Earth Syst. Dynam. Discuss.,
https://doi.org/10.5194/esd-2019-42-AC2, 2019

Earth System Dynamics Discussions

Open Access

Interactive comment on "Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: A Theory-Guided Causal Effect Network approach" *by* Giorgia Di Capua et al.

Giorgia Di Capua et al.

dicapua@pik-potsdam.de

Received and published: 12 November 2019

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

Printer-friendly version

Discussion paper

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. Please refer to the attachment for the new Figures cited 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. Fol-

Printer-friendly version

Discussion paper

lowing 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, character-

ized 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 ($r = 0.57 \pm 0.05$), 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 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

Printer-friendly version

Discussion paper

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 PËĘ{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.

Please also note the supplement to this comment:
https://www.earth-syst-dynam-discuss.net/esd-2019-42/esd-2019-42-AC2-supplement.pdf

---

Interactive comment on Earth Syst. Dynam. Discuss., https://doi.org/10.5194/esd-2019-42, 2019.