Dear Editor,

We were pleased to receive the acceptance for the revised version of our manuscript "Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: A Theory-Guided Causal Effect Network approach" by Di Capua et al. to the Copernicus Journal *Earth System Dynamics*.

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

Please note that during the revision process, a technical problem in the code has been corrected. This has *not* affected any of the results nor the main message of the manuscript. However, in Fig. 7 in the main manuscript a minor change in the significance of one results is visible (this difference is commented in lines 484-485).

Yours sincerely,

Potsdam, 10.12.2019

Giorgia Di Capua on behalf of all authors

--
Giorgia Di Capua
Potsdam Institute for Climate Impact Research - Earth System Analysis

Telegrafenberg A62, 14473 Potsdam, Germany
email: dicapua@pik-potsdam.de
phone: +49-331-288-2514

Response to review #1

The paper is an important contribution to the literature and deserves to be published after some minor alterations. The paper could do more to mention previous literature such as that of Rodwell and Hoskins, which is important in explaining the mechanism for the link between the monsoon, central Asia (including the Ding and Wang region) and further afield to the Mediterranean. In addition, the treatment of what the paper describes as the internal dynamics of the monsoon is confusing. The authors should perhaps re-examine this explanation or make a judgement as to whether this can be better described as part of the monsoon intraseasonal variability. Note that in boreal summer, it would be more appropriate to refer to the monsoon ISO/BSISO rather than the MJO. The paper generally has some very well produced figures. Specific comments can be found below.

- Answer -

We thank the anonymous reviewer for his/her very positive overall feedback and useful suggestions on missing literature, on how to better discuss and compare the existing literature in the context of our work and on how to improve the clarity of the paper. We have included all suggestions in our revised version of the paper, with a particular attention on discussing the analogies and differences that our results show if compared to previous work. We now describe the BSISO in the introduction and discussion section, with a more specific attention to its relation to MJO in summer. We have included the suggested literature in the introduction and we now compare our results in light of previous contributions to this specific topic. A fully detailed answer to each comment is reported below.

Specific comments

Line 35: It is not clear what is meant by the "ISM convective cell".
Is this some sort of mesoscale convective system or an individual cloud? I suggest that some alternative terminology is found. (See also later on line 39.) –
Answer – We thank the reviewer for pointing out this unclear terminology. It is indeed more appropriate to substitute the term "convective cell" with the more general term "circulation system" (of the ISM). We have thus made this substitution in the text.

Line 36: "Periods with strong updraft lead to strong rainfall one week later". Please clarify if this means locally or acting at some distance along the teleconnection. –
Answer – We have further specified the location where the updraft is located, see lines 36-37: "*Moreover, we identify a negative feedback between strong updraft located over India and the Bay of Bengal and the ISM rainfall acting at a biweekly timescale, with enhanced ISM rainfall following strong updraft by one week.*".

Line 37: In your statement, "internal ISM dynamics has the strongest CE of 0.5", what does this mean in the context of the similar statement earlier that explains the meaning of these value? Effectively you are saying (unless I am misinterpreting the purpose of the CE value), "A one standard deviation shift in the internal ISM dynamics causes a 0.5 standard deviation shift in ISM rainfall one week later". It is not clear what you are meaning by this. –
Answer – The meaning of the CE is correctly understood. We have however specified that here by internal dynamics it was actually meant the updraft identified over the Indian subcontinent and the Bay of Bengal. The new sentence now reads (lines 35-38) "*Moreover, we identify a negative feedback between strong updraft located over India and the Bay of Bengal and the ISM rainfall acting at a biweekly timescale, with enhanced ISM rainfall*

following strong updraft by one week. This mechanism is possibly related to the Boreal Summer Intra-seasonal Oscillation. In our analyses, the updraft has the strongest CE of 0.5, while the Madden-Julian Oscillation variability has a CE of 0.2-0.3.".

Line 43: In the introduction it may be worth also citing the work of Stephan et al. (2019, https://doi.org/10.1175/JCLI-D-18-0405.1) who use the CEN technique to examine the CGT/SRP link to the ISM, albeit in the context of decadal variability. –
Answer – We thank the anonymous reviewer for suggesting to include this work in our paper. We now refer to this reference in the introduction to highlight how CENs can be applied to study this type of problem, see line 148-151: "*CENs can be used to test whether hypothesized links or teleconnections are likely to represent true physical pathways or rather artefacts due to spurious correlations. CEN have been applied to the study stratospheric polar vortex variability (Kretschmer et al., 2016), multi-decadal North Atlantic overturning circulation (Schleussner et al., 2014) and to study the causal interactions between the ISM, the Silk Road Pattern and the monsoon-desert mechanism (Stephan et al., 2019), showing their usefulness in testing existing hypothesis by eliminating those that are not supported by causality .*"". We discuss these results more in detail in the discussion section, and compare their outcomes with our analysis, see lines 624-629: "*Therefore, our work is in agreement with previous findings that show an influence of the ISM latent heat release on north-east Africa arid conditions via the excitation of Rossby waves to the west. However, here we focus on the relationship between the MT rainfall and the circumglobal teleconnection pattern, thus a link from the MT rainfall towards the Saharan region cannot be inferred from this analysis. Moreover, the causal relationship between the north-eastern Indian rainfall and the downdraft over the north-eastern Sahara related to the monsoon-desert mechanism has already been shown (Stephan et al., 2019).*".

Lines 63-64: In the statement, ". . .this thermodynamic perspective [cloud cover etc. acting to cool the surface] is useful to understand the quasi-biweekly variations of the

ISM elements locally", why is the quasi-biweekly time scale of particular interest? In a monsoon regime wouldn't we expect CAPE to build up and be destroyed much more regularly than this, e.g. on a daily basis, given that the surface forcing is strong and there is a plentiful supply of moisture?

Answer - We agree. Following the also the comments below, we now introduce here the BSISO, and refer to the above described mechanism as a potential local driver of the ISM intraseasonal variability. See lines 64-68: "*While this thermodynamic perspective is useful to understand the quasi-biweekly variations of the ISM elements locally, the spatio-temporal variations in the evolution of active and break phases over the Indian monsoon region are known to involve interactions between the wind anomalies and the northward propagation of the major rain band anomalies of the Boreal Summer Intraseasonal Oscillation (Chattopadhyay et al., 2009; Shige et al., 2017; Wang et al., 2006). ".*

Line 77: Rather like the earlier comment, what is meant by a "convective cell" of the MJO here? –
Answer – Corrected sentence lines 84-85: "*Normally, only one region of strong convective motions related to the MJO is present in these regions.*".

Line 87: Regarding "downstream", careful to specify what is meant. Do you mean downstream with respect to the jet, i.e. further east? This seems to be in the same direction as given in the previous sentence, rather than "on the other hand" as the sentence starts. –
Answer – We thank the anonymous reviewer for carefully reading the manuscript. We have changed the structure of the paragraph and we now discuss separately the DW2005 and

DW2007 related mechanisms in lines 90-107 and 116-131 respectively, this should improve the readability of the paper.

Line 166 onwards: the methods in this section are explained well given the complex techniques involved and the referencing is done very well.

Figure 2 and others in the paper are very inventive and generally of very good quality.

Line 205: How are the "northern mid-latitudes" used for the EOF calculation defined? –
Answer – We have clarified this point as follows (see lines 268-269):"*Panels (a) and (b): EOF1 and EOF2 expressed as covariance for the JJAS weekly Z200 field in the Northern Hemisphere (0°-90°N, 0°-360°E) for the period 1979-2017.*".

Line 205: "lag = -1 week". It would be better to clarify exactly which variable is leading the other, to avoid ambiguity. –
Answer - We have clarified this point as follows (see lines 269-271): "*Panel (c): correlation between weekly MT rainfall (lag = 0) and Z200 (lag = -1 week). Panel (d): the CGTI region (white box) and the correlation between CGTI and Z200 (both at lag = 0), which forms the circumglobal teleconnection pattern.*".

Line 224: That the first three EOFs are not mutually separable is somewhat of a mathematical interpretation, but what does it mean in terms of any physical explanation of the EOFs? –
Answer - We have clarified this point as follows (lines 287-292): "However, by design EOFs do not necessarily reflect physical patterns (irrespective of whether the individual EOFs are separable or not). They only capture dominant patterns of variability, and here EOF2 is useful since it resembles the wave-pattern in the correlation map and thus it likely results from the circumglobal teleconnection pattern related physics. In general, *we cannot a priori exclude an influence of EOF1 on the analysed system (while an influence from EOF2 is expected), therefore we will test whether this is the case in the following part of this sub-section.*".

Line 232: Perhaps it would be a good idea to add a further physical interpretation of the Z200 leading MT rainfall by 1 week. Presumably this is a west to east propagation of the signal. –
Answer – we have included this comment as follows (lines 298-300): "*The arch-shaped structure which is visible in Fig. 2c over Europe and central Asia suggests that there could be a wave pattern propagating eastward from Western Europe and affecting the MT rainfall with a 1-week lag. This hypothesis will be further tested in the next section.*".

Line 247: Does the global scale mentioned here imply that the correlation was performed over all global grid-points? It seems a bit excessive and could probably just be done over the hemisphere. –
Answer – We thank the anonymous reviewer for this remark, we did mean that the spatial correlation and the EOFs are calculated only over the Northern Hemisphere and we have corrected this ambiguity (see lines 277-279).

Line 275: Here (and also for the benefit of later), please clarify that the lags of 1-week act in the direction of the arrows (e.g. CGTI leads MT by one week, and then MT feeds back on CGTI one week later). –
Answer – We have clarified this point including the text the following sentence, (see lines 356-358) :"*Note that the causal effect here is always acting in the direction of the arrow with a lag of one week, e.g. the arrow from CGTI towards MT rainfall represents a causal effect of xx (given by the color) from CGTI to MT rainfall with a lag of one week.*".

Line 281: Consistency with Ding and Wang is mentioned here, but ultimately this also supports Rodwell and Hoskins (1996) on monsoon-desert coupling (https://doi.org/10.1002%2Fqj.49712253408). This seems to be a surprising omission from the paper given that it helps explain the relationship between the monsoon and Ding & Wang region, and ultimately further afield to the Mediterranean. It would be good to discuss this work in the introduction and perhaps see also the works of Cherchi et al. (2014; https://doi.org/10.1175/JCLI-D-13-00530.1, which examines the issues in coupled models, and also see the discussion of the Rodwell and Hoskins mechanism in relation to Ding & Wang in Beverley et al. (2019; https://doi.org/10.1007/s00382-018-4371-4). –
Answer – We thank the anonymous reviewer for suggesting to include the discussion of the monsoon-desert mechanisms in our work, we now discuss all the suggested references in the introduction, lines 108-115 "*The latent heat released via strong convection in the ISM region has also been shown to influence regions which are located upstream (i.e. eastward). The so-called monsoon-desert mechanism involves long Rossby waves to the west of the ISM region generated by ISM latent heating. These waves enhance the downward flow over the eastern Mediterranean and north-eastern Sahara Desert suppressing precipitation in these dry regions (Rodwell and Hoskins, 1996). A subset of CMIP5 models is able to capture this mechanism (Cherchi et al., 2014) and CMIP5 scenarios for the 21st century project increased ISM precipitation, despite a decrease in the strength of the ISM circulation. Thus the monsoon-desert mechanism could contribute to drying and warming trends projected for the eastern Mediterranean region, exacerbating the desiccation conditions in these regions (Cherchi et al., 2016)"* and lines 103-107 "*Seasonal forecast models, e.g. from the European Centre for Medium-Range Weather Forecasts (ECMWF), tend to have difficulties reproducing this pattern correctly: the circumglobal teleconnection pattern is too weak in models and one of the possible causes could be a too weak interaction with the north-western India rainfall (Beverley et al., 2019). "* as well as in the discussion on the paper lines 622-625 "*Moreover, the larger CGTI region as defined in Fig. 4c, while showing the strongest correlation over the CGTI region as defined in Fig. 2d, also stretches south-westwards towards north-east Africa, to the area that features the downdraft related to the monsoon-desert mechanism (Rodwell and Hoskins, 1996). Therefore, our work is in agreement with previous findings that show an influence of the ISM latent heat release on north-eastern Africa arid conditions via the excitation of Rossby waves to the west.*" and lines 679-685 "*While previous studies that have analysed the relationships between the ISM and the mid-latitude circulation have often considered the rainfall over north-western India (Beverley et al., 2019; Ding and Wang, 2005; Stephan et al., 2019), here we take into account the MT rainfall, showing that connection between active and break phases of the ISM (by definition identified over this region, Krishnan et al., 2000) and the circumglobal teleconnection pattern. The circumglobal teleconnection pattern is an important source of variability for European summer weather, thus improving its representation in seasonal forecasting models could in turn improve seasonal forecast in boreal summer (which generally show lower skill than those for boreal winter) (Beverley et al., 2019).*".

In figure 4, panel (d) seems rather pointless. Can a composite difference of ISM rainfall not be given based on the precursors illustrated in the earlier panels of the figure? –
Answer – The purpose of this panel in Fig. 4 is only to help the interpretation, as it helps to visualize which region is acting on each response variable at any given lag. lags

Line 326: What do the black contours represent in the figure? –

Answer – We have clarified this point as follows for both Fig. 4 and Fig. 6 (lines 275 and 409): "*Regions with correlation values with a p-value of $p < 0.05$ (accounting for the effect of serial correlations) are shown by black contours.*".

Lines 330-332: It may be worth citing some earlier works linking NAO and monsoon, e.g. Goswami et al. (2006, https://doi.org/10.1029/2005GL024803). –
Answer – We thank the anonymous reviewer for suggesting this useful reference. We now cite this work in the introduction, see lines 117-121: "*At inter-decadal and interannual timescales, SSTs related to the Atlantic Multi-decadal Oscillation (AMO) index have been shown to modulate the strength of the ISM by an atmospheric bridge involving the North Atlantic Oscillation (NAO) index: positive (negative) NAO phases modify westerly winds and associated storm tracks in the North Atlantic/European area, and modify tropospheric temperatures over Eurasia, thus enhancing (weakening) strength of the ISM rainfall (Goswami et al., 2006).*". We have also included this reference in the discussion section, see lines 619-621: "*This wave pattern is visible in geopotential height fields, temperature and precipitation anomalies, and acts on MT rainfall via the CGTI with a 1-2 week lead time, in agreement with the DW2007 hypothesis and with previous studies showing that there is a connection from the North Atlantic toward the ISM system (Goswami et al., 2006).*".

Line 348: The "Himalayan plateau" is not appropriate as it does not exist. Do you mean the Himalayas or the Tibetan Plateau (or both)? Better to think of a more appropriate term. –
Answer – We thank the anonymous reviewer for noticing this mistake. We indeed meant the Tibetan Plateau and we have now corrected this mistake (see line 432).

Line 351-355: Rather than the "ITCZ", isn't an interpretation of this that the one week earlier than strong rain over the trough, we have rain further south over India, such that we have a northward propagation of the BSISO? The rainband looks slightly titled rather than entirely zonal. That would be why you see a causal relationship of MJO2 to W1. –
Answer – We thank the anonymous reviewer for this insightful suggestion. We know refer to the BSISO instead of to the ITCZ, see lines 434-436: "*OLR1 spatially overlaps with W1, the largest causal precursor in the vertical wind field, representing the north-west/south-east tilted rainfall band related to the BSISO over the northern Indian Ocean and western Pacific Ocean (Fig. 6b, top panel).*".

Lines 369-370: There was an explanation on the use of OLR earlier, so it isn't needed again. –
Answer – We thank the anonymous reviewer for pointing out this repetition, which we have now removed.

Line 372: Here and elsewhere we are referred to the internal dynamics of the monsoon. Is this not perhaps better explained as the intraseasonal variability, in other words related to the BSISO of active and break phases? –
Answer – We thank the anonymous reviewer for this insightful suggestion. We know refer to the interannual variability and the BSISO instead, see lines 458-460: "*The CEN built with OLR1, W1 and the MT rainfall at weekly timescale (Fig. 6c) represents the intraseasonal variability of the monsoon cell and its relation with the BSISO and the initiation of active and break phases. ".*

Lines 373-374: The stronger MT rainfall followed by weaker ascending motions one week later is probably consistent with a monsoon active phase moving into a break. –
Answer – We have now included this suggestion in our manuscript, see lines 466-469: "Thus, from a physical point of view, this can be explained in two ways: (1) as a seesaw representing

the switch of the ISM system between an active and a break phase and (2) via an increase of atmospheric static stability due to latent heat release in the higher layers of the troposphere."

Line 377-380: The negative feedback described here is logical in the sense that convective rainfall acts to stabilize the atmospheric column (and destroy CAPE). But in the monsoon regime, CAPE will quickly be reinvigorated given the surface forcing and good availability of moisture, possibly within a day. Rather than a negative feedback, couldn't one also argue that (as in Gill's off-equatorial heating), the LH release of the monsoon convection leads to a feedback and strengthening of the flow. What you are seeing here may instead be better explained as part of the migration between states of the BSISO. –
Answer – We thank the reviewer for this insightful comment, here (and later) we now discuss the BSISO as a possible explanation of the presence of a negative feedback between W1 and the MT rainfall, see lines 357-360: "*W is calculated at 500 hPa and W1 thus represents the ascending branch of the ISM circulation cell and nearby BSISO. The CEN built with OLR1, W1 and the MT rainfall at weekly timescale (Fig. 6c) represents the intraseasonal variability of the monsoon circulation and therefore its relation with the BSISO and the initiation of active and break phases.*".

Lines 454-455: Are the internal variability and the MJO-related part not somehow related, i.e. they are intraseasonal variability of the monsoon. –
Answer – We thank the reviewer for this insightful comment, here (and later) we now discuss the intraseasonal variability instead of the internal dynamics, see lines 470-472: "*In this study, we apply causal discovery algorithms to analyse the influence of global middle and upper tropospheric fields on the ISM rainfall and study the two-way causal links between the mid-latitude circulation and ISM rainfall, together with tropical drivers and ISM intraseasonal variability.*".

Lines 461-462: Monsoon-desert coupling could be mentioned here. –
Answer – We thank the reviewer for this useful suggestion, we now include the monsoon desert mechanisms in the discussion section, see lines 620-623: "*Moreover, the larger CGTI region as defined in Fig. 4c, while showing the strongest correlation over the CGTI region as defined in Fig. 2d, stretches south-westwards towards north-east Africa, to the area that features the downdraft related to the monsoon-desert mechanism (Rodwell and Hoskins, 1996; Stephan et al., 2019). ".*

Line 478: The description here of internal variability dominating over interannual variability might be better explained in terms of intraseasonal versus interannual. –
Answer – We have included this suggestion in lines 644-626: "*The reported findings are in good agreement with the existing literature. It is well known that internal intraseasonal variability dominates ISM inter-annual variability (Goswami and Xavier, 2005; Suhas et al., 2012).*".

Lines 485-486: It would have been useful to have these discussions about the ITCZ migration earlier. This might be better explained in any case in terms of the switch between break and active phases of the BSISO, as the region of maximum rainfall propagates northward from the equatorial position during a break to Indian latitudes during an active phase. –
Answer – We have included this suggestion and we now discuss the BSISO link to the MT rainfall already in lines 605-610 "*Next to the influence of the MT rainfall, our analysis also shows that the link between the circumglobal teleconnection pattern and the ISM can be seen in the wider perspective of the BSISO variations. Our OLR1 and W1 regions (see Fig. 6a,b)*

show a north-west/south-east tilted rainfall band that shows great similarities with the BSISO rainfall band (Wang et al., 2018). In light of this relationship, the negative feedback that characterizes the causal links between W1 and the MJO2 (see Fig. 6d) and the MT rainfall and W1 (see Fig S18, in the SI) can be interpreted as the shift of the ISM system between a break and an active phase and to the north-eastward propagation of BSISO convective anomalies.".

Line 489: Here and elsewhere, the MJO is discussed but it may be better to think in terms of the BSISO, which is the summer manifestation of the MJO with northward propagation.
Answer – We have included this suggestion throughout the whole manuscript and we now discuss the BSISO together with the MJO. Moreover, we also discuss evidence that the OMI index we use to describe the MJO, is also useful to describe the BSISO, see lines 87-89 "*However, during boreal summer, the MJO strength is reduced as compared to boreal winter, and both the MJO and the BSISO propagation and phases can be well described by the outgoing longwave radiation MJO index (OMI) (Wang et al., 2018).*". In the discussion section, we now also include the discussion of the BSISO next to the MJO, see previous comments for an example.

Line 499 (and elsewhere): It would be better to expand acronyms such as RG-CPD or explain again what they are in the conclusions, for the benefit of the reader that goes straight to the conclusion section. –
Answer – We have included this suggestion in the discussion and conclusion section of the paper.

Spelling, grammar & other trivia

Line 34: Change "influences back" to "feeds back on" – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 63: "in support of suppressing" is rather contradictory English. I suggest changing it to something like, "which tends to suppress convection". – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 76: Insert "The" before "MJO". – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 103: It would be customary for a few sentences here at the end of the introduction listing what sections are to follow in the remainder of the paper. – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 125: missing space before "algorithm". – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 188 and elsewhere: Perhaps put the year of the Pai et al. reference here and elsewhere. – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 208: In describing panel (e) it would be clearer to express this as a composite difference, e.g., "Composite temperature difference between weeks with...". – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 232: timescale –> timescales – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 240: patter –> pattern – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 243: llink –> link – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 262: As in the earlier comment, it might be best to explain as a composite difference. – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 318: "figure" for the new sentence should be capitalized. – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Line 479: Enables to –> enables us. – **Answer** – We thank the reviewer for carefully reading the manuscript, we have now included this suggestion in the paper.

Response to review #2

The paper addresses an interesting topic from a network point of view, confirming previous results with this new data-driven learning methodology. However, I think that the authors should make more clear which is the added value of this methodology (with respect to using more traditional methodologies) and highlight which results are different from what was already known. In particular, the determination of the causality of the links seems to be the most important strength of the method, but this is not sufficiently stated. In addition, I find that the selection of some parameters (weekly time-span, CGTI region, 1 week time-lag, precipitation only over MT and not over all the Monsoon area, etc.) are poorly justified and some discussion on how the results are modified with a different selection is necessary. Finally, some word on why the linear framework is adequate for the study if these mechanisms is missing. –

Answer – We thank the anonymous reviewer for his/her suggestions to improve the clarity of the paper. We have taken all suggestions into account and we now provide a clearer explanation on why causal tools generate better interpretable results as compared more basic statistical approaches like correlation. We also address the issues on the linear framework and show that the results are robust when different regions are selected. A detailed answer to each point is provided below.

Major questions/comments:

1. Which is the added value of the methodology with respect of using just correlations and partial correlations? I think that probably the values of the links could be reproduced just with more simple statistics but the causality determination is the highlight of the proposed methodology. This should be more clearly emphasized and showing the values of the correlations / partial correlations among all the selected variables (MT precipitation, MJO2, PC2, CGTI, etc.) is advised for comparison. -

Answer – We thank the reviewer for pointing out that the description of the advantages of using causal discovery tools instead of simple correlation needs to be emphasized more. As a clarification, the causal algorithm we use is based on partial correlation, calculated on the residuals obtained after regressing two selected variables on a set of third variables (conditions). This is a lengthy iterative process whereby one conditions on all possible variable combinations for each individual link. We explain this in full length in the SI of the paper and shorter in the main text in lines 141-158. Following the suggestions of the reviewer, we have added some clarifying examples, and expanded the paragraph that describes the advantages of using the causal effect over simple correlation, see introduction section, lines 132-151: "*Lagged correlation and regression analysis are commonly used to assess the relationship between two or more climate variables (Ding and Wang, 2005; Lau and Kim, 2011; Vellore et al., 2014). Such an analysis is useful as it gives a first information on the association of two or more variables, but it can easily lead to erroneous interpretations. For instance, two non-causally related variables can be significantly correlated, due to strong autocorrelation and common driver effects (McGraw and Barnes, 2018; Runge et al., 2014). In lead-lag regression analyses, often the causal direction is assumed to be from the variable that leads to the variable that lags. However, in complex dynamical systems there is no solid basis for such assumptions. For example, linear regressions alone would suggest that surface temperatures over North and South America lead ENSO variability while the opposite causal relationship is generally accepted (McGraw and Barnes, 2018). When controlling for the autocorrelation of ENSO, this spurious correlation vanishes (McGraw and Barnes, 2018). However, due to the numerous (possible) linkages in the climate system, it is usually not obvious for which variables to control for when studying the dependence of two processes.*
To overcome these issues, causal discovery algorithms such as Causal Effect Networks (CEN) have been recently developed and subsequently applied to gain insights into the physical links of the climate system (Kretschmer et al., 2016, 2018; Runge et al., 2015a). For a given set of time-series, CEN reconstructs the likely underlying causal structure, by iteratively testing for conditional independent relationships among the input time-series. CEN have been applied to the study of stratospheric polar vortex variability (Kretschmer et al., 2016, 2018), multi-decadal North Atlantic overturning

circulation (Schleussner et al., 2014) and to study the causal interactions between the ISM, the Silk Road Pattern and the monsoon-desert mechanism (Stephan et al., 2019). Although shown to be a useful statistical tool to study teleconnection pathways, a successful application of CEN requires (such as for any data-driven method), expert knowledge of the underlying physical processes, including relevant variables, time-scales and temporal resolution (Runge, 2018). ". Further, following the reviewer's suggestion, we have added the simple correlation between the most important links for comparison, see SI, Table S1 and a relative comment in lines 559-564: "*The path coefficients for all the identified causal links are reported together with the Pearson correlations of each pair of variables (see SI, Table S1). In general the values of the causal effect and the simple correlation do not differ greatly, despite a few exceptions where linear correlation is not significant even though there exists a causal link, or when the sign of the causal effect and the simple correlation differ. However, one should not forget that correlation does not give a sign for causality, nor can identify chains of causality between different variables.*".

2. Which is the time-decay of the normalized causal effect among the different links?
The authors only mention the results for 1-week lag but I think it's interesting to comment on how the intensity of the connections decays (or does not) with time. -
Answer – We thank the reviewer for this suggestion. We have calculated the causal effect also for lags -2 and -3 and we show the results in figure S11 in the Supplementary material. We have inserted a corresponding comment to these results in the main text, see lines 494-496: "*We checked how these links decay over time and found that the causal links from W1 to the CGTI and from the CGTI to the MT rainfall remain stable at lag -2 weeks, but then decay to zero at lag -3 weeks. The causal effect from W1 on the MT rainfall and from MJO2 to W1 drop to values close to zero at lag -2 and -3 (see SI, Fig. S11).*"

3. How do the results change when seasonal or monthly time-scales are selected?
As mentioned, the original DW2005 hypothesis was originally defined for the seasonal/monthly time-scales but this time-scales are avoided in the current manuscript. Why? How do the results change?
Answer – Using seasonal (JJAS averaged) values, though in principle feasible, would not give a insightful information since in this case lag -1 (season) would go in the previous spring and thus not capture the distinct relationship that exists in boreal summer. Moreover, DW2005 define the circumglobal teleconnection pattern based on summer and analyse link within the summer season. While the correlation map between Z200 and the CGTI region at monthly time scale remains very similar to that for weekly time scale, the correlation map between MT rain and Z200 at lag -1 month for monthly data is not qualitatively similar to the circumglobal teleconnection pattern anymore (see Fig. S11 in the SI). One reason for this behaviour could be that the characteristic time scale of detected causal links is biweekly (as shown also in fig. S11, see earlier comment) and thus the relationship disappears at monthly timescale. Moreover, bi-weekly time scale is also a typical time scale for quasi-stationary waves in boreal summer (Kornhuber et al. 2016). We have clarified these points as follow in the main text in lines 497-504: "*This result is further visible if the timescale is changed from weekly to monthly: in this case, the spatial correlation between the circumglobal teleconnection pattern and the correlation map between MT rainfall and Z200 at lag 0 is still high (r = 0.79). However, when the correlation is calculated between MT rainfall and Z200 at lag -1 (month), the map becomes unsignificant (see SI, Fig. S12). This result further suggests that these causal relationships have a characteristic timescale shorter than one month and of about 1-2 weeks. Moreover, OLR shows no significant correlation patterns with the MT rainfall at lag -1 (month) and no significant causal links are detected between MT rainfall and the CGTI at monthly timescale (not shown).*"

4. The DW2005 is the base for this study, however it is only briefly discussed. I suggest to include a whole paragraph of the Introduction to discuss their findings more deeply. Also, a deeper discussion of DW2007 is missed. Those 2 studies are mentioned but together with other studies and, thus, their relevance and main results are difficult to identify.
Answer – We thank the reviewer for highlighting that these two references need to be described in more details in the introduction section. We now dedicate to each of them a full paragraph, see lines 90-104 "*Next to tropical drivers, mid-latitude circulation, the North Atlantic variability and mid-latitude wave trains have been proposed to modulate the occurrence of active and break phases of*

ISM (Ding and Wang, 2005, 2007; Kripalani et al., 1997). A circumglobal teleconnection pattern, characterized by a wave number 5 has been associated with seasonal and monthly rainfall and surface air temperature anomalies across the northern hemisphere in summer (Ding and Wang, 2005). One way to identify this circumglobal teleconnection pattern is via point-correlation maps of the 200hPa geopotential height with a location directly east of the Caspian Sea. An two-way interaction between ISM circulation system and the circumglobal teleconnection pattern is hypothesized: while the diabatic heat sources associated with ISM convection can reinforce the circumglobal teleconnection pattern propagating downstream (i.e. moving from west to east following the mid-latitude westerlies), the circumglobal teleconnection pattern itself may modulate the ISM rainfall, with enhanced rainfall associated with the positive phase of the circumglobal teleconnection pattern (Ding and Wang, 2005). The circumglobal teleconnection pattern also shows interdecadal variations, with a general weakening of its major centres of action over the last three decades, which has been attributed to weakening of the ISM precipitation and to the relation of the ISM with the El Nino-Southern Oscillation (ENSO) (Wang et al., 2012). Seasonal forecast models tend to have difficulties reproducing this pattern correctly: the circumglobal teleconnection pattern is too weak in models and one of the possible causes could be a too weak interaction between the north-western India rainfall (Beverley et al., 2019)." and lines 121-131 "*At intraseasonal timescales, a wave train originating from the north-eastern Atlantic and propagating with an arch-shaped trajectory into central Asia may influence the intraseasonal variability of the ISM and modulate the intensity of the northern ISM rainfall at a bi-weekly timescale, linking the latter with mid-latitude circulation features (Ding and Wang, 2007; Krishnan et al., 2009). An important feature of this wave train is the 200 hPa central Asian High, located to the east of the Caspian Sea, i.e. over the same region used to define the circumglobal teleconnection pattern, which may trigger positive rainfall anomalies over the northern ISM region by modifying the easterly vertical shear that drives the ISM circulation and its related effect on moist dynamic instability in the ISM region. Thus, this wave train generated in the North Atlantic might aid in modulating the alternating active and break conditions over central India (Ding and Wang, 2007; Krishnan et al., 2009; Saeed et al., 2011). A positive feedback mechanism between the northern ISM and the central Asia High is also hypothesized, with enhanced ISM precipitation acting to reinforce the positive anomaly in the central Asia High via a Rossby wave response related to the ISM heating source (Ding and Wang, 2007).*" in the main text.

5. The authors focus on mid-latitude and tropical links to the Indian Monsoon. Can the authors identify any high-latitude link with the MT precipitation?
Answer – We thank the anonymous reviewer for bringing up this topic. The proposed analysis does not highlight any influence of high--latitudes on the MT rainfall. However, this result is likely a consequence relatively short time scale of the analyses. Some studies indicate that Arctic regions can have an influence on all-India rainfall on seasonal timescales. A statement addressing this topic is available in lines 641-644: "*Though in this framework a direct influence of higher latitudes on the MT rainfall it is not identified, this may depend on the choice of the analysed (sub-seasonal) time scale. However, an influence from the Arctic on the ISM rainfall has been identified at longer (interseasonal) timescales and has shown to provide some forecast skill for seasonal all-India rainfall at 4- and 2-month lead times (Rajeevan, 2007; Di Capua et al., 2019).*"

6. Why is only the MT region selected? Which are the results when selecting, for example the maximum precipitation over western India or eastern of 87E? How about selecting the precipitation over all the Indian Monsoon region? -
Answer – We choose the MT regions as in this area defines the breaks and active phases of the ISM (see introduction, lines 48-55). However, following the reviewer's suggestion, we now provide in the SI an analogous figure as Fig. 7 in the main text but for all-India rainfall (AIR), eastern (EHF) and western Himalayan foothills (WHF) rainfall (see Fig. S20 and S21 in the SI). Results are commented in lines 558-568: "*Finally, we test for the robustness of the causal links when different regions other than the MT area are selected. In general, analysing all-India rainfall (AIR), western Himalayan foothills (WHF, defined over 26°-35°N and 70°-83°E) or eastern Himalayan foothills (EHF, defined over 20°-30°N and 87°-97°E) rainfall does not affect the strength or the sign of the causal links (see SI, Fig. S20-S21), thus showing that our results are robust and can actually represent the dynamics of the entire ISM basin. This is likely a consequence of the fact that AIR strongly reflect the behaviour of*

the MT rainfall (r = 0.57 ± 0.05), as here the most abundant rainfall takes place. While the correlation between AIR and WHF rainfall is lower (r = 0.35 ± 0.06), the same causal links as for the MT rainfall are observed, suggesting a strong similarity in the dynamical mechanisms that govern MT and WHF rainfall. However, when EHF rainfall is taken into account, the correlation with the AIR is low and not significant (r = 0.03 ± 0.05). As a consequence, the links from the CGTI and W1 to EHF disappear (though all other links are left unchanged). This suggests that the dynamical mechanisms that bring abundant rainfall to this region are different than those that determine the MT rainfall (see SI, Fig. S21).". This findings are further discussed in lines 596-604 in the discussion section "*We test the sensitivity of the monsoon-circumglobal teleconnection hypothesis to changes of the selected rainfall region, showing the general robustness of the identified causal links. Both all-India rainfall (AIR) and western Himalayan foothills (WHF) rainfall behave consistently with the MT rainfall (see SI, Fig. S20-21). These similarities are likely due to the strong correlation that exists between the MT rainfall and AIR, being the MT rainfall one of the dominant features of the ISM intraseasonal variability (Krishnan et al., 2000). However, eastern Himalayan foothills (EHF) rainfall seems to behave differently, and does not show any connection with the updraft region identified by W1. This is likely to be related to the fact that the EHF region receives heavy rainfall amounts during the early stage of a break event and thus it is likely to be governed by different circulation pattern than those that govern the MT rainfall (Vellore et al., 2014). Nevertheless, also in this case, all other causal links in the CEN are retained*.".

7. In all the network figures it would be useful to have the numbers indicating the path coefficient and auto-corr. path coefficient over the arrows and inside the circles, respectively. Absence of arrows indicates 0 path. Coefficient?
Answer – We thank the reviewer for improving the layout of the figures. We have included the numbers for the path coefficient and auto-corr. path coefficient in all figure as suggested. Moreover, we now specify in the caption of each figure that only significant links are shown.

8. Why linearity is a good framework here? Please cite works to justify this.
Answer – DW2005 used linear correlations to define and analyse the circumglobal teleconnection pattern and its interaction with ISM. Moreover, Ding et al. 2011 further show that the circumglobal teleconnection pattern has a linear behaviour ... (EXPAND). Therefore we also assume in our analyses that individual links are to first order linear. We provide an additional supportive figure that shows that temperature and precipitation anomalies for high and low CGTI states behave close to linearly. See SI Fig. S4 and the related comment in the main text in lines 339-341: "*Though these plots are obtained by plotting composites of weeks with CGTI > 1 CGTIstd minus composites of weeks with CGTI < -1 CGTIstd, we also shown that these composites behave close to linearly when plotted separately (see SI, Fig. S4). Thus, this further supports the assumption that a linear framework is a reasonable choice in this context.*"

Minor comments:

The terms "mid-latitudes" and "extra-tropics" are used indistinctly. This is problematic and in "extra-tropics" the high-latitudes are also included. Please only use 1 term to avoid confusing the reader. –
Answer - We thank the anonymous reviewer for pointing out this discrepancy, we now use the term "mid-latitudes" throughout the whole manuscript.

line 125: space missing ")algorithm" -
Answer – we thank the reviewer for carefully reading the manuscript. We have corrected this typo.

line 129: space missing ":It should" -
Answer – we thank the reviewer for carefully reading the manuscript. We have corrected this typo.

line 130: what is an actor? A variable? -
Answer – We thank the anonymous reviewer for pointing out that the definition of actor was misplaced compared to its first appearance. We have now moved the first reference to the word

"actor" in the place of its first appearance. See lines 183-185: "*A CEN detects and visualizes the causal relationships among a set of univariate time series of variables (here referred to as actors, (Kretschmer et al., 2016).*"

line 133: what is near-linear? Define. -
Answer – We have clarified the sentence, see lines 189-191: "*Other important assumptions are stationarity of time series and near-linear interactions between actors, i.e. the selected actors should have a linear behaviour at least in a first order approximation*."

line 134: PC algorithm means PC-MICI algorithm or other? Also the name is confusing as you used the term for Principal component before. –
Answer – We thank the anonymous reviewer for pointing out the double meaning of this abbreviation. We have specified in the sentence that PC algorithm stands for Peter and Clark algorithm as correctly pointed out by the reviewer, see lines 192: "*The PC-MCI algorithm is a two-step algorithm based on a modified version of the Peter and Clark (PC) algorithm (Runge et al., 2014; Spirtes et al., 2000).*"

line 149: which are the 2 conditions? More, generally: which are the n-conditions? –
Answer – We thank the anonymous reviewer for helping to improve the clarity of the methods section. We have improved the definition of the term condition, see lines 201-204: "*Then, partial correlations between the ith actor and each element jth in Pi0 wher $i \neq j$, are calculated, conditioning first on one condition, i.e. the first kth actor in Pi0 that has the strongest correlation with the ith actor and $i \neq j \neq k$*" and lines 151-152 "*This process leads to a reduced set of parents Pi1. In the next step, the process is repeated but conditioning on two conditions, i.e. calculating the linear regression on a set of two actors, leading to a second set of parents Pi2.*".

line 150: "parents contained in Pˆ{n}_i" is n the same as before? I guess not, change for m -
Answer – We thank the reviewer for carefully reading the manuscript. We have corrected this typo.

line 161: why is only tau=1 selected? Justify, why not look at other taus? –
Answer – We have now included a Fig. that shows how the strength of the causal links decays with time. See comment #2.

Line 176-180: why is the correction needed? –
Answer – The explanation for the corrected values is given in lines 180-182: "*All p-values are corrected using the Benjamini and Hochberg false discovery rate (FDR) correction to address the variance inflation due to serial correlations (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001).*".

Line 183: how is "circumglobal wave train" defined? –
Answer – We meant here the circumglobal teleconnection pattern, we have corrected this phrase.

Line 186: why is the NAO influence included in this subsection? The title only talks about ISM and circumglobal wave train. –
Answer – We have now changed the subtitle of this section to the following : "Causal testing of the two-way ISM-circumglobal teleconnection mechanism and the influence of NAO", see line ??.

fig 1: panel c: add in the title that it's over the MT region. Y-axis should be 10 to 10. Xaxis: it would be easier if it indicates the years. –
Answer – We thank the reviewer for his/her suggestions. We have modified the first figure accordingly.

line 197: why start on the 2rd week? Justify and only show fig 1c starting at this time. –
Answer – We have improved the explanation of why the first two time-steps need to be skipped, see line 238-240: "*Skipping the two time steps of each year is necessary since the first two time slots allow*

detecting lagged relationships, and the PC-MCI algorithm requires to add twice the maximum time lag explored (here a maximum lag of 1 week is chosen)."

Line 199: justify why max lag 1 week is selected. –

Answer – We now show also lag -2 and -3 weeks for the main causal relationships in the SI (see previous comment #2). Moreover, the choice to present results for a one 1 week lag also reflect the interest on analysing the active and break phases dynamics, which has a time period that goes from less than one up to three weeks (see introduction, lines 48-55).

line 243: typo "llink" –
Answer – we thank the reviewer for carefully reading the manuscript. We have corrected this typo.

line 248-249: why the uncertainty in r? isn't it just the pattern correlation number?
Please explain. –
Answer – The uncertainty in r refers to the confidence interval of the correlation coefficient, to show whether it is significantly different from zero at a confidence level of 0.05. We now specify this the first time that the confidence interval appears (line 314).

Line 246-255: re-arrange paragraph to talk first about EOF1 and then about EOF2.
Also, mention definition of Eurasian sector the first time it appears. –
Answer – The apparent mixing is due to the fact that we choose to describe first the global spatial correlation and then the regional one. However, within this framework, EOF1 is always treated first. We now defined the Eurasian sector in line 316 as it first appears in the text.

Line 261-266: in Precipitation there is no signal over Western Europe. –
Answer – We thank the reviewer for kindly pointing out this mistake, which we have now corrected.

line 269: Careful! How can you compare variables of different magnitude????!! you
can't say the precipitation is weaker than temperature. –
Answer – We intended to refer to the spatial extension of the precipitation anomalies, which are less spread than the temperature ones. We have now corrected the sentence as follow (lines 338-339): "*Precipitation anomalies are more spatially confined than those found for 2m-temperature. However, a clearly defined wave pattern appears over the Eurasian sector.* ".

Fig 2: CGTI has not been defined up to this point in the text. Why not show also lag +
-1? seems important later. Why show temperature anomalies when it's not the focus of
the paper? Panels e and f: subtitles are misleading, indicate it's anomalies associated
with extreme CGTI. –
Answer – We have corrected the subtitles of Fig. 2, panels (e) and (f), as suggested by the reviewer. We decided to show temperature and precipitation anomalies since we consider important to know what the effect of the circumglobal teleconnection pattern phase is on other relevant atmospheric variables for summer weather in the mid-latitude, as this can have consequences for daily life in the affected regions. Moreover, we also double-check that high CGTI values are connected with positive rainfall anomalies over India. Lag -1 weeks results for both MT rainfall and CGTI are shown in Figure 4. Positive lags are not included in the current analysis.

Line 227-241: AT this point CGTI seems important to the paper, however it is not shown.
I suggest to include its time-series. Also a justification on why such a small region is
selected is needed. –
Answer – As reported in lines 300-302, the CGTI regions is defined following DW2005. We define our CGTI in the same way, since they already show that small latitudinal variations do not affect the results (see Fig. 2 in DW2005). Following the reviewer's suggestion, we have now added a figure in the SI to show all time-series used in Fig. 7 (see SI, Fig. S19).

Line 316: OLR1 is not indicated in the figure 4b section 3.3: why add MJO in this section when it is about internal feedbacks? The sub-title is misleading –
Answer – We thank the reviewer for pointing us to this discrepancy. Section 3.3 is meant to describe both tropical and internal variability. We have thus changed the subtitle accordingly to "*Intraseasonal variability and tropical influence on the monsoon circulation*".

fig 7: why is the network overlapped on a lon-lat map? No overlapping was done before. For example, why is EOF2 located over east Asia or W1 over west India? –
Answer – This choice is made to help the reader to easily identify each actor with its corresponding geographical location. We have now included an explanation of this choice in lines 508-511: "*We include the most important identified regions from both the tropics and the mid-latitudes together in a single CEN (Fig. 7) and plot the corresponding CEN over a map to help the reader to associate each actor with its corresponding region (though in cases when the index is defined over all longitudes, such as EOF2, it is only possible to associate the actor with its average latitude).*"

Line 446-447: you only mention mid-latitudes even though later tropics and internal feedbacks are analysed. –
Answer – We have corrected the sentence accordingly to the reviewer's suggestion (lines 570-572): "*In this study, we apply causal discovery algorithms to analyse the influence of global middle and upper tropospheric fields on the ISM rainfall and study the two-way causal links between the mid-latitude circulation and ISM rainfall, together with tropical drivers and ISM internal variability.*".

Line 459: how does your results have implications for the interannual time-scales? I think this sentence is misleading. –
Answer – It was meant "seasonal", as explained in lines 584-586. We thank the reviewer for pointing out this mistake, which has been now corrected.

Line 464-465: how does this pattern compare with the regression of MT precipitation on the Z200 field? Any substantial difference? –
Answer – Following the reviewer's suggestion, we now provide the linear regression of the MT rainfall in the CGTI index, and show that the strongest relationship is found at lag -1 week, with the CGTI leading the Mt rainfall by one week: see SI Fig. S5 and the related comment in the main text in lines 342-345 "*Moreover, when regressing the MT rainfall on the CGTI index at different lags (from lag 0 to lag -2 weeks), the strongest relationship between the CGTI and the MT rainfall is found at lag -1 week, with the CGTI leading a change in MT rainfall by one week (see SI, Fig. S5). This information also further supports the choice of analysing the relationships between these two variables at lag = -1 week.*".

Line 486-488: is it possible to confirm this with your results? –
Answer – Yes, at least partly. We now discuss the relationship between the BSISO, the initiation of active and break phases and the results obtained by our analysis in more details in the text (see comments to review #1). However, we do not analyse directly the excitation of Rossby waves, and in this sense we cannot confirm this hypothesis within the scope of the present study.

Lines 502-508: why is the linear framework adequate for studying these mechanisms? Can you cite any modelling work implying linearity of these type of interactions? –
Answer – see previous answer for comment #8.

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

Giorgia Di Capua[1,2], Marlene Kretschmer[1], Reik V. Donner[1,3], Bart van den Hurk[2,4], Ramesh Vellore[5], Raghavan Krishnan[5], 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]Magdeburg-Stendal University of Applied Sciences, Magdeburg, Germany
[4]Deltares, Delft, Netherlands
[5]Indian Institute for Tropical Meteorology, Pune, India

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

Abstract.

The alternation of active and break phases in the Indian summer monsoon (ISM) rainfall at ~~sub-seasonal~~intraseasonal timescales characterizes each ISM season. ~~Tropical~~Both tropical and mid-latitude drivers influence this ~~sub-seasonal~~intraseasonal ISM variability. The circumglobal teleconnection observed in boreal summer drives ~~sub-seasonal~~intraseasonal variability across the mid-latitudes and a two-way interaction between the ISM and the circumglobal teleconnection pattern has been hypothesized. We use causal discovery algorithms to test the ISM-circumglobal teleconnection hypothesis in a causal framework. ~~Our analysis shows a~~A robust causal link from the circumglobal teleconnection pattern and the North Atlantic region to ISM rainfall~~. We~~ is

identified and we estimate the normalized causal effect (CE) of this link to be about 0.2 (a one standard deviation shift in the circumglobal teleconnection causes a 0.2 standard deviation shift in the ISM rainfall one week later). ~~In turn, the~~The ISM rainfall ~~influences~~feeds back on the circumglobal teleconnection pattern, however weakly. Moreover, we identify ~~causal links that represent the internal dynamics~~a negative feedback between strong updraft located over India and the Bay of Bengal and the ISM ~~convective cell~~rainfall acting at ~~weekly timescales: Periods~~a biweekly timescale, with ~~strong updraft lead to~~ enhanced ISM rainfall following strong ~~rainfall~~updraft by one week ~~later, but the resulting increase in static stability suppresses convection again. In our analyses, this internal ISM dynamics~~. This mechanism is possibly related to the Boreal Summer Intraseasonal Oscillation. The updraft has the strongest CE of 0.5~~. Tropical~~, while the Madden-Julian Oscillation variability has a CE of 0.2-0.3. Our results show that ~~the~~ most of the ISM variability on weekly timescales ~~is due to internal dynamics of convective cell, but both~~ comes from these tropical ~~and~~ drivers, though the mid-latitude ~~teleconnections have~~teleconnection also exerts a substantial influence. Identifying these local and remote drivers paves the way for improved ~~sub-seasonal~~subseasonal forecasts.

1 Introduction

The Indian summer monsoon (ISM) is crucial for the Indian society, which receives 75% of its total annual rainfall during the summer months June through September (JJAS). The ISM rainfall variability at ~~sub-seasonal time-scales~~intraseasonal timescales is characterized by periods of enhanced and reduced rainfall activity over the monsoon-core region of central India. These periods are usually referred to as active (wet) and break (dry) phases, respectively, and have a duration that spans from a few days up to three weeks. Prolonged active and break spells in the ISM can lead to floods or droughts and consequently have severe socio-economic implications for the Indian subcontinent. A salient semi-permanent feature of the ISM is the “monsoon trough” (MT) which manifests as a low pressure zone extending from north-western India into the Gangetic plains and the Bay of Bengal ~~(Rao, 1976; Krishnamurti and Sugi, 1987; Choudhury and Krishnan, 2011).~~(Rao, 1976; Krishnamurti and Sugi, 1987; Choudhury and Krishnan, 2011). The rainfall amount over the MT region is generally used to define dry and wet spells within the ISM season (Krishnan et al., 2000; Gadgil and Joseph, 2003). The position and strength of MT significantly influences the spatial distribution of monsoon precipitation and associated agricultural productivity on the Indian subcontinent, and the internal dynamics of the ISM circulation itself provides a first mechanism for ~~sub-seasonal~~intraseasonal rainfall variability (Pathak et al., 2017).

The land-sea temperature difference and the mid-tropospheric thermal forcing over the Tibetan Plateau are the prime drivers for the monsoon circulation (Yanai and Wu, 2006). Ascending motions over the Indian subcontinent enhance the northward air flux from the ocean toward the land thereby bringing moisture from the ocean and fuelling rainfall. The latent heat released by strong convective rainfall is important for sustaining the ISM circulation ~~(Levermann et al., 2009)~~(Levermann et al., 2009). However, it has two opposing effects: on the one hand, the latent heat release in the early stage of an active phase enhances ascending motions by heating the mid-to-lower troposphere ~~(Levermann et al., 2009)~~(Levermann et al., 2009). On the other hand, latent heat

release, that propagates upward and heats the upper layers of the air column, tends to increase the static stability and inhibits further ascending motions ~~(Saha et al., 2012).~~(Saha et al., 2012). Also, rainy weather and cloudy skies can have a cooling effect on the surface ~~in support of suppressing~~, which tends to suppress convection (Krishnamurti and Bhalme, 1976). While this thermodynamic perspective is useful to understand the quasi-biweekly variations of the ISM elements locally, the spatio-temporal variations in the evolution of active and break phases over the Indian monsoon region are known to involve interactions between the wind anomalies and the northward propagation of the major rain band anomalies of the ~~monsoon intraseasonal oscillation~~Boreal Summer Intraseasonal Oscillation (BSISO) ~~(Chattopadhyay et al., 2009; Shige et al., 2017; Wang et al., 2006).~~(Wang et al., 2006; Chattopadhyay et al., 2009; Shige et al., 2017).

The Boreal Summer Intraseasonal Oscillation represents a characteristic feature of the atmospheric circulation over the northern Indian Ocean and South Asia (Goswami and Ajaya Mohan, 2001; Saha et al., 2012; Suhas et al., 2012). The BSISO is identified as a rainfall band that stretches from the Indian sub-continent to the Maritime Continent, featuring a north-west/south-east tilt and propagating north-eastward from the equatorial Indian Ocean towards South East Asia with a timescale of about one month (Wang et al., 2006; Webster and Lawrence, 2002). The BSISO can influence the oscillation between break and active phases typical for the ISM intraseasonal variability, with favorable conditions for the initiation of an active phase when the BSISO rainfall band reaches the Indian sub-continent during its northeastward propagation (Goswami and Ajaya Mohan, 2001; Suhas et al., 2012). Krishnan et al. (2000) hypothesized that the triggering of Rossby waves by suppressed convection over the Bay of Bengal initiates ISM breaks through northwest propagation of high pressure anomalies from the central Bay of Bengal into northwest India. They noted that the initiation of suppressed convection and anticyclonic anomalies over the equatorial Indian Ocean and central Bay of Bengal occurred a week prior to the commencement of a monsoon break over India followed by the traversing of suppressed anomalies from the central Bay of Bengal to northwest India in about 2-3 days.

At ~~sub-seasonal~~intraseasonal timescales, ~~both tropical and mid-latitude drivers have been proposed to influence the active and break phases of the ISM rainfall. For example,~~ the Madden-Julian Oscillation (MJO~~)~~), which governs the ~~sub-seasonal~~intraseasonal tropical climate variability operating at 30-90 day timescale, represents an important tropical driver of the ISM ~~sub-seasonal~~intraseasonal variability (Zhang, 2005)~~. MJO consists of a transient region of strong convective motions and enhanced precipitation, which propagates eastward from the tropical Indian Ocean to the tropical western Pacific. Normally, only one MJO convective cell is present in these regions. The MJO influences the ISM system with enhanced convective rainfall activity during active MJO phases and negative rainfall anomalies during suppressed MJO phases (Anandh et al., 2018; Mishra et al., 2017; Pai et al., 2011).~~

. The MJO consists of a transient region of strong convective motions and enhanced precipitation, which propagates eastward from the tropical Indian Ocean to the tropical western Pacific. Normally, only one area of strong convective motions related to the MJO is present in these regions. The MJO influences the ISM system with enhanced convective rainfall activity during active MJO phases and negative rainfall anomalies during suppressed MJO phases (Anandh et al., 2018; Mishra et al., 2017; Pai et al., 2011). However, during boreal summer, the MJO strength is reduced as compared to boreal winter, and both the MJO and the BSISO propagation and phases can be well described by the outgoing longwave radiation MJO index (OMI) (Wang et al., 2018).

~~N~~ext to tropical drivers, mid-latitude circulation, the North Atlantic variability and mid-latitude wave trains have been proposed to modulate the occurrence of active and break phases of ~~ISM (Ding and Wang, 2005, 2007; Kripalani et al., 1997). A circumglobal teleconnection pattern, characterized by a wave number 5 that encircles the northern hemisphere, has been associated with anomalous monthly rainfall and surface air temperature across the entire mid-latitude range~~the ISM (Kripalani et al., 1997; Ding and Wang, 2005, 2007). A circumglobal teleconnection pattern, characterized by a wave number 5 has been associated with seasonal and monthly rainfall and surface air temperature anomalies across the Northern Hemisphere in summer (Ding and Wang, 2005). ~~On the one hand, wave trains originating from the North Atlantic may influence the sub-seasonal variability of the ISM and modulate the intensity of the ISM rainfall (Ding and Wang, 2007; Krishnan et al., 2009). On the other hand, the diabatic heat sources in association with ISM convection might reinforce the circumglobal wave train propagating downstream~~One way to identify this circumglobal teleconnection pattern is via point-correlation maps of the 200 hPa geopotential height with a location directly east of the Caspian Sea. A two-way interaction between ISM circulation system and the circumglobal teleconnection pattern has been hypothesized: while the diabatic heat sources associated with ISM convection can reinforce the circumglobal teleconnection pattern propagating downstream (i.e. moving from west to east following the mid-latitude westerlies), the circumglobal teleconnection pattern itself may modulate the ISM rainfall, with enhanced rainfall associated with the positive phase of the circumglobal teleconnection pattern (Ding and Wang, 2005). ~~Another recurrent coupled pattern of sub-seasonal variability between mid-latitude circulation and the ISM is the wave train originating from the north-eastern North Atlantic and propagating with an arch-shape trajectory through the western Siberian Plains into central Asia~~The circumglobal teleconnection pattern also shows interdecadal variations, with a general weakening of its major centres of action over the last three decades, which has been attributed to weakening of the ISM precipitation and to the relation of the ISM with the El Niño-Southern Oscillation (ENSO) (Wang et al., 2012). Seasonal forecast models, e.g. simulations based on seasonal forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF), tend to have difficulties reproducing this pattern correctly: the circumglobal teleconnection pattern is too weak in models and one of the possible causes could be a too weak interaction with the north-western India rainfall (Beverley et al., 2019).

The latent heat released via strong convection in the ISM region has also been shown to influence regions which are located upstream (i.e. eastward). The so-called monsoon-desert mechanism involves long Rossby waves to the west of the ISM region generated by ISM latent heating. These waves enhance the downward flow over the eastern Mediterranean and north-eastern Sahara Desert suppressing precipitation in these dry regions (Rodwell and Hoskins, 1996). A subset of CMIP5 models is able to capture this mechanism (Cherchi et al., 2014), and CMIP5 scenarios for the 21st century project increased ISM precipitation, despite a decrease in the strength of the ISM circulation. Thus, the monsoon-desert mechanism could contribute to drying and warming trends projected for the eastern Mediterranean region, exacerbating the desiccation conditions in these regions (Cherchi et al., 2016).

Variability of sea surface temperatures (SSTs) and mid-tropospheric variables in the North Atlantic region has been shown to influence the ISM at a wide range of timescales. At inter-decadal and interannual timescales, SSTs related to the Atlantic Multi-decadal Oscillation (AMO) index have been shown to modulate the strength of the ISM by an atmospheric bridge involving the North Atlantic Oscillation (NAO) index: positive (negative) NAO phases alter westerly winds and associated storm tracks in the North Atlantic/European area, modify

tropospheric temperatures over Eurasia and thus enhance (weaken) the strength of the ISM rainfall (Goswami et al., 2006). At intraseasonal timescales, a wave train originating from the north-eastern Atlantic and propagating along an arch-shaped trajectory into central Asia may influence the intraseasonal variability of the ISM and modulate the intensity of the northern ISM rainfall at a bi-weekly timescale, linking the latter with mid-latitude circulation features (Ding and Wang, 2007; Krishnan et al., 2009). An important feature of this wave train is the 200 hPa Central Asian High, located to the east of the Caspian Sea, i.e. over the same region used to define the circumglobal teleconnection pattern, which may trigger positive rainfall anomalies over the northern ISM region by modifying the easterly vertical shear that drives the ISM circulation and its related effect on moist dynamic instability in the ISM region. Thus, this wave train generated in the North Atlantic might (Ding and Wang, 2007). ~~An important feature of this wave train is the 200 hPa central Asian High which may be able to trigger positive rainfall anomalies over the ISM region. This wave train generated in the North Atlantic region might also~~ aid in modulating the alternating active and break conditions over central India (Ding and Wang, 2007; Krishnan et al., 2009; Saeed et al., 2011). A positive feedback mechanism between the northern ISM and the Central Asian High has also been hypothesized, with enhanced ISM precipitation acting to reinforce the positive anomaly in the Central Asian High via a Rossby wave response related to the ISM heating source (Ding and Wang, 2007).

~~Plain~~

Lagged correlation and ~~composite~~regression analysis ~~are~~have been commonly used to assess the relationship between two or more climate variables (Ding and Wang, 2005; Lau and Kim, 2011; Vellore et al., 2014). ~~However, correlation alone cannot distinguish whether the detected relationships represent actual causal connections or are only spurious links, due to autocorrelation, indirect links or common drivers. Recently, causal discovery algorithms have been~~Such an analysis is useful as it gives a first information on the association of two or more variables, but it can easily lead to erroneous interpretations. For instance, two non-causally related variables can appear significantly correlated, due to strong autocorrelation and/or common driver effects (McGraw and Barnes, 2018; Runge et al., 2014). In lead-lag regression analyses, the causal direction is assumed to be from the variable that leads to the variable that lags. However, in complex dynamical systems there is often no solid basis for such assumptions. For example, linear regressions alone would suggest that surface temperatures over North and South America lead ENSO variability while the opposite causal relationship is generally accepted (McGraw and Barnes, 2018). When controlling for the autocorrelation of ENSO, this spurious correlation vanishes (McGraw and Barnes, 2018). However, due to the numerous (possible) linkages in the climate system, it is usually not obvious which variables to control for when studying the dependence of two processes.

~~T~~o overcome these issues, causal discovery algorithms such as Causal Effect Networks (CEN) have been recently developed and subsequently applied to gain insights into the physical links of the climate system ~~(Kretschmer et al., 2016; Runge et al., 2015a; Schleussner et al., 2014). One can use such tools to test whether hypothesized links or teleconnections are likely to represent true physical pathways or rather artefacts due to spurious correlations. A careful analysis in the context of the ISM thus starts from physical theory and, hence, requires domain knowledge of ISM dynamics. Using this so-called "theory-guided causal discovery analysis", we here study the two-way causal links connecting both tropical and mid-latitude regions with the ISM.~~(Kretschmer et al., 2016, 2018; Runge et al., 2015, 2019). For a given set of time series, a CEN reconstructs the likely underlying causal structure, by iteratively testing for conditional independent relationships among the input time-series. CEN have been applied

to the study of stratospheric polar vortex variability (Kretschmer et al., 2016, 2018), multi-decadal North Atlantic overturning circulation (Schleussner et al., 2014), the intraseasonal stratosphere-troposphere coupling in the Southern Hemisphere (Saggioro and Shepherd, 2019) and to study the causal interactions between the ISM, the Silk Road Pattern and the monsoon-desert mechanism (Stephan et al., 2019). Although shown to be a useful statistical tool to study teleconnection pathways, a successful application of CEN requires (such as any data-driven method) expert knowledge of the underlying physical processes, including relevant variables, time-scales and temporal resolution (Runge, 2018).

Here we study the two-way interactions of tropical and mid-latitude regions with the ISM by applying such a theory-guided causal effect analysis. First, we assess whether the connection between the ISM and the mid-latitude wave trains can be considered causal in one or both directions. Next, we quantify the relative causal effect of tropical, ~~extra-tropical and internal drivers on the ISM.~~ mid-latitude and internal drivers on the ISM. The remainder of this paper is organized as follow: Section 2 describes the data and methods that have been applied. Section 3 presents the results obtained by applying causal discovery tools to the analysis of the ISM mid-latitude and tropical drivers. In section 4, these results are discussed in the context of the existing literature and our conclusions are presented.

2 Data and Methods

2.1 Data

We define the monsoon trough (MT) region as the region between 18°-25°N and 75°-88°E. We analyse weekly rainfall ~~sums~~averages over ~~the MT~~these region from the CPC-NCEP (0.25°x0.25°) observational gridded global rainfall dataset over the period 1979-2016 (Chen et al., 2008) and from the Pai et al. (2015) (0.25°x0.25°) observational gridded Indian rainfall dataset over the period 1979-2017 (Pai et al., 2015). In the remainder of this paper, we will mainly focus on the results obtained for the latter data set, while those for the former are provided as parts of the Supplementary Material. Using data taken from the ERA Interim reanalysis (Dee et al., 2011) for the period 1979-2017, precursor regions are calculated from global weekly averaged gridded (1.5°x1.5°) fields including outgoing longwave radiation (OLR) at the top of the atmosphere ~~(OLR),~~, vertical velocity at 500 hPa (~~W~~W500) and geopotential height at 200 hPa (Z200). The NAO weekly index is obtained by averaging daily data from NOAA (available at ftp://ftp.cpc.ncep.noaa.gov/cwlinks/norm.daily.nao.index.b500101.current.ascii). To identify MJO phases, we use the OLR MJO Index (OMI) provided by NOAA (https://www.esrl.noaa.gov/psd/mjo/mjoindex/). This metric features the first and second principal components obtained by the empirical orthogonal function (EOF) analysis of OLR in the tropical belt (between 30°N and 30°S) filtered to remove influences outside the MJO ~~time scale~~timescale (30-90 days). OMI PC2 corresponds to the first principal component of the Real-Time Multivariate MJO index (RMM1), which is widely used in the literature ~~(Kiladis et al., 2014; Pai et al., 2011; Wheeler and Hendon, 2004).~~ (Wheeler and Hendon, 2004; Pai et al., 2011; Kiladis et al., 2014). Moreover, the OMI index has also been proven useful in describing the BSISO behaviour, which is relevant for the ISM break and active phases in summer (Wang et al., 2018). All time series of MT rainfall, Z200 and all datasets analysed in this work are detrended and anomalies are calculated at weekly time-steps. Thus, both the climatological and seasonal cycle are removed. Since the interannual variability may affect the analysis,

we follow the approach proposed by Ding and Wang (2007) and filter the data by removing from each JJAS season its seasonal average.

2.2 Causal effect networks

~~W~~e apply Causal Effect Networks (CEN) and the Response-Guided Causal ~~Precursors~~Precursor Detection (RG-CPD) tool, two recently developed applications of the so-called Peter and Clark – Momentary Conditional Independence (PC-MCI)~~algorithm (Runge et al., 2014; Spirtes et al., 2000).~~ algorithm (Spirtes et al., 2000; Runge et al., 2014, 2019). A CEN detects and visualizes the causal relationships among a set of univariate time series of variables ~~(Kretschmer et al., 2016)~~(here referred to as *actors*, Kretschmer et al., 2016). The network is constructed using the PC-MCI algorithm, which is a causal discovery algorithm able to distinguish between spurious and true causal links for different time lags of interest (Runge, 2018). The term "causal" rests on several assumptions (Spirtes et al., 2000; Runge, 2018):~~It~~ Here, it should be understood as *causal relative to the set of analysed precursors*, meaning that the identified causal links are valid ~~in~~relative to the selected set of actors. Adding additional actors may change the structure of the causal network. It is therefore crucial to combine CEN with theoretical domain knowledge (i.e. our "theory-guided causal discovery analysis" approach). Other important ~~assumption~~assumptions are stationarity of time series and near-linear interactions between actors, i.e., the selected actors should have a linear behaviour at least in a first order approximation.

~~The PC-MCI algorithm is a two-step algorithm based on a modified version of the PC algorithm (Runge et al., 2014; Spirtes et al., 2000). In the first step, the PC-step, identifies the relevant conditions for each variable by iterative independence testing.~~The PC-MCI algorithm is a two-step algorithm based on a modified version of the Peter and Clark (PC) algorithm (Spirtes et al., 2000; Runge et al., 2014, 2019). In the first step (the PC-step), the relevant conditions for each variable are identified by iterative independence testing. The following MCI-step tests whether the link between two actors can be considered causal. The false discovery rate (FDR) approach, as described by Benjamini and Hochberg, is applied to correct for inflated *p*-values due to multiple significance testing (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001). Each step is further described below.

In a variable set P containing n univariate detrended anomaly time series (the actors), the PC-step identifies the *causal parents* of each i^{th} element in P, among all the remaining elements in P. First, the PC algorithm calculates plain correlations between each i^{th} actor in P and all the remaining elements at a certain time lag τ. Those actors that significantly correlate with the i^{th} actor form the set of its ~~initial~~potential parents P_i^0 at lag τ. Then, partial correlations between the i^{th} actor and each element j^{th} in P_i^0 (where $i \neq j$), are calculated, conditioning first on one condition, i.e., the first k^{th} actor in P_i^0 that has the strongest correlation with the i^{th} actor and $i \neq j \neq k$. If x, y and z are elements in P, the partial correlation between x and y conditioned on z is calculated by first performing linear regressions of x on z and of y on z and then calculating the correlation between the residuals. If the resulting partial correlation between x and y is still significant at a certain significance threshold α, x and y are said to be *conditionally dependent* given variable z, i.e., the correlation between x and y cannot be (exclusively) explained by the influence of variable z. When the opposite happens, the link is spurious and therefore filtered out, and x and y are called *conditionally independent*. This process leads to a reduced set of parents P_i^1. In the next step, the process is repeated but conditioning on two conditions, i.e. calculating the linear regression on a set of two actors, leading to a second set of parents P_i^2. The algorithm converges when the number of causal parents contained in ~~P_i^n~~ P_i^m is equal or greater than the number of conditions needed to calculate the next partial correlation. At the end of the PC-step, each element in P has its own set of parents, which then enters the MCI-step. In the MCI-step, the

Formatted: Font: (Default) Times New Roman, 10 pt, Italic, English (United Kingdom)

Formatted: Font: (Default) Times New Roman, 10 pt, Not Italic, English (United Kingdom)

partial correlation between an actor and its initial set of potential parents is calculated again but conditioned simultaneously on both the set of parents of the i^{th} actor and the sets of parents of each of the parents of the i^{th} actor. Those parents that pass the MCI test will then form the final set of parents for the i^{th} actor (Runge et al., 2017). A numerical example of these steps is given in the SI. , together with the values for each parameter used for PC-MCI. For this study, the python package TIGRAMITE version 3.0 is used (https://github.com/jakobrunge/tigramite_old).

The causal links detected via the PC-MCI algorithm are visualized in terms of a causal effect network (CEN). Each CEN is composed by circles, representing the various actors, and by arrows, with the colour indicating the strength and the arrow the direction of the detected causal links. The strength is expressed by its associated path coefficient, defined as following Runge et al (2015a. (2015), as the "expected change in X^j (in units of its standard deviation [s.d] and relative to the unperturbed regime) at time t if X^i was perturbed at time $t-\tau$ by a one s.d. delta peak". To give an example, a path coefficient of 0.5 means that a change in the causal parent of 1 s.d. corresponds to a change in 0.5 s.d. in the response variableanalysed actor. Here, due to the fact that only lags at $\tau=1$ are accounted for, the path coefficients also correspond to the total causal effect (CE). The path coefficient of each variable on itself is here referred to as the autocorrelation path coefficient. This should not be confused with the usual definition of the lag-1 autocorrelation coefficient. The autocorrelation path coefficient is the same as the path coefficient, where with $i = j$, and it represents the causal influence of an actor on itself.

2.3 Response-guided causal precursor detection

RG-CPD identifies the causal precursors of a response variable based on multivariate gridded observational data (Kretschmer et al., 2017).(Kretschmer et al., 2017; Di Capua et al., 2019). It combines a response-guided detection step (Bello et al., 2015)(Bello et al., 2015) with the PC-MCI causal discovery step (Runge et al., 2014; 2015a; 2015b; Spirtes et al., 2000).(Spirtes et al., 2000; Runge et al., 2014, 2015, 2019). Without requiring an a priori definition of the possible precursors, RG-CPD searches for spatially contiguous regions in multivariate gridded data that are significantly correlated precursor regions of with a variable of interest (i.e. , the *response variable*) in multivariate gridded data at a given lag and then detects causal precursors by filtering out spurious links due to common drivers, autocorrelation effects, or indirect links. Using correlation maps, an initial set of precursors is identified in relevant meteorological fields by finding regions in which the recorded variability precedes changes in the response variable at some lead time. Adjacent grid points with a significant correlation of the same sign at a level of $\alpha=0.05$ are spatially averaged to create single one-dimensional time series, called *precursor region* (Willink et al., 2017). Correlation values are calculated with a two-sided p-value for a hypothesis test whose null hypothesis is that there is no correlation, using the Wald Test with a t-distribution of the test statistic. All p-values are corrected using the Benjamini and Hochberg false discovery rate (FDR correction) approach to address the variance inflation due to serial correlations (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001). Then, adjacent grid points with a significant correlation of the same sign at a level of $\alpha=0.05$ are spatially averaged to create single one-dimensional time series, characterising the dynamics of the considered field in the so-called *precursor region* (Wilks, 2016; Willink et al., 2017). In the second step, PC-MCI identifies the set of *causal*

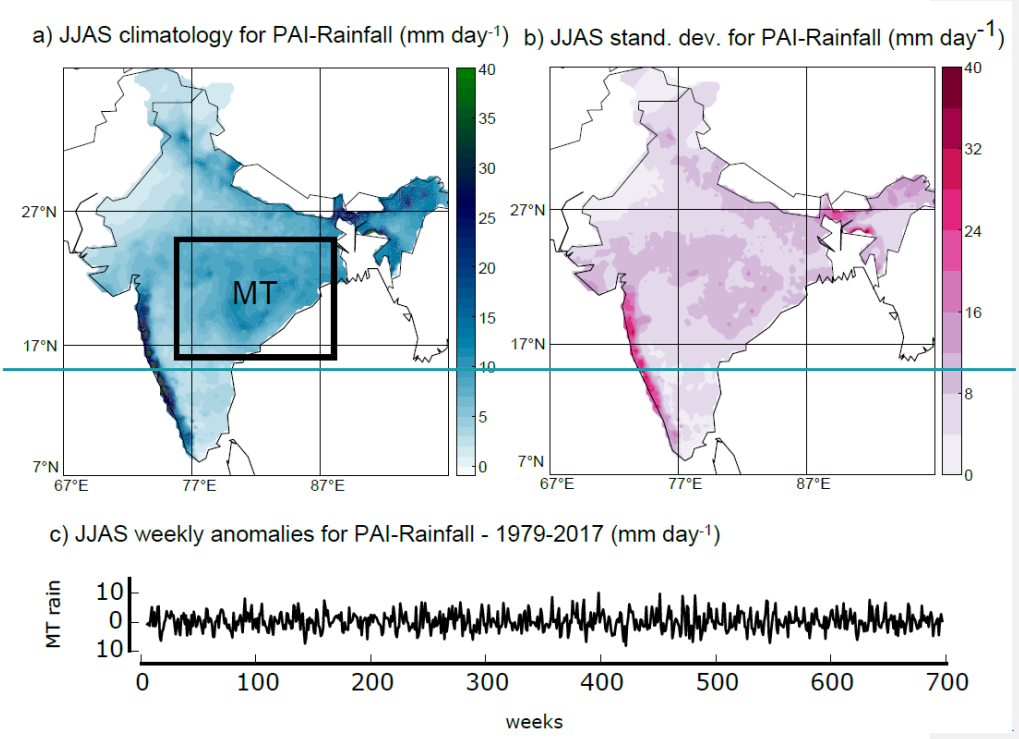
precursors for the response variable. The results present those precursor regions on a global map that are identified to be causally linked with the response variable (here, the MT rainfall).

3 Results

3.1 Causal testing of the two-way ISM-circumglobal teleconnection mechanism and the influence of NAO

First, we assess whether the two-way interaction between the circumglobal ~~wave train~~teleconnection pattern (that characterizes the boreal summer circulation) and the MT rainfall, as hypothesized by Ding & Wang (2005, hereafter DW2005), is reproduced using our CEN analysis. We will refer to this theory as the monsoon-circumglobal teleconnection mechanism. We also analyse the influence of the North Atlantic on the MT rainfall, as hypothesized by Ding & Wang (2007, hereafter DW2007).

~~F~~igures 1a,b show the JJAS climatology and the s.d. of weekly ISM rainfall from the Pai et al.



~~(~~2015)**Figure 1. Rainfall climatology over India.** Panel (a): JJAS rainfall climatology over the 1979-2017 period from the Pai et al. dataset. The black box identifies the MT region. Panel (b): standard deviation (s.d.) of weekly JJAS rainfall over the

1979-2017 period from the Pai et al. dataset. Panel (c): time series of weekly MT JJAS rainfall over the period 1979-2017; each year contains 18 weeks, with the first week starting on the 27th of May.

dataset for the period 1979-2017. We average the rainfall over the MT region and identify a univariate time series, which represents the weekly variation of the ISM during JJAS over the MT region (Fig. 1c). This time series contains 18 weeks for each of the 39 analysed years, each year starting on the 27th of May. The analysis of the MT rainfall starts on the 3rd week (10th of June) for a total of 16 7-day time slots per year. Skipping the two first time steps (weeks) of each monsoon season is necessary since they allow detecting lagged relationships, and the PC-MCI algorithm requires to add twice the maximum time lag explored (here a maximum lag of 1 week is chosen). Information from the previous year does not interfere with the following year. The weekly timescale is considered to represent the relevant interaction between ISM rainfall and Northern Hemisphere atmospheric circulation at intraseasonal timescales (Krishnamurti and Bhalme, 1976; Ding and Wang, 2007; Vellore et al., 2014). Moreover, a weekly timescale is also adequate to represent the propagation of different BSISO phases and the switch between active and break phases of the ISM (Goswami and Ajaya Mohan, 2001).

~~F~~igures 1a,b show the JJAS climatology and the s.d. of weekly ISM rainfall from the Pai et al. dataset for the period 1979-2017. We average the rainfall over the MT region and identify a univariate time series, which represents the weekly variation of the ISM during JJAS over the MT region (Fig. 1c). This time series contains 18 weeks for each of the 39 analysed years, each year starting on the 27th of May. The analysis of the MT rainfall starts on the 3rd week (10th of June) for a total of 16 7-day time slots per year. ~~The first two time slots allow detecting lagged relationships, and the PC-MCI algorithm requires to add twice the maximum time lag explored (here a maximum lag of 1 week is chosen). Thus, information from the previous year does not interfere with the following year. The weekly time scale is considered to represent the relevant interaction between ISM rainfall and Northern Hemisphere atmospheric circulation at a sub-seasonal timescale (Ding and Wang, 2007; Krishnamurti and Bhalme, 1976; Vellore et al., 2014).~~

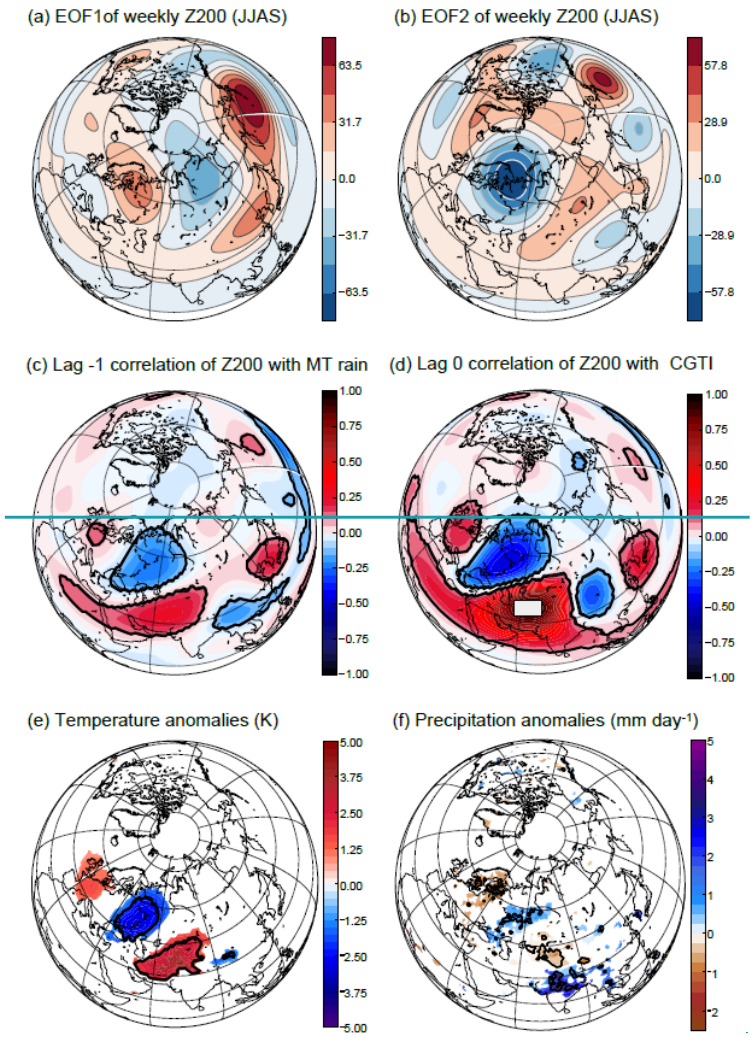


Figure 2. Mid-latitude variability associated with the ISM. Panels (a) and (b): EOF1 and EOF2 for the JJAS weekly Z200 field in the northern mid-latitudes for the period 1979-2017. Panel (c): correlation between weekly MT rainfall and Z200 (lag = -1 week). Panel (d): the CGTI region (white box) and the correlation between CGTI and Z200 (lag = 0), which forms the circumglobal teleconnection pattern. In panels (c) and (d), correlation coefficients and anomalies with a p-value of $p < 0.05$ (accounting for the effect of serial correlations) are shown by black contours. Panel (e): mean temperature anomalies over the Northern Hemisphere during weeks with CGTI > 1 $CGTI_{std}$ minus weeks with CGTI < -1 $CGTI_{std}$. Panel (f): as panel (e) but for rainfall anomalies from the CPC-NCEP dataset (as rainfall data for the Pai dataset are available over India only) for the period 1979-2017. In panels (e) and (f), anomalies with a p-value of $p < 0.05$ (accounting for the effect of serial correlations) are shown by black contours, while grid points which are found significant only with non-corrected p-values are shaded.

To analyse the major mode of Z200 variability in the Northern Hemisphere mid-latitudes, following DW2005, we calculate the first and second empirical orthogonal functions (EOFs) of weekly averaged Z200 fields over the

Northern Hemisphere (0°-90°N, 0°-360°E) for the summer months coinciding with the ISM season (June to September, weeks 21 to 38). Figures 2a,b show ~~the~~these first and second EOFs. EOF1 represents the dominant component of ~~sub-seasonal~~intraseasonal variability of the Z200 field, and qualitatively resembles the s.d. of Z200 (see Fig. S1 in the SI). EOF2 represents the second dominant pattern of ~~sub-seasonal~~intraseasonal variability of the Z200 field. However, EOF1 and EOF2 explain respectively only 6.4% and ~~5.9~~6.2% of the ~~sub-seasonal~~intraseasonal variability, i.e~~.~~., after removing mean annual cycle and interannual variability. The first 5 EOFs explain together ~~27.9~~28.5% of the total variability. These low values are not surprising if we consider that the mean annual cycle and interannual variability have been removed, thus leaving only the disturbances from the year-specific mean state. We apply a test for pronounced separation of EOFs following North et al, (1982). This test is based on the eigenvalues of the EOF analysis and provides a rule of thumb to determine whether two EOFs are degenerate, i.e. "indistinguishable between their uncertainties" (for further details see North et al. 1982 and ~~Hannachi et al. 2007)~~ Hannachi et al. 2007). The ~~North~~ test shows that the first three EOFs are not well separated mutually, while EOF1 and EOF2 are well separated from EOF4 and EOF5 (see SI, Fig. S2). Similarly, DW2007 also found a twin EOF mode in their study, ~~however their work focuses on the Eurasian sector only.~~ although their work focuses on the Eurasian sector only. However, by design EOFs do not necessarily reflect physical patterns (irrespective of whether the individual EOFs are separable or not). They only capture dominant patterns of variability, and here EOF2 is useful since it resembles the wave-pattern in the correlation map and, thus, likely results from the circumglobal teleconnection pattern related physics. In general, we cannot a priori exclude an influence of EOF1 on the analysed system (while an influence from EOF2 is expected). Therefore, we will test whether this is the case in the following part of this sub-section.

DW2005 define the ISM-circumglobal teleconnection mechanism at interannual and monthly ~~time scale~~timescale, DW2007 uses daily data (after removing the interannual variability) to analyse the influence of the Eurasian wave train initiated from the North Atlantic on the north-central Indian rainfall. Thus, first we need to show that the definition of the circumglobal-teleconnection pattern as in DW2005, also applies meaningfully at weekly timescales, with the interannual variability ~~is~~ filtered out. Figure 2c shows the spatial correlation structure of MT rainfall with the variability of Z200 in the Northern Hemisphere at weekly timescale with Z200 leading the MT rainfall by one week (lag = -1). The arch-shaped structure which is visible in Fig. 2c over Europe and Central Asia suggests that there could be a wave pattern propagating eastward from Western Europe and affecting the MT rainfall with a 1-week lag. This hypothesis will be further tested in the next section. Following DW2005, we define a circumglobal teleconnection index (CGTI) as the weekly-mean Z200 spatially averaged over the region 35°-40°N, 60°-70°E (white box in Fig. 2d). The contemporaneous (lag~~=~~ = 0) spatial correlation structure between the CGTI and Z200, i.e~~.~~., the circumglobal teleconnection pattern, is shown in Fig. 2d. A large region of strong positive correlation surrounds the Caspian Sea, while downstream, ~~i.e. moving from west to east following the mid-latitude westerlies,~~ a circumglobal wave train is shown with 4 positive centres of action positioned over ~~east~~East Asia, the North Pacific, North America and ~~western~~Western Europe. Despite different temporal averaging, using different datasets and the fact that we remove the interannual variability, our results with 5 centres of positive correlation strongly resemble the circumglobal teleconnection pattern described by DW2005, in terms of sign and the geographical position of its centres of action. Thus, we conclude that the ~~circumgobal~~circumglobal teleconnection ~~patter~~pattern defined by the CGTI region manifests also at weekly ~~time scale~~timescale. Moreover, the wave pattern shown over Eurasia in Fig. 2c and 2d also resembles the wave pattern identified by DW2007. Thus, this resemblance represents the first hint that the circumglobal teleconnection pattern and the Eurasian wave

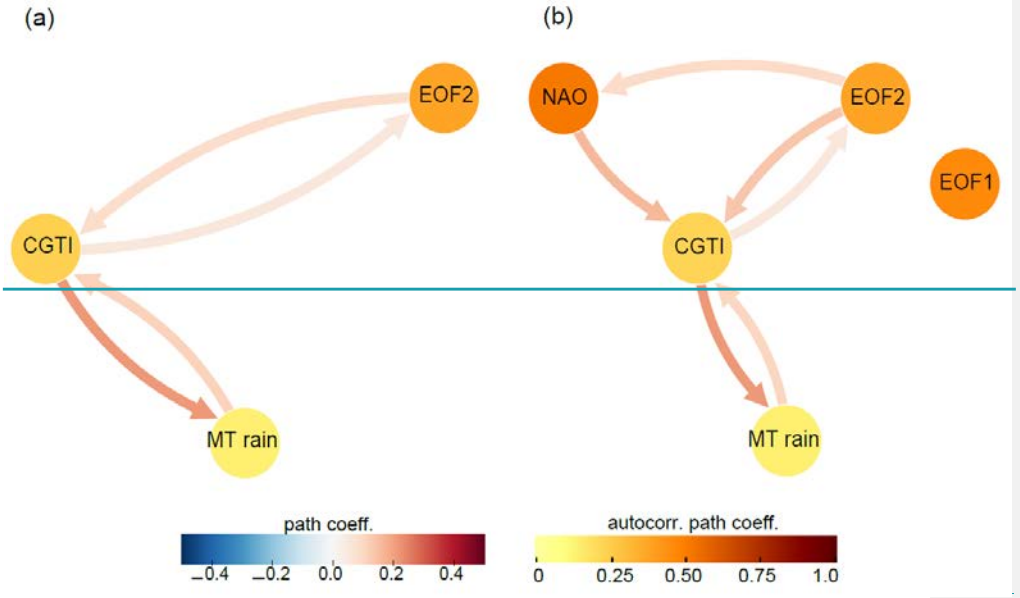
train analysed by DW2007 share common elements, and ~~llink~~link the North Atlantic ~~variavility~~variability to the ISM region. In both Fig. 2c and 2d correlation values with corrected *p*-values $p < 0.05$ are highlighted with black contours.

The circumglobal teleconnection pattern (Fig. 2d) shows a ~~close to zero and non~~ very low though significant negative spatial correlation with the EOF1 pattern~~.~~ (r = -0.09 ± 0.02; here and later, r is shown together with its confidence interval at $\alpha = 0.05$). The spatial correlation of EOF2 (Fig. 2b) with the circumglobal teleconnection pattern (Fig. 2d) ~~at a global scale~~over the Northern Hemisphere is $r = 0.32\,35 \pm 0.01\,02$, and the spatial correlation with the circumglobal teleconnection pattern increases when limited to the Eurasian sector only (0°-90°N – 0°-150°E, $r = 0.46\,52 \pm 0.02$). Contrarily, the spatial correlation with EOF1 and the circumglobal teleconnection is also low when only the Eurasian sector is taken into account ($r = -0.05\,16 \pm 0.03$). When EOFs are calculated only on the Eurasian sector ~~(0°-90°N – 0°-150°E),~~ the order of the first and second EOFs is reversed, but the spatial patterns are very similar: $EOF1_{Eurasia}$ is strongly spatially correlated with EOF2 ($r = 0.93\,89 \pm 0.01$) and with the circumglobal teleconnection pattern ($r = 0.44 \pm 0.02$; see SI, Fig. S3). Thus, the choice of the region to calculate the EOFs does not strongly affect the results. Moreover, since we are interested in the two-way effects of the entire mid-latitude circulation and the MT rainfall, we decided to use the EOFs defined over the entire Northern Hemisphere (as shown in Figs. 2a,b).

The time series for the principal component of EOF2 also significantly correlates with the CGTI time series ($r = 0.26\,30 \pm 0.07$) while the correlation for the first principal component is low and not significant ($r = -0.01\,06 \pm 0.08$). The circumglobal teleconnection pattern (Fig. 2d) also strongly resembles the correlation structure between MT rainfall and Z200 at lag -1 (Fig. 2c) with a spatial correlation of 0.69 ± 0.01. These findings are consistent with those of DW2005, and thus illustrate a likely interaction between MT rainfall and Z200 variability.

T~~emperature~~Composite 2m-temperature and precipitation ~~anonalies during~~differences between weeks with strong CGTI ($CGTI > 1\ CGTI_{s.d.}$, where $CGTI_{s.d.}$ is the s.d. of the CGTI index) minus weeks with weak CGTI anomalies ($CGTI < -1\ CGTI_{s.d.}$) also exhibit a clear wave train pattern originating from Western Europe~~. We calculate 2m-temperature and precipitation anomalies of the composite of weeks with CGTI > 1 $CGTI_{std}$ minus composites of weeks with CGTI < -1 $CGTI_{std}$ (where $CGTI_{std}$~~ and central Europe respectively.is the s.d. ~~of the CGTI index).~~ Figures 2e,f present the corresponding 2m-temperature and precipitation anomalies from the ~~Pai et al.~~ERA-Interim and CPC-NCEP dataset, respectively, showing anomalies with *p*-values $p < 0.05$ shaded and highlighting anomalies that have corrected *p*-values $p < 0.05$ by black contours. In both variables, a wave train from Western/Central Europe to India via European Russia is detected~~.~~ (in Fig. 2e,f highlighted by a black arrow). Wet and cold anomalies appear over central India and European Russia, while warm and dry anomalies are found over Western/Central Europe and east of the Caspian Sea. Warm ~~and dry~~ anomalies appear together~~,~~ with the high and low features shown in the circumglobal circulation pattern over Europe and central Asia (Fig. 2d), however precipitation anomalies show a slight eastward shift with respect to temperature anomalies. Precipitation anomalies are ~~weaker than those found for 2m-temperature, however a clearly defined wave pattern appears over the Eurasian sector.~~ more spatially confined than those found for 2m-temperature. However, a clearly defined wave pattern appears over the Eurasian sector for both variables. Though these plots are obtained by plotting composites of weeks with $CGTI > 1\ CGTI_{s.d.}$ minus composites of weeks with $CGTI < -1\ CGTI_{s.d.}$, we have found that these composites behave close to linearly when plotted separately (see SI, Fig. S4). This further supports the assumption that a linear framework is a reasonable choice in this context. Moreover, when regressing the MT rainfall on the CGTI index at different lags (from lag 0 to lag -2 weeks), the strongest relationship between the CGTI and the

CPC-NCEP rainfall is found at lag -1 week, with the CGTI leading a change in MT rainfall by one week (see SI, Fig. S5). This information also further supports the choice of analysing the relationships between these two variables at lag = -1 week.



~~Figure 3. Causal mid-latitude interactions with the ISM.~~ ~~Panel (a): Causal Effect Network (CEN) built with CGTI, the PC of EOF2 and MT rainfall for the period 1979-2017. Panel (b): Same as panel (a) but with the addition of the PC of EOF1 and NAO. The strength of the causal links expressed by the standardized regression path coefficients and autocorrelation path coefficients are shown. All links have a lag of 1 week~~. See the main text for discussion.

Based on the DW2005 hypothesis, we build a CEN with CGTI, MT rainfall and the principal component of EOF2 (Fig. 3a). This CEN depicts a positive two-way connection both between CGTI and MT rainfall and between CGTI and EOF2. This implies that anomalously high CGTI values (with relatively high Z200 east of the Caspian Sea) enhance MT rainfall at a 1-week lag, while lower CGTI values have the opposite effect. Note that the causal effect here is always acting in the direction of the arrow with a lag of one week, e.g., the arrow from CGTI towards MT rainfall represents a causal effect of 0.22 (given by the colour) from CGTI to MT rainfall with a lag of one week. The link directed from MT rainfall to CGTI illustrates a reverse influence, creating a positive feedback between CGTI and MT rainfall, ~~consistent with DW2005~~.supporting the monsoon-circumglobal teleconnection pattern hypothesis. The strength of the causal link between the CGTI and MT rainfall is expressed by the path coefficients (see Methods section).

The causal link strength of the CGTI acting on the MT rainfall is $\beta_{CGTI \rightarrow MT}$ ~0.2 (meaning that a change of 1 s.d. in CGTI leads to a change in 0.2 s.d in MT rainfall, while the reverse link is weaker ($\beta_{MT \rightarrow CGTI}$~0.1) but still significant. EOF2 shows a two-way link with the CGTI. The links between EOF2 and CGTI support the DW2005 hypothesis that wave trains in the mid-latitudes (represented by EOF2) affect the MT rainfall via the CGTI.

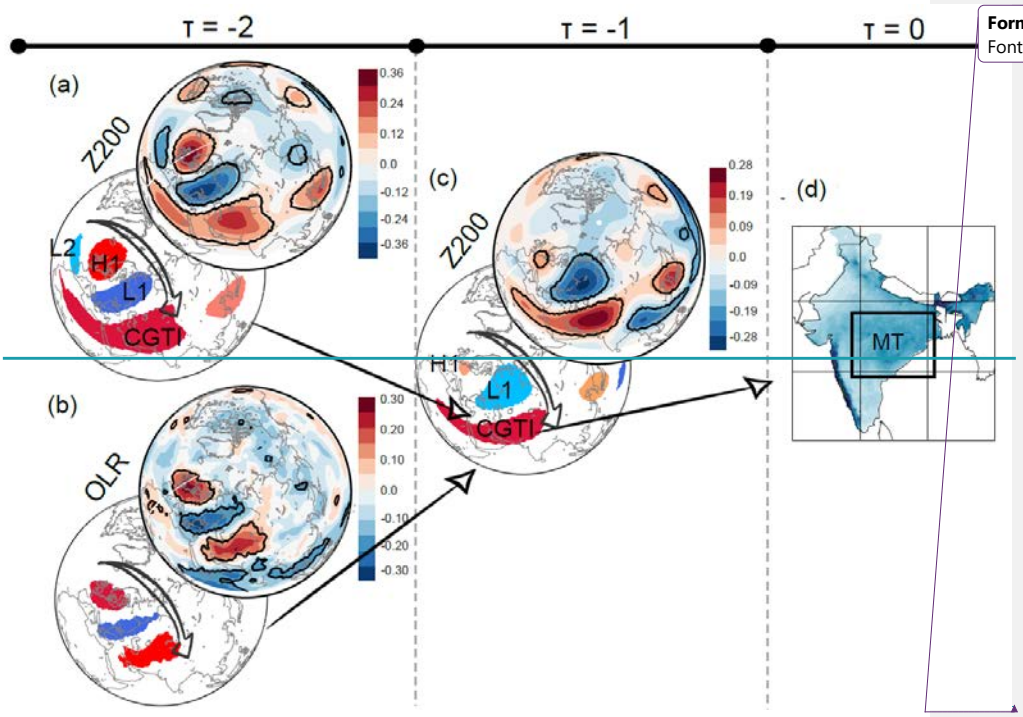
To assess whether the North Atlantic variability affects the MT rainfall (as hypothesized by DW2007), we add the NAO index to our original CEN. In order to check whether the first mode of variability in the Northern Hemisphere may also play a role in shaping MT rainfall variability, we additionally include EOF1. Figure 3b shows the resulting CEN: the causal links identified in the previous CEN (Fig. 3a) remain unaltered, and two additional positive links from NAO to CGTI and from the EOF2 to the NAO emerge. A positive NAO phase will strengthen the CGTI at 1-week lag ($\beta_{NAO \rightarrow CGTI}$~0.2~~) with a reverse influence from~~), while the NAO is influenced by EOF2 ~~on NAO~~ ($\beta_{EOF2 \rightarrow NAO}$~0.1), though ~~weaker~~weakly. EOF1 does not show a causal connection with CGTI or with any other actor, showing that in this context, EOF1 and EOF2 present a different behaviour, though in principle not statistically separated following the North test. This CEN unveils both an influence of the mid-latitude atmospheric dynamics (EOF2) and North Atlantic variability (NAO) on MT rainfall and back, with the CGTI acting as a gateway, thus supporting both the DW2005 and DW2007 hypotheses. Substituting EOF2 with its corresponding EOF calculated over the Eurasian sector, $EOF1_{Eurasia}$, shows consistent results (see SI, Fig. ~~S4~~S6).

3~~.2 Regional extratropical~~3.2 Eurasian and North Atlantic mid-latitude features affecting the Indian summer monsoon

The EOF-based CEN analysis depicts the total hemispheric response of Z200 without differentiating among the influences of different geographical regions. To detect influential regions, we next apply ~~RG-CPD~~the Response-guided Causal Precursors Detection (RG-CPD) scheme to search for causal precursors of both the CGTI and MT rainfall at 1-week lead time in weekly OLR and Z200 fields. In the tropical belt, OLR is often used as a proxy of rainfall and convective activity due to its relation with the temperature at the top of the clouds. Deep convection is characterized by high altitude cloud tops and low emission temperatures (and thus low OLR values), while at clear sky conditions, the emission temperature of the land surface is higher and leads to larger OLR values (Krishnan et al., 2000).

Figure 4 summarizes our corresponding findings. The right column (Fig. 4) shows the MT region and represents lag 0. Moving towards the left, the second column shows the correlation maps (top) and causal precursors (bottom) of the MT rainfall identified at 1-week lead time in Z200 (Fig. 4c). Causal precursors describe an arch-shaped wave train from Western Europe to India. The wave train features one low-pressure region (L1 over European Russia) and two high-pressure regions (the CGTI and H1 over Western Europe). The leftmost column in Fig. 4 shows the correlation maps (top) and causal precursors (bottom) of the CGTI identified in the Z200 (Fig. 4a) and OLR (Fig. 4b) fields at 1-week lead time with respect to the CGTI (2-week lead time with respect to the MT rainfall). Again, both Z200 and OLR correlation maps show an arch-shaped wave pattern emanating from the North Atlantic and propagating towards the Caspian Sea via European Russia. The associated Z200 causal precursors for the CGTI clearly depict this mid-latitude wave train both in Z200 and OLR (Figs. 4a,b). In the Z200 field (Fig. 4a), the wave train features two lows (L1 over European Russia and L2 over the eastern North Atlantic) and two highs (the CGTI and H1 over Western Europe). OLR causal precursors (Fig. 4b) depict only the L1 and H1 components of the wave, as the correlation over L2 is not significant. Moreover, the prominent influence of the tropical belt on the CGTI is also detected (OLR1, Fig. ~~4b~~4b), in agreement with previous findings (se Fig. 3). This result further supports the hypothesis that a wave train coming from the mid-latitudes influences the ISM circulation system via the CGTI region as already shown in Fig. 2, while strong temperature and precipitation anomalies shown in Figs. 2e,f coincide markedly with the regions H1, L1 and CGTI identified in Figs. 4a,c. ~~figure~~Figure 4a shows a North Atlantic pattern that resembles a negative NAO, whereas Fig. 3b indicated a positive

causal link from NAO on CGTI. This ~~seaming~~seeming mismatch is explained by the difference in pressure level and lead-time with a more positive NAO pattern at lag -2 (see SI, Fig. ~~S5~~S7).

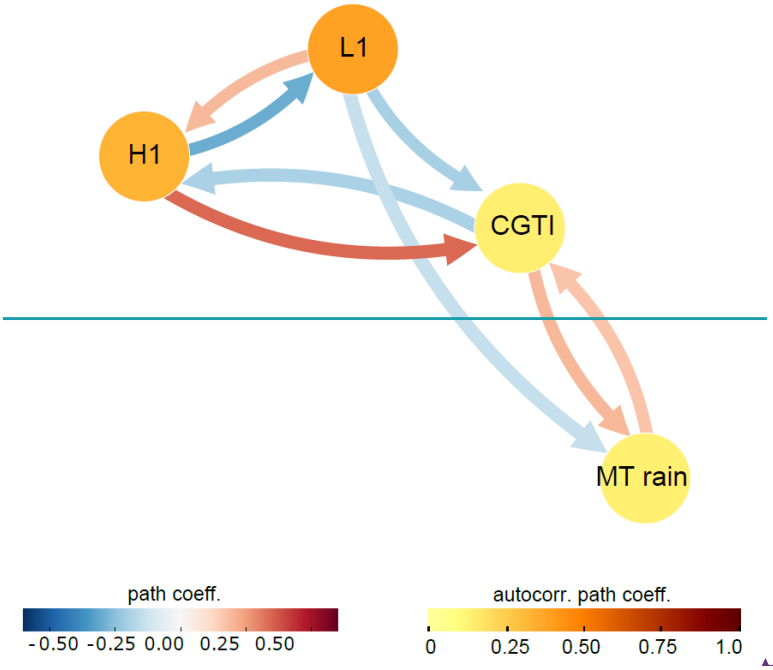


Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Red, English (United Kingdom)

~~F~~igure 4. Mid-latitude causal precursors of ISM. Panel (a): correlation of CGTI with Z200 at 1-week lead time (top), and the causal precursors of CGTI identified via RG-CPD (bottom) for the period 1979-2017. Panel (b): as for panel (a) but for OLR fields. Panel (c): correlation map for weekly MT rainfall and Z200 field at 1-week lead time (top) and the causal precursors identified via RG-CPD (bottom). Panel (d): ISM rainfall over the MT region from the Pai et al. ~~dataset (reproduced from Fig.~~ 1a).

The CEN built with MT rainfall, CGTI and the upstream part of the mid-latitude wave train, i.e~~.~~., L1 and H1, is shown in Fig. 5. The causal links between the CGTI and the MT rainfall remain unaltered (see Fig. 3), with the CGTI mediating the connection between the mid-latitude wave (represented by H1 and L1) and the MT rainfall. Further, the obtained CEN can be interpreted as a wave train that propagates eastward from the ~~east~~East Atlantic towards the Indian monsoon region, as postulated by DW2007. Let us focus on the connections between H1 and L1. If H1 increases, then L1 decreases at 1-week lag. The backward link is positive: thus, if L1 increases H1 also increases at 1-week lag. This dampening loop is easiest explained by an eastward propagating wave. To illustrate this, we design a simple experiment using synthetic time series representing a simple cosine-wave propagating eastward with added noise sampled at two locations. We estimate the wavelength, wave speed and amplitude based on the observed properties of H1 and L1 (see SI, Fig. ~~S7~~S8-S9). The CEN resulting from these time series is equivalent to that of H1 and L1 in Fig. 5. ~~With~~with a negative forward link and a positive backward link. Physically, this can be understood ~~in that~~by a traveling wave ~~amplifies~~amplifying in the forward direction, but at the same time ~~dampens~~dampening in the backward direction.

Formatted: Font: (Default) Times New Roman, 9 pt, English (United Kingdom)



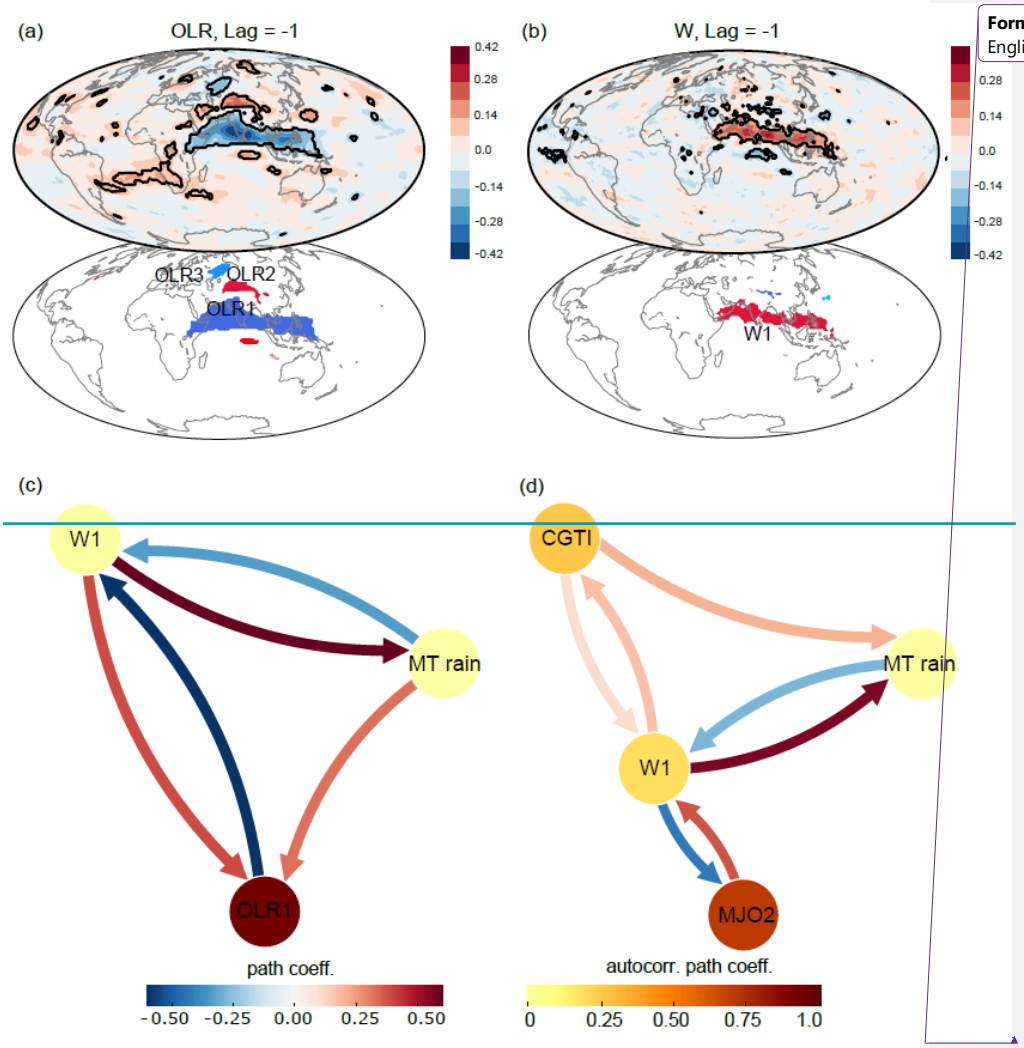
Formatted: Font: (Default) Times New Roman, 9 pt, Bold, English (United Kingdom)

~~F~~igure 5. Mid-latitude wave train. CEN built with the MT rainfall, CGTI, L1 and H1 (as identified in Fig. 4a) for the period 1979-2017. The strength of causal links expressed by the standardized regression path coefficients and autocorrelation path coefficients are shown. All links have a lag of 1 week. See the main text for discussion.

3.3 ~~Internal feedbacks in~~Intraseasonal variability and tropical influence on the monsoon circulation

Next, we perform a similar analysis ~~using~~applying RG-CPD to fields of vertical wind velocities (~~W~~W500) and OLR to capture the internal feedbacks and dynamics of the ISM convective updraft. Their correlation maps and detected causal precursors are shown in Figs. 6a,b. MT rainfall has four causal precursors in OLR. A large region ~~located over India and~~extending from the Arabian Sea towards the Maritime Continent via the Indian sub-continent (OLR1, Fig. 6a) shows a negative causal link, while another region covering parts of the ~~Himalayan plateau and extending toward the Arabian Peninsula~~Tibetan Plateau shows a positive causal link (OLR2, Fig. 6a). A third region is found north-east of the Caspian Sea, over western Russia (OLR3) with a negative causal relationship with MT rainfall. ~~Finally, a last region located over north/central Europe showing a negative causal relationship with MT rainfall is found~~A fourth, though small, region of positive correlation is found over the equatorial Indian Ocean (OLR4, Fig. 6a). OLR1 spatially overlaps with W1, the largest causal precursor in the vertical wind field, representing the ~~summer branch of~~north-west/south-east tilted rainfall band related to the ~~ITCZ~~BSISO over the northern Indian Ocean and western Pacific Ocean (Fig. 6b, top panel). The positive correlation of W1 and negative correlation of OLR1 with MT rainfall indicate that ascending motions and associated high-level cloud formation

~~(reducing OLR)~~ are followed by enhanced rainfall over the MT region with a lag of 1 week. OLR2~~,~~ and OLR3 ~~and OLR4~~ largely overlap with the regions ~~H1~~L1 and ~~L1 as well as~~ the CGTI region identified in Fig. 4c, further supporting the importance of this mid-latitude wave pattern in modulating the rainfall over the MT region also when a different variable is selected. OLR4 hints to the presence of a seesaw of enhanced and suppressed convective activity alternating between the equatorial Indian Ocean and the Indian Peninsula, likely related to the northward propagation of the BSISO.



Formatted: Font: (Default) Times New Roman, 9 pt, English (United Kingdom)

~~F~~**igure 6. Tropical causal interactions of ISM.** Panel (a) shows the correlation map of weekly MT rainfall with the global OLR field at 1-week lead time (top panel) and the causal precursors identified via RG-CPD (bottom panel) for the period 1979-2017. ~~Panels (b): as for panel (a) but for the W field. Panels (c) and (d) show the CENs build with W1, OLR1 and MT rainfall and MT rainfall, W1, CGTI and MJO2, respectively. The strength of causal links expressed by the standardized regression path coefficients and autocorrelation path coefficients are shown~~. All links have a lag of 1 week. See the main text for discussion.

Figures 6c,d show two CENs constructed from the MT rainfall causal precursors in OLR and ~~W~~W500 (region OLR1 as defined in Fig. 6a and W1 as defined in Fig. 6b), and including CGTI and MT rainfall itself. Since the OLR and ~~W fields are more noisy~~W500 fields present less organized large-scale spatial patterns than Z200, we use only the two dominant causal precursors. Although the regions OLR1 and W1 show a large spatial overlap, they represent two different components of the ISM system. OLR is calculated at the top of the atmosphere ~~and~~with low OLR values ~~represent high clouds that reach the top of the troposphere. Thus, OLR is a proxy for deep convection and clouds. W~~representing strong convective motions. W500 is calculated at 500 hPa and W1 thus represents the ascending branch of the ISM circulation cell and nearby ~~ITCZ~~BSISO. The CEN built with OLR1, W1 and the MT rainfall at weekly timescale (Fig. 6c) represents the ~~internal dynamics~~intraseasonal variability of the monsoon ~~cell.~~circulation and therefore its relation with the BSISO and the initiation of active and break phases. Enhanced vertical motions (W1) precede an increase in the MT rainfall by one week~~, while stronger MT rainfall leads to weaker ascending motions one week later. Meanwhile~~; seemingly, more clouds (lower OLR1) are associated with enhanced W1 one week later (thus enhancing the ISM circulation). Contrarily, enhanced rainfall shows a negative feedback on both OLR1 and W1: stronger MT rainfall leads to reduced vertical motions and clearer skies (lower W1 and higher OLR1, respectively). ~~This~~Hence, this CEN ~~thus~~ depicts a negative feedback intrinsic to the ISM ~~internal~~intraseasonal variability: while enhanced ascending motions and thus strengthened ISM circulation lead to stronger MT rainfall, enhanced rainfall leads to weaker ISM circulation and in turn diminished MT rainfall~~. From~~, hinting to an alternation of rainy and dry periods over the MT region, likely related to the alternation of active and break phases. Thus, from a physical point of view, this can be explained in two ways: (1) as a seesaw representing the switch of the ISM system between an active and a break phase and (2) via an increase of atmospheric static stability due to latent heat release in the higher layers of the troposphere. ~~This~~The latter mechanism is ~~shown~~further analysed in Fig. ~~S9~~S10 (see SI), where we build a CEN with MT rainfall, sea level pressure (SLP) over the Bay of Bengal (SLP_BOB), temperature at 400hPa (Tp4_MT) and temperature at 600 hPa (Tp6_MT). The resulting causal links show that while a decrease in SLP over the Bay of Bengal is followed by an increase in MT rainfall and both Tp4_MT and Tp6_MT, an increase in Tp4_MT (thus enhanced static stability) leads to a decrease in MT rainfall. The described mechanism is in agreement with what was proposed by ~~Saha et al. (2012)~~Saha et al. (2012) and Krishnamurti and Bhalme (1976).

Next, we test the causal relationships identified between MT rainfall and W1 when adding the CGTI and MJO variability to the CEN (Fig. 6d). MJO variability is expressed by the OMI PC2 index (here referred to as MJO2), the second EOF of OLR in the tropical belt (Kiladis et al., 2014). OMI PC2 corresponds to the RMM1 index, widely used in previous work on the relationship between the MJO and the ISM ~~(Kiladis et al., 2014; Mishra et al., 2017; Pai et al., 2011). Positive RMM1 values correspond to MJO phases 3-6 (also referred to as the active phases of MJO) and physically represent the presence of strong convection activity propagating eastward from the Indian Ocean towards the western Pacific. Several studies show that MT rainfall is enhanced during the active MJO phase (Anandh et al., 2018; Mishra et al., 2017; Pai et al., 2011). In this CEN, the links between the MT rainfall and W1 remain unaltered, while the CGTI shows a positive feedback with W1: an increase in the CGTI causes stronger ascending motions, while stronger ascending motion over the Indian region strengthens the CGTI. The CGTI has a direct positive causal link to MT rainfall (Fig. 6d). Even though the direct link from MT rainfall to CGTI has now disappeared (likely because this link was already relatively weak in a simpler CEN configuration,~~

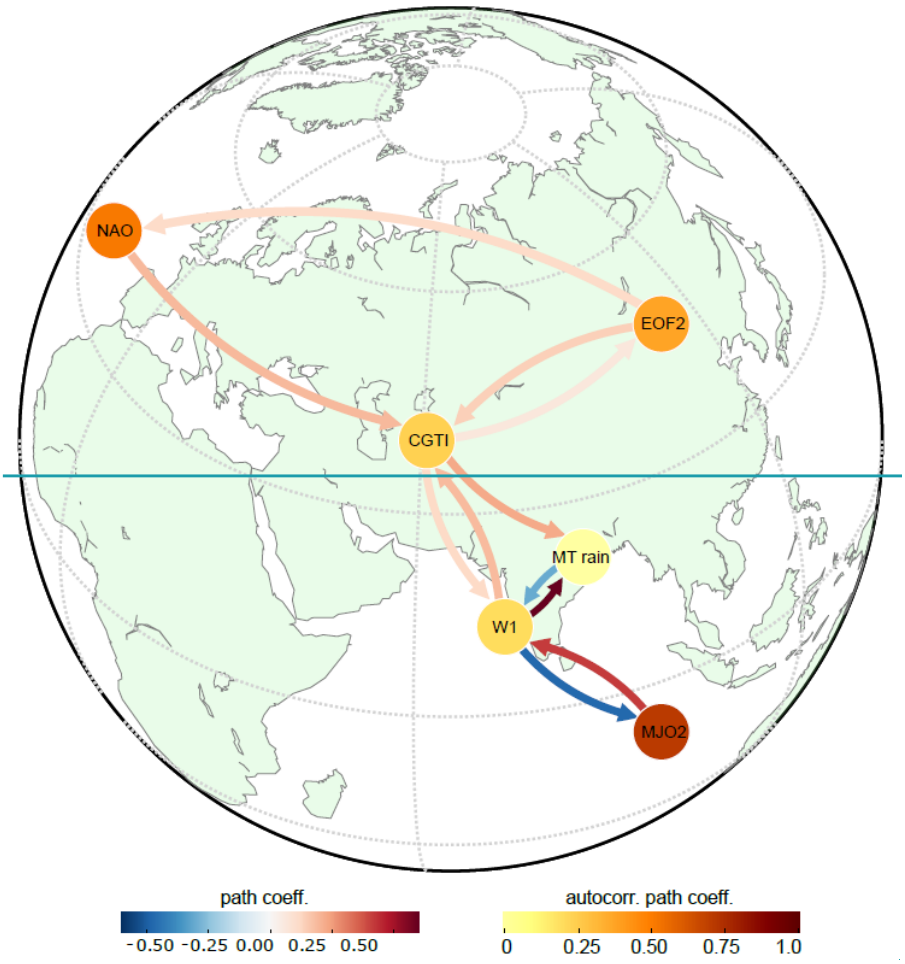
~~see Fig. 3a), the link between the ISM circulation and the mid-latitudes remains via W1. Stronger ISM circulation (higher W1) leads to increased CGTI and vice versa as seen in Fig. 3, where higher CGTI leads to enhanced MT rainfall and vice versa. MJO2 displays a positive causal link with W1, meaning that OMI PC2 positive values lead to enhanced vertical motions with 1-week lead time and, as a consequence, enhanced MT rainfall with 2-week lead time. However, stronger W1 leads to decreased MJO2 (negative causal link from W1 to MJO2). Thus, MJO2 shows exactly the same causal relationships (but opposite signs) with W1 as shown for OLR1 in Fig. 6c, likely because the OMI index is also defined based upon OLR fields in the tropical belt and the OLR pattern shown by the second EOF of the OMI index largely overlaps with our OLR1 region (Kiladis et al., 2014).~~(Pai et al., 2011; Kiladis et al., 2014; Mishra et al., 2017). Positive RMM1 values correspond to MJO phases 3-6 (also referred to as the active phases of MJO) and physically represent the presence of strong convection activity propagating eastward from the Indian Ocean towards the western Pacific. Several studies show that MT rainfall is enhanced during active MJO phases (Pai et al., 2011; Mishra et al., 2017; Anandh et al., 2018). Moreover, in summer the OMI index also represents the BSISO, further linking OLR1 and W1 to the BSISO related convective activity (Wang et al., 2018). In this CEN, the link from the MT rainfall towards W1 remains unaltered (while the link in the opposite direction disappears, indicating that this link is less robust than the others), while the CGTI shows a positive feedback with W1: an increase in the CGTI causes stronger ascending motions (W1), while stronger ascending motion over the Indian region strengthens the CGTI. The CGTI has a direct positive causal link to MT rainfall (Fig. 6d). Even though the direct link from MT rainfall to CGTI has now disappeared (likely because this link was already relatively weak in a simpler CEN configuration, see Fig. 3a), the link between the ISM circulation and the mid-latitudes remains via W1. Stronger updraft over the ISM and the Maritime Continent regions (higher W1) leads to increased CGTI and vice versa as seen in Fig. 3, where higher CGTI leads to enhanced MT rainfall, both directly and via W1. MJO2 displays a positive causal link with W1, meaning that OMI PC2 positive values lead to enhanced vertical motions with 1-week lead time and, as a consequence, enhanced MT rainfall with 2-week lead time. However, stronger W1 leads to decreased MJO2 (negative causal link from W1 to MJO2). Thus, MJO2 shows exactly the same causal relationships (but opposite signs) with W1 as shown for OLR1 in Fig. 6c, likely because the OMI index is also defined based upon OLR fields in the tropical belt and the OLR pattern shown by the second EOF of the OMI index largely overlaps with our OLR1 region (Kiladis et al., 2014).

We checked how these links decay over time and found that the causal links from W1 to the CGTI and from the CGTI to the MT rainfall remain stable at lag -2 weeks, but then decay to zero at lag -3 weeks. The causal effect from W1 on the MT rainfall and from MJO2 to W1 drop to values close to zero at lag -2 and -3 (see SI, Fig. S11). This result is further visible if the timescale is changed from weekly to monthly: in this case, the spatial correlation between the circumglobal teleconnection pattern and the correlation map between MT rainfall and Z200 at lag 0 is still high ($r = 0.79$). However, when the correlation is calculated between MT rainfall and Z200 at lag -1 (month), the map becomes non-significant (see SI, Fig. S12). This result further suggests that these causal relationships have a characteristic timescale shorter than one month and of about 1-2 weeks. Moreover, OLR shows no significant correlation patterns with the MT rainfall and no significant causal links are detected between MT rainfall and the CGTI at monthly timescale (not shown).

3.4: Combining local, tropical and mid-latitude causal interactions

Finally, we bring together the findings obtained with CEN and RG-CPD throughout this study and summarize them in a single CEN to provide an overall picture and test the consistency of the results. We include the most

important identified regions from both the tropics and the mid-latitudes together in a single CEN (Fig. ~~7)~~.7) and plot the corresponding CEN over a map to help the reader to associate each actor with its corresponding approximate region (though in cases when the index is defined over all longitudes, such as EOF2, it is only possible to associate the actor with its average latitude). Specifically, this CEN is built with elements that come from both, the DW2005 and DW2007 hypotheses (Fig. 3) and from our RG-CPD analysis (Figs. 4 and 6). Moreover, to test the consistency of the results under a change of rainfall dataset, we also performed the same analysis using the CPC-NCEP rainfall dataset over the period 1979-2016 (see SI, Figs. S13-S19) and found that our results are robust when using the CPC-NCEP rainfall dataset. In Fig. 7, all causal links are reproduced with the same directions and only the magnitudes of the causal effects exhibit minor changes with respect to CEN shown in Fig. 3 and 6. Our results show that the influence of both, the mid-latitude circulation (EOF2) and the North Atlantic (NAO), on the MT rainfall is mediated via the CGTI and is robust: the structure and the direction of the causal links are retained. The backward influence of the MT rainfall on the mid-latitude circulation is weaker and more complex, as shown already in Fig. 6d. CGTI has both a direct causal link to MT rainfall and an indirect one via W1.



~~F~~igure 7. Combined mid-latitude and tropical causal interactions of ISM. CEN built with W1, MJO2, MT rainfall, NAO, CGTI and EOF2 for the period 1979-2017. ~~The strength of causal links expressed by the standardized regression path coefficients and autocorrelation path coefficients are shown~~MT rainfall and W1 show a negative feedback (though the link from the MT rainfall towards W1 is only significant when CPC-NCEP rainfall data are used, see SI, Fig. S19), with increased W1 leading to enhanced MT rainfall but stronger MT rainfall leading to weaker W1. The MT rainfall is also influenced by MJO2 via W1 with a two-way connection. Higher MJO2 values (inked to the MJO phases 3 to 6) are followed by stronger upward motions (increased W1) one week later, which in turn causes enhanced MT rainfall two weeks later (with respect to MJO2). In the opposite direction, suppressed or weakened ascending motions promote lower MJO2 one week after, thus promoting the switch toward MJO phases 7-8 and 1-2, also known as suppressed MJO phases. Suppressed MJO phases are in turn linked with the onset of break phases of the Indian monsoon (Pai et al., 2011).

~~T~~o quantify the relative influence of tropical and mid-latitude teleconnections on MT rainfall, we report the causal effect (CE) strength of each link, which is given by the path coefficient in the CEN (see Fig. 7 and Table S1 in the SI). . All links have a lag of 1 week. See the main text for discussion.

~~MT rainfall and W1 show a negative feedback, with increased W1 leading to enhanced MT rainfall but stronger MT rainfall leading to weaker W1. The MT rainfall is also influenced by MJO2 via W1 with a one-way connection. Suppressed or weakened ascending motions promote lower MJO2 one week after, thus promoting the switch toward MJO phases 7-8 and 1-2, also known as suppressed MJO phases. Suppressed MJO phases are in turn linked with the onset of break phases of the Indian monsoon (Pai et al., 2011).~~

T~~o quantify the relative influence of tropical and mid-latitude teleconnections on MT rainfall, we report the causal effect (CE) strength of each link, which is given by the path coefficient in the CEN.~~W1 has the strongest causal effect on MT rainfall with $\beta_{W1\rightarrow MT}$ = 0.~~52~~54, implying that a one s.d. shift in W1 causes about a half s.d. change in MT rainfall after one week (under the previously mentioned conditions, see Methods section). The CGTI influences MT rainfall directly and indirectly via W1 and the total causal effect is given by $\beta_{CGTI\rightarrow MT} + \beta_{CGTI\rightarrow W1} \cdot \beta_{W1\rightarrow MT} = 0.~~24~~23$, where $\beta_{CGTI\rightarrow MT} = 0.~~19~~18$ and $\beta_{CGTI\rightarrow W1} = 0.~~10~~09$ The causal effect of CGTI is thus roughly half as strong as that of the ~~internal~~intraseasonal variability of the Indian monsoon system as represented here by W1. MJO2 has a causal effect on W1 of $\beta_{MJO2\rightarrow W1}$ = 0.~~35~~49 and is influenced by W1 with $\beta_{W1\rightarrow MJO2}$ = -0.39. ~~Thus, taking both W1 and MT as representatives of internal ISM dynamics, the effects of external drivers from both tropics and mid-latitudes on this internal dynamics are both of the order of half as strong as that of W1 on MT.~~ Looking more specifically at the causal effect of MJO2 on MT rainfall mediated via W1, we find $\beta_{MJO2\rightarrow W1}*\beta_{W1\rightarrow MT}$ = 0.~~18~~26, i.e., the tropical driver effect on MT rainfall itself is similar in magnitude as that of CGTI as for the key extratropical driver. ~~The~~Thus, taking both W1 and MT as representatives of ISM intraseasonal variability, the effects of external drivers from both tropics and mid-latitudes on the ISM circulation are both about half as strong as that of W1 on MT. For comparison, the path coefficients for all the identified causal links are reported ~~in~~together with the corresponding Pearson correlations of each pair of variables (see SI, Table S1~~ (see SI).~~). In general the values of the causal effect and the simple correlation do not differ greatly, despite a few exceptions where linear correlation is not significant even though there exists a causal link, or when the sign of the causal effect and the simple correlation differ. However, one should not forget that plain correlation does not indicate causality, nor can identify chains of causality between different variables.

Moreover, we calculate the average causal effect (ACE) and average causal susceptibility (ACS) for each actor. While ACE gives a measure of the causal effect that each actor has on the rest of the network, ACS measures the sensitivity of each actor to perturbations entering in any other part of the network ~~(Runge et al., 2015a). In this CEN, W1 has the highest ACE (~0.21) and both CGTI and MJO2 the second highest ACE (~0.07). MT rainfall and W1 show the strongest ACS (~0.1). ACE and ACS values for each actor are further summarized in Table S2, see SI. These values again stress the importance of both, ISM internal~~(Runge et al., 2015). In this CEN, W1 has the highest ACE (~0.22) and both CGTI and MJO2 the second highest ACE (~0.10). MT rainfall and W1 show the strongest ACS (~0.1). ACE and ACS values for each actor are further summarized in Table S2, see SI. These values again stress the importance of both, ISM intraseasonal variability and CGTI, in mediating mid-latitude waves towards the ISM.

~~We performed the same analysis using the CPC-NCEP rainfall dataset over the period 1979-2016 (see SI, Figs. S10-S16) and find that our results are robust when using this different rainfall dataset. All causal links are reproduced with the same directions and only the magnitudes of the causal effects show minor changes.~~

Finally, we test for the robustness of the causal links when different regions other than the MT area are studied. In general, analysing all-India rainfall (AIR), western Himalayan foothills (WHF, defined over 26°-35°N and 70°-83°E) or eastern Himalayan foothills (EHF, defined over 20°-30°N and 87°-97°E) rainfall does not affect the strength or the sign of the causal links (see SI, Fig. S20-S21), thus showing that our results are robust and can actually represent the dynamics of the entire ISM basin. This is likely a consequence of the fact that AIR strongly reflect the behaviour of the MT rainfall ($r = 0.57 \pm 0.05$), as here the most abundant rainfall takes place. While the correlation between AIR and WHF rainfall is lower ($r = 0.35 \pm 0.06$), the same causal links as for the MT rainfall are observed, suggesting a strong similarity in the dynamical mechanisms that govern MT and WHF rainfall. However, when EHF rainfall is taken into account, the correlation with the AIR is low and not significant ($r = 0.03 \pm 0.05$). As a consequence, the links from the CGTI and W1 to EHF disappear (though all other links are left unchanged). This suggests that the dynamical mechanisms that bring abundant rainfall to this region are different from those that determine the MT rainfall (see SI, Fig. S21).

4. Discussion and conclusions

In this study, we apply causal discovery algorithms to analyse the influence of global middle and upper tropospheric fields on ~~the~~weekly ISM rainfall and study the two-way causal links between the mid-latitude circulation and ISM rainfall, together with tropical drivers and ISM intraseasonal variability. We perform a validation of both the monsoon-circumglobal teleconnection hypothesis proposed by DW2005 and the North Atlantic-monsoon connection proposed by DW2007 using causal discovery tools. We use the Response-guided Causal precursor detection (RG-CPD) scheme to detect causal precursors from both mid-latitude and tropical regions and then apply Causal Effect Networks (CEN) to assess the influence of different drivers of MT rainfall and their relative contribution to the MT rainfall ~~sub-seasonal~~intraseasonal variability.

The new findings of this work can be summarized in two main aspects: first, we prove the DW2005 hypothesis from a causal point of view, showing that the hypothesized two-way link between the MT rainfall and the mid-latitude circulation ~~is detected~~can be demonstrated in a causal framework. Second, we quantify the relative importance of a) the mid-latitude circulation, b) the ~~internal dynamics of the convection cell~~ intraseasonal variability and c) the tropical drivers of the ISM circulation, showing a link that connects the Boreal Summer Intraseasonal Oscillation (BSISO) and the Madden-Julian Oscillation (MJO) ~~on~~to the ISM ~~subseasonal~~intraseasonal variability (see Figure 7 and Table S1 in the SI). Moreover, we also show that the circumglobal teleconnection hypothesis by DW2005, initially defined at seasonal/monthly ~~time scale~~timescale, holds also at ~~sub-seasonal~~intraseasonal (weekly) ~~time scale~~timescale and it can be unified with the hypothesis by DW2007, which analysed wave trains propagating from the North Atlantic toward the MT region at 2-week ~~time scale~~timescale. Thus, we argue that the CGTI region and the mid-latitude circulation are important for both the ~~sub-seasonal and interannual variability of the MT rainfall~~intraseasonal and seasonal variability of the MT rainfall. An explanation for this could be that the seasonal/monthly representation of the monsoon-circumglobal teleconnection interaction is actually strongly influenced by mechanisms that act on intraseasonal timescales, similarly to what has been proposed by Stephan et al. (2018), who come to a similar conclusion but analysing the

causal interaction between the Silk Road pattern and the monsoon-desert mechanism. Here, a hint in this direction is provided by the analysis of the monsoon-circumglobal teleconnection link at monthly timescale: while the spatial correlation between the MT rainfall and the Z200 field at lag 0 strongly resembles the circumglobal teleconnection pattern (when monthly averages are analysed), this similarity disappears when Z200 leads the MT rainfall by 1 month (see SI, Fig. S12). Therefore, we suggest that the mechanism responsible for the monsoon-circumglobal teleconnection is acting at a timescale shorter than one month.

~~Our causal analyses confirm the influence of the mid-latitude circulation on MT rainfall via the CGTI, as hypothesized by DW2005. We also confirm that MT rainfall forces the mid-latitude circulation via the CGTI but this link is weaker. Moreover, we use a causal precursor identification tool and find a wave train that emanates from the eastern Atlantic stretching towards India. This wave pattern is visible in geopotential height fields, temperature and precipitation anomalies, and acts on MT rainfall via the CGTI with a 1-2 week lead time, in agreement with the DW2007 hypothesis.~~

Our causal analyses confirm the influence of the mid-latitude circulation on MT rainfall via the CGTI, as hypothesized by DW2005. We also confirm that MT rainfall forces the mid-latitude circulation via the CGTI but this link is weaker (see SI, Fig. S18). We test the sensitivity of the monsoon-circumglobal teleconnection hypothesis to changes of the selected rainfall region, showing the general robustness of the identified causal links. Both all-India rainfall (AIR) and western Himalayan foothills (WHF) rainfall behave consistently with the MT rainfall (see SI, Figs. S20-21). These similarities are likely due to the strong correlation that exists between the MT rainfall and AIR, with the MT rainfall being one of the dominant features of the ISM intraseasonal variability (Krishnan et al., 2000). However, eastern Himalayan foothills (EHF) rainfall is found to behave differently, and does not show any connection with the updraft region identified by W1. This is likely related to the fact that the EHF region receives heavy rainfall amounts during the early stage of a break event and thus it is likely to be affected by different circulation patterns than those that govern the MT rainfall (Vellore et al., 2014). Nevertheless, also in this case, all other causal links in the CEN are retained.

Next to the influence of the MT rainfall, our analysis also shows that the link between the circumglobal teleconnection pattern and the ISM can be seen in the wider perspective of the BSISO variations. Our OLR1 and W1 regions (see Fig. 6a,b) show a north-west/south-east tilted rainfall band that shows great similarities with the BSISO rainfall band (Wang et al., 2018). In light of this relationship, the negative feedback that characterizes the causal links between W1 and the MJO2 (see Fig. 6d) and the MT rainfall and W1 (see Fig. S18 in the SI) can be interpreted as the shift of the ISM system between a break and an active phase and to the north-eastward propagation of BSISO convective anomalies. Therefore, while on the one hand we analyse MT rainfall (which is directly linked to active and break phases at intraseasonal timescales), on the other hand our W1 region also represents the BSISO influence on the monsoon-circumglobal teleconnection mechanism, thus linking the mid-latitude circulation not only to local latent heat release connected to the MT rainfall, but to the wider updraft region related to the South Asian monsoon in its broader definition that closely resembles the BSISO rainfall band. This connection is supported by the fact that the direct link from the MT rainfall to the CGTI disappears when W1 is added to the CEN (see Fig. 6). While an indirect link from the MT rainfall to the CGTI remains via W1, the disappearance of the direct link indicates that the influence of W1 on the CGTI is stronger than the direct link from the MT rainfall.

Applying RG-CPD, we find a wave train that emanates from the eastern North Atlantic stretching towards India via Europe and western Russia. This wave pattern is visible in geopotential height fields, temperature and

precipitation anomalies, and acts on MT rainfall via the CGTI with a 1-2 week lead time, in agreement with the DW2007 hypothesis and with previous studies showing that there is a connection from the North Atlantic toward the ISM system (Goswami et al., 2006). Moreover, the larger CGTI region as defined in Fig. 4c, while showing the strongest correlation over the CGTI region as defined in Fig. 2d, also stretches south-westwards towards North-east Africa, to the area that features the downdraft related to the monsoon-desert mechanism (Rodwell and Hoskins, 1996). Therefore, our work is in agreement with previous findings that show an influence of the ISM latent heat release on North-east Africa arid conditions via the excitation of Rossby waves to the west. However, here we focus on relationship between the MT rainfall and the circumglobal teleconnection pattern, thus a link from the MT rainfall towards the Saharan region cannot be inferred from this analysis. Moreover, the causal relationship between the north-eastern Indian rainfall and the downdraft over the north-eastern Sahara related to the monsoon-desert mechanism has already been shown (Stephan et al., 2019).

Our results show that RG-CPD can detect ~~the~~ well-known circulation features of the MT rainfall with 1-week lead time, without using any a priori theoretical or geographical constraint to select the causal precursors among all precursor regions demonstrating the efficacy of the presented method. Moreover, causal discovery tools can quantify the causal influence of tropical drivers versus ~~extra-tropical~~mid-latitude influences and ~~internal dynamics of~~BSISO on the ISM ~~sub-seasonal~~intraseasonal circulation dynamics.

~~Adding internal dynamics helps to further understand the mechanisms that underlie this mid-latitude-ISM relationship. Internal variability of the ISM system has the strongest effect on MT rainfall. The influence from the extratropics on MT rainfall, as mitigated by CGTI, is about half of the magnitude of that of internal dynamics, while the influence of MJO as the key tropical driver on MT rainfall is only slightly weaker (Fig. 7). However, when taking MT rainfall and vertical wind field over the Indian subcontinent together as two interdependent yet different facets of the ISM, we find that the general effect of tropical drivers on the system is slightly stronger than that of the extratropical drivers, while looking on the one hand on the effect of MJO on the circulation and on the other hand on the total effect of CGTI on MT rainfall via both, directed linkages and through a parallel influence on the vertical wind field over India.~~

Intraseasonal variability of the ISM system, here represented by the updraft region identified by W1 has the strongest effect on MT rainfall. The influence from the mid-latitudes on MT rainfall, as mediated by CGTI, is about half of the magnitude of that of the internal dynamics, while the influence of MJO as the key tropical driver on MT rainfall is only slightly weaker (Fig. 7). However, when taking MT rainfall and vertical wind field over the Indian subcontinent together as two interdependent yet different facets of the ISM, we find that the general effect of tropical drivers (MJO) on the system is slightly stronger than that of the extratropical drivers, while looking on the one hand on the effect of MJO on the circulation and on the other hand on the total effect of CGTI on MT rainfall via both, directed linkages and through a parallel influence on the vertical wind field over India. Though in this framework a direct influence of higher latitudes on the MT rainfall it is not identified, this may depend on the choice of the analysed (intraseasonal) timescale. However, an influence from the Arctic on the ISM rainfall has been identified at longer (inter-seasonal) timescales and has shown to provide some forecast skill for seasonal all-India rainfall at 4- and 2-month lead times (Rajeevan et al., 2007; Di Capua et al., 2019;).

~~T~~he reported findings are in good agreement with the existing literature. It is well known that ~~internal~~intraseasonal variability dominates ISM inter-annual variability ~~(Goswami and Xavier, 2005).~~(Goswami and Xavier, 2005; Suhas et al., 2012). Our causal approach enables us to quantify the relative importance of local internal dynamics, separate it from the influence of remote actors, and remove spurious factors. The negative

feedback in the ISM ~~internal~~intraseasonal variability, here represented by the opposite relationship between MT rainfall and W1 (see Figs. 6d and 7), ~~further supports the hypothesis that the internal dynamics of the ISM itself provides a mechanism to~~features a switch from an active to a break phase~~. The~~ inside the ISM system, likely linked to the BSISO. Other physical ~~mechanism~~mechanisms can include both radiative effects (Krishnamurti and Bhalme, 1976) and local changes in static stability due to the latent heat release that follows convective precipitation ~~(Saha et al., 2012).~~(Saha et al., 2012). While strong upward motions precede strong MT rainfall, enhanced rainfall over the ~~ITCZ~~W1 region is known to lead the initiation of breaks by 7-10 days (Krishnan et al., 2000). Moreover, suppressed convection over the Bay of Bengal initiated over the tropical Indian Ocean and associated westward propagating Rossby waves ~~bring~~cause break conditions over the monsoon trough (Krishnan et al., 2000).

M~~oreover, our~~ Our results also support previous findings that suggest a link between the active MJO phases (3-6, corresponding to positive RMM1 values) and enhanced ascending motions over the MT region and adjacent Maritime Continent which in turn promote enhanced MT rainfall (Fig. 6d and 7) (~~Anandh~~Pai et al., ~~2018~~2011; Mishra et al., 2017; ~~Pai~~Anandh et al., ~~2011~~2018).

Field Code Changed

Our theory-guided causal effect network approach, i.e. creating CENs starting from physical hypotheses, enables us to: (1) test those hypotheses in a causal framework, removing the influence from spurious correlations, and (2) quantify the relative strength (i.e.~~.~~, the causal effect) of different local and remote actors. With this approach, one can gain insight in which role each part of a complex system such as the ISM circulation plays in relation to the other components. However, domain knowledge is essential to be guided by known physical processes and associated timescales. By combining RG-CPD and CEN, one can test initial hypotheses and perform further more explorative causal analyses to identify new features. For example, in this study, we first identify our initial actors based on the literature. Then, we increase the pool of actors by searching for causal precursors using RG-CPD. Finally, we reconstruct a CEN that combines those findings and helps to put them into a broader context. This approach can be applied to both observational data (as done here) and climate model data to validate the underlying processes behind ~~sub-seasonal~~intraseasonal variability, which might pave the way for improved forecasts.

The described identification and quantification of causal dependences is based on linear statistical models between the different considered variables quantified in terms of partial correlations. While such linear models can provide useful approximations of real-world climate processes, there could be cases in which they miss other existing linkages that are not described by linear functional relationships. In turn, extending the present analysis to a fully nonlinear treatment is straightforward but would come on the cost of much higher computational demands, which is why we have restricted ourselves in this work to the linear case. Nevertheless, accounting for possible nonlinearities may add further information on the inferred mechanisms and should therefore be undertaken in future research.

O~~ur~~In conclusion, our results indicate that, on weekly timescales, the strength of the influence from the mid-latitudes on MT rainfall itself is as large as that from the tropics (MJO) but about a factor of two smaller than the ISM ~~internal dynamics.~~intraseasonal variability. However, the tropical (MJO) effect on the associated vertical wind speed over the MT region is larger than that of extratropical drivers on MT rainfall. ~~Related to the confirmed relevance of extratropical drivers for ISM variability at weekly~~While previous studies that have analysed the relationships between the ISM and the mid-latitude circulation have often considered the rainfall over north-western India (Ding and Wang, 2005; Beverley et al., 2019; Stephan et al., 2019), here we take into account the MT rainfall, showing that connection between active and break phases of the ISM (by definition identified over

this region, Krishnan et al., 2000) and the circumglobal teleconnection pattern. The circumglobal teleconnection pattern is an important source of variability for European summer weather, thus improving its representation in seasonal forecasting models could in turn improve seasonal forecast in boreal summer (which generally show lower skill than those for boreal winter) (Beverley et al., 2019). Related to the confirmed relevance of extratropical drivers for ISM variability at weekly time scales, we emphasise that there exists a substantial body of literature suggesting that the influence from the mid-latitudes is particularly important for extremes (Lau and Kim, 2011; Vellore et al., 2014, 2016). Future work should therefore aim to further disentangle the specific mechanisms that particularly act in the context of extremes.

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Red, English (United Kingdom)

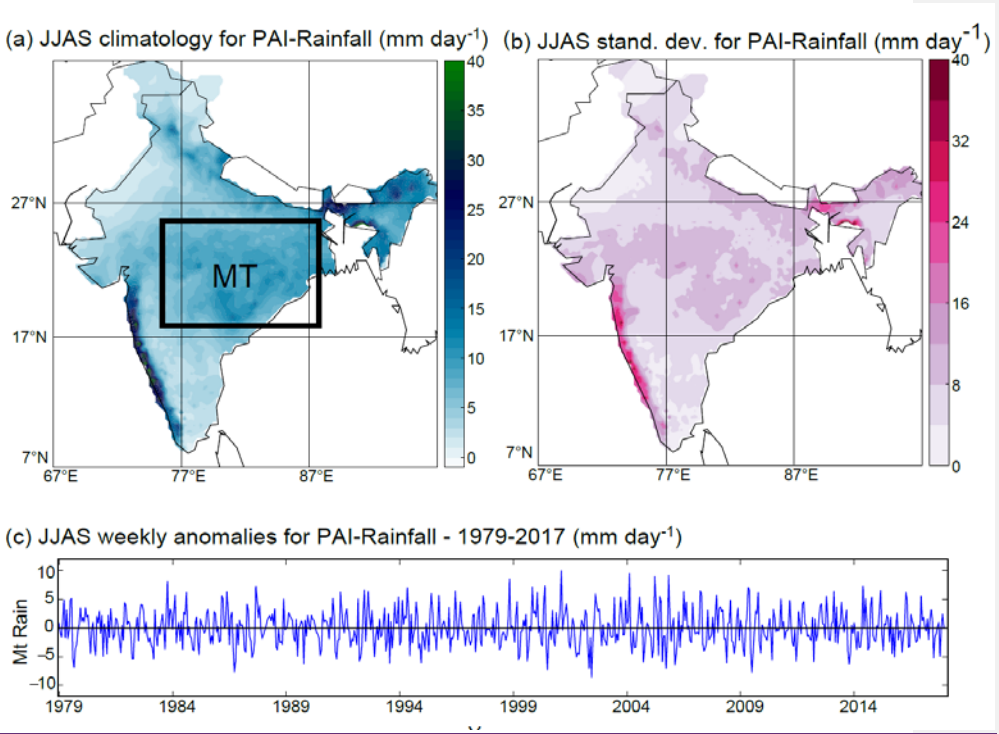


Figure 1. Rainfall climatology over India. Panel (a): JJAS rainfall climatology over the 1979-2017 period from the Pai et al. (2015) dataset. The black box identifies the MT region. Panel (b): standard deviation (s.d.) of weekly JJAS rainfall over the 1979-2017 period from the Pai et al. (2015) dataset. Panel (c): time series of weekly MT JJAS rainfall over the period 1979-2017; each year contains 18 weeks, with the first week starting on the 27th of May.

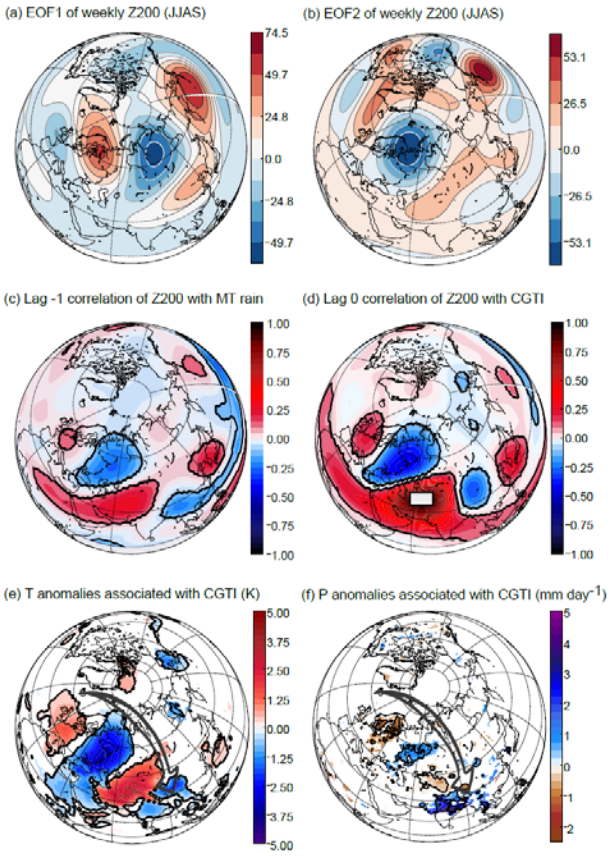


Figure 2. Mid-latitude variability associated with the ISM. Panels (a) and (b): EOF1 and EOF2 expressed as covariance for the JJAS weekly Z200 field in the Northern Hemisphere (0°-90°N, 0°-360°E) for the period 1979-2017. Panel (c): correlation between weekly MT rainfall (lag = 0) and Z200 (lag = -1 week). Panel (d): the CGTI region (white box) and the correlation between CGTI and Z200 (both at lag = 0), which forms the circumglobal teleconnection pattern. In panels (c) and (d), correlation coefficients with a *p*-value of $p < 0.05$ (accounting for the effect of serial correlations) are shown by black contours. Panel (e): Composite temperature difference between weeks with CGTI > 1 $CGTI_{s.d.}$ minus weeks with CGTI < -1 $CGTI_{s.d.}$ over the Northern Hemisphere. Panel (f): as panel (e) but for rainfall anomalies from the CPC-NCEP dataset (as rainfall data for the Pai et al. (2015) dataset are available over India only) for the period 1979-2017. In panels (e) and (f), anomalies with a *p*-value of $p < 0.05$ (accounting for the effect of serial correlations) are shown by black contours, while grid points which are found significant only with non-corrected *p*-values are shaded.

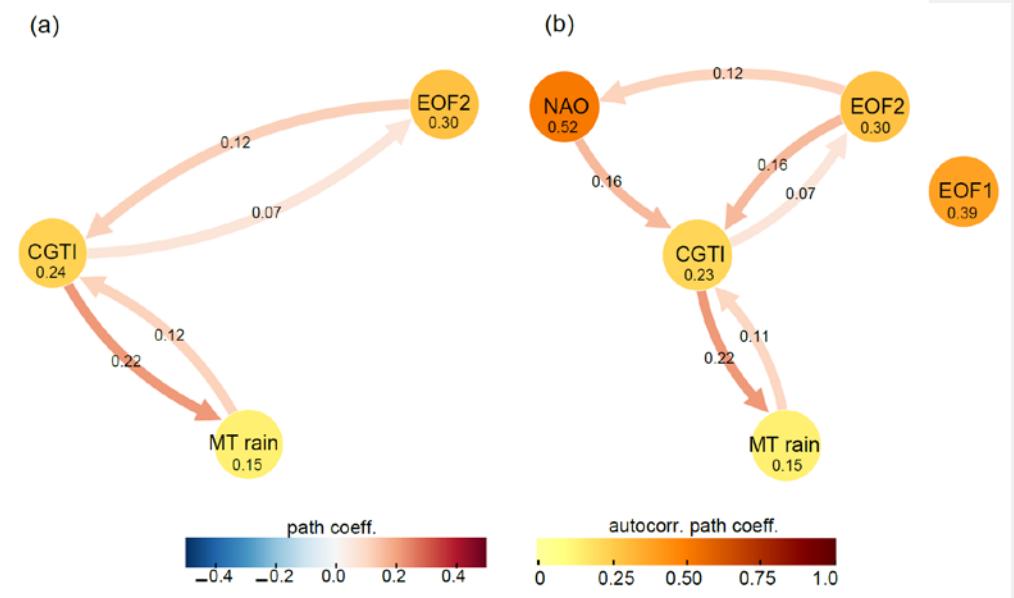


Figure 3. Causal mid-latitude interactions with the ISM. Panel (a): Causal Effect Network (CEN) built with CGTI, the principal component of EOF2 and MT rainfall for the period 1979-2017. Panel (b): Same as panel (a) but with the addition of the principal component of EOF1 and NAO. The strength of the causal links expressed by the standardized regression path coefficients and autocorrelation path coefficients are shown over the arrows and inside the circles respectively. All links have a lag of 1 week. Only causal links with corrected $p < 0.05$ are shown. See the main text for discussion.

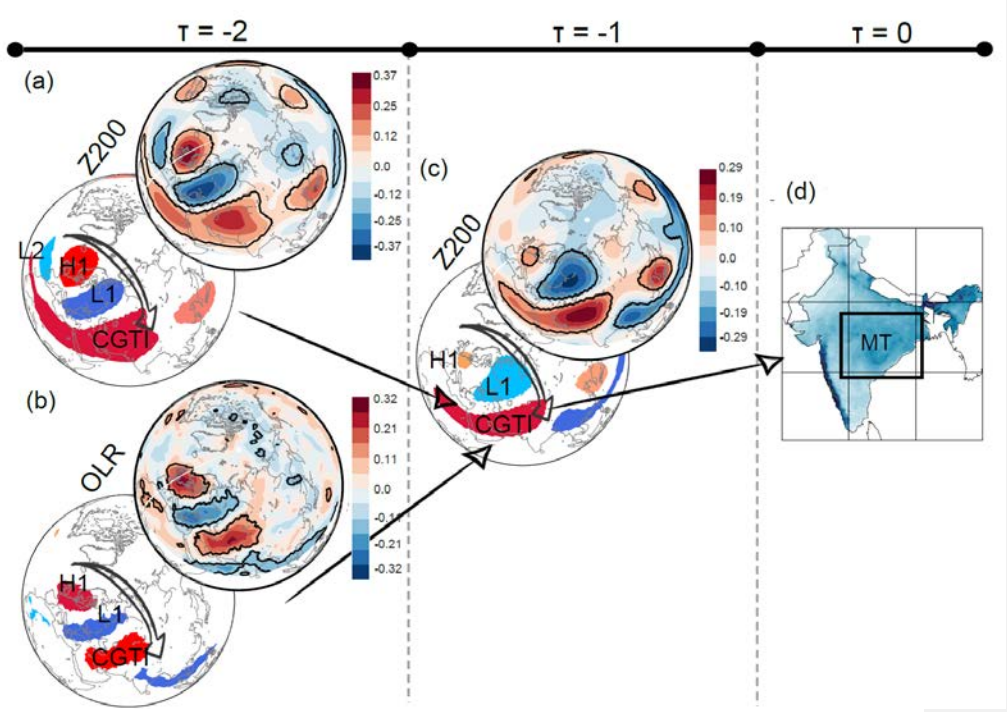


Figure 4. Mid-latitude causal precursors of ISM. Panel (a): correlation of CGTI with Z200 at 1-week lead time (top), and the causal precursors of CGTI identified via RG-CPD (bottom) for the period 1979-2017. Panel (b): as for panel (a) but for OLR fields. Panel (c): correlation map for weekly MT rainfall and Z200 field at 1-week lead time (top) and the causal precursors identified via RG-CPD (bottom). Panel (d): ISM rainfall over the MT region from the Pai et al. (2015) dataset (reproduced from Fig. 1a). Regions with correlation values with a p-value of $p < 0.05$ (accounting for the effect of serial correlations) are shown by black contours.

Formatted: Font: (Default) Times New Roman, 10 pt, Font color: Red, English (United Kingdom)

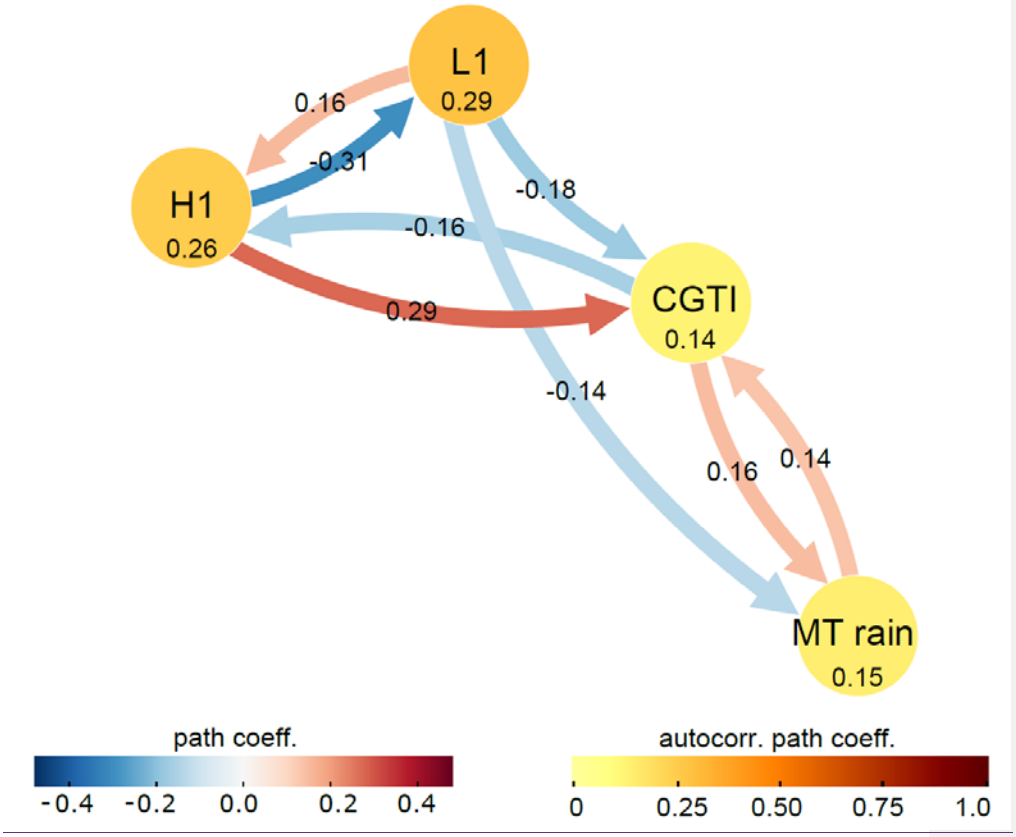


Figure 5. Mid-latitude wave train. CEN built with the MT rainfall, CGTI, L1 and H1 (as identified in Fig. 4a) for the period 1979-2017. The strength of causal links expressed by the standardized regression path coefficients and autocorrelation path coefficients are shown above the arrows and inside the circles respectively. All links have a lag of 1 week. Only causal links with corrected $p < 0.05$ are shown. See the main text for discussion.

Formatted: Font: (Default) Times New Roman, 9 pt, English (United Kingdom)

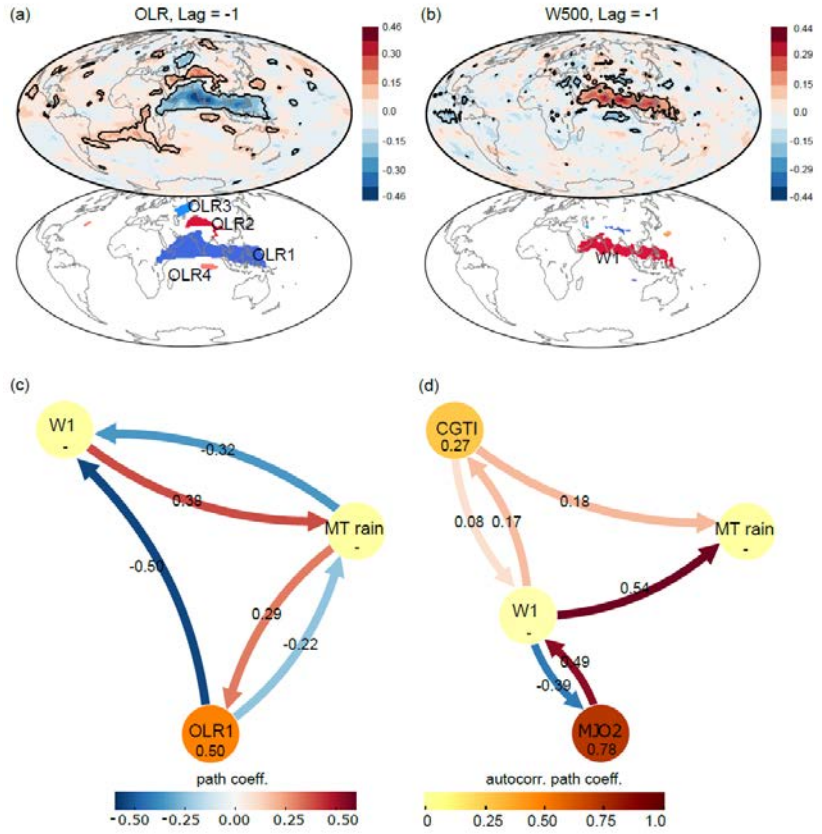


Figure 6. Tropical causal interactions of ISM. Panel (a) shows the correlation map of weekly MT rainfall with the global OLR field at 1-week lead time (top panel) and the causal precursors identified via RG-CPD (bottom panel) for the period 1979-2017. Regions with correlation values with a *p*-value of $p < 0.05$ (accounting for the effect of serial correlations) are shown by black contours. Panel (b): as for panel (a) but for the W500 field. Panels (c) and (d) show the CEN build with W1, OLR1 and MT rainfall and MT rainfall, W1, CGTI and MJO2, respectively. The strengths of causal links expressed by the standardized regression path coefficients and autocorrelation path coefficients are shown above the arrows and inside the circles respectively. All links have a lag of 1 week. Only causal links with corrected $p < 0.05$ are shown. See the main text for discussion.

Formatted: Font: (Default) Times New Roman, 9 pt, English (United Kingdom)

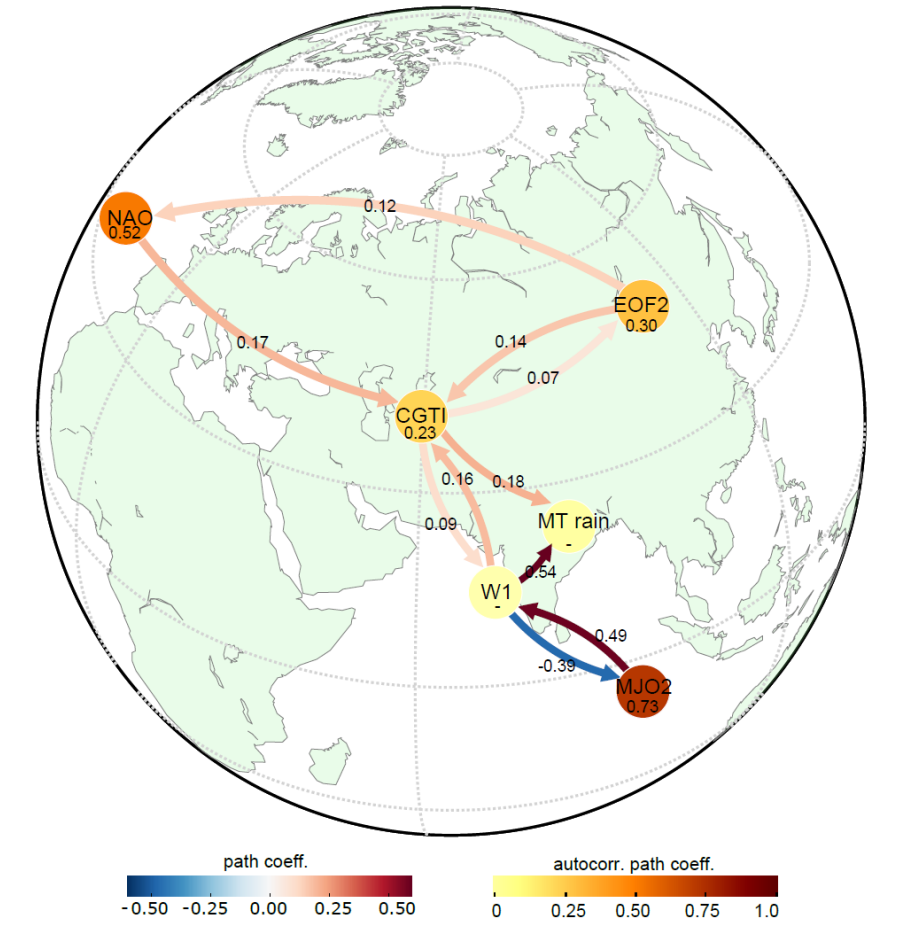


Figure 7. Combined mid-latitude and tropical causal interactions of ISM. CEN built with W1, MJO2, MT rainfall, NAO, CGTI and EOF2 for the period 1979-2017. The strengths of causal links expressed by the standardized regression path coefficients and autocorrelation path coefficients are shown above the arrows and inside the circles respectively. All links have a lag of 1 week. Only causal links with corrected $p < 0.05$ are shown. See the main text for discussion.

Acknowledgments

We thank ECMWF and NCEP for making the ERA-Interim and CPC-NCEP data available. ~~This work was supported by~~G.D.C., D.C., and R.V.D. acknowledge cofunding from the German Federal Ministry of Education and Research (~~projects~~BMBF Grant 01LP1611A) under the auspices of the Belmont Forum and JPI-Climate project GOTHAM~~, Sacre-X and~~ . R.V.D and M.K. acknowledge cofunding from the German Federal Ministry of Education and Research (BMBF Grants 01LN1306A (CoSy-CC^2) and 01LN1304A (Sacre-X). Code for the causal discovery algorithm is freely available as part of the Tigramite

Python software package at ~~https://github.com/jakobrunge/tigramite~~https://github.com/jakobrunge/tigramite. We thank the anonymous reviewers for helping us improving the content of this manuscript and for their encouraging words.

References

Anandh, P. C., Vissa, N. K. and Broderick, C.: Role of MJO in modulating rainfall characteristics observed over India in all seasons utilizing TRMM, Int. J. Climatol., 38(5), 2352–2373, doi:10.1002/joc.5339, 2018.

Bello, G. A., Angus, M., Pedemane, N., Harlalka, J. K., Semazzi, F. H. M., Kumar, V. and Samatova, N. F.: Response-Guided Community Detection: Application to Climate Index Discovery, in Machine Learning and Knowledge Discovery in Databases, pp. 736–751., 2015.

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.

Benjamini, Y. and Yekutieli, D.: The control of the false discovery rate in multiple testing under dependency, Ann. Stat., 29(4), 1165–1188, doi:10.1214/aos/1013699998, 2001.

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.

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.

Chattopadhyay, R., Goswami, B. N., Sahai, A. K. and Fraedrich, K.: Role of stratiform rainfall in modifying the northward propagation of monsoon intraseasonal oscillation, J. Geophys. Res., 114, D19114, doi:10.1029/2009JD011869, 2009.

Chen, M., Shi, W., Xie, P., Silva, V. B. S., Kousky, V. E., Higgins, R. W. and Janowiak, J. E.: Assessing objective techniques for gauge-based analyses of global daily precipitation, J. Geophys. Res. Atmos., 113(4), 1–13, doi:10.1029/2007JD009132, 2008.

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.

Cherchi, A., Annamalai, H., Masina, S., Navarra, A. and Alessandri, A.: Twenty-first century projected summer mean climate in the Mediterranean interpreted through the monsoon-desert mechanism, Clim. Dyn., 47(7–8), 2361–2371, doi:10.1007/s00382-015-2968-4, 2016.

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.

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. and Xavier, P. K.: Dynamics of "internal" interannual variability of the Indian summer monsoon in a GCM, J. Geophys. Res., 110, D24104, doi:10.1029/2005JD006042, 2005.

Goswami, B. N., Madhusoodanan, M. S., Neema, C. P. and Sengupta, D.: A physical mechanism for North Atlantic SST influence on the Indian summer monsoon, Geophys. Res. Lett., 33(2), 1–4, doi:10.1029/2005GL024803, 2006.

Hannachi, a., Jolliffe, I. T. and Stephenson, D. B.: Empirical orthogonal functions and related techniques in atmospheric science: A review, ,Int. J. Climatol., 713(November 2006), 697–713, doi:10.1002/joc, 2007.

Houze Jr, R. A., McMurdie, L. A., Rasmussen, K. L., Kumar, A. and Chaplin, M. M.: Multiscale Aspects of the Storm Producing the June 2013 Flooding in Uttarakhand, India, Mon. Weather Rev., 145, 4447–4466, doi:10.1175/mwr-d-17-0004.1, 2017.

Kiladis, G. N., Dias, J., Straub, K. H., Wheeler, M. C., Tulich, S. N., Kikuchi, K., Weickmann, K. M. and Ventrice, M. J.: A Comparison of OLR and Circulation-Based Indices for Tracking the MJO, Mon. Weather Rev., 142, 1697–1715, doi:10.1175/MWR-D-13-00301.1, 2014.

Kretschmer, M., Coumou, D., Donges, J. F.,. and Runge, J. and Donges, J. F.: 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., Coumou, D., Agel, L., Barlow, M., Tziperman, E. and Cohen, J.: More-persistent weak stratospheric polar vortex States linked to cold extremes, Bull. Am. Meteorol. Soc., 99(1), 49–60, doi:10.1175/BAMS-D-16-0259.1, 2018.

Kripalani, R. H. Ã., Kulkarni, A. and Singh, S. V.:.: Association of the Indian Summer Monsoonsummer monsoon with the Northern Hemisphere Mid-Latitude Circulationnorthern 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.

Krishnamurti, T. N. and Bhalme, H. H.: Oscillations of a monsoon system Part I: Observational apects, J. Atmos. Sci., 33, 1937–1954, doi:10.1175/1520-0469(1976)033<1937:OOAMSP>2.0.CO;2, 1976.

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.

Krishnan, R., Kumar, V., Sugi, M. and Yoshimura, J.: Internal Feedbacks from Monsoon – Midlatitude Interactions during Droughts in the Indian Summer Monsoon, J. Atmos. Sci., 66, 553–578, doi:10.1175/2008JAS2723.1, 2009.

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.

Levermann, A., Schewe, J., Petoukhov, V. and Held, H.: Basic mechanism for abrupt monsoon transitions ~~Anders~~, Proc. Natl. Acad. Sci., 106(49), 20572–20577, doi:10.1073/pnas.0901414106, 2009.

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.

Mishra, S. K., Sahany, S. and Salunke, P.: Linkages between MJO and summer monsoon rainfall over India and surrounding region, Meteorol. Atmos. Phys., 129(3), 283–296, doi:10.1007/s00703-016-0470-0, 2017.

North, G. R., Bell, T. L., Cahalan, R. F. and Moeng, F. J.: Sampling Errors in the Estimation of empirical Orthogonal Functions, Mon. Weather Rev., 110, 699–706, doi:https://doi.org/10.1175/1520-0493(1982)110<0699:SEITEO>2.0.CO;2, 1982.

Pai, D. S., Bhate, J., Sreejith, O. P. and Hatwar, H. R.: Impact of MJO on the intraseasonal variation of summer monsoon rainfall over India, Clim. Dyn., 36(1), 41–55, doi:10.1007/s00382-009-0634-4, 2011.

Pai, D. S., Latha, D. S. P., Badwaik, S. M. R. and Rajeevan, M.: Analysis of the daily rainfall events over India using a new long period (1901–2010) high resolution (0.25° × 0.25°) gridded rainfall data set, Clim. Dyn., 45, 755–776, doi:10.1007/s00382-014-2307-1, 2015.

Pathak, A., Ghosh, S., Kumar, P. and Murtugudde, R.: Role of Oceanic and Terrestrial Atmospheric Moisture Sources in Intraseasonal Variability of Indian Summer Monsoon Rainfall, Sci. Rep., 7(1), 12729, doi:10.1038/s41598-017-13115-7, 2017.

Rajeevan, M., Pai, D. S., Anil Kumar, R. and Lal, B.: New statistical models for long-range forecasting of southwest monsoon rainfall over India, Clim. Dyn., 28(7–8), 813–828, doi:10.1007/s00382-006-0197-6, 2007.

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

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.

~~R~~unge, J., ~~Heitzig, J., Marwan, N. and Kurths, J.: Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy, Phys. Rev. E, 86(6), 061121, doi:10.1103/PhysRevE.86.061121, 2012.~~

R~~unge, 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~~2015.

~~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., Sejdinovic, D. and Flaxman, S.: Detecting causal associations in large nonlinear time series datasets, ,

doi:arXiv:1702.07007, 2017.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. and Sejdinovic, D.: Detecting causal associations in large nonlinear time series datasets, Sci. Adv., 5, eaau4996 [online] Available from: http://arxiv.org/abs/1702.07007, 2019.

Saeed, S., Müller, W. A., Hagemann, S., Jacob, D., Mujumdar, M. and Krishnan, R.: Precipitation variability over the South Asian monsoon heat low and associated teleconnections, Geophys. Res. Lett., 38, L08702, doi:10.1029/2011GL046984, 2011.

Saggioro, E. and Shepherd, T. G.: Quantifying the timescale and strength of Southern Hemisphere intra-seasonal stratosphere-troposphere coupling, Geophys. Res. Lett., 2019GL084763, doi:10.1029/2019GL084763, 2019.

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.

Schleussner, C. F., Runge, J., Lehmann, J. and Levermann, ~~a~~A.: The role of the North Atlantic overturning and deep ocean for multi-decadal global-mean-temperature variability, Earth Syst. Dyn., 5(1), 103–115, doi:10.5194/esd-5-103-2014, 2014.

Shige, S., Nakano, Y. and Yamamoto, M. K.: Role of Orography , Diurnal Cycle , and Intraseasonal Oscillation in Summer Monsoon Rainfall over the Western Ghats and Myanmar Coast, J. Clim., 30, 9365–9381, doi:10.1175/JCLI-D-16-0858.1, 2017.

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

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.

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., Webster, P., Kikuchi, K., Yasunari, T. and Qi, Y.: Boreal summer quasi-monthly oscillation in the global tropics, Clim. Dyn., 16, 661–675, doi:10.1007/s00382-006-0163-3, 2006.

Wang, H., Wang, B., Huang, F., Ding, Q. and Lee, J. Y.: Interdecadal change of the boreal summer circumglobal teleconnection (1958-2010), Geophys. Res. Lett., 39(12), 1–6, doi:10.1029/2012GL052371, 2012.

Wang, S., Ma, D., Sobel, A. H. and Tippett, M. K.: Propagation Characteristics of BSISO Indices, Geophys. Res. Lett., 45(18), 9934–9943, doi:10.1029/2018GL078321, 2018.

Webster, P. J. and Lawrence, D. M.: The Boreal Summer Intraseasonal Oscillation: Relationship between Northward and Eastward Movement of Convection, J. Atmos. Sci., 59(9), 1593–1606, doi:10.1175/1520-0469(2002)059<1593:TBSIOR>2.0.CO;2, 2002.

Wheeler, M. C. and Hendon, H. H.: An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction, Mon. Weather Rev., 132(8), 1917–1932, doi:10.1175/1520-

0493(2004)132<1917:AARMMI>2.0.CO;2, 2004.

Wilks, D. S.: The Stippling Shows Statistically Significant Grid Points, Bull. Am. Meteorol. Soc., 97(December), 2263–2274, doi:10.1175/BAMS-D-15-00267.1, 2016.

Willink, D., Khan, V. and Donner, R. V.: Improved one-month lead-time forecasting of the SPI over Russia with pressure covariates based on the SL–AV model, Q. J. R. Meteorol. Soc., 143(707), 2636–2649, doi:10.1002/qj.3114, 2017.

Yanai, M. and Wu, G.: Effects of the tibetan plateau, in The Asian Monsoon, pp. 513–549, Springer., 2006.

Zhang, C.: Madden-Julian Oscillation, Rev. Geophys., 43, RG2003, doi:10.1029/2004RG000158, 2005.