Final response to Anonymous Referee #2

We thank the Anonymous reviewer #2 for their positive and encouraging review. Below we explain in detail how we intend to address the reviewer's comments during the revision process. Our replies are highlighted in italics.

General comments

This paper represents a crucial effort in identifying potential shortcomings in model representation of recent Greenland blocking trends and is therefore an important contribution to the literature. I found the authors' methods to be well-suited to the objectives of the paper and thought that the conclusions were mostly well-founded. My main critique is that more careful consideration and in-depth discussion of the implications of the authors' choice of blocking detection method are warranted. Given that much of the work demonstrating a positive trend in Greenland blocking – including the Preece et al. 2023 hypothesis that this work tested – was based on a field-departure-based blocking index, how might the application of a reversal-based detection method in this study impact the results presented herein and how they compare to previous works?

We thank the reviewer for their positive outlook on the manuscript. Considering that the reviewer rises a similar point to that by anonymous reviewer #1, i.e. the possible dependence of the results on the chosen metric, we plan to address this point in the revised version of the manuscript by including the GBI index as defined in Preece et al. (2023), i.e. the area averaged 500 hPa geopotential height within the domain of 60-80°N and 20-80°W. We might expect some difference in trends, with the reversal index being strong, but the overall findings should be similar (Luu et al. 2024). However, this addition should shed light on whether the conclusions we draw from this analysis are robust when different indices are employed.

Luu, L.N., Hanna, E., de Alwis Pitts, D. et al. Greenland summer blocking characteristics: an evaluation of a high-resolution multi-model ensemble. Clim Dyn 62, 10503–10523 (2024). https://doi.org/10.1007/s00382-024-07453-2

The authors' use of the deconstructed components of the reversal-based blocking index in their causal discovery method was particularly novel and interesting; however, I do wonder if too much emphasis was placed on the GHGN criterion in distinguishing between anticyclonic conditions and Greenland blocking. For example, L441 states, "Thus, in ERA5 our findings generally support the first part of Preece23's hypothesis, showing that T2M-Arctic and May snow cover may influence MSLP over North America, which in turn favours pressure highs over Greenland (GHGS>0). However, this does not consistently lead to blocking, as MSLP also contributes to GHGN>0, reducing the likelihood of blocking." I'm not sure that this distinction is quite so definitive. For example, Tyrlis et al. (2021) argue that high-latitude blocks such as those that impact Greenland are distinct in that they shift the jet stream to the south and, consequently, requiring strong westerly flow to the north may not be appropriate. They argue that the poleward geopotential height gradient criterion should be relaxed to 0 m per degree latitude for locations north of 60◦N latitude. How might this argument impact the interpretation of the seemingly contradictory links with the GHGN index revealed by the authors?

We thank the reviewer for their insightful comment. In the revised version of the manuscript, we will reassess our interpretation of the GHGN behavior considering the Tyrlis et al. (2021) results. We will also explore the possibility to apply their detection method, even if this will not affect the specific results obtained with GHGN and GGI. More in general, checking for the robustness of the results depending on different blocking indices (following the previous point), will help us understand whether blocking an absolute-field index like the GBI used by Preece et al. (that do not depend on the

reversal-based detection) shows consistent results with GGI or GHGS, or both. If consistent results are found with GHGS, but not GHGN, this would be further evidence that GHGS may be a better proxy for blocking over Greenland than GHGN, and that GHGN may not be as important in this region to define blocking, as suggested in Tyrlis et al. (2021).

Specific comments

We thank the anonymous reviewer for their thorough read of our manuscript. Below we reply to all comments, other than those highlighting a typo or a suggestion to rephrase a sentence (which will be dealt with in the point-by-point reply following the revision).

L184: What is the reason for extending the domain as far east as the prime meridian? Why start the southern bound of the domain at 67N?

The domain of the GBI region is identified based on where a significant increase in GBI index is detected (Fig. 1a, currently described in lines 253-255). In the revised version of the manuscript, we will make sure that this key information is better highlighted when the definition of the domain first appears.

L185: The Greenland Blocking Index, or GBI, has already been well established with a specific definition of the average 500 hPa height within the domain of 60-80N and 20-80W. I strongly suggest that the name here is altered to distinguish the index defined herein from the established GBI. Perhaps something as simple as the reversal-based Greenland blocking index (rGBI).

As state earlier, we plan to address the effect of different blocking indices as highlighted by the reviewers. Thus, following the reviewer's suggestion, we will name identify the reversal-based Greenland blocking index as "rGBI" in the revised version of the manuscript.

L322: Here you note that 33% of the GGI>1 s.d. T2m-G fall above the 90th climatological quantile in SEAS5.1-03; however, Figure 3f indicates 35.3% fall above the 90th climatological quantile. Which is correct?

We thank the anonymous reviewer for pointing out this discrepancy. There is indeed a mistake in the text, and we will update the revised version of the manuscript with the correct percentage (35%).

Figure 3: The meaning of the red shading and the text annotations in panels (e) and (f) should be noted in the figure caption.

Following the reviewer's suggestion, we will add the meaning of the red shading and the text annotations in panels (e) and (f) in the caption of Fig. 3.

Figure 4: The caption title is a bit confusing. Do the time series in the right column show the 11-year running mean of index values or an 11-year moving window trend of monthly-mean index values? The units at the top of each plot would suggest the latter, but the caption title suggests the former.

Figures 4i-l show the 11-year moving window trend of monthly-mean index values. In the revised version of the manuscript we will revised the caption title to avoid confusion.

L408: GBI is repeated here. Should one of these be GGI?

We thank the anonymous reviewer for spotting this mistake, the sentence should indeed read "GBI and GGI". We will correct this mistake in the revised version of the manuscript.

L418-421: The stationary wave response should increase as the background westerly flow weakens due to Arctic amplification. This could explain why the relationship with NA snow cover anomalies is stronger in the ERA5-81 record.

We thank the anonymous reviewer for point out this relevant literature, which we will include in the discussion of our results in the revised version of the manuscript.

L462: FDR has not been defined

We thank the anonymous reviewer for point out this missing information. FDR stands for "false discovery rate", which is described in lines 219-220. However, the definition of the acronym was missing. We will correct this discrepancy in the revised version of the manuscript.

Figure 6: More explanation is needed in the figure caption. What is the meaning of the numbers on the linkage arrows? Why do some connecting lines not include an arrow head? Why are there two color bars included at the bottom of the figure (i.e., what does each bar correspond to?) I see that this information is given on L232, but it would be helpful to have it in the caption as well.

We will revise the caption of Fig. 6 following the reviewer's suggestion.

Figure 7a: Where does this CEN diagram come from? Is this based on the analysis summarized in Figure 6? If so, why is the lag-0 linkage between MSLP-NAm and GGI positive?

The definition of the CEN in Fig. 7a is currently described in lines 521-523: "To provide a fair comparison between ERA5-81 and SEAS5.1-03, we adopt the concept of causal inference (see Methods Section 2.3). Based on the results shown in Fig. 6, we now assume that the selection of causal links shown in Fig. 7a is found in both ERA5-81 and SEAS5.1-03, and also for GHGS and GHGN indices.". We agree with the reviewer's comment that this panel requires more explanation. Here, we construct the CEN network by imposing a sub-selection of all the links that PCMCI detects in Fig. 6a and 6b. Links are sub-selected to best represent the Preece et al. (2023) hypothesis and to balance between links found in ERA5 and those found in SEAS5.1. Another difference between Fig. 6a and 7a is that in Fig. 7a the maximum lag used is -1. The sign of the MSLP-NAm -> GGI link at lag 0 changes from negative (Fig. 6b) to positive (Fig. 7a) because Fig. 7a show the CEN of ERA5, while Fig. 6b shows the CEN network for SEAS5.1. However, changes in sign between Fig. 6a and 7a (both showing ERA5 CENs) are also seen. For example, the T2m-Arctiv -> GGI link goes from positive in Fig. 6a to negative in Fig 7a. This does not however represent a contradiction since the lag of the link in Fig. 6a is -2, while in Fig. 7a we restrict it to -1 (for consistency with SEAS5.1).