Review of „Network connectivity between the winter Arctic Oscillation and summer sea ice in CMIP6 models and observations“ by Gregory et al.

General comments

This study by Gregory et al. investigates the spatio-temporal covariability of the winter Arctic Oscillation (AO) and summer Arctic sea ice and its representation in CMIP6 models with a complex network approach. This is interesting for the predictability of summer sea ice in the light of a suggested physical mechanism between the winter AO to summer sea ice, supported by the observations. The authors point out discrepancies between the covariability of winter AO and summer sea ice in observations/reanalyses and CMIP6 models, particularly also in the transition from earlier to more recent years, and hypothesize that these could be due to regional biases in sea ice thickness.

This is a carefully performed and well-presented study and a valuable addition to previous literature on mechanisms for sea ice predictability. The paper is clearly structured and well written. The methods are well explained, so that also someone who is not familiar with complex networks can easily follow and understand the results. Previous literature is thoroughly cited. Overall, I recommend the paper for publication after consideration of the comments outlined below.

Specific comments

- As you write in L66 ff., climate network analysis is an useful addition to more conventional statistical approaches. I find the climate network analysis very intriguing, but (being new to the method) it is not entirely clear to me why you chose it here and what is the added value to more conventional statistics. Would you expect to arrive at the same conclusions with e.g. an EOF analysis?
- To L124 ff.: Is the number of network nodes N determined by the clustering algorithm or is it prescribed? What would be a typical number for N?
- Equation 2: I understand that the grid cells need to be weighted by their grid cell area. However, I find it a bit puzzling that for the polar grid the weights should be the square-root of the cell area, while for a regular grid the weights are a non-dimensional number between 0 and 1. Would it be possible to introduce normalized area-weights, such that there is no [km] in the unit and the anomaly time series (and with that also the covariance and node strengths) are not dependent on the node size? This could then maybe also resolve the issue that a high node strength does not necessarily coincides with large explained variance, as you explain in L252 ff.?
- Relating to what you write in L175-L187 (which is a very helpful paragraph), I am still wondering how much agreement/disagreement between model simulations (even if they had perfect model physics) and observations you can expect, simply because of internal variability and the time series of 42 years being relatively short, even more so when you look at the first and second half of the time series separately. What could be helpful in this regard is a comparison of the observations to a single-model initial-condition large ensemble, which can give you a good indication of the range of internal variability within the model. The CanESM5 model with its 20 members comes closest to a large ensemble. Looking at the ARI and D values of the CanESM5 ensemble members gives an idea of the expected spread just due to internal variability. I don’t have a concrete suggestion for changing the manuscript, but I think this point should be kept in mind when describing and discussing the results.
- In all Figures showing covariance and node strengths, please add the units. Moreover, in the Figures for SLP (Figs. 1 and 5), I don’t see how the orders of magnitude for covariance (10) and node strength (10^{10}) fit together. As the node strength is the sum of the absolute value of all its associated link weights, it shouldn’t be orders of magnitude higher than the individual link weights/covariance, I would assume?
- Figure 8: The two sub-figures are a bit difficult to compare with the different axes ranges/color scales. If you keep the different ranges (which I can see is also useful for comparing Fig. 8a) with the observations), I still think that the range for Fig. 8b) could be

capped at lower values, as right now the colors are very light (which is misleading at first glance).

- In the caption of Figure 12, the values of ARI and D for the individual model simulations don't seem plausible/don't fit to what you write in the main text.
- Figure 14: I don't understand the meaning of the numbers in the legend(?) in the lower right panel of the figure. A proper legend with the names of individual models, the model mean and PIOMAS would be helpful.

Technical corrections

- L1: Suggest changing "proceeding" to "following", "subsequent" or similar
- L9: Suggest adding comma before "respectively" (here and at other occasions in the manuscript)
- L14: "the north Africa" → "the north of Africa", "north Africa", or "northern Africa"
- L56: When reading, I was wondering why only 31 of the CMIP6 models, which you explain later in the Data section. Very optional, but you could consider omitting the number of 31 in the introduction to avoid questions at this place.
- L109: Remove comma: "sea ice concentration and mean sea level pressure"
- L117, Table 1 and other occasions: "ensembles" → "ensemble members". In my understanding, a single simulation is denoted as an ensemble member, while the individual simulations together form a (single-model or multi-model) ensemble. If you change this, please be sure to stay consistent throughout the manuscript.
- Eq. 2: Suggest writing down that E[...] is the expected/mean value.
- L245: Remove comma: "north Africa and southern Europe"
- L298: Suggest changing heading from "Observations" to "Observations/Reanalyses"
- L430: JC → JS