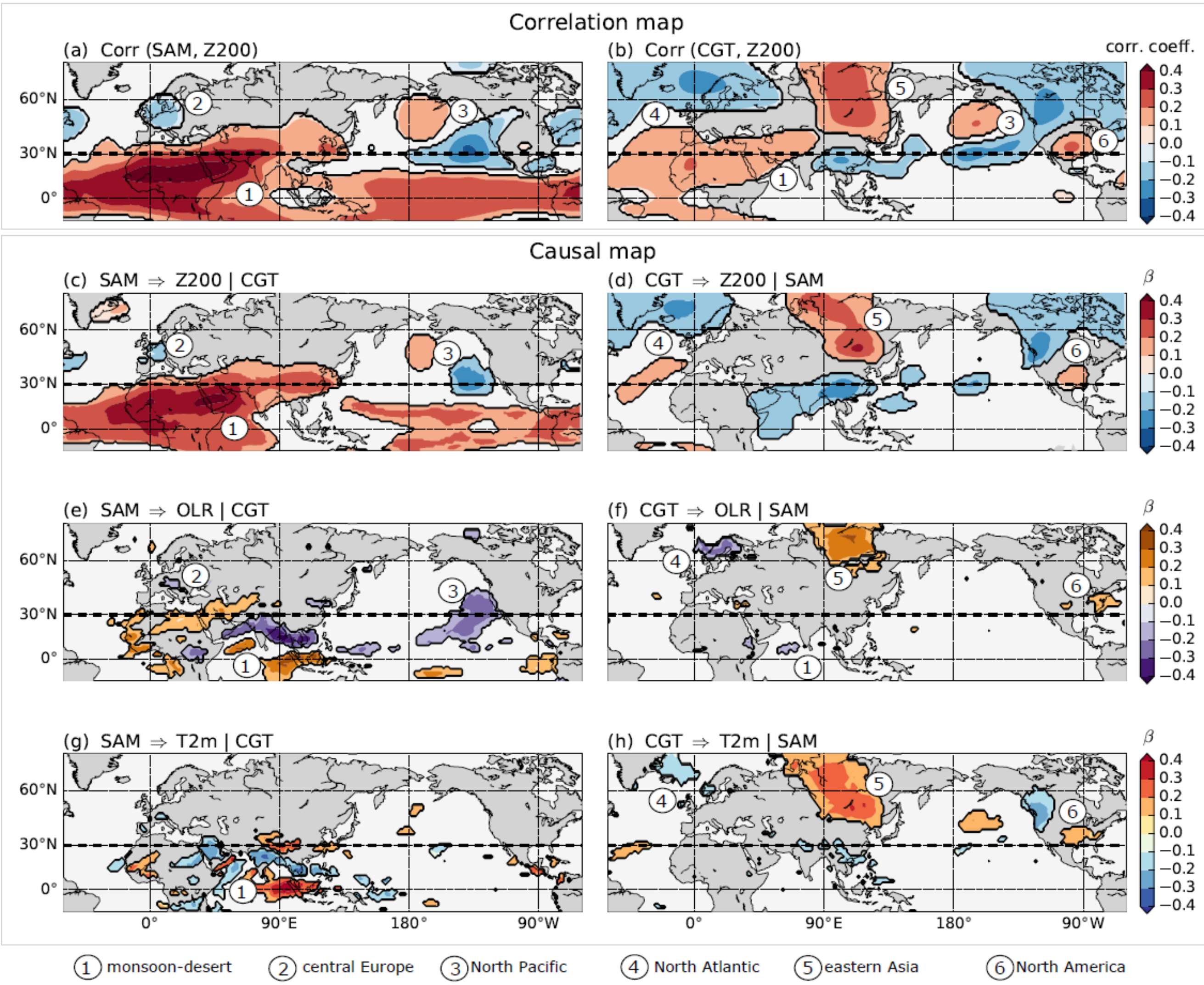

Figure 3: Influence of MCA mode 1 on Northern Hemisphere circulation. Panel (a): ~~correlation~~ Correlation map between the weekly SAM time series and the Z200 field. Panel (b): Same as panel (a) but for the correlation between weekly CGT time series and the Z200 field. Panel (c): ~~path~~ Path coefficient β for link SAM $_{\tau=-1} \rightarrow$ Z200$_{\tau=0}$ for a 3-actors CEN built with SAM, CGT and Z200. Here, the "|" denotes the conditioned-out actor: CGT. Panel (d): Same as panel (c) but for the link CGT $_{\tau=-1} \rightarrow$ Z200$_{\tau=0}$. The "|" denotes the conditioned-out actor: SAM. Panels (e) and (g): Same as panel (c) but for the influence of SAM on OLR and T2m fields respectively. Panels (f) and (h): Same as panel (d) but for the influence of CGT on OLR and T2m fields respectively. Only path coefficients β with $p < 0.05$ (accounting for the effect of serial correlations) and the robustness mask (see Fig. S8 in the Supplementary Material) are shown. ~~by black contours, while grid points which are found significant only with non-corrected p-values are shaded.~~ The dashed black line located at 30°N shows the border between the tropical and the mid-latitude belt which separates OLR and Z200 analysis.

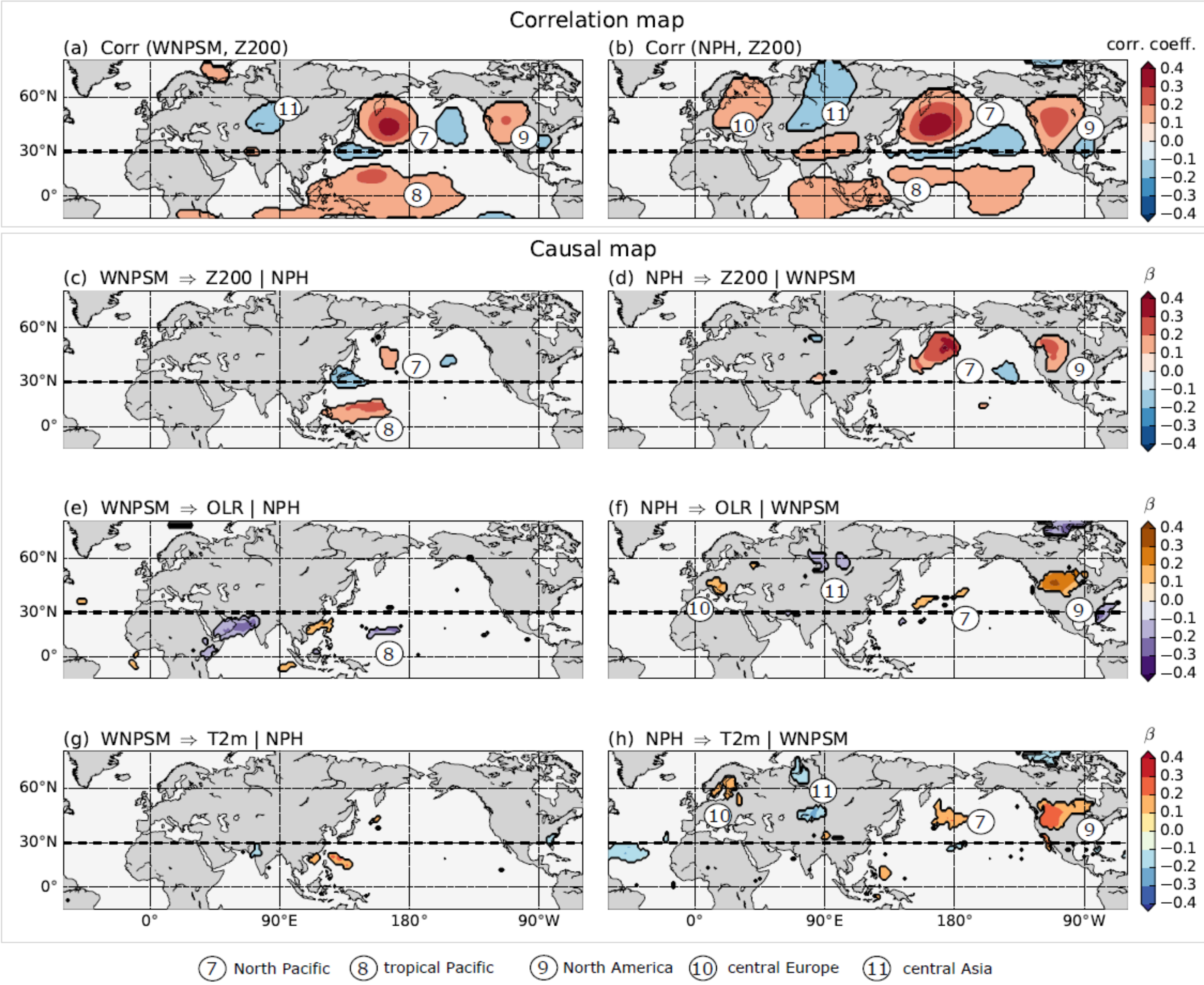

Figure 4: **Influence of MCA mode 2 on Northern Hemisphere circulation.** Panel (a): ~~correlation~~ Correlation between the weekly WNPSM time series and the Z200 field. Panel (b): Same as panel (a) but for the correlation between weekly NPH time series and the Z200 field. Panel (c): ~~path~~ Path coefficient β for the link WNPSM $_{\tau=-1} \rightarrow$ Z200$_{\tau=0}$ in a 3-actors CEN built with WNPSM, NPH and Z200. Here, the "|" denotes the conditioned-out actor: NPH. Panel (d): Same as panel (c) but for the link NPH $_{\tau=-1} \rightarrow$ Z200$_{\tau=0}$. Here, the "|" denotes the conditioned-out actor: WNPSM. Panels (e) and (g): Same as panel (c) but for the influence of WNPSM on OLR and T2m fields respectively. Panels (f) and (h): Same as panel (d) but for the influence of NPH on OLR and T2m fields respectively. Only path coefficients β with $p <$ 0.05 (accounting for the effect of serial correlations) and the robustness mask (see Fig. S9 in the Supplementary Material) are shown ~~by black contours, while grid points which are found significant only with non-corrected *p*-values are shaded~~. The dashed black line located at 30°N shows the border between the tropical and the mid-latitude belt which separates OLR and Z200 analysis.

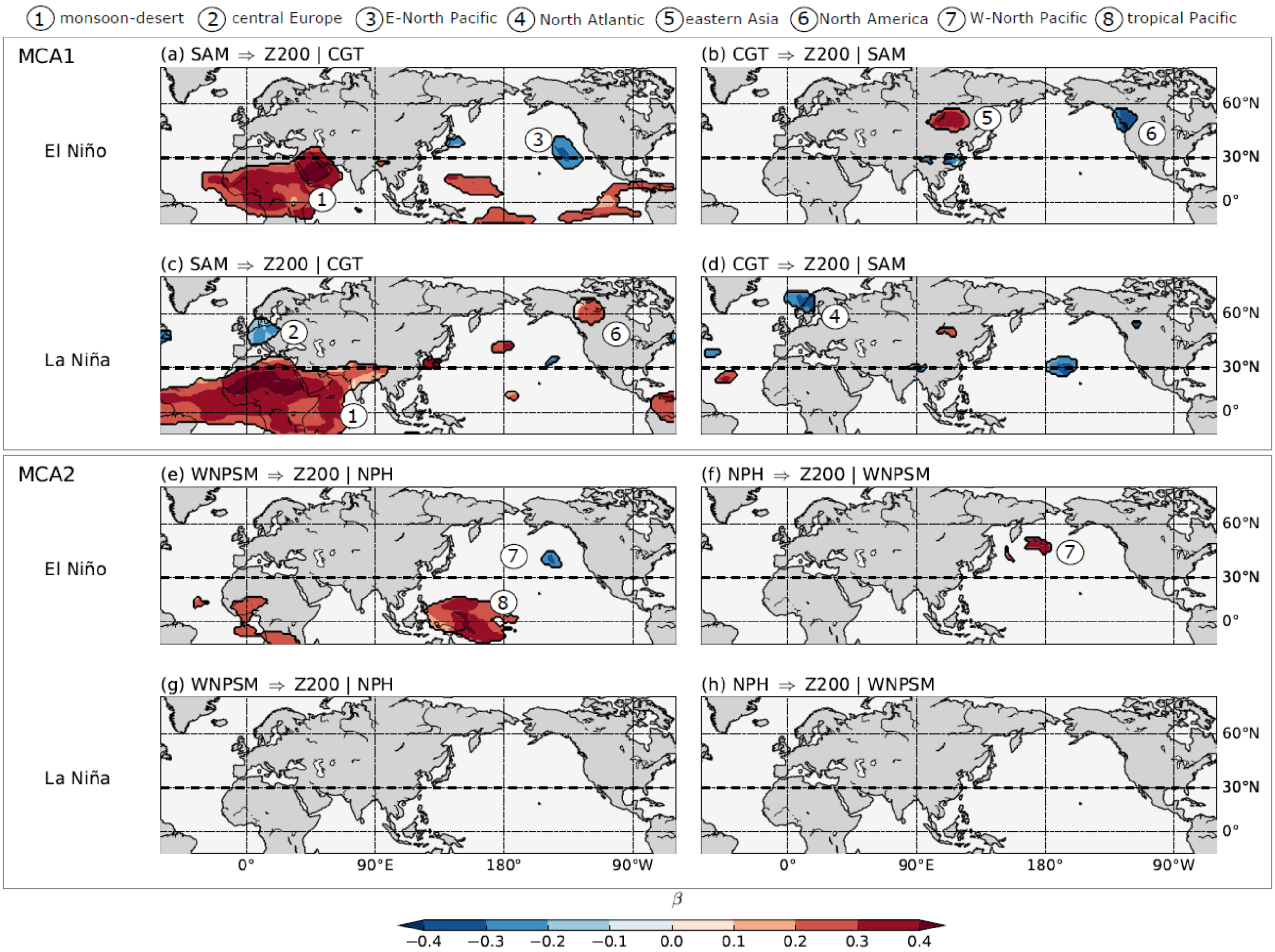

Figure 5: Causal maps and ENSO influence. Panel (a) shows the β for link SAM $_{\tau=-1}$ → Z200$_{\tau=0}$ a 3-actors CEN built with SAM, CGT and Z200 during El Niño years. Here, the "|" denotes the conditioned-out actor: CGT. Panel (b)~~ and (d)~~: Same as panels (a) ~~and (c) ~~but for the link CGT $_{\tau=-1}$ → Z200$_{\tau=0}$. The "|" denotes the conditioned-out actor: SAM. Panel (c): Same as panel (a) but for La Niña years. Panel (d): Same as panel (c) but for the link CGT $_{\tau=-1}$ → Z200$_{\tau=0}$. Panel (e) and (g): Same as panels (a) and (c) but for the link WNPSM $_{\tau=-1}$ → Z200$_{\tau=0}$ from a 3-actors CEN built with WNPSM, NPH and Z200. Panel (f) and (h): Same as panels (e) and (g) but for the link NPH $_{\tau=-1}$ → Z200$_{\tau=0}$. Only path coefficients β with $p < 0.05$ (accounting for the effect of serial correlations) and the robustness mask (see Fig. S12 in the Supplementary Material) are shown~~ by black contours, while grid points which are found significant only with non-corrected p-values are shaded~~. The dashed black line located at 30°N shows the border between the tropical and the mid-latitude belt which separates OLR and Z200 analysis.

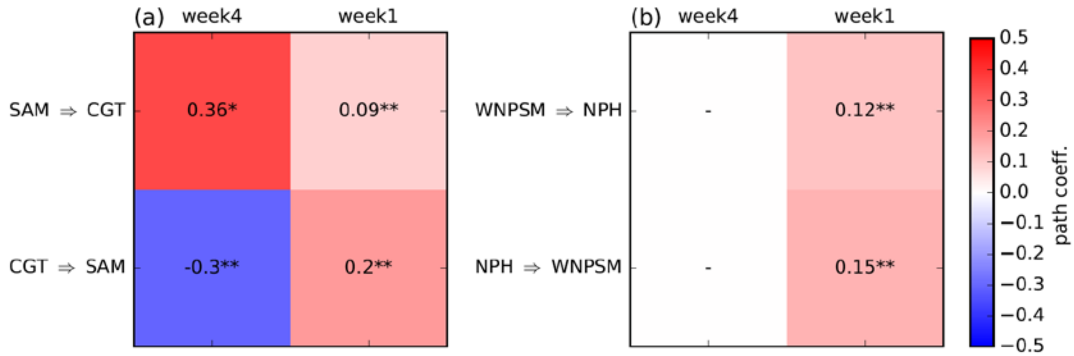

Figure 6: Two-way causal link between tropical OLR and mid-latitude Z200. Shown is the path coefficient for pairs of MCA time series. The CGT is studied along with the SAM, while the NPH is analysed together with the WNPSM. Panel (a) shows the path coefficient β for the link SAM $_{\tau = -1}$ → CGT $_{\tau = 0}$ over the 1979-2018 period (first row), and path coefficient β for the link CGT $_{\tau = -1}$ → SAM $_{\tau = 0}$ (second row). 4-weekly β are shown in the left column, weekly β values are shown in the right column. Panel (b): Same as for panel (a) but for WNPSM $_{\tau = -1}$ → NPH $_{\tau = 0}$ and NPH $_{\tau = -1}$ → WNPSM $_{\tau = 0}$ links respectively. β values with $p < 0.1$ (0.05) are identified with one (two) asterisks.

Acknowledgments

This work has been financially supported by the German Federal Ministry for Education and Research of Germany (BMBF) via the Belmont Forum / JPI Climate project GOTHAM (grant no. 01LP1611A) and the BMBF Young Investigators Group CoSy-CC^2: Complex Systems Approaches to Understanding Causes and Consequences of Past, Present and Future Climate Change (grant no. 01LN1306A). The contribution of AGT was supported by NERC in the UK under the BITMAP project (grant number NE/P006795/1) and in the STIMULATE project through the Weather and Climate Science for Service Partnership (WCSSP) India, a collaborative initiative between the Met Office, supported by the UK Government's Newton Fund, and the Indian Ministry of Earth Sciences (MoES).

Data availability. The data used in this article can be accessed by contacting the corresponding author.

Author contributions. GDC, DC, BvdH, JR and AGT designed the analysis. GDC performed the analysis and wrote the first draft of the paper. All authors contributed to the interpretation of the results and to the writing of the paper.

Competing interests. The authors declare that they have no conflict of interest.