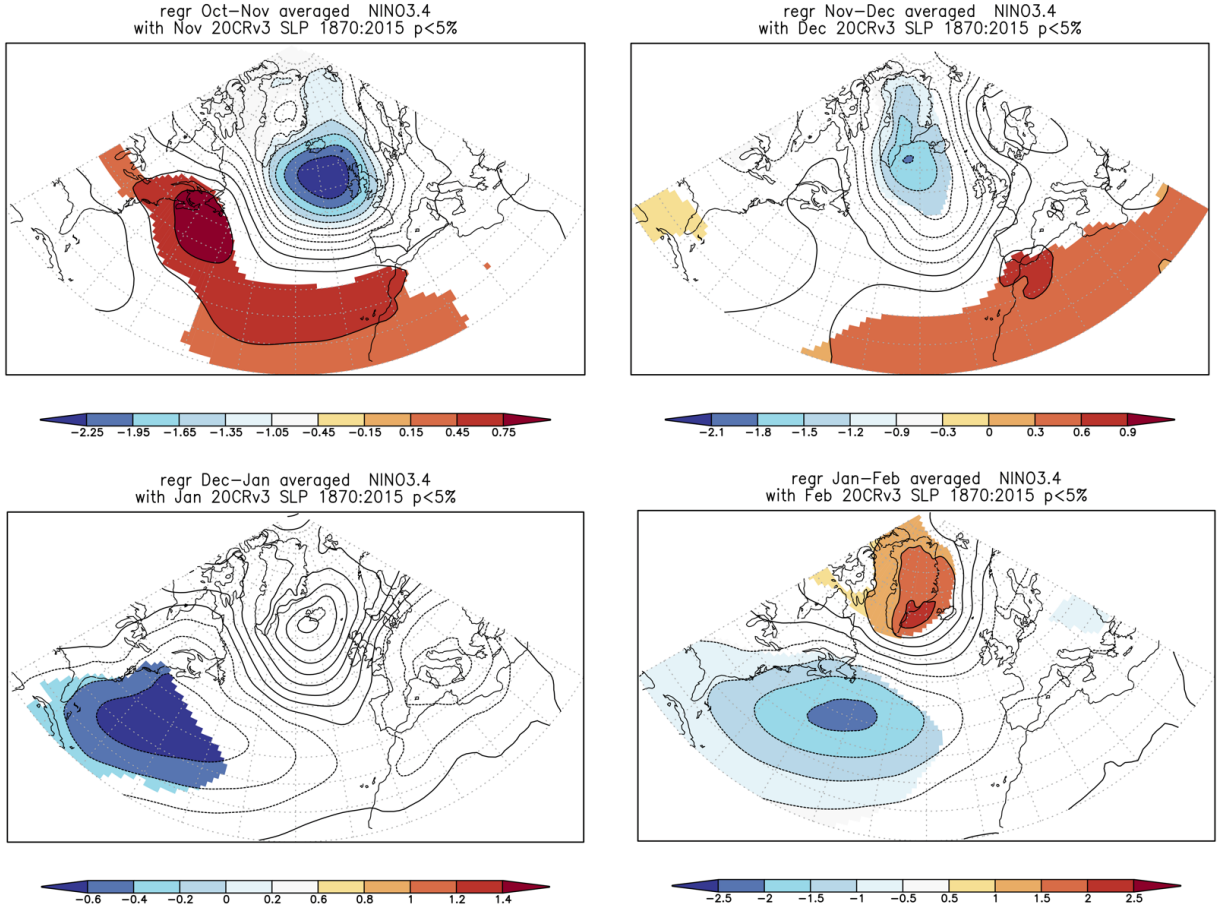
Thank you for the review of our manuscript. In this final authors' comments, we give a few immediate/early replies to selected comments. All comments will be given full attention and point-by-point replies will be provided with the revised manuscript later.

Reviewer 1.

General comments:

1. Given that the authors argue that splitting ND from JF is important, I think it is important to show each of these 4 months individually in a version of figure 2 with four columns and three rows, before then showing the ND and JF composites. This additional figure would help make the point to the reader that ND are more alike, and JF are more alike, before the authors then go on to combine ND together and JF together.

We show below the regressions (not composites) of SLP in Oct, Nov, Dec, and Jan separately on HadISST Nino3.4 plotted quickly using the KNMI Climate Explorer web tool (https://climexp.knmi.nl/start.cgi).



It is seen here that there is an evolution through these months. In particular, the signs in the two major centres of action change from November/December to January/February.

In the revision, we will carry out additional analysis for these months using the method consistent with the rest of the manuscript and document the result appropriately.

2. There is one recent paper the authors seem to have missed that is very relevant and complementary to the present study. The size of a composite necessary for the stratospheric route in late winter to become robust was discussed in detail in Weinberger et al 2019. They concluded that the EN-weak vortex route is robust with around 15 events, though the impact of LN on the vortex was only apparent with more than 25 events (see their figure 5) as compared to climatology. They don't show the difference between EN and LN, but clearly the difference will be significant with less.

The polar cap SLP response was also found to be significant after around 15 events are taken (their figure 7) both for EN and LN. This is consistent with the results shown in the present paper.

Weinberger et al also found robust impacts on subpolar Eurasian temperatures (see their figure 7) that are robust for EN and LN individually with composite sizes similar to that used in the present paper. Weinberger et al didn't consider early winter and they analyzed historical data for a much shorter period than analyzed here, and so there are clearly new results here. But as best as I can tell the results are consistent when the late winter and the stratospheric route is considered.

Weinberger, I., Garfinkel, C.I., White, I.P. and Oman, L.D., 2019. The salience of nonlinearities in the boreal winter response to ENSO: Arctic stratosphere and Europe. Climate dynamics, 53(7), pp.4591-4610.

Thank you for bringing our attention to this study. We will add discussion related to it. The question of statistical significance and its dependence on sample size (i.e., number of events) is an interesting one. We have not studied nor reported this thoroughly in our submitted manuscript (although we have done some exploratory tests previously). As reported in the manuscript, increasing the sample size to include events in the 19th century, compared to the post-1920 events used by Deser et al. 2017 (see Table 1), does not change the findings.

As we are using reanalysis data, we have much smaller number of original (the complete non-resampled) events compared to Weinberger et al. with their analysis on a model ensemble. However, bootstrapping does permit us to re-sample the available events to smaller sample sizes (as well as to larger ones if certain assumptions are made), and therefore the effect of sample sizes in observation/reanalysis can also be estimated.

We are interested to investigate this more thoroughly and will decide on an appropriate method to document the result without changing the focus or essence of the current manuscript.