Review of "Regime-oriented causal model evaluation of Atlantic-Pacific teleconnections in CMIP6" by Karmouche et al.:

The authors utilize the PCMCI causality approach to investigate the interanual Atlantic-Pacific teleconnections, then they compare the causal relations in observations with that in CMIP6 model simulations. I think that there are two advantages of this work: (1) Analyzing interanual Atlantic-Pacific teleconnections is an important task in climate science, and it is novel to include the state-of-art PCMCI causality method. Additionally, when interpreting the causal detection results, the authors cite and combine enough previous studies to explain the mechanisms within Atlantic-Pacific teleconnections. (2) The authors present an example for using the PCMCI to evaluate the CMIP6 model simulations, and they demonstrate that this way can help to extract more useful information than the conventional spatial-pattern-based evaluation method.

Overall, this work is scientific and worth to be published. But it is necessary for the authors to carefully revise the manuscript before it is accepted for publication on Earth System Dynamics. Please find my comments and suggestions below.

1. The Atlantic-Pacific teleconnections are dependent on the specific time scales, e.g., the MJO-NAO teleconnection on intraseasonal time scale, and the PDV-AMV teleconnection on interanual or decadal time scale. The current manuscript may only focus on the interanual time scale, such that the authors may consider to confine the title of the article into interanual Atlantic-Pacific teleconnections.

2. As the authors mention, the length of the time series can influence the performance of causal detections. The authors should provide a table to list the used length of time series for every cases which they analyzed in the manuscript, including the observation and model data under different cases. Especially, when analyzing the same causal link such as the PDV-AMV link under out-of-phase regime, please put the used time series lengths of observation and model data together, which is benefit for comparisons. This supplementary is necessary because the readers can use this information to judge how convincing the results are, and whether their data lengths are enough if they would also utilize the PCMICI to analyze their own work.

3. The authors focuses on the regime-dependent causal interactions. This is a very good attempt because the nonlinearity nature of the climate system determines that the climatic interactions are varied with the different temporal regimes. However, I am a few concerned whether their technical processing on the time series will influence the accuracy of causal inference. My concern is as following.

As previous studies suggested (Smirnov and Bezruchko, 2012; Smirnov, 2013), the low temporal resolution or resampling of the time series can lead to spurious causalities within the underlying variables. The authors divide the PDV/AMV/PNA/PSA time series into several segments within different regimes, then combine the time series segments within in/out-of-phase regimes. This is actually a resampling processing on the time series, thus is it possible to lead to spurious causalities as Smirnov and Bezruchko (2012) found? I agree that this issue is very challenging in techniques, but at least this should be mentioned and discussed in the manuscript. Otherwise the readers who are not expect in causal detection may neglect this issue, and this is not conductive to the method development of data-driven causal detections in climate science.

Reference

[1] Smirnov D., Bezruchko B. 2012. Spurious causalities due to low temporal resolution: Towards detection of bidirectional coupling from time series. EPL, 100, 10005. https://iopscience.iop.org/article/10.1209/0295-5075/100/10005

[2] Smirnov D. 2013. Spurious causalities with transfer entropy. Phys. Rev. E, 87, 042917. https://journals.aps.org/pre/abstract/10.1103/PhysRevE.87.042917

4. In line 252, it read as "This is to quantify the similarity between the observed and simulated spatial patterns for the four modes and build credibility that the CMIP6 simulated indices we use in the regime-oriented model evaluation have spatial expressions that resemble those of indices calculated from reanalysis datasets." This sentence is too long such that I cannot understand its meaning. Please rephrase it.

5. In line 352, it reads as "The limitation presented by the length of unmasked time series during specific short regimes is eliminated when combining them." I would like to know more technical details about "combining them". Do you glue the time series segments during specific regimes and then calculate the lagged correlations and partial correlations? Please provide a little more explanations of your processing here.

6. In figure 6, there is no unit for the horizontal axis, is it [years]?

7. In line 494, it reads as "The model doesn't only have the lowest level ...". "doesn't" should be corrected as "does not".