Reviewer #1

This manuscript analyzes the stratospheric wave reflection events and the following tropospheric/surface weather evolutions in North America using reanalysis data. The authors start with a regional stratospheric wave reflection detection metric and 44 events have been detected since 1980. By comparing with the North American weather regimes, the wave reflection events show a transition from a regime characterized as a trough anomaly in North Pacific and a ridge anomaly in North America to a regime with a reversed pattern. The surface air temperature changes from warm anomaly to cold anomaly in about two weeks, presenting a large-scale decrease evolution.

I think the two mechanisms are in fact different aspects of the same planetary wave-zonal flow interaction rather than totally different dynamical mechanisms. Both wave-mean flow interaction and wave reflection are useful for our understanding of the influence of the stratosphere on the troposphere. While the wave-mean flow interaction has been well studied, relatively less attention has been paid to wave reflection. The current study is a systematic analysis of the tropospheric weather evolution following the detected wave reflections. The manuscript is straightforward and well-organized. I believe this manuscript could be a helpful reference for the research community of stratosphere-troposphere coupling. Thus, I think it could be acceptable after minor revision.

Thank you for your positive outlook on our submission and suggestions for further improvement. We agree that both wave-mean flow interactions and wave reflection are both a form of wave-flow interaction, and have added a brief comment to that effect in the introduction of our manuscript. We have copied your comments below in *Italics*, with our replies in Blue.

Specific comments:

1. What's the spatial pattern of the stratospheric polar vortex associated with the detected wave reflection events? Stretching, shifting, or zonally symmetric weakening? The authors may add an Appendix figure at least.

We thank the reviewer for this suggestion. In the revised manuscript we have included the new Fig. A3 (also copied below) showing the mean evolution of Z10 for the same lags as the Z500 anomalies. This shows an anomalously strong vortex (negative Z10 anomalies over the central Arctic) on average during the reflection event. This is consistent with the expected favorable condition for wave reflection. At the onset of the reflection event, there is a small signature of an Aleutian anticyclone, which intensifies at positive lags. This is somewhat associated with stretching the vortex along 90°W by D10-15 (reminiscent of Fig. 2 in Cohen *et al.* 2021), although not notably so. To enable a better comparison with Cohen *et al.* 2021, we also include below the Z100 composites as FIg. R1. We now briefly comment on this at the end of Sect. 3 and in Sect. 6.

Judah Cohen, Laurie Agel, Mathew Barlow, Chaim I Garfinkel, and Ian White. Linking Arctic variability and change with extreme winter weather in the United States. *Science*, 373(6559):1116–1121, 2021

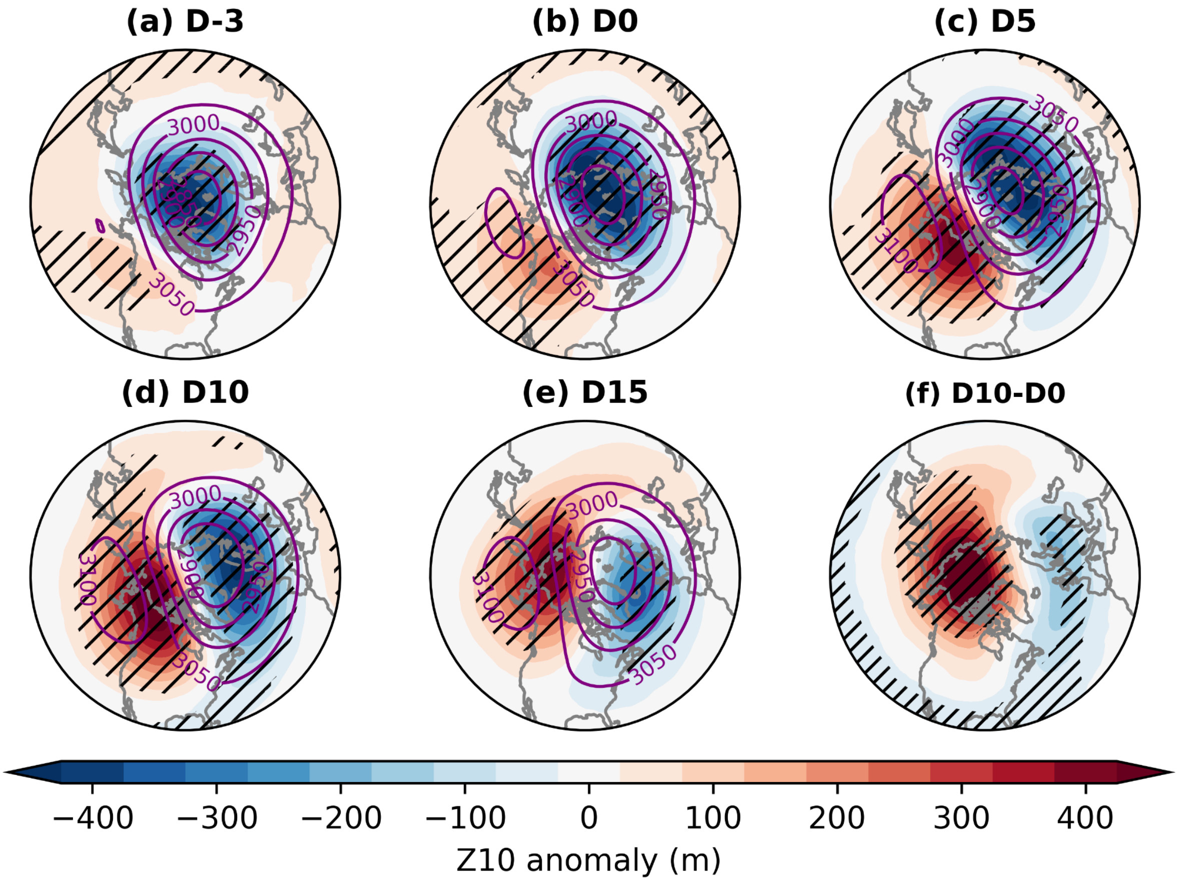

FIg. A3: *(a–e) Composite-mean 10 hPa geopotential height (Z10, contours, dam) and Z10 anomalies (shading, m) at various lags relative to the reflection event onset. (f) Average difference between the Z10 anomalies on day 10 and day 0 (i.e., (d)-(b)). Hatching denotes statistically significant anomalies, assessed as described in Sect. 2 in the main text.*

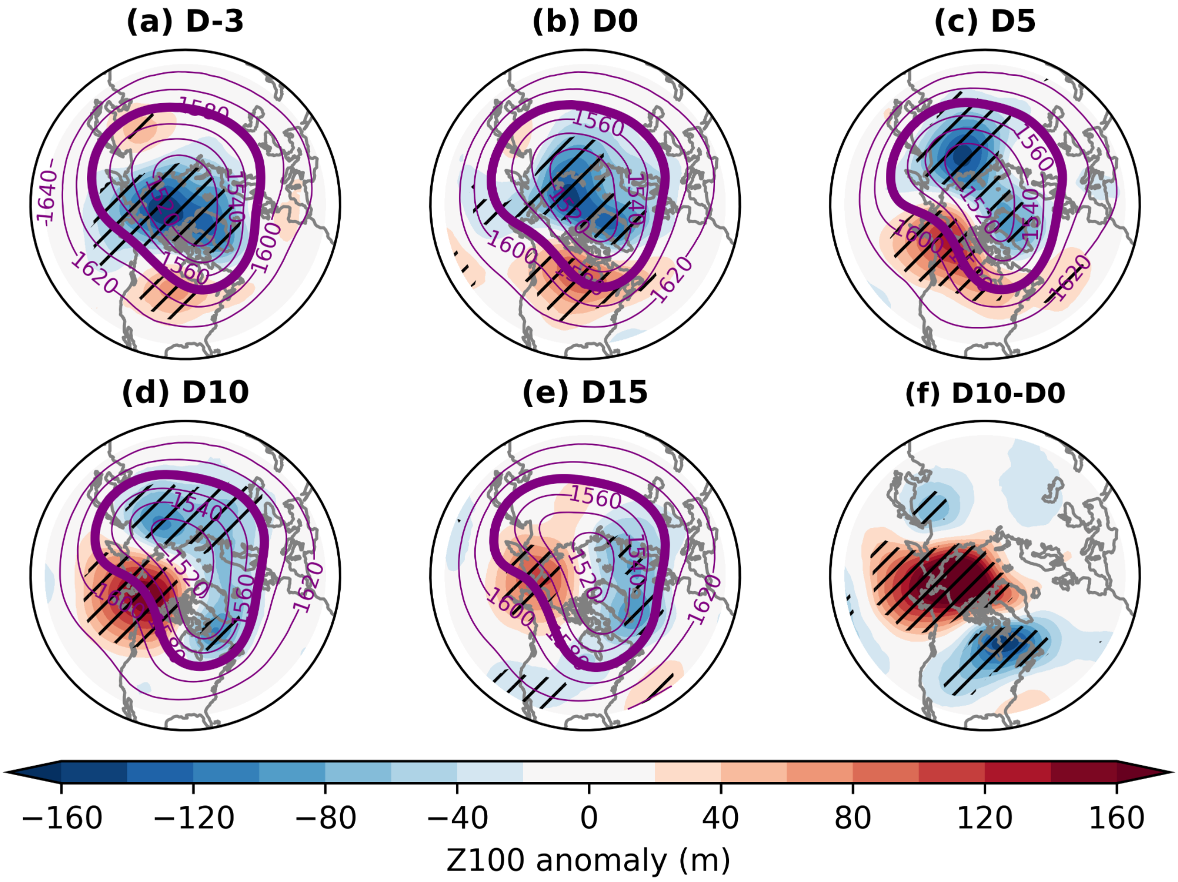

Fig. R1: (a–e) Composite-mean 100 hPa geopotential height (Z100, contours, dam) and Z100 anomalies (shading, m) at various lags relative to the reflection event onset. (f) Average difference between the Z100 anomalies on day 10 and day 0 (i.e., (d)-(b)). The 1580 dam contour is bolded. Hatching denotes statistically significant anomalies, assessed as described in Sect. 2 in the main text.

2. Line 115-117: I'm wondering why the authors employed a lower threshold (1.5->1). To allow more samples? I'd guess the conclusions still hold with different thresholds, such as 1.5 or 2, but the authors may double-check and make a clear statement.

The threshold of 1 was chosen here because the associated events already fulfill the assessed criteria (as shown in Fig. 1 and Fig. 2). We have tested using higher thresholds of RI > 1.5 and RI > 2, which naturally reduces our sample size (35 and 27 events respectively). The qualitative results concerning the surface anomalies associated with the reflection events are nonetheless consistent with those obtained for RI > 1. We further note that while there are some regions in which the temperature anomalies for the higher RI thresholds are larger than those for RI > 1, the difference is generally small. This points to RI>1 being a threshold which selects reflection events that are sufficiently intense to be associated with a large surface impact. We show in Figs. R2-R5 below the results corresponding to Figs. 4c, d and Fig. 5 in the main text. We now mention this in the revised text in Sect. 3.

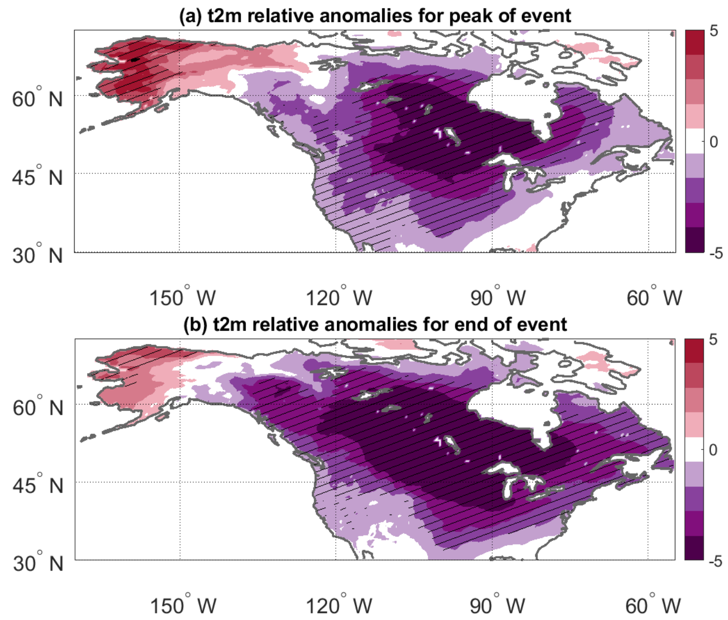

Fig. R2. Same as Fig. 4c, d in the main paper, but for RI > 1.5 (35 Events).

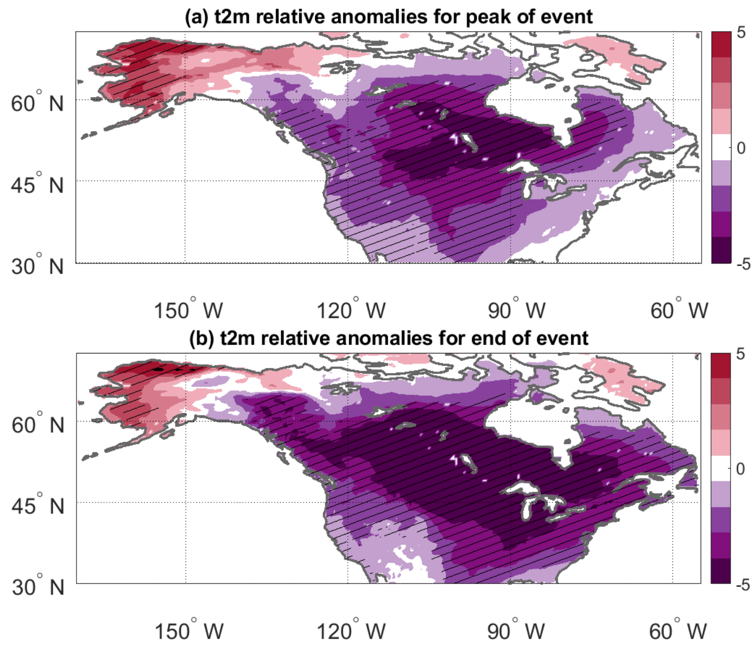

Fig. R3. Same as Fig. 4c, d in the main paper, but for RI > 2 (27 Events).

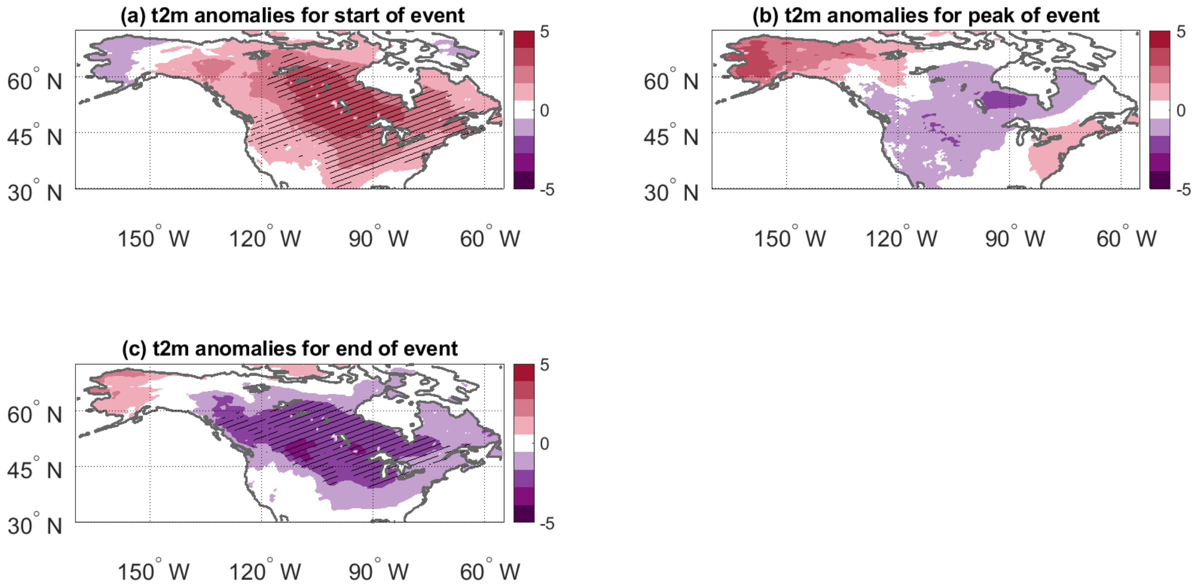

Fig. R4. Same as Fig. 5 in the main paper, but for RI > 1.5.

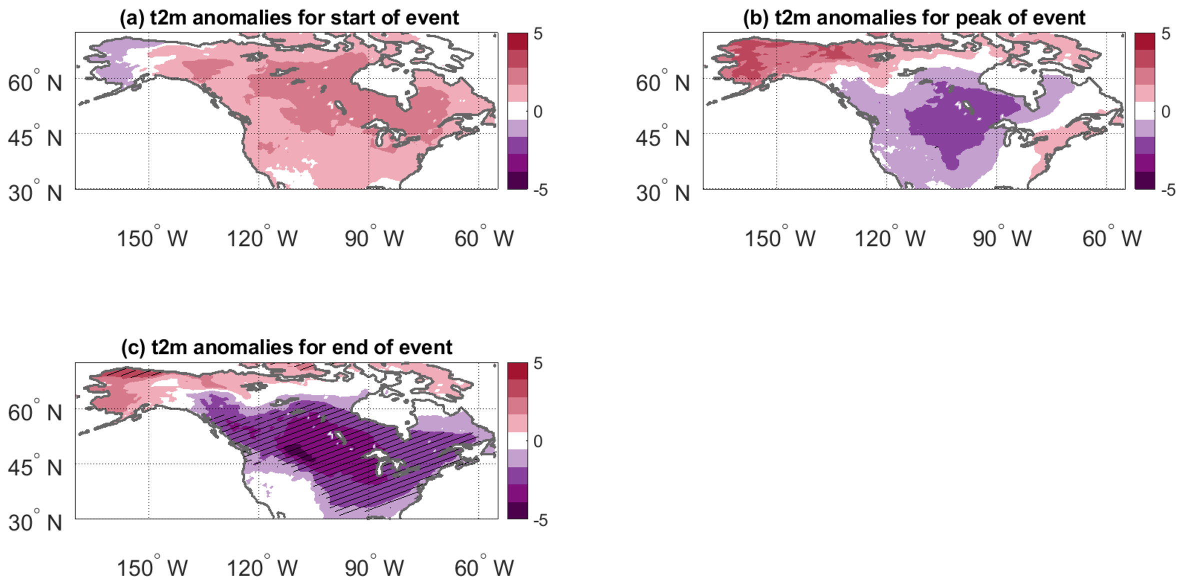

Fig. R5. Same as Fig. 5 in the main paper, but for RI > 2.

3. Table A1: event#42 the event detected in the 2017-2018 winter is different from their previous study (MK20). The authors may explain why.

MK20 used the term “events” to refer to three mid-latitude cold spells that occurred in the winter 2017/18. In contrast, we use it here to refer to the occurrence of wave reflection. Fig. 10 in MK20 shows the evolution of the reflection index in that winter (red line) which shows that the index was high (>1.5) during the second half of January, which is consistent with event#42 from this study (13. Jan 2018 - 2 Feb 2018). The reflection event preceding the cold spell at the end of December 2017 (“event 1” in MK20) is indeed not included in this study as its start date is in November which we excluded here (see also our answer to the next comment). We have now clarified this point in Sect. 2 of the main text.

4. Table A1: Is it possible for a reflection event to occur in November?

It is indeed possible (see our answer to the previous comment), but we decided to focus on DJFM only, primarily because it is the period with the largest stratosphere-troposphere coupled variability (e.g. Fig 2a in Baldwin *et al.* 2003). Furthermore, the regimes analysis was previously defined only for DJFM (Lee *et al.* 2019, 2022), and cold extremes are typically colder during this period due to the seasonal cycle of temperature. We also note that it is often DJFM which is considered 'extended winter' for analysis of wintertime climate modes (e.g. https://climatedataguide.ucar.edu/climate-data/hurrell-wintertime-slp-based-northern-annular-mode-nam-index). In the revised manuscript we nonetheless now mention the possibility of the occurrence of wave reflection events in October and November not covered by our study in Sect. 2.

SH Lee, JC Furtado, and AJ Charlton-Perez. Wintertime North American weather regimes and the Arctic stratospheric polar vortex. *Geophysical Research Letters*, 46(24):14892–14900, 2019

Simon H Lee, Andrew J Charlton-Perez, Steven J Woolnough, and Jason C Furtado. How do stratospheric perturbations influence North American weather regime predictions? *Journal of Climate*, 2022. https://doi.org/10.1175/JCLI-D-21-0413.1

5. Table A1: It would be helpful for readers if the authors can add a column to show whether the event is associated with a major SSW.

Thank you for this very good suggestion. We have added this information to the table, basing it on NOAAs Sudden Stratospheric Warming Compendium data set (https://csl.noaa.gov/groups/csl8/sswcompendium/majorevents.html) with the addition of the January 2021 SSW described in Lee (2021). An SSW occurs during 9 of our 44 events; however, only 3 of the SSW events occur within 15 days of the onset of a reflection event, which is the period our analysis focuses on. None occur within 9 days. The total number of SSWs considered is 27. We have further added some discussion of the correspondence or lack thereof between our reflective events and SSWs to the text in Sect. 6, also in reply to several of the comments of Reviewer #2.

Baldwin, M. P., Stephenson, D. B., Thompson, D. W., Dunkerton, T. J., Charlton, A. J., & O'Neill, A. (2003). Stratospheric memory and skill of extended-range weather forecasts. *Science*, *301*(5633), 636-640.

Lee, S.H. (2021), The January 2021 sudden stratospheric warming. *Weather*, **76**: 135-136. https://doi.org/10.1002/wea.3966

6. Line 133 "except for a couple of days": I don't expect much difference, but just wondering whether the authors removed these cases in the following analysis, given they are not strictly downward? If a higher threshold (e.g., RI>1.5) is used, is it possible (v't')Canada is completely lower than 0?

During our 44 reflection events (defined as in the paper with a threshold of RI>1.0 and minimum 10 days in duration; total number of reflection event days: 1012), there are in total 21 days for which v'T' averaged over the Canadian sector is greater than or equal to zero. This corresponds to only 2.7% of all days during reflection events. When using a threshold of RI>1.5 the corresponding value is 9 days. Since the fraction of positive v'T' over the Canadian sector is very small for both thresholds, we do not expect this to affect our results – as also supported by the similarity between the surface anomalies for RI > 1.5 and for RI > 1 shown in Figs. R2–R5.

Due to the small fraction of v'T' reflective days (RI > 1) showing positive values over the Canadian box, on average during reflective days almost all grid-points in this region show negative values (see Figure R6a below). The few days with domain-averaged positive v'T' values still display regions of negative values, yet these are limited in their spatial extent. Figure R6b shows an example of a reflective event with positive average v'T' over Canada.

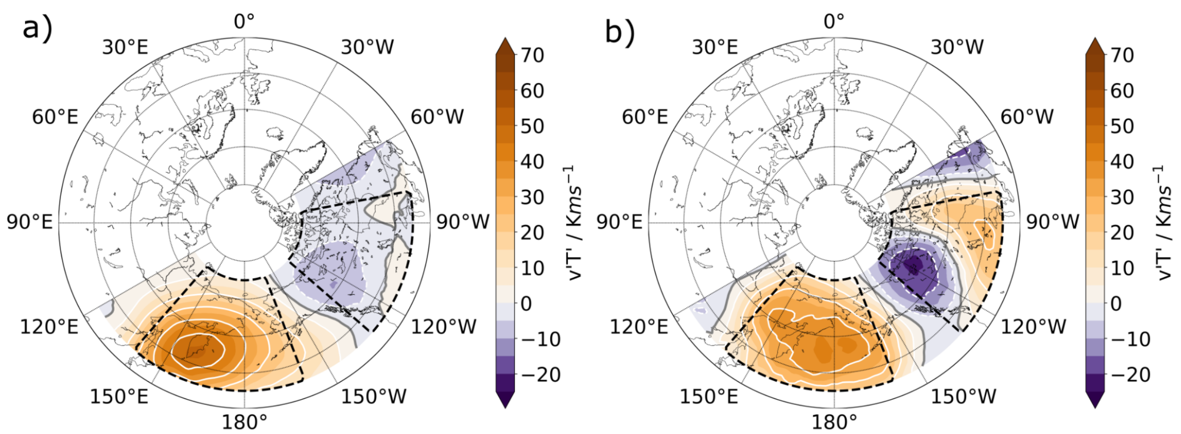

Fig. R6: a) Average over all reflection days (RI>1.0); b) example of an event with domain-averaged positive v'T' over Canada on 12.02.1980.

We believe that these very few cases have a negligible effect on our analysis. In the revised version of the manuscript we now state the exact number of cases when v'T' over Canada is positive.

Technical comments:

1. Line 45: Zhang et al. 2020 discussed the North American cold spells following stratospheric anomalies mainly through the lens of wave reflection by giving the pre-existing Arctic surface conditions, which is more relevant to the discussion in Line 47, 50, and 320.

We appreciate that in Zhang *et al.*, 2020, the authors discuss both the regional effect of planetary waves and the role of vertical propagation of wave anomalies in engendering North American cold spells. However, their analysis is framed around the role of SSWs, which we found have only a minimal link to wave reflection events leading to surface cold spells (only 4 of the "cold" reflection events correspond to a SSW). Moreover, there is little resemblance between the mid-tropospheric anomalies shown in Fig. 4 in Zhang *et al.*, and the anomalies we observe in our Fig. 7. We do nonetheless agree that the point Zhang *et al.* make concerning the role of surface pre-conditioning is highly relevant to the discussion of the origin of the mechanistic chain of events in Sect. 6. We now refer to Zhang *et al.*, (2020) in that paragraph. We have additionally modified the passage where Zhang *et al.* (2020) are cited in the introduction to acknowledge the reflection-like mechanisms discussed in that study.

2. Line 123: It would be better to state the term1 in Eq.1 (v't')sib >0?

As suggested, we have added the respective mathematical formulations to our list.

3. Line 124: similar comment here, (v't')can <0?

As suggested, we have added the respective mathematical formulations to our list.

4. Line 221 "Fig. A1": This is the first time Fig. A1 is called while Fig. A2-A4 are called earlier. The authors may adjust the order of the Appendix figures.

We mistakenly omitted referring to Fig. A1 in Sect. 2. We have now added this reference and the Appendix figures should appear in the order they are referred to. Note that due to new analysis performed during the revision the numbering of figures in the Appendix has changed.

5. Line 139 "vertical wind shear": Do the authors mean "vertical shear of zonal wind"?

Yes, that is indeed what we meant, and have now corrected it in the text both here and at a later instance.

6. Line 189 "the southernmost portion of the USA": Although this is not a piece of key information, one cannot see that in the figures. The authors may slightly enlarge the domain.

This was bad phrasing on our part. In Fig. 4d, which is the panel we are commenting on in that passage, there is actually not much signal south of the chosen domain (see Fig. R7 below showing a domain extended by 10 degrees to the South). What we meant to point to was actually the bottom right-hand corner of the shown domain. We have now edited this in the text: “Only Alaska and the south-eastern corner of the domain show neutral or weakly positive anomalies (Fig. 4d).”

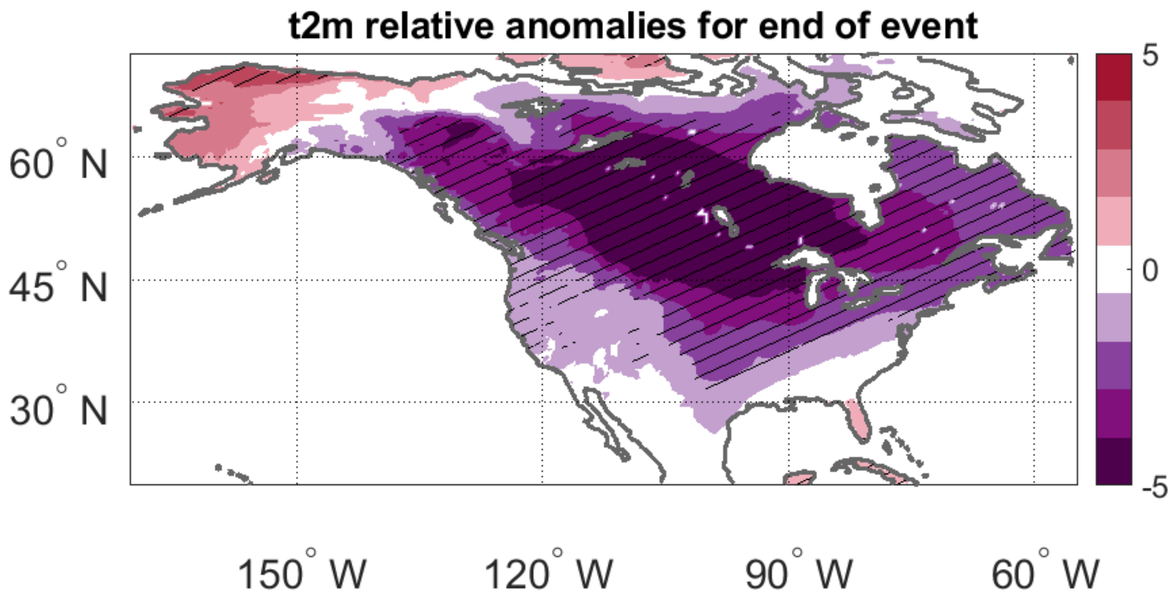

Fig. R7: Same as Fig. 4d, but for an extended meridional domain (20–70 °N).

7. Line 227-228 “using k-mean clustering of the leading 12 PCs of the daily Z500 anomalies”: I’m not sure my understanding is correct. Did the authors reconstruct new Z500 daily anomalies using the leading 12 EOFs and their time series, and then assign this new reconstructed daily field to 4 groups using clustering?

We followed a relatively standard method whereby we first computed the timeseries of the leading 12 PCs of the daily Z500 anomalies, then performed k-means clustering of these PCs to obtain the centroid coordinates in PC space, and then assigned each day to a ‘regime’ by determining the nearest cluster centroid by Euclidean distance in 12-dimensional PC space. We have added some clarifying discussion, although we emphasize that the method here is by no means novel. Some more detailed discussion can be found in Robertson et al. (2020), which we cite in the main text.

Robertson, A. W., Vigaud, N., Yuan, J., & Tippett, M. K. (2020). Toward identifying subseasonal forecasts of opportunity using North American weather regimes. *Monthly Weather Review*, *148*(5), 1861-1875.

8. Line 236-237 “before … than climatology”: Are the authors talking about the evolution from day 0 to day 5 and then to day 8 in Fig.8c?

We have rephrased this passage, which now reads: “The AkR regime is very unlikely immediately prior to the onset of the reflection event. It then approximately triples in frequency within the first five days after the event onset to become slightly more frequent than climatology and peaks in frequency around days 9 to 12. At all subsequent positive lags shown here, the regime is more frequent than in the days before the event onset (Fig. 8c).”

9. Figure A4: the authors may show the number of cases used to calculate each line in the figure legend.

We have added these numbers to the legend as suggested.