General comments.

The manuscript addresses the relationship and causal connection between spring North Atlantic sea surface temperature and the summer East Atlantic pattern in reanalysis and MPI-ESM-MR model simulations during the 20th century, which fits the scope of the WCD journal. This study uses correlation and linear regression maps to primarily estimate the relationships among various variables, and then uses causal discovery to detect cause-effect relationships. First, such analysis is performed for ERA-20C reanalysis, and then the results are compared to MPI-ESM-MR simulations. In my opinion, the authors address the novelty of the application of causal discovery in their research area, and clearly present their results in clear and fluent language, however, several major issues need to be addressed.

Methodology. The authors provide a brief overview of the used methods, in particular causality, via referencing Runge et al., 2015, Kretschmer et al.,2016, Di Capua et al., 2020b. The authors mention the "causal inference-based tool", but it is not clear which tool is meant. The authors do not mention particular settings used in the causal discovery algorithm, such as τ_{min}, τ_{max}, pc_alpha, alpha_level, conditional independence test etc. Without these parameters, it is very hard to reproduce particular results.

While, for example, in L193 the authors indicate that their analysis focuses on 3 and 4 months lag only (guessing τ_{min}=3, τ_{max} =4), it is not clear from the context of the paper why contemporaneous links were not included.

The authors used different variables for reanalysis and model simulation to construct their causal graphs/CENs and then compared outcomes. I address this issue in more detail in "Specific comments", but I highly recommend that the authors stay consistent in their analysis, especially in drawing proper conclusions.

Naming of variables. The authors use SST to indicate the North Atlantic extratropical surface temperate (P1L2) or SST index (P1L4) for the meridional SST gradient over the North Atlantic region. I highly recommend that the authors use the SST term to indicate SST and include a geographical indicator to their variable, e.g. NA-SST, when they refer to the North Atlantic extratropical surface temperate.

The term "causal associations", apart from the title, is used twice in the manuscript. I suggest the terms "relationships/links/connections".

Specific comments.

Abstract.

P1L6 "We only find this link to be causal, however, during the period 1958 - 2008." → the authors did not mention the analyzed period for ERA-20C reanalysis therefore "however" in L6 sounds odd. Moreover, later in the manuscript the authors provide the reasoning and explanations why there are no detected causal links for specific periods. Therefore, I suggest to keep the sentence simple "We find the causal link during 1958-2008".

P1L9 "we find that spring SST" → do the authors mean spring North Atlantic SST? "NA-SST" would be helpful to follow the text. In P1L4 the authors introduced the SST index, so do the authors talk about SST index here as well?

P1L14 What is the meaning of "moderately" here?

P1L13 "We find that while MPI-ESM-MR…"→ both pre-industrial and historical simulations?

Introduction.

P2L26 What is this causal inference-based tool?

P2L40 SST is already introduced in P2L27. I refer the authors to my comment to improve the notation of the North Atlantic extratropical surface temperate instead of simply using SST.

P3L55 Runge et al., 2015 does not use term “Causal Effect Networks”. It was Kretchmer et al., who was one of the first authors to use this term.

P3L58 regarding overcoming spurious correlations: see also Runge et al., 2014; Runge, Bathiany, et al., 2019.

P3L59 please add more examples for the application of causal discovery for other teleconnections. For example, Atlantic-Pacific teleconnections: Karmouche et al., 2023; Arctic-midlatitude teleconnections: Kretschmer et al., 2020, Siew et al., 2020; Galytska et al., 2023, marine cold-air outbreaks: Polkova et al., 2021, Walker circulation: Runge, Bathiany, et al., 2019 and others.

Methodology.

P3L73 Here SST stands just for Sea Surface Temperature. That is confusing with the North Atlantic SST introduced before.

L106 EOF is already introduced. The calculation of EA index is already mentioned in the Introduction.

L114. The reference to Fig. 1b comes before Fig. 1a. What is the meaning of the colors in Fig. 1b? Please fix this. Generally, I suggest that the authors insert the Figure and/or Table, which would summarize/show the indices used in their studies and explain them in Sect. 2.2. In that case panels c-f from Fig. 1 would solely address the discussions from Sect. 3.1

L117 The authors explain how they calculated an SLP index, however the introductions lacks the explanation of the impact and significance of the pressure over this region on East Atlantic Pattern and North Atlantic SST.

L121 “wherever useful”→ where exactly?

L127 Cite Spirtes et al., 2000 for PC part of the algorithm.

L134 “circles” → nodes?

L137 The authors should also explain the meaning of the color of the nodes.

L141-142 I assume the authors used Tigramite for their research. I find it important that the authors stay transparent on the software that they used as well as the settings that were applied, τ_{max}, τ_{min}, pc_alpha, alpha_level. I am wondering if the authors already tested PCMCI+ algorithm for their study, which is able to capture causal contemporaneous connections. Why authors did not include contemporaneous connections?

Results

P6L146-147 If the authors follow my suggestion and introduce a separate Figure for Sect. 2.2 and summarize the used indices, I suggest to move this sentence to Sect. 2.2.

P6L147-149 Fig. 1 c and d do not show “below average temperature/precipitation”, they show correlation between EA-index and temperature/precipitation, which is associated with below-average temperatures/precipitation over this region.

P6L149-150 This sentence should be moved to Sect. 2.2. The simple usage of term “SST index” is very confusing.

P6L153 “correlation reaches significant values” → what is the definition of significance here? Did the authors use significance test here?

Figure 2. Caption. I suggest: “Distinct spatial characteristics of the spring SST influence on the summer circulation over the 20th century (for ERA-20C) for early (1908-1957, left), late (1958-2008, middle) and full periods (1908-2008, right column). …” …”Box in Fig.2i illustrates…”. To label y-axes the authors use “Influence of AM SST index”, which implies the direction of the impact, however by this point it was not yet discussed. I suggest terms “relationship/connection” instead. Another minor suggestion: if the authors draw coastlines in gray, then SLP contours in panels d-f could be plotted in black and become more visible. Further, since Fig. 5 shows only correlations (models vs reanalysis), I suggest that the authors similarly restructure the panels in Fig. 2, e.g. first two rows showing correlations, the third raw showing regression.

L166 Is it about panels a and b?

L166-167 Not shown in this paper? Or which Figure/panels?

L171 But not in panel d?

L173-174 In which geographical region?

P9L191 “correlation” → anticorrelation

L193 I assume $\tau_{min} = 3, \tau_{max} = 4$.

L200-201. In regard to the full periods, the authors already stated that the spring SST -summer EA relationship is nonstationary for the full period, which would lead to the non-realistic causal graph/CEN (also see Runge 2018).

L208 “actor” → variable. The authors did not define the term “actor”, but used term “variable” throughout the manuscript.

L208 excluded → ”not included”.

Figure 3. Given that panel a consists of two subfigures that aim to show the same result, I strongly recommend that the authors combine them into one figure. For example, in plot_graph function, the authors can use node_pos.

Figure 5. Since panels g-h are the same as Fig. 2a-c, the authors can either remove these panels completely, or show the differences instead.

P9 L222 I do not recommend to comparing causal graphs/CEN between reanalysis and model simulations while using different sets of variables. The assumption of causal sufficiency says that *Measured variables include all of the common causes.* For the reanalysis, the authors motivated the usage of 2m Temperature. It is important to explain, why authors did not include this variable for model simulations. For example, the correlation of AM SST index and 2m Temperature is not that strong in the model. But this might lead to the other discussion, e.g, low correlation could be a consequence of different state(s) of AM SST index in the model vs reanalysis. Currently the results from Fig3a and Fig.6a,c are not directly comparable.

P9 L222-223 Here the authors indicate that for the model simulations the time lag was 3 and 4 months, however Fig. 6a also shows two causal lagged links with 2 months delay. Please, clarify this.

P10 L224 “suggesting an atmospheric forcing into ocean” → please, also specify exact links. This will improve readability and comparison with the results from Fig. 6. Same for L225.

P10 L227-230 I would like to highlight that differences between reanalysis and model simulations were expected, since the authors already showed that the timeseries (Fig.4) as well as spatial patterns differ quite a lot between different data sources. I would appreciate if the authors addressed this issue here (as well as in the Discussion).

L232 There is no Fig. 6e

Figure 6 a and c. Node “EAP”→ EA.

Figure 7a. It is not clear why the authors did not keep SLP-ind in the causal graph? Is there a reason why the colormap for the β -coef. is changed?

Discussion

P15LL288 For which time period?

Conclusions

P16 L323 “we find that”… but in L303 the authors state that the nonstationarity has been previously reported by other studies.

L326 Please, specify if this conclusion is for the reanalysis. Same for L330.

L328 “that an external physical mechanism not included in our analysis” → the usage of PCMCI implies that the user makes an assumption that all variables are included to represent a specific mechanism. If user does assume that all variables are included in the causal graph/CEN, then LPCMCI should be rather used.

L333. Please, also summarize why the model does not reproduce the links from the reanalysis.

L334 “weakly”→ weak

L341 “limited performance in reproducing”→ I suggest to highlight the reasons while giving such a statement.