Final response to Anonymous Referee #1

We thank the Anonymous reviewer #1 for their insightful 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.

Major comments

1. Observations and reanalysis are repeatedly conflated in the paper. Reanalysis is still a model-derived product, and its snow cover is biased (e.g. Mudryk et al 2015, Mortimer et al 2020) when comparing to in-situ and observation-derived gridded products. This makes me wonder how dependent the results in this paper are on the use of ERA5 as 'observations', and I'd recommend first that the authors are more careful about their use of the word observations, and second that some discussion around how ERA5's biases could be impacting the results. I also wonder whether other reanalysis products would be able to reproduce the same causality? Or whether a different metric for GB would yield similar results, both for causality and for how unusual the reanalysis trend is. I wonder as well where there is a state-dependence and how that might come in to play, for example a non-linearity when future snow cover over North America is much lower on average?

We agree that – despite being a common terminology in climate modelling - using the word "observations" when reanalysis data are used can generate confusion. Therefore, we will refer to "reanalysis" only in the revised version of the manuscript. Following the reviewer's suggestions, we intend to check whether a different GB metric yields to the same results. Specifically, we plan to use the GBI index defined as the area averaged 500 hPa geopotential height within the domain of 60-80°N and 20-80°W., as described in Preece et al. (2023). We might expect some difference in trends, with the reversal index being strong, but the overall findings should be similar (Luu et al. 2024). Moreover, to address potential state dependence of our analysis, we plan to check whether the causal effect networks change during years with high or low snow cover over North America. This technique has been used in the past in Di Capua et al. (2023) and Tian et al. (2024).

Tian, Y., Giaquinto, D., Di Capua, G. et al.: Historical changes in the Causal Effect Networks of compound hot and dry extremes in central Europe. Commun Earth Environ 5, 764 (2024). [*https://doi.org/10.1038/s43247-024-01934-2*](https://doi.org/10.1038/s43247-024-01934-2)

Di Capua, G., Coumou, D., van den Hurk, B., Weisheimer, A., Turner, A. G., and Donner, R. V.: Validation of boreal summer tropical–extratropical causal links in seasonal forecasts, Weather Clim. Dynam., 4, 701–723, https://doi.org/10.5194/wcd-4-701-2023, 2023.

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

2. Despite the title, quite a lot more time is spent on the idea that there is a forced positive trend in GB, driven by the Preece et al 2023 mechanism, rather than the idea that natural variability (in particular anything other than the AMV), or even a forced increase in variability, has caused the trend in reanalysis. Evidence from CMIP6 is that the forced trend is negative with a lot of variability super-imposed, and so even if the Preece et al 2023 mechanism is correct and is missing from models, it's not obvious to me that that means the models are wrong in the direction of their trend. Perhaps the forced trend for GB is not driven from the pole, but rather from the lower latitudes (on balance) and that's the source of the decline in future GB? I do agree, however, that a missing mechanism that increases GB variability on an interannual timescales could still be important for

future Greenland melt, and I do think that the results here are useful science, I'm just not sure about the way it has been framed.

We agree with the anonymous reviewer that, while our analysis suggests that the snow cover mechanisms is missing in SEAS5.1, and potentially also in CMIP models – which are not part of this study -, this does not rule out that other mechanisms are at play and that (natural) tropical variability could be influencing the observed trend. While we do not directly focus on tropical forcings in this manuscript, we will address this comment in the revised version of the manuscript, to better highlight in the discussion section that while we show that the snow cover mechanism is important for GBI variability and is too weak/missing from seasonal forecasts, climate models showing a GBI decrease in future projections could still be correct if natural variability, or other mechanisms not addressed here, prevail over the snow cover mechanism.

3. The intro and the conclusions are both long and meandering at times between forcing of GB between the tropics, midlatitudes and poles, and between climate models and observations. Please consider re-writing to make it clearer.

Following the reviewer's comment, we plan to shorten and simplify both the introduction and discussion sections, to improve the readability and clarity of the manuscript.

4. I don't think using T2m-Arctic as an indicator for Arctic amplification is sufficient. A difference between the Arctic and some mid-latitude band would probably be better, as a year with high T2m Arctic could also have high temperatures in general, i.e. T2m Arctic is highly correlated with T2m global. In general, I think the term Arctic amplification is used when the authors intend to say Arctic warming, so I'd recommend more careful wording.

We agree with the reviewer's suggestion and, in the revised version of the manuscript, we plan to redefine our Arctic amplification index as the ratio between T2m Arctic and Global T2m.

Minor comments

L143: Is the mean of each month for the entire period removed from that month? Following sentence is obvious and need not be included.

We will remove this sentence following the reviewer's suggestion.

L150 Why isn't April one of the initialisations for SEAS5.1?

We chose to analyse both 1. March and 1. May initialisation dates for SEAS5.1 to assess the dependency of the results on the length of the forecasts prior to the target season, e.g on the forecast lead time. It would be reasonable to assume that forecasts initialized on 1. April, may show a mixed signal of both the earlier and later initialised runs. However, since we do not detect significant differences between SEAS5.1-03 and SEAS5.1-05, it is reasonable to assume that an initialisation date in between would not diverge from the obtained results.

L155: Everything after 'Linear correlation should be moved to the section 2.2

Lines 155-158 will be moved to section 2.2 in the revised version of the manuscript.

Figure 2: It's interesting that there's a reversal in the positions of ERA-40 and ERA-81 in terms of their percentile between GBI and GGI. The red lines do not look to be correlated in (c) and (d), as in Figure 1(c). Is there is a mistake in the plot or in the caption? Why is GHGS and GHGN written on panels (c) & (d)?

We thank the reviewer for point to this mistake. We will correct the figure text in the revised version of the manuscript.

Figure 3: I wonder if a difference plot of (b)-(a) would be helpful for visualising where ERA5 and SEAS5.1 differ

In the revised version on the manuscript, following the reviewer's suggestions, we will provide the additional panels showing the difference between ERA and SEAS5.1-03 for both T2m and Z500 fields.

Figure 4: (j) It's interesting that all the members are so tightly constrained for Snow-Nam compared to other fields, and I wonder why that might be, and if it's showing a related issues, whereby the seasonal model is not simulating variability in snow cover properly?

Being the seasonal forecast simulations initialised for a state as much close as possible to observations, they tend to diverge with increasing forecast lead time. However, fields such as sea surface temperature (SST) or snow cover are characterized by larger inertia and have a slower variability than atmospheric fields such as T2m or Z500, so it is absolutely expected that they diverge from the initialization state more slowly. Therefore, both AMV (which is derived from SST) and snow cover time series in Figs. 4i and 4l show smaller spread around the average values when compared to T2m and MSLP fields. Moreover, for snow cover (Fig. 4i), only May is considered, and because these plots are obtained using SEAS5.1-05, the divergence from the initial state is minimal. Therefore, this behaviour should not highlight an underlying issue, but rather an expected behaviour of the seasonal forecast fields. We will make sure that this point is clearly explained in the revised version of the manuscript. (double-check whether variability increases for SEAS5.1-03, figure should be in the SI)

Paragraph L396: non-significant correlations can't support a relationship, the only thing that's been shown there is that Arctic temp and GB are correlated.

While we think worthwhile to report non-significant correlations (given that the significance is also affected by the shortness of the time series), we agree with the reviewer's comment that the paragraph needs to be revised to make clear that we do not infer any relationship from non-significant correlations.

L429: Why does a seasonal forecast model have lower signal-to-noise ratios?

We thank the reviewer for raising this question. While we cannot define signal-to-noise in ERA5 in the same way (given that we have a single realisation), we will consider removing this sentence if we are not able to fully justify it, or, alternatively refer to the signal-to-noise paradox.

Technical comments:

There are quite a few typos, missing words, and instances of poor grammar throughout the paper. I will highlight a few examples here but there are far too many and I would recommend a more thorough edit and grammar check before re-submission.

We thank the anonymous reviewer for carefully reading the manuscript and for spotting grammatical errors and typos. In the revised version of the manuscript, we will make sure to improve this aspect of the manuscript and address of the highlighted issues.