Review of "Causal associations and predictability of the summer East Atlantic teleconnection" by Carvalho-Oliveira et al.

The paper adopts the quite novel technique of Causal Effect Networks (CEN) to analyze the link between spring SSTs in the North Atlantic (northwest-southeast difference) with the East Atlantic (EA) pattern in SLP in summer. This is done both for the ERA20C reanalysis and for the MPI model, in various configurations. Moreover, it is analyzed whether the representation of such causal link influences the model forecast skill on a specific region.

The work is well framed in the introduction and the proposed CEN also quite well justified in terms of possible physical pathways in the introduction, though not much in the rest of the paper. The methodology is quite complex and involves a large amount of work for a challenging analysis. The results are interesting, but I think they could be presented in a clearer way, at least in some parts. Also, the nomenclature is quite complex to follow due to the large number of variables, experiments, techniques considered: some more effort on this issue would help the reader better grasp the main results of the work.

Main issues, mostly related with the presentation of the results:

- Period. MA, AM, MAM? It is quite difficult to follow the various spring selections and to understand why one is chosen with respect to the other. As long as I understood, most results refer to AM, but the other seasons are cited here and there. I think this is quite confounding for the reader, and a single spring window would help, maybe just stating that changing it does not change the results.
- The methods section 2.2 could be expanded to include a more detailed description of the bootstrap ensemble, which is currently just in the main text. It is unclear how the authors refer to the bootstrap ensemble throughout the text. Linked to this, the term "causal ensemble" or "causal timeseries" is not explained in the text.

  Also, since many indexes are considered in the text, a point-by-point list of all indexes considered with a unique and identifying name would help. For example the SST_Ridge just appears at the end of the results section, but would be useful to have it here for quick reference.
- The end of the results section could possibly make for a separate subsection: see comment at lines 256-281. I think this is one of the most interesting results of the paper, but is currently difficult to grasp and could be easily lost in the main text.
- Physical pathway. The CEN framework is a powerful tool but should be used with caution. In particular two things are necessary, following Kretschmer et al. (2021):
  - the CEN should only be used to "measure" a causal link for which there is already an hypothesized physical pathway;

- the causality of the link is always conditional to the choice of the variables included in the model, meaning that if a relevant variable is missing, the CEN result may be wrong.

I think the authors respect both "requirements", but I suggest to:

- recall the possible physical mechanism behind the link when presenting the CEN. It is currently only cited in the introduction.
- put more emphasis and discuss on the possible impact of a missing process in the CEN.

Specific points.

L94. The ocean state is derived ~~in~~ *from?*

L94-97. It is not completely clear to me how the assimilation experiment is performed for the ocean. If is said at line 94 that an ocean-only simulation with MPI-OM forced with ERA-20C is performed. Is this ocean state then used for the nudging of the 30 ensemble members? I think this choice should be (quickly) motivated in the text.

L102. ..at lead times *of* 3-4 months..

L106. I would say: "... the second principal component (PC) of the ~~leading~~ empirical orthogonal function (EOF) *decomposition* of ..."

L131. Since the technique is pretty novel and can easily generate misunderstandings, I suggest the authors to add some further disclaimer for the reader. In particular, as implicitly stated a few lines above, a limitation of the Causal effect network analysis is that the choice of the variables to be considered is crucial for determining the causality of the link. In this sense, the possibility of a spurious correlation can never be completely excluded. I think this should be made clear in the text, expanding the sentence at L131.

Figure 1 caption. Specify the period considered for the SST anomaly. From the text it is spring SSTs and summer EA (L151), but this is not clear for the caption. Also, it is not clear what period panels c, d, e are referring to. In general, since different periods are considered for different variables, a suggestion would be to put the period as a subscript: SST_{MA}.

L151. Linked to the above. Is the spring SST referred to MAM, MA or AM?

L166. Why not use MAM? Is there some process changing significantly between early and late spring?

L200. The $T2m_{CE}$ looks like potentially correlated with EA with no lag (from the composite in Fig. 1).. isn't this correlation appearing in the CEN?

L207-8. I appreciate that the author acknowledge that something could be missing from the CEN. Also, could this mean that the observed causal link may be to some extent spurious, since some key process is missing from the CEN?

L212. I wouldn't call this "skill in reproducing the summer EA" since Fig. 4a is evaluating the pdf of the EA index, so just from a statistical/climatological point of view.

L216. In what sense is Fig. 4b showing "MR-30 capturing the temporal variability of the relationship in the early period" ? I think the only information is about the spread of the relationship, but I do not see a tendency for a negative correlation in the early period as observed for ERA20C.

L221. I agree MR-30 looks slightly better for the early period, but still is quite far from ERA20C (the positive correlation in the southern North Atlantic is not significant and does not extend so much North). This is true for the ensemble mean, but have you checked whether some individual member is getting a response closer to the observed relation? This could possibly inform on the "missing" process in the chain.

L222. I do not understand "first" in the sentence.

L228. 0.03 seems very small. Is it significant?

L245. What do you mean by "MR-30 causal timeseries"? If referred to the "MR-30 causal ensemble" in Fig. 6 caption, I would change the wording to something different.. e.g. MR-30 bootstrap ensemble

L245. How do you perform the "predictive skill assessment"? By selecting random members with beta close to beta_1 and checking the skill only for those?

L249. How rare is this? The information might be relevant.

L257. I can't easily see the contours in Fig. 7b. Would be better a separate figure, or black contours.

L256-L281. I would separate this last part, since it is focussing on a different topic: how does the existence of a causal link between SST spring and T2m Ridge (a different predictand from the rest of the paper) influence the forecast skill on the Ridge region in the random MR-30 ensemble?

Also, I had a hard time following the section, which I think should be separated from the rest, better framed and possibly expanded. I say this because the result looks interesting but is quite difficult to grasp from the current text. Also, the choice of the new predictand might look like a cherry pick, but I think it could be better motivated with the fact that it is the only causal link reproduced in the random MR-30 ensemble. The question "what happens to the skill when a causal link is reproduced?" seems relevant.

L312-316. This part looks a bit technical for the discussion, I suggest to remove it.