Comment on wcd-2021-52
Anonymous Referee #3

This paper examines relationships between autumn Eurasian snow conditions and subsequent winter NAO development and associated climatic conditions over a 110-year span in the ERA-20C reanalysis, the ECMWF ASF-20C seasonal hindcasts, and a tailored hindcast set in which land surface initial conditions, including snow, are sampled from 20 adjacent years. It is found that differences in the longitudinal gradient of Eurasian snow at the beginning of November have discernable influences on the subsequent winter NAO, but that this relationship is not stationary over the 110-year period and is weaker in the hindcasts than in the reanalysis. Anomalies composited on extreme values of the longitudinal snow gradient point to roles played by the Ural ridge and wave fluxes influencing the stratosphere.

Overall, the paper makes interesting contributions to efforts to unravel the hypothesized causal connection between autumn Eurasian snow distribution and winter climate in the Northern Hemisphere. Although the methodologies and conclusions drawn appear generally robust, the paper could benefit from improvements to the presentation and some additional discussion of certain points as recommended below.

Main comments:

1) In the first paragraph of the introduction at line 57, regarding the NAO and its impacts in the winter of 2020/21 the authors could cite https://doi.org/10.1175/2021BAMSStateoftheClimate.1 which touches on this on p. S73 and in Fig. 2.40.

2) The following cited papers (may not be a complete list) are missing from the references: Meehl et al., 2021; Diro and Lin, 2020; Wegmann et al., 2017; Wegmann et al., 2016; Jones et al., 1997; Koster et al. 2011.

3) Near line 143, please say a few words about why these particular eastern and western domains were chosen so that the reader does not have to refer to Wegmann et al., 2017 (which is not in the reference list as commented above). Presumably their latitudinal range is intended to encompass variations in the November 1 snow line? (This is explained finally around line 209, so maybe could add “whose choices are motivated below” or something similar at line 145.) Also, is it possible to represent the domains in Fig. 3.
accurately, with borders along latitude circles rather than inaccurately as line segments?

4) Inconsistent terminology is used to describe the index for Eurasian snow distribution. For example, in section 2a the eastern minus western difference in Eurasian snow depths is described as the “west-east snow cover gradient”. However, snow cover as conventionally defined refers to the presence or absence of snow (irrespective of depth), and snow cover extent to area blanketed by snow. To avoid potential confusion, and for consistency with the discussion on lines 150-154, I suggest referring here to the “west-east snow depth gradient” or simply “west-east snow gradient”.

5) Although the derived index is consistently and appropriately called the Eurasian snow dipole index or simply dipole index, the sub-ensembles of high and low dipole index hindcasts are referred to the “high-snow ensemble” and “low-snow ensemble”. However, it’s not obvious what “high snow” and “low snow” refer to, and calling these the “high-dipole ensemble” and “low-dipole ensemble” would be clearer and more consistent.

6) At lines 156-157, it’s not entirely clear whether using the “same definition for the NAO DJF index in seasonal prediction runs” means that the predicted NAO index is obtained using the first EOF of ERA-20C SLP, or the first EOF of predicted SLP.

7) It should be clarified in section 2b that these are not, strictly speaking, seasonal retrospective predictions since the ocean and sea ice boundary conditions after initialization are based on observations (same as ERA-20C) rather than being predicted using damped anomaly persistence or some other means.

8) In my opinion Fig. 2 would be easier to process if the three “LSxx” labels in panel (a) and associated dashed lines were removed, since their meaning isn’t explicitly defined and it’s not obvious exactly what the dashed lines are connecting to. (Although their meaning can be distilled from the main text, I found these features to be more distracting than illuminating, e.g. one wonders why there are three and not some other number, why those particular years, etc. all of which are irrelevant to the point of the figure.) More broadly, it might be commented that it may take most readers longer to understand the figure than the descriptions of the two experiments in section 2b which are straightforward, although the figure does nicely schematize what was done.

9) Considering that the western domain typically has little snow on November 1 (lines 212-213), is the dipole index time series, and by extension the overall results of this study, much different if only variability in the eastern domain is considered? (The discussion of Fig. 11 on page 12 also suggests this might be the case.)

10) In the paragraph beginning at line 222, regarding the range of variability in the two forecast experiments, how do the NAO standard deviations for the individual ensemble members shown in Fig. 4a compare to the observational values? Also, is there any detectable difference in the ensemble spreads between the CTL and EXP forecasts, considering that the latter start with considerable ensemble spread in the land initial conditions?

11) The legends in Fig. 4 say AFS instead of ASF.

12) At line 242, please include a reference in order to provide some context about the early twentieth century Arctic warming, such as Polyakov et al. 2003, https://doi.org/10.1175/1520-0442(2003)016%3C2067:VATOAT%3E2.0.CO;2

13) The placement of “as depicted in Figure 4” in line 244 suggests that what is being referred to is the depiction of the positive dipole snow pattern in Fig. 3a. If indeed this is referring to Figure 4 then it would be better placed at the end of the sentence.
14) One or more citations should be added to the sentence in 254-257 (even if cited previously).

15) At line 281, regarding “significantly reduced geopotential heights over the extratropics”, according to Fig. 7b these changes are only statistically significant over the Mediterranean region. Please reword accordingly.

16) Line 301 should state e.g. that anomalies between the two ensemble means in Fig. 7h are “less than” 1 hPa rather than “around” 1 hPa, considering that they never exceed the 0.83 hPa contour level. (Similarly in line 403.)

17) More attention should be drawn to the different color scales in Figs. 6 vs 7, and 5 vs 8 in order to keep the differences in the magnitudes of the observed and modeled show dipole responses in perspective.

18) It seems noteworthy that the stratospheric evolution in Fig. 8 is similar to, but delayed with respect to that in Fig. 5. This merits at least mentioning, as would any hypotheses the authors might have for the origins of the delay.

19) Although the caption to Fig. S2 says “Shading indicates 90% significance level”, there is in fact no shading. Does this mean that none of the anomalies are statistically significant?

20) Is the heading in line 332 intended to be there (and to be labeled “e.”)?

21) The term “resulted in” on lines 352 and 354 may imply causality more strongly than intended considering the context of opposite NAO responses to the same sign of snow anomalies. Suggest changing to “preceded”.

22) The terminology describing the snow anomalies relating to Fig. 10 is inconsistent and somewhat confusing in that the main text in lines 351-352 refers to “high-minus low snow cover anomalies” and the Fig. 10 caption to “positive snow dipole forcing”. It would be preferable if consistent terminology describing snow anomalies were used, as addressed also in comments (4) and (5).

23) The composite November 1 SST and sea ice concentration differences shown in Fig. S4 which are around 0.1C and 0.02 respectively seem unlikely to have any major impact of DJF NAO. Also, do the authors have confidence that pre-satellite sea ice concentrations during 1901-1978 in ERA20C are sufficiently accurate for such an analysis?

24) Do the authors have any hypotheses for why the lagged atmospheric responses to autumn snow dipole index differences are so much weaker in the hindcasts than in ERA-20C? In particular, could this be related to the “signal-to-noise paradox” whereby circulation responses to radiative and surface forcings, as well as the predictability of the NAO, appear to be much weaker in models than is observed as argued by Scaife and Smith 2020, https://doi.org/10.1038/s41612-018-0038-4?

Minor:

lines 141 and 170: a-forementioned -> aforementioned

line 153 identically -> identical

line 211: computes -> “represents” or “indicates”
line 297: mirrors -> mirror
line 298: increased -> positive
line 336: pre-ceding -> preceding
line 379: is -> does ?
line 383: remove 2nd "the" ?
line 412: "re-enforced" and "re-inforced" -> reinforced (also line 419)
line 413: pre-cedes -> precedes
line 423: mi- -> mid-
line 424: remain -> remains
line 424: suggest "dominant" -> "prominent" (or else simply "remains")
line 447: "On the opposite" -> "Oppositely" or "On the contrary"
line 482: could remove "additional"
line 488: suggest adding "as in this study" after "years"
line 740: heighst -> heights

Caption to Fig. S3: "< 1 stand. dev" -> "< -1 stand. dev."