Dear editor and reviewers,

Thank you very much for the detailed comments on the original manuscript and for considering publication after revision of the manuscript. Please find below the answers to the comments of all three reviewers on our initial manuscript with the title "How do different pathways connect the stratospheric polar vortex to its tropospheric precursors?". For your convenience, we cited the reviewers' comments in separate sections, and responded in bold font and indent.
Additionally, at the end of this document, we included a detailed description of the adapted analysis of vertical wave propagation (pages 22-23), which was conducted based on the reviewers' comments.

We believe that the revised manuscript will strongly improve due to the constructive comments of the reviewers.

Reviewer 1

The manuscript presents a quantitative and mechanistic analysis of mean-sea-level-pressure precursor patterns observed before sudden stratospheric warmings, comparing results between ERA5 and ICON model simulations. One important contribution of the study is therefore to demonstrate the potential of ICON for studying stratosphere-troposphere coupling and for seasonal forecasting in general. It is noteworthy that the results indicate anomalous surface weather to be present for up to three months after the precursor patterns first occur. Additionally, the relatively large sample size obtained from the model simulations allows the authors to resolve the results for different months, indicating the need for careful interpretation when composites are based on events from an entire winter season.

However, parts of the mechanistic explanations were unconvincing to me in the current form. A discussion of how the applied localized wave activity analyses relate to planetary wave diagnostics would be beneficial (please see my comments below).

Overall, I found the paper well-written and interesting to read. Therefore, I recommend that the paper be published once the major concerns are addressed.

Major comment 1:

You have demonstrated that a strong positive difference of MSLP over the Ural region and the Pacific region, which is defined as the "precursor index", is followed by "significant tropospheric NAM signals for up to 3 months after the original anomaly in the precursor index" (ll. 184f). I agree that this is remarkable and that it can likely represent an important source of predictability for seasonal forecasting.

However, it would be beneficial to address or discuss the potential role of any confounding processes. Specifically, it would be worthwhile to explore whether ENSO teleconnections could project on both the identified precursor regions, particularly over the Pacific, and anomalies of the polar vortex.

> **We want to thank Reviewer 1 for this comment, as we agree that it is important to discuss potentially confounding processes in the manuscript. Different studies have connected El Niño with a deepened Aleutian low and a ridge downstream over North America (e.g.**

O'Reilly, 2018; Soulard et al., 2019). To further investigate how ENSO relates to anomalies in the precursor regions and of the stratospheric polar vortex, we calculated the ENSO3.4 index and correlated it to the Precursor index, MSLP in the Aleutian and Ural area as well as the NAM at 10hPa. Please find our results for the ERA5 reanalysis in Figure 1. The only significant correlation between the different indices is found for ENSO3.4 and the Aleutian region, where El Niño is connected to a deepened Aleutian low in December. The Ural region does not exhibit any significant correlation to the ENSO index, so that we don't assume that ENSO is a confounding process. The non-significant correlation between ENSO and NAM at 10hPa in January seems to be modulated via the Aleutian low, i. e. El Ni Niño is connected to a deepened Aleutian low in December, which then favours a weaker stratospheric polar vortex in January. A further confounding factor could be global warming, but by detrending the data before calculating the correlations we can exclude this factor.

We will include a short discussion of this analyses into the manuscript.

O'Reilly, C.H. Interdecadal variability of the ENSO teleconnection to the wintertime North Pacific. *Clim Dyn* 51, 3333–3350 (2018). https://doi.org/10.1007/s00382-018-4081-y

Soulard, N., Lin, H. & Yu, B. The changing relationship between ENSO and its extratropical response patterns. *Sci Rep* 9, 6507 (2019). https://doi.org/10.1038/s41598-019-42922-3

Major comment 2:

I have some concerns about the Plumb flux analysis in the second part of the study. It would be helpful if the authors provided more information on the formulation used for the vertical Plumb flux and referenced previous successful applications.

We are happy to provide more information on the Plumb flux. Nevertheless, based on the comments of the other reviewers, we will change the manuscript so that it is no longer used. From the perspective of Eurasian surface anomalies and their influence on stratosphere-troposphere coupling, Cohen et al. 2007 was potentially one of the first influential applications of the vertical component of the Plumb 1985 flux. We take the freedom to cite from there:

The vertical component of the WAF is given by

$$F = \Omega p \sin(2\phi) S^{-1}\{v'T' - [2\Omega a \sin(2\phi)]^{-1}\partial(T'\Phi')/\partial\lambda\},$$

where Ω is the angular velocity of rotation, p is pressure, φ is latitude, v is meridional velocity, T is temperature, a is the mean radius of the earth, Φ is geopotential, λ is longitude, and primes indicate the departure from the local climatological mean. The static stability $S = \langle -(p/H)\partial T/\partial p + \kappa T/H \rangle$, where angle brackets indicate the 20°–90°N polar-cap and climatological average, $\kappa = R/cp$ and H is the scale height. The field F is useful for localizing in longitude and latitude the source of vertically propagating stationary planetary Rossby waves. The zonal mean of F is proportional to the meridional eddy heat flux and the vertical component of the quasigeostrophic Eliassen–Palm flux.

Cohen, J., Barlow, M., Kushner, P. J., & Saito, K. (2007). Stratosphere–troposphere coupling and links with Eurasian land surface variability. Journal of Climate, 20(21), 5335-5343.

Our formulation is basically the same, while we further describe the basic state to calculate the perturbations (primed quantities) in the manuscript: We use a 91-day rolling mean annual cycle of the full ERA5 data or full ICON ensemble, respectively. We consider only low-frequency or large-scale contributions to the Plumb Flux and therefore apply a low-pass filter sensitive to timescales above 10 days to the perturbations.

Additionally, we provide a list of a few very recent studies where Plumb 1985 flux has been applied in the context of troposphere-stratosphere coupling, showing that it is widely and actively used by reputable authors in the field:

Finke, Kathrin, Hannachi, Abdel, and Hirooka, Toshihiko, 2023, "Exceptionally persistent Eurasian cold events and their stratospheric link" Asia-Pacific Journal of Atmospheric Sciences 1976-7951

Ding, Xiuyuan, Chen, Gang, Zhang, Pengfei, Domeisen, Daniela I. V., and Orbe, Clara, 2023, "Extreme stratospheric wave activity as harbingers of cold events over North America" Communications Earth & Environment Vol. 4, No. 1, 2662-4435

Gastineau, Guillaume, Frankignoul, Claude, Gao, Yongqi, Liang, Yu-Chiao, Kwon, Young-Oh, Cherchi, Annalisa, Ghosh, Rohit, Manzini, Elisa, Matei, Daniela, Mecking, Jennifer, Suo, Lingling, Tian, Tian, Yang, Shuting, and Zhang, Ying, 2023, "Forcing and impact of the Northern Hemisphere continental snow cover in 1979–2014" The Cryosphere Vol. 17, No. 5, pp 2157, 1994-0424

Zhang, J., Orsolini, Y.J., Limpasuvan, V. *et al.* Impact of the Pacific sector sea ice loss on the sudden stratospheric warming characteristics. *npj Clim Atmos Sci* 5, 74 (2022). https://doi.org/10.1038/s41612-022-00296-w

Unlike the Eliassen-Palm flux (divergence), I am unaware of how localized Plumb flux theoretically relates to changes in the zonal mean flow, which could be discussed more in detail (e.g., see ll. 234f: "The connection between wave activity in 100 hPa and the stratospheric polar vortex is not as straightforward as one might expect, as increased Plumb flux does not necessarily imply a weakening of the stratospheric polar vortex.").

In our understanding, Plumb 1985 flux (as well as the flux defined by Takaya and Nakamura (2001)) is a generalization of the original Eliassen-Palm relation with the aim to be zonally resolved. These zonally resolved quantities describe a more generalized form of wave activity and therefore do not have a direct relation to the zonal wind via their divergence. Nevertheless, authors of these studies describe that these kinds of wave activity can be related to westerly momentum that adds to a zonal mean basic flow. Thus, additional wave activity can have an influence on the westerly flow (in our case the stratospheric polar vortex). We intended to build on that to find and/or explain physical pathways of our precursor index.

I wonder if the presented precursor geopotential anomalies in Figure 1 might reflect anomalous planetary wave activity, given their projection on the climatological wave-1. An intensified wave-2 pattern could possibly explain why anomalies are larger over the Ural region than over the Pacific, potentially questioning if it were in fact two distinct mechanisms that relate the observed precursor signals over the Ural and over the Pacific to the stratospheric anomalies.

Thank you very much for this comment. Based on it, we calculated the wave amplitudes related to our Precursor index. As expected, we see that the Precursor index strongly projects onto the climatological wave-1, with increased wave-1 amplitudes during high Precursor index states and decreased wave-1 amplitudes during low index states (cf. Table 1 and 2). The overall wave-2 amplitude changes are small for the Precursor index. However, as you suspected wave-2 amplitude differences could explain, why the Ural region affects the stratosphere stronger than the Aleutian region, although the Aleutian region is connected to larger wave-1 amplitude changes (cf. Table 3,4 and 5,6). Whereas a strong (weak) Ural high is linked to increased (decreased) wave-1 and wave-2 activity (Table 3 & 4), a low (high) Aleutian low is connected to increased (decreased) wave-1 amplitudes, but decreased (increased) wave-2 amplitudes (Table 5 & 6). Hence, for MSLP anomalies in the extended Aleutian region wave-2 can counteract wave-1 activity. This can explain why climatologically the Ural region is the main tropospheric driver of the stratospheric polar vortex.

As we think it is very beneficial for the understanding of the involved mechanisms, we will also include this discussion into our revised manuscript.

In addition, I have two suggestions regarding the Plumb flux analyses. First, providing typical orders of magnitude for the Plumb flux and its variability in Figure 5 would be helpful in interpreting the observed anomalies of up to 0.02 m^2/s^2. Second, providing typical correlation coefficients associated with the plotted regression coefficients in Fig. 6 may aid in interpreting the amount of co-variability.

Thank you for pointing out this issue. Related to comments from other reviewers we decided not to show Plumb fluxes in the manuscript. It will be replaced by zonal mean Eliassen-Palm flux diagnostics as defined by the Transformed Eulerian Mean equations. A discussion can be found at the end of the response document. Nevertheless, we will provide the requested information for other plots where applicable in the updated manuscript.

Minor comments:

- ll. 39-41: maybe you could be a bit more clear about what you mean by "not predictable beyond deterministic timescales"; (e.g., in the provided reference, Karpechko 2018 concludes that "[...] days 8–12 before the events, the SSWs can be considered predictable in a probabilistic sense. After day 7 the predictability can be considered to be close to deterministic.")

  This statement is based on Domeisen et al. (2020b). To make things clearer, we reformulated the sentence to "However, Domeisen et al. (2020b) show that stratospheric extreme events themselves tend exhibit similar predictability to tropospheric weather and, in particular, SSW events tend to be less predictable."

- ll. 77: Have you tested using only winters where you have both ICON simulations and ERA5 data available (1979/80 to 2016/17)? Alternatively, you may want to add a small comment why you think this is not necessary.

  We performed some basic tests using ERA5 data only up to winter 2016/2017 and received the same qualitative results. Due to the very large variability of the stratospheric polar vortex in winter, we decided to maximise our sample size by including the whole ERA5 period available at this point in time.

- ll 86-89: You may want to consider stating which spatial domain was used to calculate the EOFs, if the data was area-weighted for the calculation of the EOFs and how ERA5 and ICON data was deseasonalized.

  We improve the description of the calculation of the NAM indices by adding information about the area, the removing of the seasonal cycle and the area weighting. The text of former l86-89 was extended accordingly:

  > **The basic mean of quantifying the strength of the stratospheric polar vortex in this study is the Northern annular mode (NAM), which is the dominant pattern of dynamic variability in the extratropical Northern Hemisphere (NH). Based on Baldwin and Thompson (2009), we used an empirical orthogonal function (EOF) analysis of the daily, zonally-averaged geopotential anomalies at each pressure level over the area 20°N to 90°N to calculate the NAM. Geopotential height anomalies were obtained by removing the mean seasonal cycle at each grid point, which is calculated for each data set separately (ERA5 and ICON) by averaging the respective geopotential heights day by day over all years. To ensure equal-area weighting, the data was weighted by the square root of the cosine of latitude before performing the EOF analysis. The standardized corresponding timeseries of the first principal components are the NAM indices.**

- Fig. 1: I'm unsure about what you mean by "Composites of [...] MSLP [...] for the same January weak vortex events in ERA5 and ICON". If I'm correct then events in ERA5 and ICON are basically independent (sorry for my ignorance if not)?

  Thanks for pointing out this misleading wording. "The same" refers to the same events as were used to generate Figure 1 a and b. The events were calculated separately for ERA5 and ICON. We will add "as in (a) and (b)" to the caption.

- Fig. 2: The presented composites are based on a strongly negative January-mean NAM. Given the relatively long timescales in the stratosphere, I would imagine that in some cases, the negative NAM in January results from a SSW that happened already in December. In that case, part of the surface signal observed in December could in principle correspond already to a surface response of the SSW. In case you agree, you could consider adding a comment. However, I see that it is not strictly necessary as you will address the time lag between tropospheric precursor and stratospheric anomaly more in detail in Fig. 2.

  We assume that the comment relates to Figure 1. We agree that SSW events in late December could potentially impact the NAM in January. However, we would like to point out, that the downward propagation of stratospheric signals into the troposphere usually happens one to two months after the stratospheric extreme event. Therefore, we think it is highly unlikely that the precursor patterns are generated by weak vortex events in December. Moreover, the surface response to SSW strongly differs from the precursor patterns visible in Figure 1 c) and d) (cf. Fig. 1a in Afargan-Gerstman, H., & Domeisen, D. I., 2020).

  > **Afargan-Gerstman, H., & Domeisen, D. I. (2020). Pacific modulation of the North Atlantic storm track response to sudden stratospheric warming events. *Geophysical Research Letters*, *47*(2), e2019GL085007. https://doi.org/10.1029/2019GL085007**

- ll. 130: A two-sample significance test could provide additional confirmation that the observed differences between ERA5 and ICON are explainable by sampling.

When we look into the single ICON ensemble members, we see that signals are also noisier in ICON. Ensemble approaches usually make the overall signal smoother due to the larger sample size. Therefore, we note that a qualitative agreement between ICON and ERA5 is given. Accordingly, we adjusted the original statement to: "The larger sample size in ICON results in a smoother image, while the qualitative agreement between ICON and ERA5 is given. "

- Fig. 2: I expect "@ 100hPa" to be wrong.

    Yes, this is a mistake and will be corrected in the new manuscript.

- Fig. 2: Is it surprising that ICON shows significant regression coefficients even after 60 days and longer? (How much longer?)

    We agree with the reviewer's comment - this is surprising. Nevertheless, the regression coefficients become gradually weaker indicating its actual very low influence. The statistically significant relation ends at about 80 days. Furthermore, we note that this particular relation appears only in December. Other months are more similar to the results of ERA5.

- Fig. 4: How many events are shown in each composite?

    For ERA5 there are 9 events for each month and for the ICON ensemble 40 events each. We will include this information in the revised manuscript.

- ll. 181-182: Is this because the precursor pattern directly projects onto the NAM or is it because the precursor pattern connects with a polar vortex anomaly which then leads to an altered NAM index?

    This is a good remark, thank you. In our opinion, the precursor pattern partly projects onto the tropospheric NAM. This is particularly the case for the Aleutian region (cf. Fig. 1c and e.g., https://www.cpc.ncep.noaa.gov/products/precip/CWlink/daily_ao_index/ao.loading.shtml). Hence, a deepened Aleutian low is connected to a negative NAM in the troposphere and vice versa. Moreover, the Ural region partly overlaps with the low-pressure centre of the surface NAM. Thus, high pressure in the Ural region also favours a negative NAM and vice versa. Overall, this seems to suggest that the precursor pattern directly projects onto the tropospheric NAM, with a high (low) precursor index favouring a negative (positive) NAM. We will mention this in the revised manuscript and furthermore adjust ll. 181-182 accordingly.

- Fig. 6: Maybe you want to add "at 100 hPa" after meridional mean Plumb flux.

    That is a good idea, thank you.

- Fig. 7: Only the 10% largest absolute correlation coefficients are shown. What is the obtained threshold for the correlation coefficient?

    The threshold values in terms of the 10th percentile of absolute correlation coefficients are listed referring to the respective figures:

    Figure 7a (ERA5, Ural): 0.19

    Figure 7b (ICON, Ural): 0.12

    Figure 8a (ERA5, Aleutian): 0.16

> **Figure 8b (ICON, Aleutian): 0.14**
>
> **These correlation coefficients are low, which is related to the high amount of noise introduced from daily values. Even the highest correlation between slp and geopotential in the origin region of either Ural or Aleutian reach only 0.82 … 0.89.**

- l. 287: "and/or": this might be beyond the scope of your study, but I imagine it would be indeed interesting to understand whether the combined pattern emerges from the composite averaging or really occurs combined during individual events

  > **As our two regions do not show any correlation, we assume that the pattern emerges from composite averaging. Yet, when we look at single events in ERA5, there can be some events where we see both, a Ural and an Aleutian precursor. This motivated us to write "and/or". As the patterns are, however, uncorrelated we decided to write only "or" in the revised manuscript.**

- l. 307: "Vertical planetary wave fluxes do not explain the link between extended Aleutian MSLP and the stratospheric NAM. Here it is a pattern that resembles the two main centers of action of the PNA which directly penetrates into the stratosphere.": Fig. 8c does show a clear displacement, i.e., a wave 1 signal in the stratosphere and to my understanding, Fig. 8b does not exclude the possibility that anomalous wave 1 activity is also present in the troposphere, even if strongly induced via the PNA. It would be useful if you could clarify/ discuss this when considering my major comment 2.

  > **As discussed in the response to major comment 2, anomalies in MSLP in the Aleutian region are strongly related to anomalous wave-1 activity in the troposphere. However, a deepened Aleutian low is additionally connected to increased wave-3 activity (cf. Table 5). Hence, we conclude that anomalous wave-1 activity is induced via the PNA. We will include this into the revised manuscript.**

- General comment: To my understanding, (quasi-)observed SSTs are prescribed in the simulations. To what extent do you think does this need to be considered when interpreting the good agreement between ERA5 and ICON results?

  > **Yes, the SSTs and sea ice cover are prescribed using CMIP6 standard protocol for AMIP simulations. More specifically, it is based on a merged observational product of the UK MetOffice HadISST and NCEP OI2. Although SSTs and sea ice are prescribed in ERA5 and ICON, one cannot expect that the coupling of stratosphere and troposphere are automatically simulated realistically in ICON. Therefore, too many atmospheric processes are involved. Additionally, it is standard practice to investigate stratospheric circulation in AMIP-style simulations.**
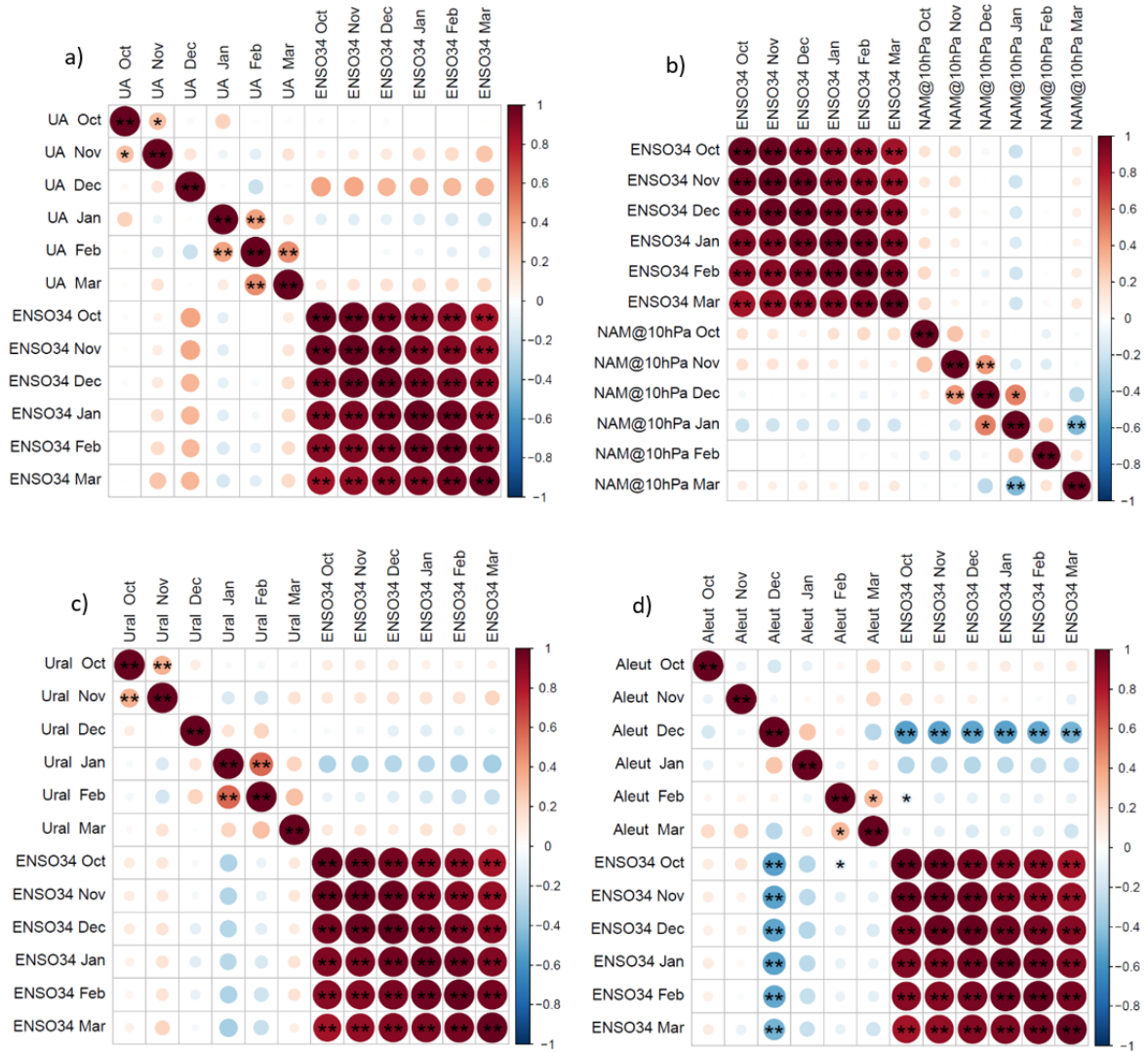
Figure 1. Spearman correlation between the detrended, monthly mean ENSO3.4 index and the Precursor index (a), the NAM at 10 hPa (b), Ural sea level pressure (c) and the Aleutian sea level pressure (d). Circle size and colour correspond to the strength of the correlation coefficients. Blue (red) colours denote a negative (positive) correlation and the stars denote significance, with one star indicating statistical significance at the 95% level and two stars statistical significance at the 99% level.

	ERA5				ICON			
k	1	2	3	4	1	2	3	4
Nov	**34.6%**	**20.4%**	**-14.2%**	-23.8%	**30.0%**	**21.9%**	**-10.5%**	48.4%
Dec	**50.9%**	**0.6%**	**-10.0%**	9.2%	**47.8%**	**8.5%**	**-15.2%**	46.7%
Jan	**67.8%**	-11.0%	**-11.7%**	**57.4%**	**57.7%**	6.6%	**-33.7%**	**59.8%**
Feb	**74.3%**	-5.4%	**-11.8%**	**117.7%**	**52.9%**	3.4%	**-7.7%**	**54.8%**
Mar	**80.6%**	**12.0%**	-2.7%	-13.3%	**75.5%**	**14.2%**	1.9%	47.1%
NDJFM	**61.6%**	**3.4%**	**-10.0%**	**28.2%**	**52.9%**	**11.0%**	**-13.2%**	**51.3%**

Table1. Wave amplitude difference in percent between months with a high Precursor index (80th percentile) and the climatology over all years for ERA5 and ICON ensemble simulations. The wave amplitudes are calculated from the monthly mean 500hPa geopotential height fields and the wave amplitudes of the different wave numbers (1-4) are averaged over the precursor core region of 50°-70°N. The colour denotes the sign of the change (red=positive, blue=negative) and bold font indicates agreement between ERA5 and ICON in terms of sign.

	ERA5				ICON			
k	1	2	3	4	1	2	3	4
Nov	**-35.7%**	**-5.1%**	**18.8%**	**31.2%**	**-18.8%**	**-31.9%**	**18.5%**	**20.3%**
Dec	**-26.1%**	**14.4%**	**29.9%**	**-19.1%**	**-35.0%**	**1.1%**	**20.2%**	**-29.5%**
Jan	**-22.1%**	3.9%	**26.6%**	-22.8%	**-23.3%**	-4.4%	**18.1%**	11.5%
Feb	6.6%	13.7%	-3.5%	**-29.6%**	-4.3%	-2.8%	13.8%	**-8.9%**
Mar	**-36.4%**	**-3.6%**	43.2%	**-14.9%**	**-36.4%**	**-3.6%**	-4.7%	**-19.2%**
NDJFM	**-23.2%**	4.6%	**23.5%**	**-11.0%**	**-23.9%**	-8.3%	**13.1%**	**-5.3%**

Table2. Same as Table 1 but for low Precursor index (20th percentile) vs. climatology

	ERA5				ICON			
k	1	2	3	4	1	2	3	4
Nov	**19.8%**	**26.5%**	**-48.9%**	**11.5%**	**16.1%**	**30.0%**	**-34.9%**	**44.5%**
Dec	**50.3%**	**12.2%**	**-18.6%**	**29.8%**	**34.7%**	**16.5%**	**-38.0%**	**15.1%**
Jan	**59.7%**	**0.9%**	**-37.2%**	**17.6%**	**57.8%**	**15.7%**	**-36.5%**	**34.7%**
Feb	**49.6%**	-1.1%	**-29.1%**	**77.2%**	**40.9%**	15.6%	**-33.3%**	**45.8%**
Mar	**73.3%**	**18.3%**	**-6.5%**	**8.4%**	**61.6%**	**18.7%**	**-17.8%**	**35.2%**
NDJFM	**50.8%**	**11.5%**	**-27.9%**	**28.1%**	**42.4%**	**19.3%**	**-32.1%**	**34.8%**

Table3. Same as Table 1 but for high Ural index (80th percentile) vs. climatology

	ERA5				ICON			
k	1	2	3	4	1	2	3	4
Nov	3.0%	**-31.3%**	**88.9%**	**64.4%**	-9.4%	**-42.4%**	**27.4%**	**24.7%**
Dec	**-4.0%**	**-5.7%**	**46.9%**	**5.0%**	**-22.9%**	**-13.8%**	**43.3%**	**14.2%**
Jan	**-4.8%**	**-13.5%**	**39.8%**	**17.8%**	**-23.4%**	**-19.2%**	**43.6%**	**21.7%**
Feb	**1.3%**	1.1%	**23.3%**	-20.5%	**3.3%**	-27.8%	**36.5%**	17.1%
Mar	**-36.1%**	2.0%	**61.8%**	-15.7%	**-16.4%**	-20.6%	**30.5%**	10.4%
NDJFM	**-8.4%**	**-9.5%**	**52.4%**	**10.4%**	**-14.1%**	**-24.6%**	**36.3%**	**17.6%**

Table4. Same as Table 1 but for low Ural index (20th percentile) vs. climatology

k	ERA5				ICON			
	1	2	3	4	1	2	3	4
Nov	**23.5%**	-4.9%	**20.5%**	**5.7%**	**25.9%**	11.1%	**17.5%**	**25.8%**
Dec	**46.3%**	**-10.4%**	**39.9%**	-15.5%	**28.7%**	**-1.3%**	**13.4%**	41.7%
Jan	**54.0%**	**-14.3%**	**15.3%**	**34.4%**	**36.2%**	**-5.2%**	**30.2%**	**46.9%**
Feb	**62.3%**	**-26.4%**	**25.8%**	**86.3%**	**40.4%**	**-12.3%**	**13.3%**	**38.2%**
Mar	**40.8%**	-14.3%	**10.6%**	-30.4%	**54.7%**	2.5%	**36.3%**	60.0%
NDJFM	**45.2%**	**-13.9%**	**22.4%**	**14.9%**	**37.2%**	**-0.9%**	**22.3%**	**42.7%**

Table5. Same as Table 1 but for low Aleutian index (20th percentile) vs. climatology

k	ERA5				ICON			
	1	2	3	4	1	2	3	4
Nov	**-36.9%**	**1.1%**	**77.7%**	43.4%	**-21.6%**	**13.1%**	**1.8%**	-40.1%
Dec	**-37.8%**	**14.6%**	2.9%	46.9%	**-30.7%**	**15.1%**	-25.9%	-56.9%
Jan	**-26.5%**	**21.3%**	-3.2%	**-33.5%**	**-12.1%**	**14.4%**	**-30.4%**	**-34.5%**
Feb	**-23.2%**	**28.7%**	**-23.3%**	**-38.2%**	**-4.3%**	**26.3%**	**-26.1%**	**-24.5%**
Mar	**-38.0%**	**20.2%**	**-12.6%**	**-21.0%**	**-40.9%**	**4.2%**	**-26.4%**	**-20.3%**
NDJFM	**-32.6%**	**17.1%**	8.4%	**-0.1%**	**-22.2%**	**14.4%**	-21.5%	**-35.4%**

Table6. Same as Table 1 but for high Aleutian index (80th percentile) vs. climatology

Reviewer 2

Summary

This analysis identified and examined the pathways of troposphere-stratosphere coupling at the seasonal timescale in ERA5 and ICON model. The results highlight tropospheric precursors of stratospheric extended winter circulation and the timescales of their influence. The authors make use of the NAM index to identify the coupling mechanisms and regions: the Ural area and the extended Aleutian area. Furthermore, the authors develop a precursor index based on the MSLP difference between these two regions and show that it is mostly correlated with the subsequent NAM in the stratosphere, which highlights the importance of coupling. Moreover, vertical Plumb fluxes are investigated to identify pathways of coupling. In the Ural area vertical planetary waves propagation transmits signal from the troposphere to the stratosphere, whereas for the Aleutian region the pathway seems to be different. The geopotential anomalies in the upper layers seem to play a greater role.

The research is timely and novel as it discusses the mechanisms of the stratosphere-troposphere coupling at the subseasonal-to-seasonal timescales which can be an important source of predictability. However, as the authors mention, there is great variability across individual events. Overall, the manuscript presents results in a well-constructed way. I have some major and minor comments presented below and I recommend this manuscript for publication after addressing the comments.

Major comments

- As the Plumb flux is formulated using a quasi-geostrophic assumption, I would suggest using the wave activity flux by Takaya and Nakamura (2001). This flux is defined for the case of a zonally varying basic flow and focuses on the wave activity associated with planetary waves, as the wavy anomalies are considered to be embedded in the basic flow that includes the climatological planetary waves. The basic state in the Northern Hemisphere in winter shows inhomogeneities that can modulate the propagation of planetary waves.

  We acknowledge the comment by the reviewer and notice a similar comment by reviewer 3. We agree, it makes sense to implement the more advanced version of a 3D wave activity flux by Takaya and Nakamura (2001), although they still use quasi-geostrophic (QG) approximations. As listed in the responses to reviewer 1, Plumb fluxes are actively implemented in other very recent studies, while references of implementations of the vertical component of Takaya Nakamura fluxes are barely seen in literature.

  More persuasive, we recognize the argument of difficulties in the definition of the basic flow. In our case, we used a basic flow defined by a 91-day rolling mean seasonal cycle derived from the mean of all available years. Varying this definition leads to different results. This means, we agree that inhomogeneities of the winter climatology can modulate the propagation of planetary waves and these inhomogeneities cannot be addressed by a climatological mean state which covers multiple decades. Regarding this, we want to cite from the original Takaya and Nakamura study:

  > **The wave-activity flux is a useful diagnostic tool for illustrating a "snapshot" of a propagating packet of stationary or migratory QG wave disturbances and thereby for inferring where the packet is emitted and absorbed, as verified in several applications to the data.**

Regarding the reviewers' comments, the cited passage from the original study and our actually much worse results when we implement the Takaya Nakamura flux, we now believe a substantial redesign of our approach would be necessary to implement such a diagnostic in a climatological way since it is only useful on "snapshots". Therefore, we decided to reject the use of 3D wave activity flux and plan to implement similar diagnostics with zonal mean Eliassen-Palm flux based on Transformed Eulerian Mean equations, which will be described at the end of this response document.

The actual implementation of Takaya Nakamura fluxes in the same fashion as in Figure 6 of the original manuscript led to degraded signals (cf. Figure 2), which we explain with the aforementioned discussion.

- Regression coefficients in Figs. 2 and 6 have very different magnitudes and provide limited information as for the meaningful effects of this regression. While the results are meaningful for the non-hatched areas, it would be good to provide the effect-size by adding R^2, which shows the explained variance ratio in the dependent variable.

  We will provide the requested metrics for corresponding figures. Please recognize that Figure 6 will change due to the first comment. In Figure 2a (ERA5) the maximum value of the correlation coefficient is 0.26, the minimum value -0.35. In Figure 2b (ICON) the maximum value of the correlation coefficient is 0.18, the minimum value -0.24. Indeed, these values are very low, which is related noise introduced by the evaluation of daily data. Nevertheless, we point out that the regression indicates that a reasonable 10hPa sea level pressure anomaly can push the NAM index by up to 0.5 index points, which is a significant change of the circulation. Furthermore, we present further evidence of the predictive skill of our precursor index in Figures 3 and 4. While the statistical explained variance is low, we still find skill in the given relation.

- In part of the analysis (Figs 4 and 5) you show the difference between high (80th percentile) and low (20th percentile) precursor index. I do not fully understand the purpose of showing the difference and not high/low index separately. It seems to me that this would be more clear for interpretation. Please, provide more explanation of your choice or maybe show the results separately for high and low index.

  We chose to show the differences between the high and the low precursor index, as the signals of the individual differences with respect to the climatology are rather similar with an opposite sign. By only showing the differences between the two extreme states of the Precursor index we reduce the number of figures without a significant loss of information. To confirm this, we attach the same figure for the individual extreme state differences with respect to the climatology (Figure 3 for high Precursor index minus climatology and Figure 4 for low Precursor index minus climatology). Figure 5 of the manuscript will be replaced by a different Figure, which does not use the composite approach.

Minor comments

L77: The length of the model simulations is a bit shorter than the length of the ERA5 period chosen for the analysis. I think it would be more straightforward to take the same periods of time for the intercomparison reasons.

We did some basic testing by using ERA5 data only up to the winter 2016/2017 and received the same qualitative results. Due to the very large variability of the stratospheric

polar vortex in winter, we, however, decided to maximise our sample size by including the whole ERA5 period available at this point in time.

L112-113: Why did you choose to limit the showed correlation coefficients to the 10% largest? Might this result in the loss of information?

We tested the approach with several thresholds. If the threshold is increased, either areas with already visible results become simply larger, or additional noise is introduced in particular at longer time shifts. The given threshold is a compromise to arrive at a clear image with distinguishable anomalies.

Fig.1 c, d: The font of the longitude labels and colorbar seems too small to me.

We will adjust this in the revised manuscript.

L129-130: Would it be better to use ensemble means for the intercomparison reasons?

When we look into the single ICON ensemble members, we see that signals are also noisier in ICON. Ensemble approaches usually make the overall signal smoother due to the larger sample size. The ensemble approach thereby minimises the internal variability and maximises the actual signal. Moreover, using single ensemble members only makes a process-based analysis more difficult as the daily development will differ due to internal variability. Overall, we note that a qualitative agreement between ICON and ERA5 is given. Accordingly, we adjusted the original statement to: "The larger sample size in ICON results in a smoother image, while the qualitative agreement between ICON and ERA5 is given. "

Fig.2: It seems that "@100hPa" at the x-axis label is wrong.

Thanks for pointing this out. This will be corrected in the revised manuscript.

L161-162: Please, consider adding that these areas are also shown in Fig.1.

That is a good idea, thank you. We will do so.

L171: It seems interesting that "the negative correlation between January precursor index and February NAM is not statistically significant in the ERA5 data, but highly significant in the ICON ensemble simulations." Do you have any explanation for this?

This is a good point. An explanation can be given with the help of Figure 4e. The NAM composite for the January Precursor extremes demonstrates that there is a strong significant link between the Precursor index in January and the strength of the stratospheric polar vortex in the following two months (February and March). However, in the upper stratosphere the signal is more short-lived. Figure 3 is based on the NAM at 10hPa and Figure 4e) exemplifies that the negative NAM signals remain at 10hPa only up to mid-February, whereas the signal is long-lived in the lower stratosphere. Therefore, we assume that this non-significance could be related to the limited sample size in ERA5.

L175: "The two components of the precursor index were also analysed individually": I guess this is not shown?

Yes, this is not shown to reduce the overall number of Figures. However, as we describe in the manuscript, the Ural component contributes more strongly to the correlation with the NAM at 10 hPa, yet the Aleutian component adds significant value. The 1-month lag correlation coefficients increase by 34% for O-N, 19% for D-J, 70% for J-F and 30% for F-M, yet decrease by 22% for N-D when comparing Ural index to the combined Precursor index.

Figure 2 also exemplifies that both regions are linked to strength of the stratospheric polar vortex.

L179: Please, provide the sample sizes of composites. Are the same samples used also in Section 3.2?

For ERA5 the sample size is 9 for each extreme event and for ICON it is 40. To make things clearer we will also add this information into the manuscript. The same samples were used in Figure 5 of Section 3.2. This Figure will, however, be replaced by a Figure using a different method in the revised manuscript.

L218: Did you test your results using the wave activity flux at different heights? Why did you choose to use 100hPa in the study?

We assess from literature that 100 hPa is a standard height to diagnose wave activity flux that reached the stratosphere. It is important to use a height level, which is safely above the tropopause. In our experience, the wave activity flux becomes relatively uniform above the tropopause, if the deviations are not too big. At lower altitudes, in particular when reaching tropospheric levels, signals become very noisy.

L255-258: "This signal reaches the tropopause without a time lag but is disconnected from the main stratospheric signals, which emerge east of the Ural region and propagate west- and eastward throughout time. Thus, a high-pressure anomaly in the Ural area is connected to positive stratospheric geopotential anomalies that are a clear indication of a weakened stratospheric polar vortex." This explanation seems confusing to me, as in the first part you mention that the signal is disconnected, but then the anomalies are connected at different heights. I suggest the authors to rewrite this part more in a more clear way.

Thanks for pointing this out. We agree that the wording is somewhat confusing. Replacing "... in the Ural area is connected to positive stratospheric geopotential .." with "… in the Ural area is related to positive stratospheric geopotential…" should make it clearer. Moreover, the explanation that follows in the original manuscript will be adjusted due to the overall changes in the revised manuscript.

L261-262: The ICON ensemble is somewhat in agreement, but it also have strong negative correlations, how would you explain this? At least, this is worth mentioning in text.

This negative correlation is a remnant of the fact that the stratospheric polar vortex tends to be stronger before it breaks down / weakens. We also see this in the ERA5 data when we go backwards in time.

Fig. 7 and 8: Please consider adding briefly heights to the b,c,e,f captions
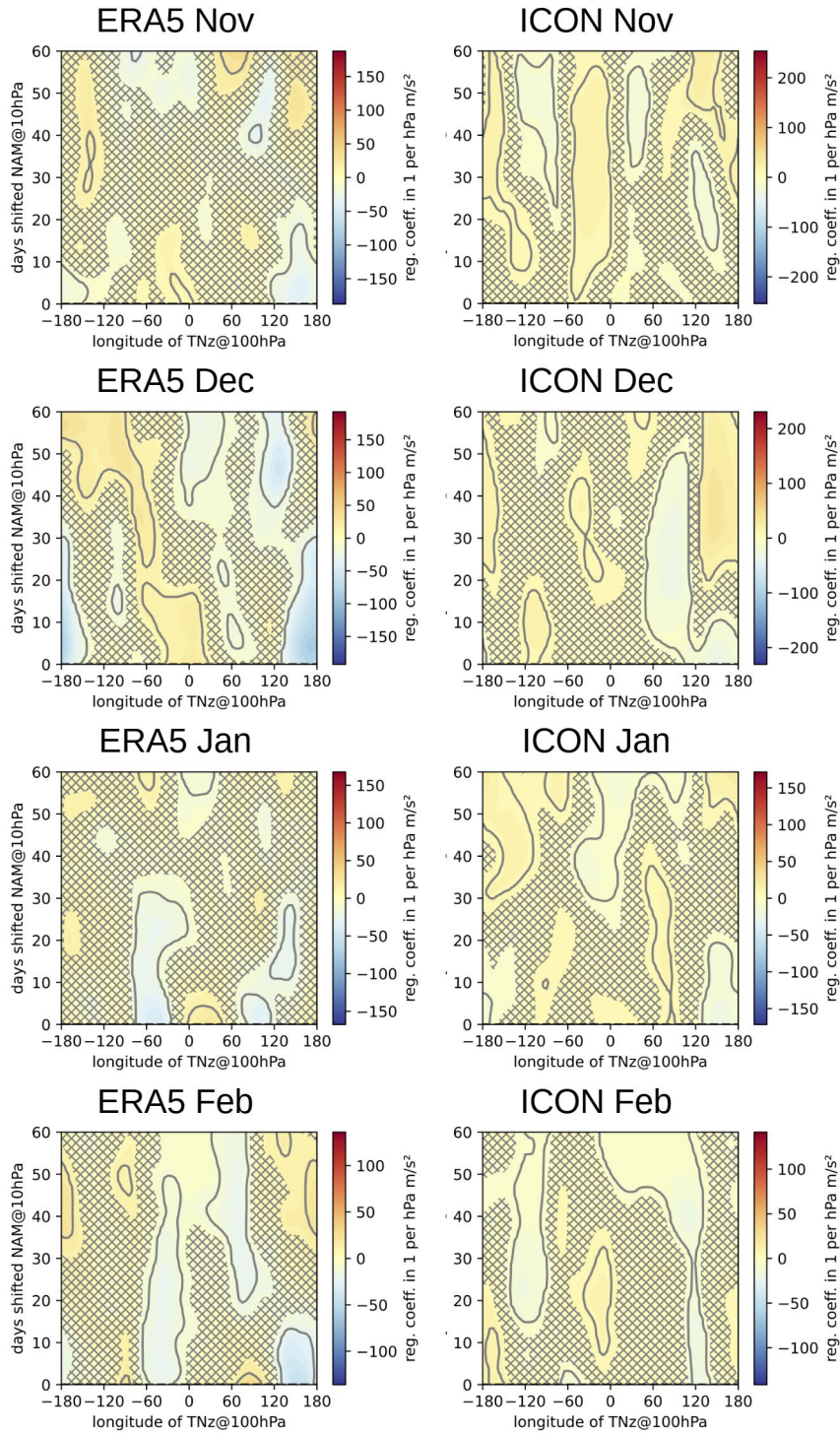
We will add this in the revised manuscript.

Figure 2. Regression 100 hPa Takaya Nakamura flux onto the 10 hPa NAM. The figure is produced in the same way as Figure 6 in the original manuscript.
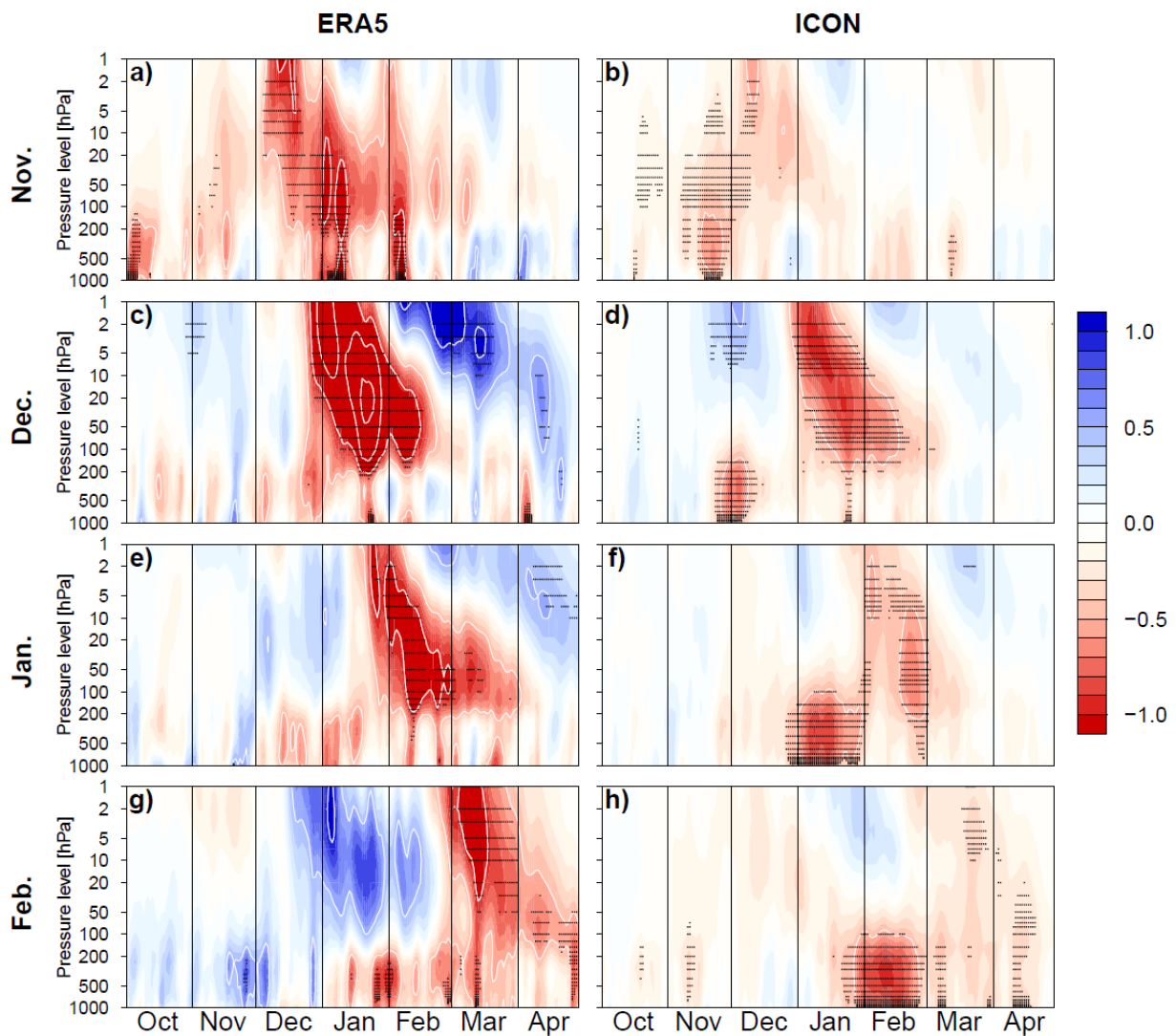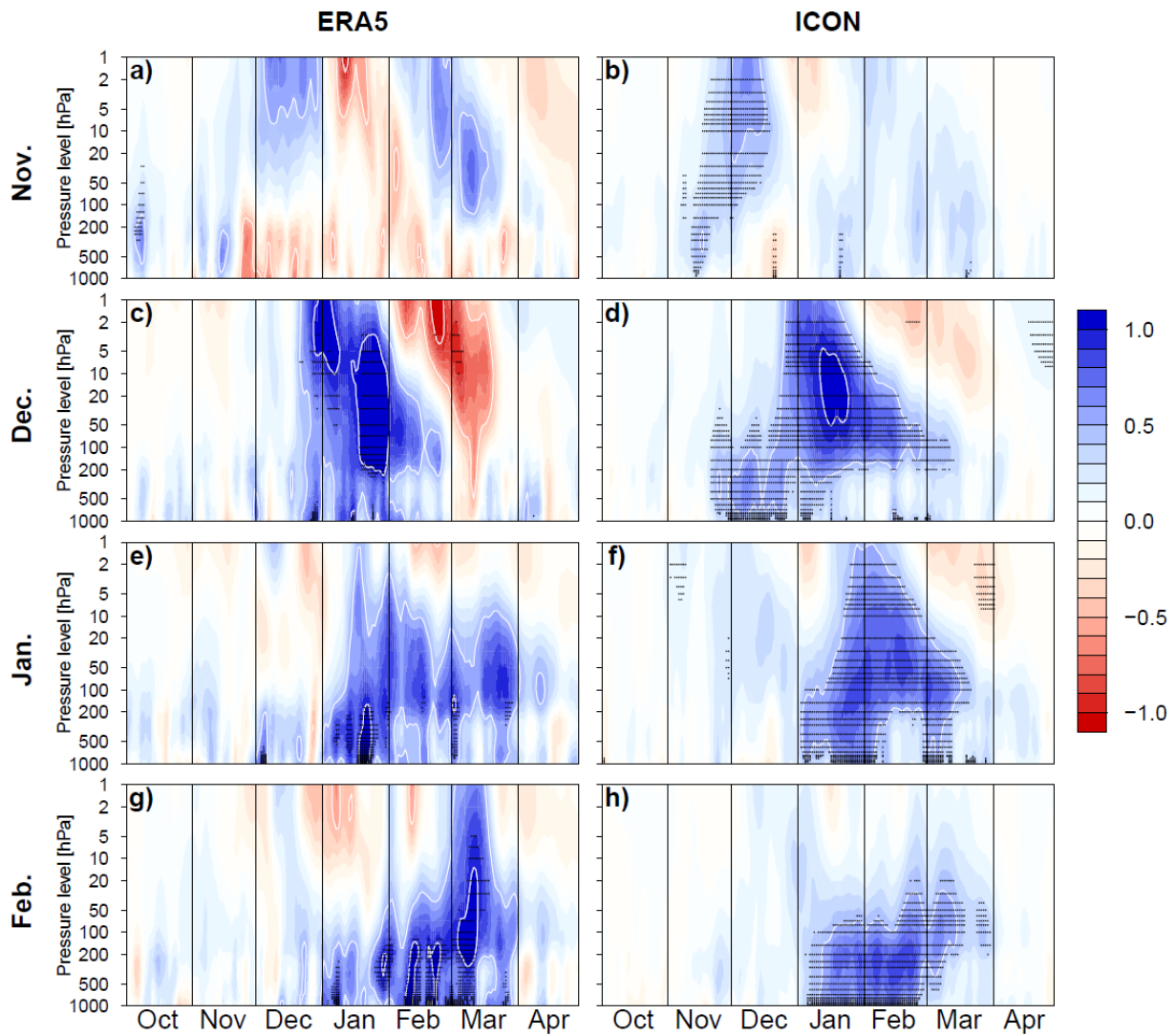
Figure 3. Composites of the time–height development of the NAM for the difference between high precursor index (80th percentile) and the climatology in ERA5 (a,c,e,g) and ICON (b,d,f,h). The composites are created based on the monthly mean precursor index for the months November (a,b), December (c,d), January (e,f) and February (g,h). The NAM index is nondimensional. The contour interval for the white contours is 0.5 and stippling indicates statistical significance at the 95% level according to a two-sided Wilcoxon-test. Note that due to common practice negative values (weak NAM) are red and positive values (strong NAM) are blue.

Figure 4. Composites of the time–height development of the NAM for the difference between low precursor index (20th percentile) and the climatology in ERA5 (a,c,e,g) and ICON (b,d,f,h). The composites are created based on the monthly mean precursor index for the months November (a,b), December (c,d), January (e,f) and February (g,h). The NAM index is nondimensional. The contour interval for the white contours is 0.5 and stippling indicates statistical significance at the 95% level according to a two-sided Wilcoxon-test. Note that due to common practice negative values (weak NAM) are red and positive values (strong NAM) are blue.

Reviewer 3

This paper revisits the role of a strengthened Ural high and Aleutian low for variability of the NH polar vortex. In agreement with previous work, the authors find that a Ural high and an Aleutian low leads to a weaker vortex. The more novel aspects of this paper are: 1) inclusion of ICON results; 2) a focus on daily timescales to better isolate the relevant timescales; 3) a discussion of which calendar month the response peaks. Finally, the authors include a discussion of mechanisms, but as discussed below I find this discussion to be limited and problematic. Major revisions are needed before I can give a final assessment of the quality of this paper/

Major comments:

1. As discussed above, I don't find the mechanism part of this paper convincing. The authors use Plumb 1985 fluxes which were derived explicitly for stationary waves, and apply this method for time varying waves! There are better fluxes available (Plumb 1986, or even better Takaya and Nakamura 2001) that are more appropriate for the application. Hence any conclusions reached using a unsuitable formulation should be treated with caution.

> **We acknowledge the comment by the reviewer and notice a similar comment by reviewer 2. We agree, it makes sense to implement the more advanced version of a 3D wave activity flux by Takaya and Nakamura (2001). As listed in the responses to reviewer 1, Plumb fluxes are actively implemented in other very recent studies, while references of implementation of the vertical component of Takaya Nakamura fluxes are barely seen in literature.**
>
> **The actual implementation of Takaya Nakamura fluxes in the same fashion as in Figure 6 led to degraded signals. The corresponding Figure 2 can be found in the responses to reviewer 2. In part, we explain reduced signals by difficulties in defining an appropriate climatological basic state. In our case, we used a basic flow defined by a 91-day rolling mean seasonal cycle derived from the mean of all available years. Varying this definition leads to different results. To discuss this further, we want to cite from the original Takaya and Nakamura study:**
>
>> **The wave-activity flux is a useful diagnostic tool for illustrating a "snapshot" of a propagating packet of stationary or migratory QG wave disturbances and thereby for inferring where the packet is emitted and absorbed, as verified in several applications to the data.**
>
> **Taking this into account, we believe a substantial redesign of our approach would be needed to implement such a diagnostic in a climatological way since it is only useful on "snapshots". Therefore, we decided to reject the use of 3D wave activity flux and plan to implement similar diagnostics with zonal mean Eliassen-Palm flux based on Transformed Eulerian Mean equations, which will be described at the end of this response document.**

Further, they appear to ignore the previously proposed mechanism. As discussed in Garfinkel et al 2010 (already cited) and Smith and Kushner 2012 (not cited) among many others, transients that are in phase with the climatological wave-1 or wave-2 field enhance it, and via constructive interference lead to an overall increase in wave-driving. It just so happens that a transient Ural high and Aleutian low are in phase with the stationary waves, and hence enhance it. The importance of wave-1 and wave-2 arises specifically because higher wavenumbers cannot propagate vertically in the presence of strong zonal wind (Charney and Drazin 1961). Do the authors think this mechanism is wrong? unimportant? They cite Garfinkel et al 2010 and Bao et al 2017, and so I assume they are aware of it.

This mechanism was found to be important for the stratospheric polar vortex response to a range of external forcings, as discussed in detail in Smith and Kushner (see references therein) and dozens since that have cited Garfinkel et al 2010 and Smith and Kushner.

The authors agree that these mechanisms are important and were so far under-represented in the manuscript. Therefore, we decided to add an extensive analysis on how the precursor index and its components relate to the amplitudes of the different wave numbers based on the 500hPa geopotential in ERA5 and ICON (cf. Tables 1-6). As expected, we see that the Precursor index strongly projects onto the climatological wave-1, with increased wave-1 amplitudes during high Precursor index states and decreased wave-1 amplitudes during low index states (cf. Table 1 and 2). The overall wave-2 amplitude changes are small for the Precursor index. However, wave-2 amplitude differences could explain, why the Ural region might affect the stratosphere stronger than the Aleutian region, although the Aleutian region is connected to larger wave-1 amplitude changes (cf. Table 3,4 and 5,6). Whereas a strong (weak) Ural high is linked to increased (decreased) wave-1 and wave-2 activity (Table 3 & 4), a low (high) Aleutian low is connected to increased (decreased) wave-1 amplitudes, but decreased (increased) wave-2 amplitudes (Table 5 & 6). Hence, for MSLP anomalies in the extended Aleutian region wave-2 can counteract wave-1 activity. This can explain why the Ural region is the main tropospheric driver of the stratospheric polar vortex. This analysis will be part of the revised manuscript. Moreover, we will shortly discuss the basics of wave-driving in the introduction, thereby including important references, such as Smith and Kushner (2012) and Charney and Drazin (1961).

That the authors appear to ignore the previous proposed mechanism is particularly grating due to statements along the likes of "there is only a very vague understanding of the localised coupling mechanisms and involved timescales, in particular when it comes to connecting tropospheric precursor patterns to the strength of the stratospheric polar vortex. " (line 3) or "pressure anomalies in the North Pacific region, has received less attention in literature focusing on precursor patterns of SSWs" (line 53). These two statements are just plain incorrect.

The authors will get rid of these statements in the revised manuscript. Instead, they will discuss the important findings from recent literature, including e.g., Garfinkel et al. (2010) and Smith and Kushner (2012) concerning the wave-driving mechanisms and e.g., Ineson and Scaife (2009) concerning the role of the Pacific for strength of the stratospheric polar vortex.

2. Why should the trop->strat coupling differ from early winter to late winter? Why would ICON have too weak coupling? These questions should be answered in the revised paper (assuming these differences across months and between data sources are statistically signififcant).

These questions will be addressed in the revised manuscript in more detail. Overall, the authors do not see any large qualitative changes in troposphere-stratosphere coupling between early and late winter (cf. Figure 3 and 4 from the manuscript). However, the strength and position of the stratospheric polar vortex will impact the amount of waves that can propagate into the stratosphere from the troposphere (Charney and Drazin, 1961). In the low-variability, build-up phase of the stratospheric polar vortex troposphere-stratosphere coupling is more direct with smaller time lags, whereas the coupling is less direct in the more variable phase of the vortex in late winter (cf. Figures 3 and 4 from the manuscript). Moreover, the tropospheric wave forcing related to the precursor patterns can differ between early and late winter (cf. Tables 1-6). In particular the wave-1

amplitudes tend to increase in mid to late winter for high Precursor index states.
In the revised manuscript the authors will also go into more detail when comparing the ICON results to the results of ERA5. The ensemble approach in ICON tends to smooth out strong peaks in the apparent signals, thus giving the impression that the overall signals are not as strong, although they agree qualitatively. When looking into individual ensemble members instead of the ensemble mean we see that amplitudes tend to be comparable to ERA5. However, one factor that could potentially contribute to differences in the coupling behaviour is the mean state of the stratospheric polar vortex in ICON, as it tends to be too weak in mid-winter (cf. Köhler et al., 2021). This temporally coincides with the somewhat smaller correlation coefficients in ICON in mid-winter (cf. Figure 3 of the manuscript).

> **Köhler, R., Handorf, D., Jaiser, R., Dethloff, K., Zängl, G., Majewski, D., & Rex, M. (2021). Improved Circulation in the Northern Hemisphere by Adjusting Gravity Wave Drag Parameterizations in Seasonal Experiments With ICON-NWP.** *Earth and Space Science*, *8*(3), e2021EA001676. **https://doi.org/10.1029/2021EA001676**

It is not clear to this reviewer how Plumb fluxes would help answer this question. In contrast, the stationary waves in the different calendar months or different data sources will differ, and hence the heat flux anomaly for a given strengthed height or MSLP anomaly will differ (Watt-Meyer and Kushner 2015).

> **We appreciate the input of the reviewer. This further motivated us to use a more established metric to diagnose the vertical propagation of planetary waves. We will provide an overview about our changed discussion implementing zonal mean Eliassen-Palm flux based on Transformed Eulerian Mean equations at the end of this response document.**

minor comments:

line 38 the most recent multi-model and multi-SSW assessment of SSW predictability is by Chwat et al 2022

> **Thanks for pointing this out. We will add this citation to the revised manuscript.**

xlabel of figure 2: MSLP is not at 100hPa.

> **Correct, we will fix this mistake in the revised manuscript.**

line 145 "conceptual" seems an odd word to choose here

> **To make it more specific we will replace "conceptual" with "time-resolved" in the revised manuscript.**

line 260: you would need to look at the zonal component of the Plumb flux to better relate the zonal position of anomalies in the troposphere with that in the stratosphere, though as discussed above I don't find Plumb fluxes useful. An alternate plot would be something similar to figure 2 of Ineson and Scaife 2009, but created for a range of lags.

> **We appreciate the idea to produce a figure that resolves wave activity fluxes in zonal and vertical direction as well as in time. This would be a similar approach as we took in our figures 7 and 8. Nevertheless, we plan to remove the application 3D wave activity fluxes from our manuscript due to the several other comments from the reviewers.**

Smith, K.L. and Kushner, P.J., 2012. Linear interference and the initiation of extratropical stratosphere-troposphere interactions. *Journal of Geophysical Research: Atmospheres*, *117*(D13).

Watt-Meyer, Oliver, and Paul J. Kushner. "The role of standing waves in driving persistent anomalies of upward wave activity flux." *Journal of Climate* 28, no. 24 (2015): 9941-9954.

Chwat, Dvir, Chaim I. Garfinkel, Wen Chen, and Jian Rao. "Which Sudden Stratospheric Warming Events Are Most Predictable?." *Journal of Geophysical Research: Atmospheres* 127, no. 18 (2022): e2022JD037521.

Plumb, R.A., 1986. Three-dimensional propagation of transient quasi-geostrophic eddies and its relationship with the eddy forcing of the time—mean flow. *Journal of Atmospheric Sciences*, *43*(16), pp.1657-1678.

Plumb, R.A., 1986. Three-dimensional propagation of transient quasi-geostrophic eddies and its relationship with the eddy forcing of the time—mean flow. *Journal of Atmospheric Sciences*, *43*(16), pp.1657-1678.

Ineson, S, and A. A. Scaife. "The role of the stratosphere in the European climate response to El Niño." *Nature Geoscience* 2, no. 1 (2009): 32-36.

Takaya, K., & Nakamura, H. (2001). A formulation of a phase-independent wave-activity flux for stationary and migratory quasigeostrophic eddies on a zonally varying basic flow. *Journal of the Atmospheric Sciences*, *58*(6), 608-627.

New zonal mean EP flux-based diagnostics

Related to the several critical points that the reviewers brought up regarding our application of Plumb 1985 fluxes, we recognized the inadequate applicability of 3D wave activity fluxes in our climatological case. In particular in the stratosphere, it is uncertain, if a zonally resolved wave flux and the attempt to connect it to a surface anomaly is reasonable. As brought up by the reviewers and shown by our newly introduced discussion of wave numbers, the surface pressure anomalies project on low wave number patterns. Therefore, it is more straight forward, if the resulting wave activity is a zonal mean measure.

We implement an analysis of conventional zonal mean Eliassen-Palm (EP) flux (Andrews and McIntyre, 1976) based on tools implemented by Jucker (2021a, b). Since we now have only zonal mean information, we decided to first analyse how the zonally resolved sea level pressure (slp) anomalies project onto the vertical component of the zonal mean EP flux during the whole winter season from November to February. The results are very similar for ERA5 in Figure 5a and ICON in Figure 5b. They show positive (negative) slp anomalies in the Ural region at 40°E inducing a positive (negative) vertical EP flux anomaly at 100 hPa with a time delay between 0 and 5 days of the maximum of the regression coefficients. The intensity of this positive relation weakens and reaches up to 20 days, while it also broadens regionally. The relation between negative (positive) slp anomalies in the Aleutian region and positive (negative) vertical EP flux anomalies is smaller. It emerges first with almost no time delay at 150°E. Then the relation gradually moves west with a secondary maximum at 120°W with a time delay of 10 days. Both anomalies indicate that a strengthened Ural blocking and a deeper Aleutian low contribute to enhanced upward EP flux emerging in the lower stratosphere. Corresponding correlation coefficients in the Ural region are 0.27 for ERA5 and 0.25 for ICON. In the Aleutian region they are -0.19 in ERA5 and -0.15 in ICON. Although these values are low in terms of explained statistical variance, the regression coefficients indicate a significant contribution to upward EP flux from a typical 10 hPa slp anomaly, that reaches 0.001 m^2/s^2, which is about one third of typical mean EP flux or half of the typical standard deviation. We plan to replace Figure 5 of the original manuscript with Figure 5 and adapt the corresponding discussion.

Next, we explore how vertical EP flux anomalies at 100 hPa influence the zonally resolved geopotential field in 10 hPa depending on time during winter from November to February. In ERA5 (Figure 6a) and ICON (Figure 6b) we see a dominating positive relationship indicating positive (negative) EP flux anomalies inducing positive (negative) geopotential anomalies. The maximum of the signal appears first around 180°W with a time delay of 0 to 5 days. This maximum then extends westwards towards 100°E where it reaches a time delay of about 10 days relative to the initial EP flux anomaly. In particular in ICON, we further see the signal also extending eastwards. Generally, this influence on the polar vortex shows that a weakening related to upward EP flux appears first over the Pacific and then develops into a full polar vortex weakening in the following 10 days. This can also be interpreted as a vortex displacement from the Pacific towards the Eurasian continent, which is a very typical behaviour. The positive correlation reaches 0.41 in ERA5 and 0.43 in ICON, which is again a relatively low explained statistical variance. Nevertheless, the regressions indicate that a 0.001 m^2/s^2 vertical EP flux anomaly can induce a shift of geopotential up to 1000 m^2/s^2, which is about one fifth of the standard deviation of geopotential in 10 hPa. The negative regressions in Figure 6 indicate the displaced more stable vortex in case of upward EP flux around 50°E in the very beginning. Furthermore, they indicate that the vortex re-emerges about a month after the initial disturbance. We plan to replace Figure 6 of the original manuscript with Figure 6 and adapt the corresponding discussion.
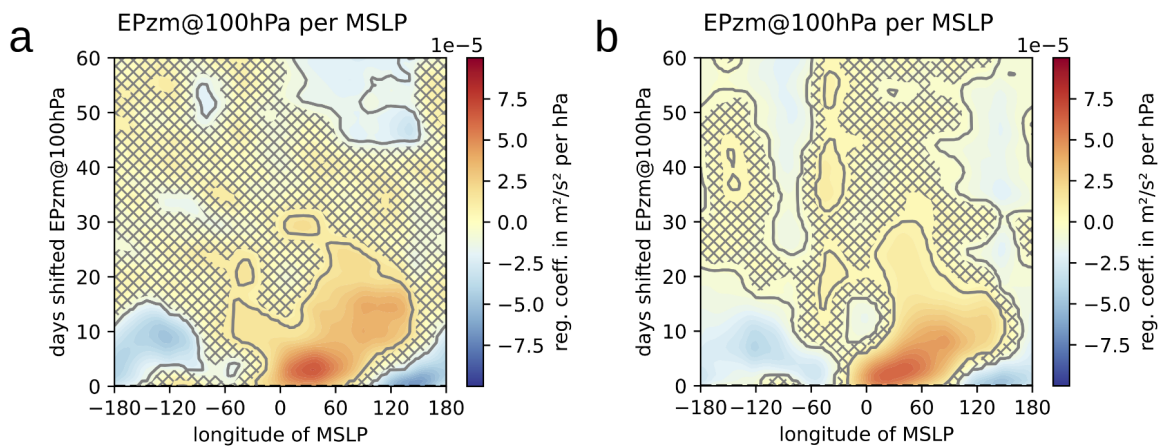
Figure 5. Regression of 45°N to 80°N meridional mean MSLP onto 40°N to 80°N meridional mean 100 hPa zonal mean vertical EP flux. MSLP is averaged over 10° longitude bins with their respective centre positions given along the x-axis. Daily values of MSLP from November to February of ERA5 data (a) and all years and ensemble members of ICON data (b) are used. Data of the 100 hPa zonal mean vertical EP flux is shifted by days given on the y-axis relative to MSLP data. Areas with regression coefficients with p-values above 0.05 are hatched, areas with p-values below 0.001 are encircled.
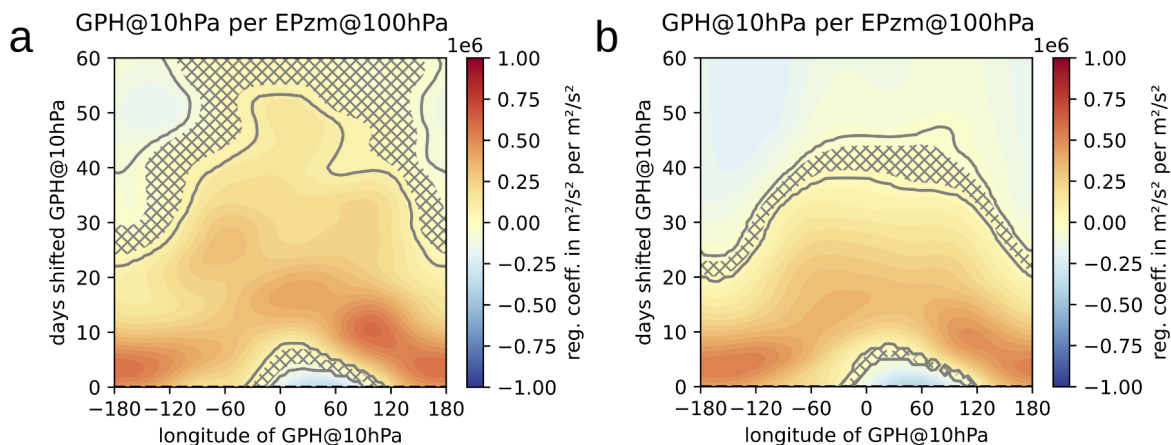

Figure 6. Regression of 40°N to 80°N meridional mean 100hPa zonal mean vertical EP flux onto 45°N to 80°N meridional mean 10hPa geopotential. Geopotential is averaged over 10° longitude bins with their respective centre positions given along the x-axis. Daily values of vertical EP flux from November to February of ERA5 data (a) and all years and ensemble members of ICON data (b) are used. Data of the 10 hPa geopotential is shifted by days given on the y-axis relative to vertical EP flux data. Areas with regression coefficients with p-values above 0.05 are hatched, areas with p-values below 0.001 are encircled.

Andrews, D. G., & Mcintyre, M. E. (1976). Planetary waves in horizontal and vertical shear: The generalized Eliassen-Palm relation and the mean zonal acceleration. *Journal of Atmospheric Sciences*, *33*(11), 2031-2048.

Jucker, M. (2021a). Scaling of Eliassen-Palm flux vectors. *Atmospheric Science Letters*, *22*(4), e1020.

Jucker, M., (2021b): mjucker/aostools: v2.3.2 (Version v2.3.2). – Zenodo. DOI:10.5281/zenodo.4588067