We would like to thank Reviewer 1 for the very helpful comments. Please find below our responses to all of the reviewer's comments.

Reviewer #1: Reviewer Comments

The manuscript investigates the role of the 2018 SSW in S2S prediction of the "Wet Iberia and Dry Scandinavia" precipitation pattern by using observations and a new S2S forecasts project SNAPSI. Some of the conclusions are - the early nudged ensembles, unlike the early free ensembles, successfully capture the "Wet Iberia and Dry Scandinavia" precipitation pattern, suggesting that an accurate forecast of stratospheric variability is helpful to improve S2S predictability of precipitation. In addition, the SSW related precipitation anomaly accounts for about 1/4 of the observed precipitation anomaly, suggesting the importance of other tropospheric variability.

The manuscript covers an important topic in assessing and quantifying the contribution of SSW on the "Wet Iberia and Dry Scandinavia" pattern in 2018. It's also overall well-written. However, I have some major comments about assessing the role of tropospheric internal variability using ensemble spread.

Major comments:

Comment 1: About the role of tropospheric internal variability. While I understand that the majority of the analysis focuses on ensemble mean to extract the SSW-induced variability, I would suggest the authors examine more on the ensemble spread since the ensemble spread includes the tropospheric variability. Some of the questions that the authors could investigate include - Can some ensemble members capture the observed magnitude of the precipitation pattern? (This seems to be yes given Fig. 9 but the question is different from Section 7) What does the ensemble mean versus ensemble spread tell about the role of tropospheric variability? (The role of tropospheric variability is assessed in the paper by using the linear regression approach. Does the model simulations tell the same attribution?) What are the model-to-model differences?

Reply:

- "Can some ensemble members capture the observed magnitude of the precipitation pattern?"

  To address this question, we show the most extreme data value for each set of forecast ensembles (blue whiskers in Fig. 1 in this reply). As can be seen, for seven of the eight SNAPSI models analyzed in this study, several individual ensemble members predict Iberian precipitation anomalies comparable to reanalysis (blue whiskers encompass ERA5 in Fig. 1). The only exception is GLOBO, for which none of the ensemble members capture the observed magnitude of the precipitation pattern, regardless of the initialization dates or the application of the nudging technique (blue whiskers remain below ERA5).

- "What does the ensemble mean versus ensemble spread tell about the role of tropospheric variability? (The role of tropospheric variability is assessed in the paper by using the linear regression approach. Does the model simulations tell the same attribution?)"

  To address this question, for each set of forecast ensembles, we show the multi-member average (color markers in Fig. 2 in this reply) and the 1.65 standard deviations around the multi-member average (black whiskers in Fig. 2). For ERA5, the predictable component of $\Delta\mathrm{PRCP}_{\mathrm{Iberia}}$ (solid black line in Fig. 2) and the 90% prediction intervals (dashed black line in Fig. 2) are shown. Here, 1.65 standard deviations is compared to the 90% prediction intervals because mean $\pm$ 1.65 standard deviations covers the middle 90% if we assume a normal distribution.

  Taking IFS as an example, the ensemble mean $\Delta\mathrm{PRCP}_{\mathrm{Iberia}}$ from the late free ensemble and the two nudged ensembles agree well with the predictable component of $\Delta\mathrm{PRCP}_{\mathrm{Iberia}}$ from ERA5. Given that the ensemble mean and the predictable component both represent SSW-induced variability, the agreement between them indicates that the role of stratospheric variability is well captured by IFS. With regard to the role of tropospheric variability, it is also well captured by IFS. In particular, the ensemble spread (black whiskers in Fig. 2) in general agrees with the corresponding prediction intervals from ERA5 (horizontal dashed lines in Fig. 2). This is, in general, the case for the rest of the models.

The only exception is GLOBO, for which the ensemble spread is much smaller than the prediction intervals from ERA5.

- "What are the model-to-model differences?"

  Please see our reply to the above two questions.

Based on the above results, we have added a discussion section (the new section 9) to the revised manuscript. Please see lines 468-506 of the revised manuscript.

Comment 2: About 2018 SSW compared to other SSWs. I would suggest the authors add some conclusions/discussion on how 2018 SSW is compared to other SSWs. As shown in Fig. 7, 2018 SSW is not particularly strong based on T100 but the SLP_Atlantic is the strongest. This suggests that the 2018 SSW is not very different from other SSWs and the SLP_Atlantic and precipitation pattern is mostly driven by other tropospheric variability. This information is here and there in the paper but would be very useful information to include in Conclusions/Discussion.

Reply: To address this question, we show a boxplot of ΔT_{100}, $\Delta \mathrm{SLP}_{\mathrm{Atlantic}}$, and $\Delta \mathrm{PRCP}_{\mathrm{Iberia}}$ for 46 historical SSW events from ERA5 (Fig. 3 in this reply). As can be seen, ΔT_{100} of the 2018 SSW is close to the median value, indicating the moderate strength of the sudden warming itself. By contrast, $\Delta \mathrm{SLP}_{\mathrm{Atlantic}}$ and $\Delta \mathrm{PRCP}_{\mathrm{Iberia}}$ following the 2018 SSW fall within the bottom and top 25th percentile of all historical events, respectively. The contrast between the moderate intensity of the case and the disproportionately strong surface anomaly indicates that the observed surface anomalies following the 2018 SSW are probably significantly driven by tropospheric variability.

Accordingly, we have made changes to the conclusion section in the revised manuscript. Please see lines 532-537 of the revised manuscript, where we write:

"This suggests that the exceptionally strong surface anomalies following the 2018 SSW are probably significantly driven by tropospheric variability. In fact, while the 2018 SSW ranks among the strongest on record in terms of the subsequent surface anomalies, the sudden warming itself is not particularly strong (Fig. S19). The contrast between the case's moderate intensity and its disproportionately strong surface impact also points to the substantial role of tropospheric variability in driving the observed surface anomalies following the 2018 SSW."

Comment 3: This might be minor but why all the free running models predict a stratospheric cooling?
Reply:

It has been pointed out by previous studies that the S2S prediction systems (from the S2S database) cannot forecast the 2018 SSW event more than 10 days in advance (Karpechko et al., 2018; Rao et al., 2020; Butler et al., 2020). This is because the 2018 SSW event was associated with primarily wavenumber-2 wave forcing that was not well predicted more than 7-10 days ahead of time (Butler et al., 2020). For example, according to Fig. 1a of Karpechko et al. (2018) (or Fig. 4 in this reply), at a lead time of 18 days (initialized from 25 January 2018), none of the S2S models from the S2S database predict an SSW (grey dots in Fig. 4). Our results show that this is also the case for SNAPSI models because none of the SNAPSI models' ensemble mean forecast from 25 January 2018 (at a lead time of 18 days) predicted an SSW.

It has also been pointed out by a previous study that, initialized at 25 January 2018, all S2S models show a positive Stratospheric Polar Vortex (SPV) error [green dots in Fig. 10b of Butler et al. (2020); or Fig. 5 in this reply]. The positive SPV error is consistent with the stratospheric cooling predicted by SNAPSI models early free ensembles. Understanding the origin of the strong vortex error is beyond the scope of this study, and here are just a couple of thoughts.

- (a) Errors in the stratosphere forecasts might have contributed to errors in the SPV forecasts.

  Butler et al. (2020) found that the SPV error (too strong polar vortex winds) correlates significantly with the QBO error (see the correlation values in the upper-right corner of Fig. 5 in this reply). This suggests that the SPV error is larger in those systems that have a larger QBO error. Hence, the SPV state can be better predicted in those systems that better predict the QBO state, highlighting the importance of reducing errors within the stratosphere (particularly in the QBO forecasts).

- (b) Errors in the troposphere (particularly in the wave-2 activity) forecasts might have contributed to errors in the SPV forecasts.

  Wu et al. (2024) compared two clusters of ensemble forecasts, one with ensemble members that successfully predict the SSW ('SSW cluster') and one that predicts a strong vortex state ('strong vortex cluster'), to understand the origin of the different predictions of the vortex strength. Both the SSW cluster and the strong vortex are initialized 16 days before the onset of the 2018 SSW event [Fig. 4a of Wu et al. (2024); or Fig. 6a in this reply], close to the initialization date of the early free ensemble forecasts analyzed in this study. It turns out that the two clusters do not differ significantly in wave-1 activity (Fig. 6b in this reply). By contrast, the two clusters differ significantly in wave-2 activity around the onset of the 2018 SSW. In particular, both clusters show an initial increase (Fig. 6c in this reply) in wave-2 activity, but the increase in the strong vortex cluster is much weaker and less persistent than that in the SSW cluster. These results suggest that the strong vortex state in the strong vortex cluster may arises from the errors in the tropospheric wave-2 activity forecasts. In particular, the amplification of tropospheric wave-2 activity in the strong vortex cluster lacks the strength and persistence required to disrupt the polar vortex state.

  In this sense, the question of why the early free ensembles predict a stratospheric cooling may be understood as follows: first of all, the early free ensembles are initialized under a strong vortex state [indicated by the positive NAM index around the orange line in Fig. 5 of Hitchcock et al. (2022); or Fig. 7 in this reply]. The strong vortex state is then maintained due to the weak and short-lived nature of the wave-2 forcing. As a consequence, a stratospheric cooling is predicted in the early free ensembles.

References

Butler, A. H., Lawrence, Z. D., Lee, S. H., Lillo, S. P., and Long, C. S.: Differences between the 2018 and 2019 stratospheric polar vortex split events, Quarterly Journal of the Royal Meteorological Society, 146, 3503–3521, https://doi.org/https://doi.org/10.1002/qj.3858, 2020.

Hitchcock, P., Butler, A., Charlton-Perez, A., Garfinkel, C. I., Stockdale, T., Anstey, J., Mitchell, D., Domeisen, D. I. V., Wu, T., Lu, Y., Mastrangelo, D., Malguzzi, P., Lin, H., Muncaster, R., Merryfield, B., Sigmond, M., Xiang, B., Jia, L., Hyun, Y.-K., Oh, J., Specq, D., Simpson, I. R., Richter, J. H., Barton, C., Knight, J., Lim, E.-P., and Hendon, H.: Stratospheric Nudging And Predictable Surface Impacts (SNAPSI): a protocol for investigating the role of stratospheric polar vortex disturbances in subseasonal to seasonal forecasts, Geoscientific Model Development, 15, 5073–5092, https://doi.org/10.5194/gmd-15-5073-2022, 2022.

Karpechko, A. Y., Perez, A. C., Balmaseda, M., Tyrrell, N., and Vitart, F.: Predicting sudden stratospheric warming 2018 and its climate impacts with a multi-model ensemble, Geophysical Research Letters, p. 2018GL081091, 2018.

Rao, J., Garfinkel, C. I., and White, I. P.: Predicting the Downward and Surface Influence of the February 2018 and January 2019 Sudden Stratospheric Warming Events in Subseasonal to Seasonal (S2S) Models, Journal of Geophysical Research: Atmospheres, 125, e2019JD031 919, https://doi.org/https://doi.org/10.1029/2019JD031919, e2019JD031919 2019JD031919, 2020.

Wu, R. W.-Y., Chiodo, G., Polichtchouk, I., and Domeisen, D. I. V.: Tropospheric links to uncertainty in stratospheric subseasonal predictions, Atmospheric Chemistry and Physics, 24, 12 259–12 275, https://doi.org/10.5194/acp-24-12259-2024, 2024.

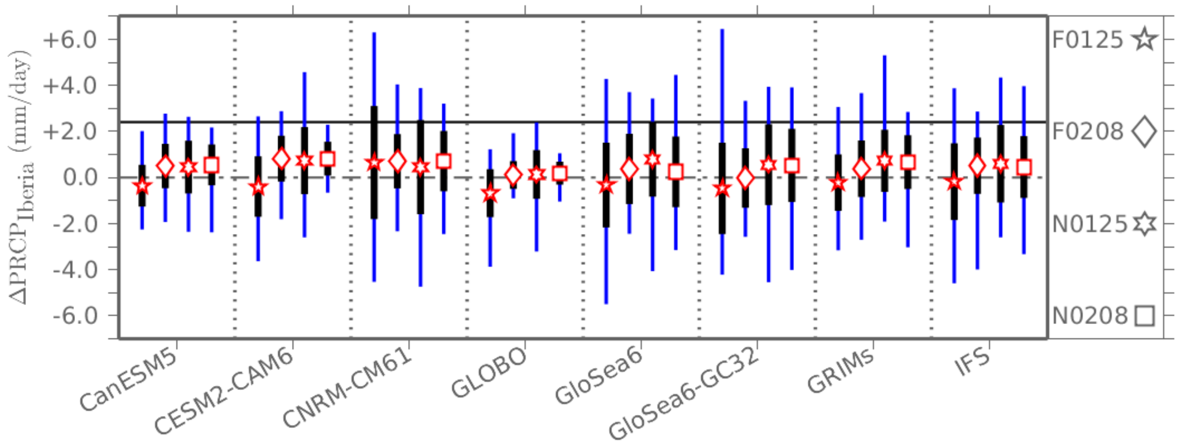

Figure 1: $\Delta\mathrm{PRCP}_{\mathrm{Iberia}}$ by individual SNAPSI models. For each ensemble, the color marker is the multi-member average, the black whiskers correspond to 1 standard deviation around the multi-member average, and the blue whiskers extend to the most extreme data value. F0125: early free ensemble; F0208: late free ensemble; N0125: early nudged ensemble; N0208: late nudged ensemble. The solid black horizontal line indicates $\Delta\mathrm{PRCP}_{\mathrm{Iberia}}$ from ERA5.

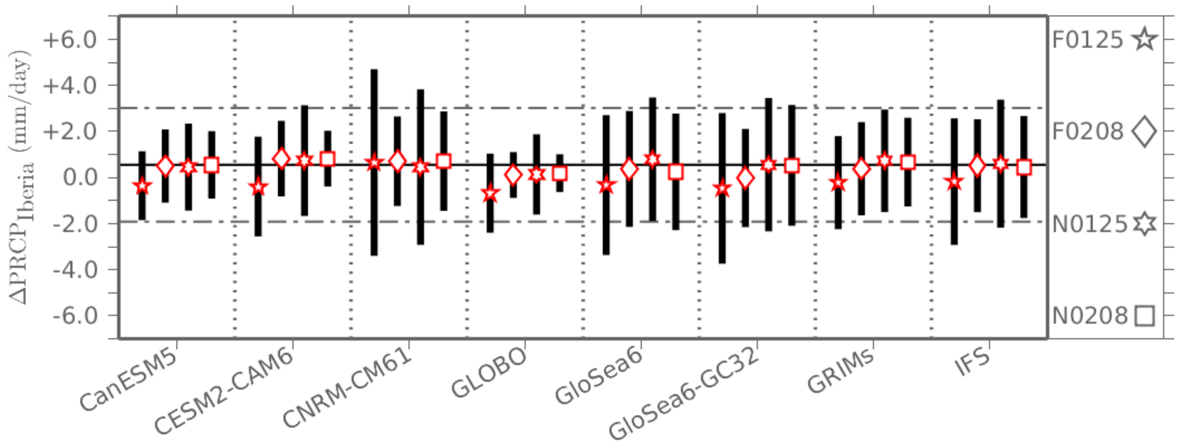

Figure 2: $\Delta\mathrm{PRCP}_{\mathrm{Iberia}}$ by individual SNAPSI models. For each ensemble, the color marker is the multi-member average, and the black whiskers correspond to 1.65 standard deviations around the multi-member average. F0125: early free ensemble; F0208: late free ensemble; N0125: early nudged ensemble; N0208: late nudged ensemble. The solid black horizontal line indicates the predictable component of $\Delta\mathrm{PRCP}_{\mathrm{Iberia}}$ from ERA5, and the dashed black lines indicate the 90% prediction intervals (the predictable component and the prediction intervals are the same with those indicated by the red dot and the corresponding blue shadings in Fig. 7b in the manuscript, which are obtained from a linear regression model built between $\Delta\mathrm{PRCP}_{\mathrm{Iberia}}$ and ΔT_{100} using historical SSW events from ERA5).

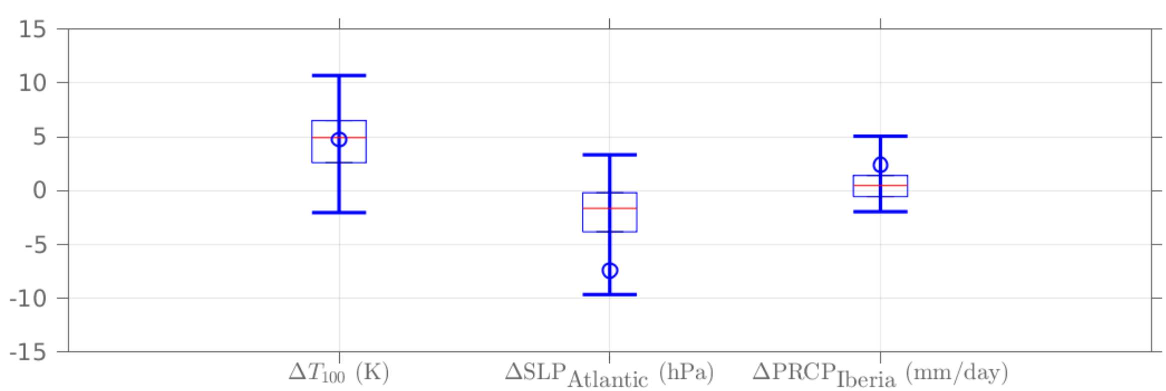

Figure 3: ΔT_{100}, $\Delta\mathrm{SLP}_{\mathrm{Atlantic}}$, and $\Delta\mathrm{PRCP}_{\mathrm{Iberia}}$ for 46 historical SSW events from ERA5. On each box, the red line is the median, the edges of the box are the 25th and 75th percentiles, and the whiskers extend to the most extreme data value. The blue circle indicates the value for the 2018 SSW event.

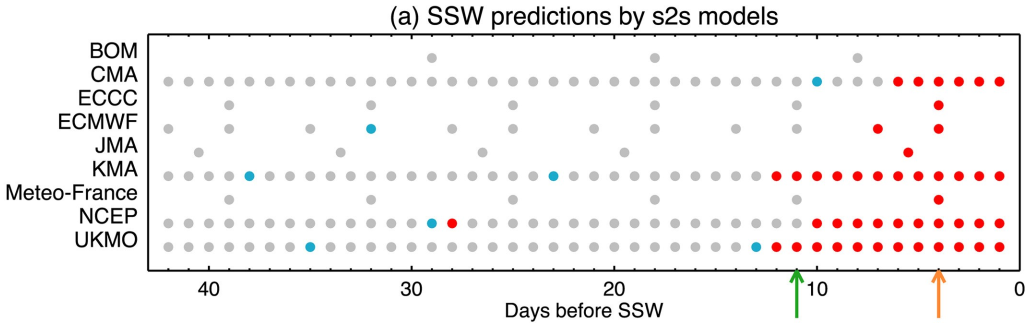

Figure 4: Predictions of the 2018 SSW by individual S2S models. Red dots indicate that an SSW is predicted within 3 days from the actually observed event. Blue dots indicate that an SSW is predicted on a different date than the actual event. Gray dots indicate that no SSW is predicted. Arrows point to forecasts used in the multimodel ensembles. Figure 1a from Karpechko et al. (2018).

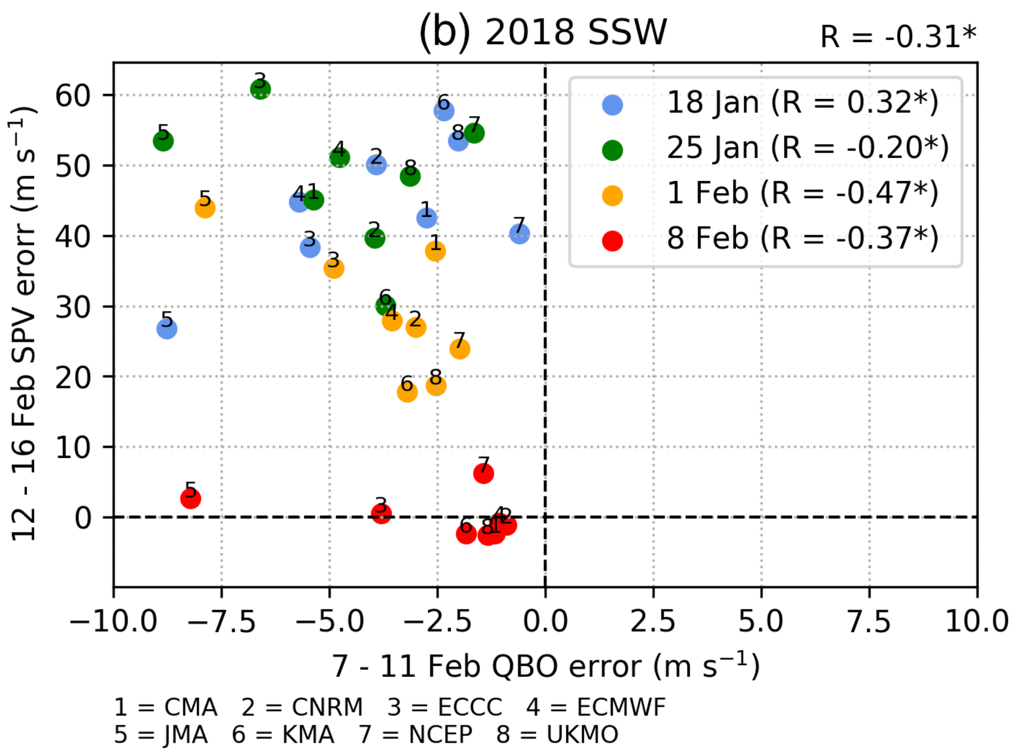

Figure 5: Scatter plots of the averaged error in the QBO forecasts versus error in the zonal-mean zonal winds at 60°N and 10-hPa (Stratospheric Polar Vortex) for the 2018 SSW. Correlation values are calculated using the multi-model ensemble of 201 members. An asterisk indicates that the correlation is significant at the 95% confidence level. Figure 10b from Butler et al. (2020).

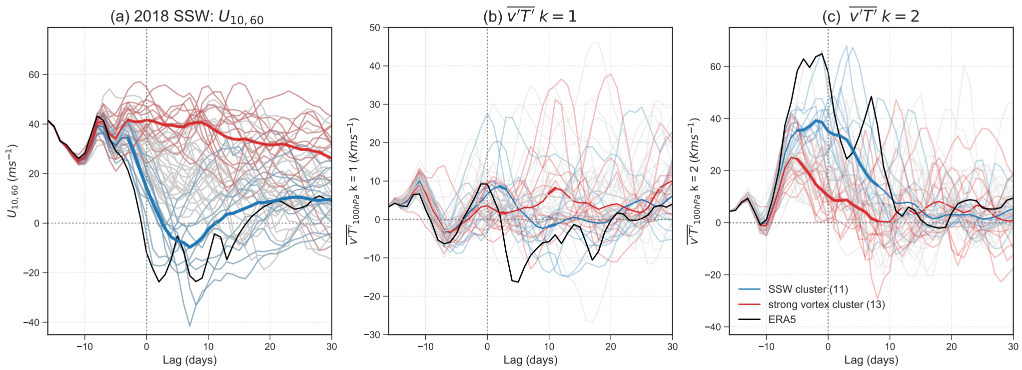

Figure 6: (a) A comparison of the zonal-mean zonal winds at 60°N and 10-hPa between strong vortex cluster (red), SSW cluster (blue), and ERA5 (black). (b and c) same as panel a but for (b) wave-1 and (c) wave-2 component of the zonal average of meridional eddy heat fluxes at 100-hPa averaged over 45-75°N, respectively. The vertical line denotes the central date of the SSW on 12 February 2018. Figure 4 from Wu et al. (2024).

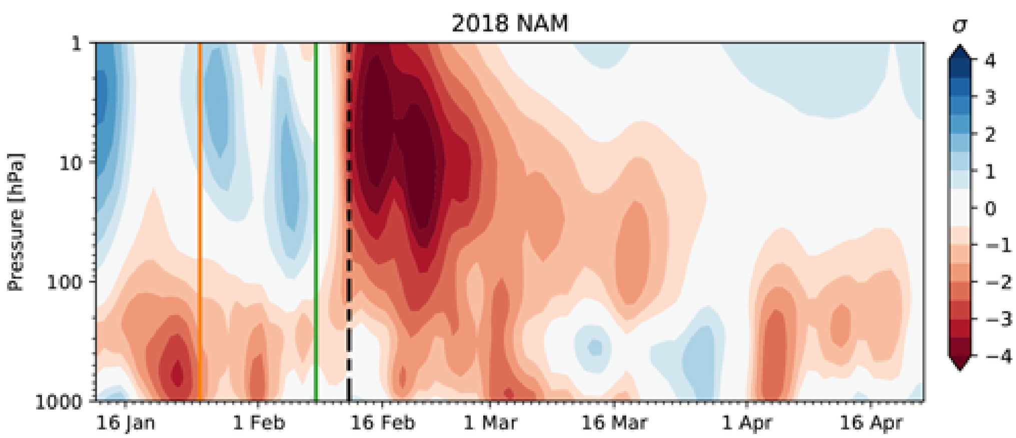

Figure 7: NAM indices during the February 2018 boreal major warming. The vertical dashed-dotted black line indicates the date of the wind reversal at 10-hPa, 60°N. The vertical green and orange lines indicate the requested initialization dates. Figure 5 from Hitchcock et al. (2022).