Answer to Comment by Referee 1

20.12.2021

We thank the Referee 1 very much for his/her time and valuable comments, which helped a lot to improve the manuscript. In the following, we provide our answers to each of the comments and corrections in sequential order. The original Referee comment is repeated in bold, and our answers and changes in the manuscript are provided in italic. Text added to the manuscript is indicated in *blue italic*.

The paper of Haenel et al. presents a comparison of observed 2D distributions of species from the GLORIA instrument with model simulations. The authors use one particular flight from winter 2015/2016 at high northern latitudes (POLSTRACC) to compare the capabilities of ICON-ART and EMAC to simulate H_2O, O_3 and HNO_3 as well as cloud occurrence in the UTLS region. The selected flight comprises very different meteorological situations which allows to evaluate different aspects of the relevant model parametrisations.

The ICON-ART data are based on a R2B6 global simulation with a R2B7 nest in the region of interest, the latter corresponding to 20 km horizontal spacing. EMAC data are available at T106 spectral resolution corresponding to a grid spacing of approx. 40 km at 70N. Data are interpolated at the tangent points of the observations and vertical cross section of relevant species are analysed.

Discrepancies are found for cloud occurrence in ICON-ART. Stratospheric water vapour is simulated too high for EMAC not too surprisingly underestimating the vertical gradients. Contrary, ozone is represented well in EMAC while ICON-ART ozone data suffer from the modified LINOZ-scheme.

The authors put a strong focus on the potential reasons for the misrepresentation of clouds in the high resolution simulation of ICON-ART and conclude on matching / timing problems. For EMAC the applied cloud mask better fits the observations, which the authors partly attribute to the lower resolution (noting fundamental model and diagnostic differences). Further, based on T42-simulations of EMAC they show, that the model resolution plays a key role for the H_2O gradients and mixing ratios as well as HNO_3 in EMAC. To check the impact of scavenging on HNO_3, which is only provided by EMAC, they conclude, that scavenging is essential to simulate HNO_3 correctly.

We thank Referee 1 for the precise summary of our study.

The paper is well written, and illustrates some problems of state-of-the-art models to simulate the composition of the challenging UTLS-region governed by strong gradients and often sub-grid processes. However, the central goal of the study is not clear, despite the authors state: " …with the goal to aid model development and improving our understanding of processes in the upper

troposphere/lowermost stratosphere…". It leaves the reader with the main key messages: Resolution matters, chemistry matters which are both not too surprising.

We agree that the goal of the study with regard to aiding model development needs further clarification. Under consideration of the comment by the Referee and comments by Referee 2, we now summarise suggestions for model improvement more clearly in the new Section 4.5 "Suggestions for model improvement". Furthermore, a corresponding summary statement has been added to the Section 5 (Discussion and conclusions).

Following the comment by the Referee given below, we furthermore put a stronger focus on the capabilities of the models of simulating dynamical structures of troposphere-to-stratosphere exchange in the presented case study. As suggested, we now investigate the development of the narrow tropospheric filaments seen in the observations and model data between 17:30 and 18:30 UTC with the help of the model data and discuss the results in Section 4.3. Using ICON-ART, we show that a larger filament in the west was transported horizontally into the Arctic LMS in connection with poleward breaking of a cyclonically sheared Rossby wave, while two weaker filaments in the east are associated with an older tropopause fold there. From our point of view, it is remarkable that the model representation of the underlying processes and the resulting modelled structures result in such a high degree of agreement with the observations.

Since the fundamental properties of the model systems are very different, but the resolution is one key aspect of the comparison results the authors should provide in addition a comparison of similar grid spacing (e.g. between T106, R2B6 or coarse graining).

We agree that a comparison of the different models in a similar resolution would allow for a more direct comparison of EMAC and ICON-ART. Following the suggestion by the Referee, we revised Figure 5 and 7 and added a new Figure to analyse differences between the ICON-ART global and nested domains. However, the different properties of the grids should be kept in mind in the comparison (cf. Fig. 2 and 3), and the EMAC grid does not converge towards the poles in meridional direction.

Specifically, we included an additional panel with the ICON-ART cloud mask at the R2B6 grid in Figure 5, which shows only small differences between the ICON-ART global and nested domains.

Furthermore, we replaced the results of the ICON-ART nested R2B7 domain in Figure 7 by the global ICON-ART R2B6 domain to allow a more direct comparison with EMAC (see below). Following the suggestion by Referee 2, we furthermore included residual plots between the corresponding model and GLORIA cross sections.

In a new Figure (Fig. 8, see below), we investigate residuals between the ICON-ART global and nested domain and provide a bridge from the ICON-ART global domain to the nested domain used in the following. We added the following discussion:

"To investigate potential differences between the global R2B6 and the nested R2B7 ICON-ART domain, differences between these grids are depicted in Fig. 8. Mesoscale patterns in the residuals of q_v (Fig. 8a) and O_3 (Fig. 8b) in the tropopause region and, in the case of q_v, in the regions where clouds were present (compare Fig. 5), are attributed to finer/coarser representation by the different

model grids and the subsequent interpolation to the GLORIA geolocations. Overall, no significant systematic biases are identified."

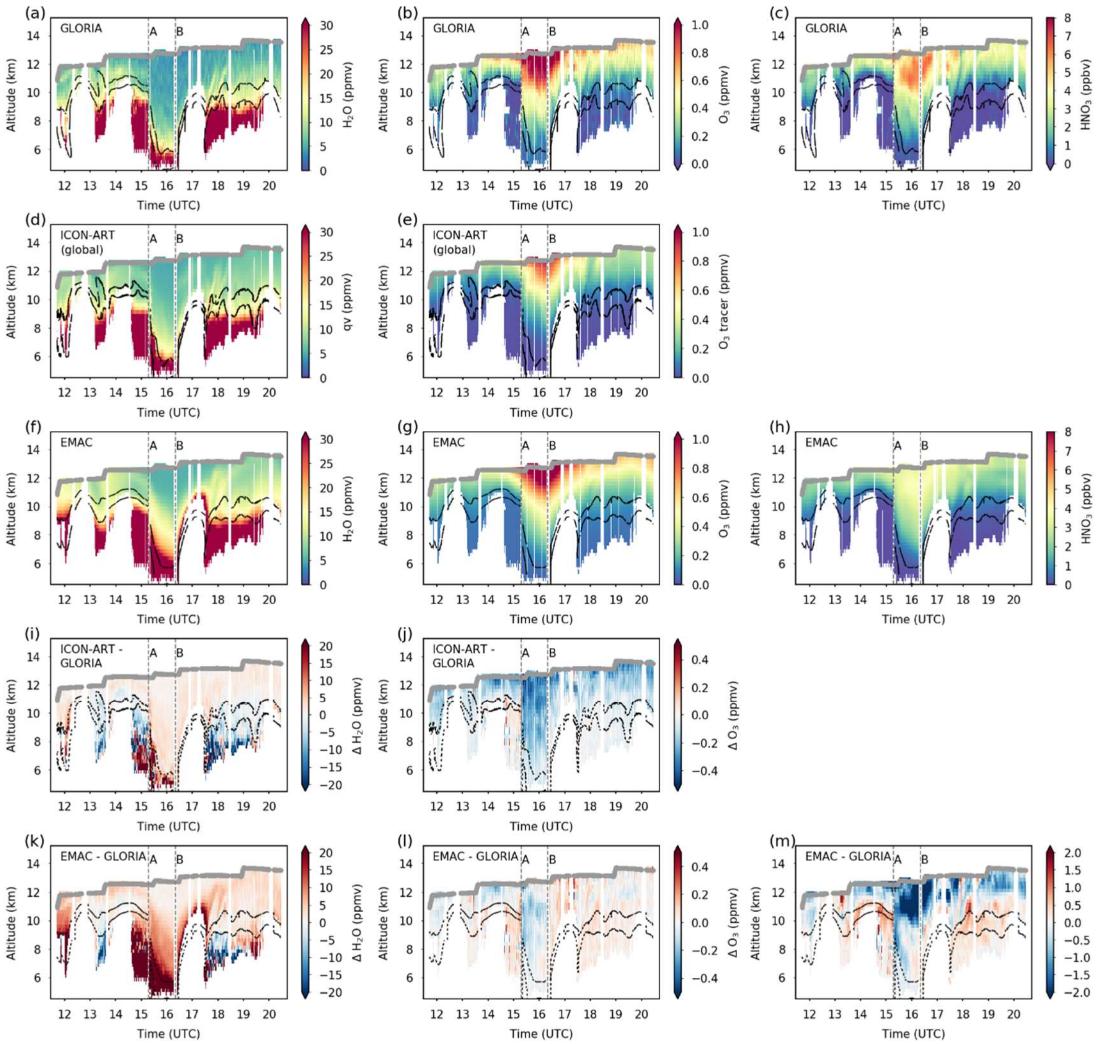

Figure 7. *Observed and modelled trace gas distributions. GLORIA observations of water vapour, ozone and nitric acid (a-c). ICON-ART (global R2B6 grid) short-term forecast of specific humidity (d) and free-running simulation of ozone using simplified ozone depletion parameterisation (e). EMAC free-running simulations of water vapour, ozone and nitric acid (f-h). Residuals between the shown model data and GLORIA observations above (i-m). Black lines: 2 PVU and 4 PVU isolines (lower and higher lines, respectively) from ECMWF reanalysis (a-c), ICON-ART (d,e) and EMAC (f-h) as indicators for the dynamical tropopause. Grey lines: HALO flight altitude.*

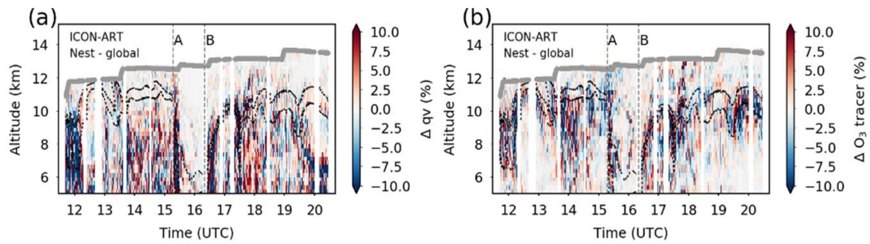

Figure 8. *Residuals between the ICON-ART nested R2B7 and global R2B6 domains of q_v (a) and O_3 (b). Black lines: 2 PVU and 4 PVU isolines as indicators for the dynamical tropopause. Grey lines: HALO flight altitude.*

Second, the paper shows a bunch of comparisons and sensitivities for different species, processes and models and partly some very nice diagnostics (e.g. the ICON-ART passive water forecast), which are – and partly have to be – model specific (e.g. scavenging and HNO_3 in EMAC), but what are the consequences e.g. for the model developments, which model parametrisations should be improved?

We appreciate the positive rating of the diagnostics used. As discussed above, we agree that more specific suggestions should be provided with regard to model development, which are provided in a new Section:

"4.5 Suggestions for model improvement

In the following, the diagnosed model biases and suggestions for model improvement are summarised:

- *ICON-ART q_v: Here, the water vapour is a short-term forecast based on ECMWF IFS data, and the moist bias found in the ICON-ART data is comparable with the same bias in ECMWF data. Therefore, no specific improvement for ICON-ART can be suggested here. Suggestions to improve the ECMWF data are provided in the literature (e.g., Dyroff et al., 2015; Woiwode et al., 2020).*
- *ICON-ART O_3: The ozone is modelled by the LINOZ-scheme, which represents a linearised ozone chemistry, and by using a cold tracer. The observed bias might be reduced by tuning of this scheme. An optimized setup may be achieved by adaptation of the main parameters threshold temperature and lifetime of the cold tracer such that agreement with observations is improved (e.g. satellite observations such as MLS or field observations with suitable coverage).*
- *EMAC H_2O: The water vapour is simulated continuously in the EMAC model, i.e. it is neither reinitialised at 0 UTC nor nudged. The moist bias found in the EMAC simulation ranging from the troposphere to the LMS suggests that the cumulative impact of drying events in the entire altitude region is underrepresented in late winter. Such drying events might be precipitation events, which are dominated by ice and snow at the latitude and season associated with our case study. The parameterisation of ice nucleation and growth of ice particles might be optimized and tuned to improve the agreement with observations (e.g. satellite observations such as MLS or field observations with suitable coverage). Since our*

results show that the UT/LMS water vapour distribution is affected by model resolution in case of EMAC, a resolution-dependent tuning might be required.

- *EMAC O_3: Ozone in the EMAC model agrees well with the GLORIA data. Therefore, no significant suggestion for improvement can be provided here.*
- *EMAC HNO_3: Nitric acid is systematically underestimated by the EMAC model in most parts of the LMS, while it is overestimated in the tropopause region and slightly above. The clearly noticeable negative bias of EMAC HNO_3 in the LMS suggests that downward transport of this species by sedimentation of NAT particles originating from polar stratospheric clouds (PSCs) with associated nitrification of the LMS is underrepresented. While considerable progress has been made in the representation of NAT in model simulations in recent years, significant uncertainties remain in the microphysical parameterisation of NAT particles in PSCs (Tritscher et al., 2021 and references therein). More field observations of NAT containing PSCs would be helpful to improve model physics including, among other factors, NAT nucleation rates, particle sedimentation characteristics and particle size distributions, and thereby simulate the associated nitrification of the LMS more realistically.*
  The positive bias of HNO_3 in the tropopause region is even larger in EMAC-NOSCAV compared to EMAC-STD, i.e. results of EMAC-STD including scavenging processes are closer to the GLORIA observations in these regions. This suggests that scavenging processes of HNO_3 by high altitude cirrus clouds are relevant and might be underestimated in EMAC. An optimisation of the microphysical parameterisation of the scavenging process in the model with the help of observations might reduce this deficiency. Thereby, it should be taken into account that an optimisation of the representation of denitrification/nitrification by NAT particles might modulate the HNO_3 distribution here, too.

We propose to consider the model biases and deficits found here and our respective suggestions for future model development. As this work represents a case study, our findings hint at model deficiencies that might also be present in different seasons or latitudes. Further observations and model validation studies are needed to investigate these issues and to pinpoint these deficiencies to the respective deficits in the parameterisations."

A corresponding summary statement has been included in Section 5.

Third, how representative are the findings based just on one individual flight? Does e.g. ozone also show discrepancies for the early winter, or is HNO_3 affected by scavenging during other months, is the cloud mismatch a general problem, etc.

We agree that comparisons for other seasons would also be really interesting. However, our goal was to present a case study with a detailed and focused analysis of a single flight, while the investigation of entire seasons is beyond the scope of our study. To address the representativeness of our study and following the related suggestion by Referee 2, we now discuss the representativeness of our results in the discussion and conclusions Section:

"The GLORIA data were measured during a single flight on 26 February 2016 with a duration of 9 hours 40 minutes and a total distance of ~8000 km. The flight covered multifaceted scenario of the UT/LMS at high latitudes performed prior to the final major warming (Manney and Lawrence, 2016,

and Matthias et al., 2016). Therefore, the presented comparisons of the GLORIA and model data can be considered representative for the polar UT/LMS at high latitudes in late winter prior to the vortex breakdown."

It is an important issue to assess the capabilities of models of different kind to represent the composition of the UTLS and thus merits publication, but the focus of the given study is difficult to find.

Since the paper is not intended to provide novel aspects of atmospheric sciences, but focuses on the capabilities of models to represent tracer fields in complex regimes, a publication in GMD should also be considered.

As discussed above, we now included a new Section with more specific suggestions for model improvement. Following the suggestion by the Referee given below and to put a stronger focus on novel aspects of atmospheric research, we now investigate the evolution of the narrow filaments from troposphere-to-stratosphere exchange observed by GLORIA with the aid of ICON-ART.

Major point: Since the two models differ fundamentally in their basic properties it would be desirable to have at least one similar set of resolutions for comparison, especially since the authors emphasize the importance of resolution for their conclusions. The R2B6 simulation would allow for direct comparisons between EMAC T106 and ICON ART or at least coarse graining of the R2B7 data to the approx. T106 grid spacing at 70N would provide more consistency between both data sets. Alternatively one could think to use a high resolution EMAC simulation corresponding to the R2B7 setting (which might, however, be too expensive...). I highly recommend to add at least one comparison at similar resolutions.

Agreed. See above: We now included the ICON-ART global domain in the discussion of the cloud masks (Fig. 5) and replaced the nested ICON-ART data by the global grid ICON-ART data in Figure 7. Differences between the global and nested ICON-ART data are analysed in a new Figure (Fig. 8). In both the cloud masks and the trace gas data, only small differences are found between the global and nested domain and are attributed to the more detailed representation by the nested grid. No systematic biases are identified.

Another principle question for the comparison with GLORIA is the use of weighting functions. Since I'm not familiar with GLORIA data, aren't kernels necessary for a quantitative comparison?

We agree with the referee that this aspect should be addressed. We added the following statement in Section 2.4:

"The vertical resolution of the GLORIA data used here is around 500 m and therefore comparable with the vertical resolution of the simulations by both models in the tropopause region. Therefore, the use of averaging kernels in the vertical domain, such as often used in context of vertical profiles retrieved from satellite limb observations (e.g. Microwave Limb Sounder (MLS)) that are characterized by notably coarser vertical resolution is not expected to improve the comparison significantly. Furthermore, 2D-effects due to the limited resolution of the GLORIA observations along the line-of-sight are expected to cancel out in the correlations due to the large amount of data. Therefore, the computationally demanding use of 2d averaging kernels (see Ungermann et

al., 2011), in particular in the case of the GLORIA high spectral resolution observations used here, is not expected to change the comparison significantly."

Specific points:

p.16., line 15 (also line 23): Why vortex remnant? Couldn't it be just stratospheric air, which descended as part of the stronger downwelling in the high latitude stratosphere outside the vortex?

We thank the Referee for pointing out this missing information. The vortex characteristics of the observed air masses were analysed by Johansson et al. (2019), who showed that large regions of the air masses covered by the flight, and particularly the descended air masses above Canada, were inside the vortex according to the dynamical vortex criterion by Nash et al. (1996). We added a corresponding statement in the text.

p.18, line 32,33: Again, if one compares both models at the same coarse grid spacing, how does this affect ICON-ART H2O gradients?

See above: The comparison of the global and nested ICON-ART data in the new Figure 8 confirms the absence of overall systematic biases.

Also: When only using stratospheric data away from the tropopause (e.g. H2O for Ozone > 400 ppbv or PV > 8 PVU): How large ist the water vapour bias away from the gradient regions?

Following the suggestion by the Referee given below, we revised Figure 9 (see below). We extended the colour scale to a range from 0 to 12 and overplotted binned data points in intervals of 1 PVU and their standard deviations. The data points confirm that the moist bias is present also above 8 PVU and decreases (but does not disappear) towards higher altitudes. We updated the corresponding discussion in Section 4.4 under consideration of the updated Figure 9.

p.18 and Fig. 4/8: The enhanced water vapour from GLORIA above the 4 PVU implies cross tropopause exchange. This is an interesting case which would be much stronger, if the authors could provide evidence on the process, by e.g. analyzing trajectories or the history of the moisture filaments before the time of flight by comparing e.g. dynamical tropopause altitude, Lagrangian cold points and moisture evolution in both models before the flight. This would also provide a strong case for publication in ACP.

We thank the Referee for this very helpful suggestion. We now analysed the evolution with the help of ICON-ART in the vertical and horizontal domain until 3 days before the flight. We added the following discussion and a new Figure 10 in Section 4.3:

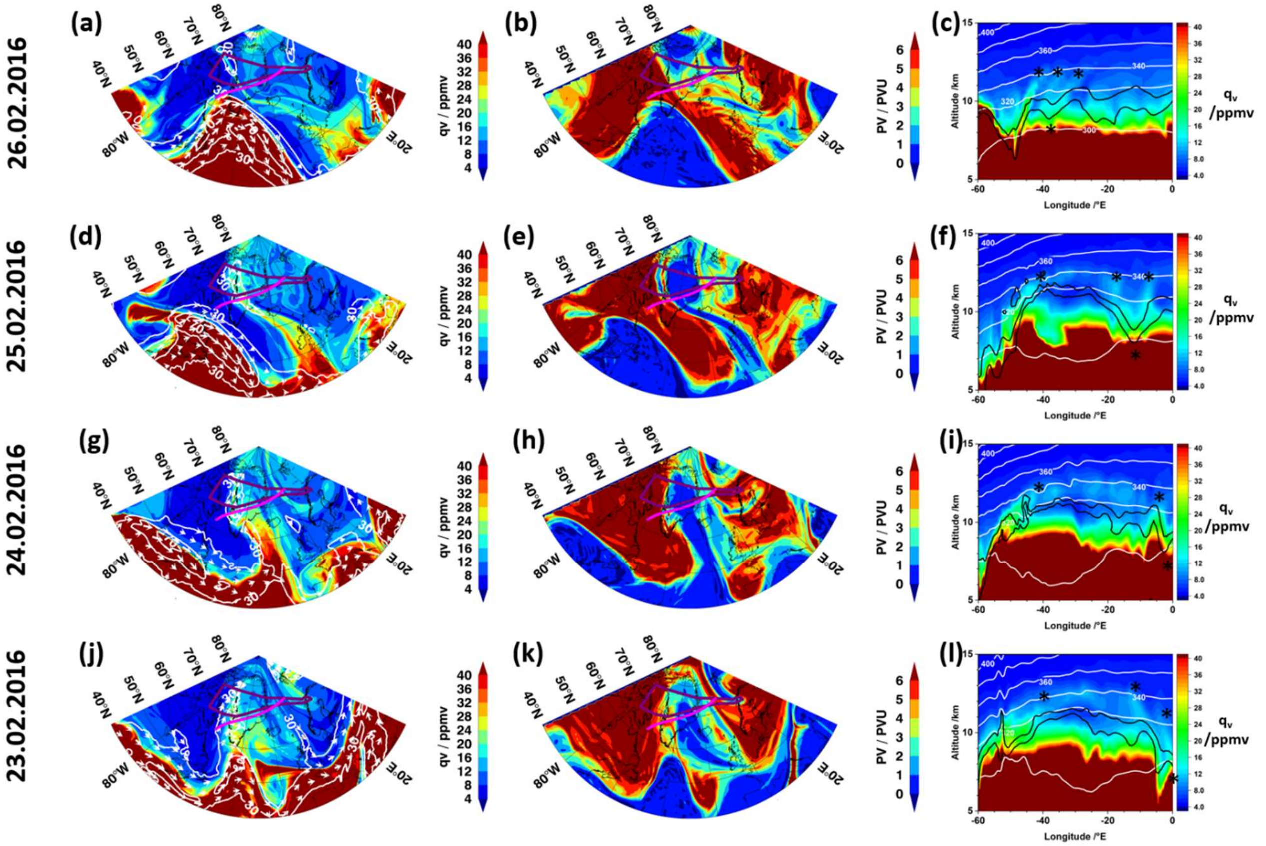

Figure 10: *Evolution of filaments in nested ICON-ART domain. (a,d,g,j) Horizontal distribution of q_v (coloured contour) and horizontal wind speed (white contour lines, in intervals of 20 m s^{-1}, and arrows) and (b,e,h,k) PV (coloured contour) at 10 km altitude. (c,f,i,l) Vertical distribution of q_v (colored contour, in ppmv), potential temperature (white contour lines, in intervals of 20 K), and 2 and 4 PVU isoline (lower and upper black line) as indicator for the dynamical tropopause. Purple lines in in the left and middle column indicate the flight track and magenta lines the location of the vertical cross sections shown in the right column. Stars (c,f,i,l) indicate features in these panels which correspond with features in the other panels out of these. The model data is shown at 12 UTC of the dates indicated in the left.*

"The evolution of the filaments seen in the GLORIA and model data is analysed with the help of ICON-ART. Figures 10a,d,g,j show the horizontal distribution of water vapour and horizontal wind from 23 until 26 February 2016 at 10 km altitude. The wind contours south of ~60°N show the polar jet with meridional undulations, characteristic of a midlatitude Rossby wave (e.g. Gabriel and Peters, 2008; Wirth et al., 2018), which also manifests in the gradients of q_v and PV (Fig. 10b,e,h,k). It separates moist upper tropospheric air masses in the south (high q_v, low PV) from dry stratospheric air masses in the north (low q_v, high PV). On 23 February 2016, the water vapour distribution in a ridge above southern Greenland is patchy, the jet is split into a northern and southern branch, with the northern branch carrying moist tropospheric air northward (Fig. 10j). The ridge formed previously in a complex Rossby wave pattern above North America (not shown). The evolving moist filament is elongated towards the pole in the following two days (Fig. 10g,d). At the same time, the moist upper tropospheric air masses in the south move on eastwards, while an occlusion forms at the Icelandic low at south-eastern tip of Greenland in front of the ridge connected with the Azores high (see Fig. 4c). The wind speeds of the resulting northward-moving jet stream

band in Fig. 10a decrease, resulting in the narrow moist filaments found at the flight day above central Greenland and a weak jet stream band in the northwest. Moist upper tropospheric air masses associated with the ridge above south of Greenland on 23 February 2016 (Fig. 10j) and the moist filament (Fig. 10g,d) are framed by strong PV gradients (compare Fig. 10k,h,e). Only a narrow filament with weak PV gradients remains at the flight day (compare Fig. 10a with Fig. 10b).

In the region of the moist upper tropospheric air masses south of Greenland and the evolving broad filament with low PV towards the pole on the following days (Fig. 10k,h,e,b), the PV distribution shows meridional overturning of the PV gradient that frames the moist upper tropospheric air masses. The pattern suggests poleward breaking of a cyclonically sheared Rossby wave (e.g. Gabriel and Peters, 2008 and references therein). Thereby, a separate isolated large patch of low PV values above west Greenland and the Atlantic on 23 February 2016 (Fig. 10k) combines with the moist upper tropospheric air masses with low PV in the south and seems to result from another Rossby wave breaking event that had previously occurred. As a consequence, a long broad filament with low PV stretches up to 80°N on the following days (Fig. 10h,e). On the flight day, a patch of low PV north of Greenland has been cut off almost completely from the moist upper tropospheric air masses in the south (Fig. 10b).

The vertical cross sections shown in Fig. 10l,i,f,c correspond with the magenta lines in the left and middle column. The locations of the cross sections were chosen with the intention to cover the area sampled by GLORIA and to capture the connected atmospheric structures in the vicinity that are discussed above. As can be seen from the vertical cross sections shown in Fig. 10l,i,f,c, the evolving filaments are framed in the west and east by steep gradients in tropopause height. The larger moist filament originates from the region around the jet stream band that branched away during the Rossby wave breaking event (compare Fig. 10j,g,d,a). It is aligned nearly parallel to the 320 and 340 K isentropic levels on 23 February 2016 (Fig. 10l). At lower altitudes, the 300 K isentropic level crosses the dynamical tropopause in the west in Fig. 10l,i,f,c. As discussed by Shapiro (1980), such regions provide suitable conditions for bidirectional cross-tropopause exchange. At higher altitudes, the 4 PVU isoline crosses the 320 K isentropic level in the same region and suggests conditions suitable for isentropic transport across horizontal PV gradients also here.

Local oscillations of the isentropic levels on 23 February 2016 between 55 and 50°W are attributed to a mountain wave above southern Greenland (Fig. 10l). During the following days, the moist filament aligns steeper across the isentropic levels (Fig. 10i,f). In the same region, oscillations of the dynamical tropopause become weaker on 24 February 2016, and patches of enhanced PV remain until 25 February 2016. On 26 February 2016, the remaining narrow moist filament is aligned along a newly formed tropopause fold in the west and reaches steeply into the LMS (Fig. 10c). Note however that the air masses seen in these panels are also modulated by horizontal transport in meridional direction and therefore have to be interpreted in combination with the maps shown in the left and middle row of Fig. 10.

The other two filaments on 23 February 2016 in the east are associated with a tropopause fold remnant in the east (Fig. 10l). The tropopause fold remnant declines during the subsequent days, moves west (Fig. 10i,f) and joins with the newly formed tropopause fold in the west on 26 February 2016 (Fig. 10c). Since these two filaments are aligned steeply across the isentropic levels already on

23 February 2016, they are interpreted as older structures that were previously formed in a similar way like the stronger filament in the west.
Overall, the vertical cross sections in Fig. 10l,i,f, c show that the filaments observed by GLORIA evolved along steep gradients of the dynamical tropopause in connection with Rossby wave breaking. The larger filament in the west evolved during a Rossby wave breaking event, where moist air tropospheric masses were transported horizontally into the Arctic LMS along the jet stream under conditions suitable for cross-tropopause exchange. The other two filaments are interpreted as older structures in connection with a tropopause fold remnant in the east that probably evolved during a previous Rossby wave breaking event."

In our new detailed analysis, we noticed that the troposphere-to-stratosphere exchange occurred in a more complex scenario, which is connected with the occlusion of the Icelandic low, but with further aspects playing an important role (i.e. Rossby wave breaking and jet stream split). Furthermore, our interpretation is that exchange mainly occurred at the days before the flight. We updated the discussion accordingly in the abstract, Section 4.3, and Section 5. We removed the term "mixing" from the abstract, since this aspect is not analysed in detail by our study.
We furthermore now mention Rossby wave breaking and associated stratosphere-tropopause exchange in the introduction and added references by Gabriel and Peters (2008), Wirth et al., (2018), and Jing et al. (2018).
Recalling that in ICON-ART the meteorological fields including q_v are reinitialized every day at 0 UTC from ECMWF IFS data, and since the EMAC meteorological fields are nudged to ECMWF data, too, we think that there is not much benefit in repeating the same analysis with EMAC. Differences in the shape of the filaments can be explained by the lower resolution of EMAC and the fact that the EMAC meteorological fields are nudged, while the EMAC H_2O data is simulated continuously. Therefore, dynamical features like tropopause gradients that are nudged can be "shifted" slightly versus the H_2O field. We added a corresponding discussion in the text.

p.22, line 12/13: "This in turn means...": I can't really follow the statement: What is meant with "this region"? Further: Why does trapping with high altitude cirrus affect the lower stratospheric data? Or do the authors refer to the upper troposphere only? Finally "... could play a significant role " for what?
We agree that the discussion and Figure B1j are difficult to follow. Our intention was to elaborate that in the EMAC standard simulation, HNO_3 is overestimated with respect to GLORIA between 2 and 4 PVU, and slightly below. Since the EMAC-NOSCAV simulation results in even more HNO_3 here than the EMAC-STD simulation, the EMAC-STD simulation including scavenging is closer to the observation here.
Following the suggestion by Referee 2, we shifted the residual plots in Fig. B1 to Figure 7. For better clarity, we have reversed these residuals and now show EMAC-STD minus GLORIA (instead of the other way around) in the corresponding panels that were added to Figure 7, such that a positive bias in the simulation is indicated by a positive sign.
Our interpretation of the effect in the lower LMS is that trapping by cirrus cloud plays a role here in the EMAC simulation. Cirrus clouds are known to occur also in the LMS (e.g. Spang et al. 2015) and

are not excluded in the EMAC simulation. Furthermore, troposphere-to-stratosphere exchange is likely to involve tropospheric air masses that were previously affected by HNO_3-trapping in cirrus clouds, thereby modulating HNO_3 in the LMS. We added the interpretation in the text. Our statement regarding the significant role refers to the tropopause region and LMS. We revised the discussion accordingly.

P.22, line 6 ff.: Why does scavenging has an effect up to 1 km above the 4 PVU surface (Fig. 11f) throughout the measurement region? Wouldn't this imply clouds in the stratosphere over the entire region? Even given the sporadic events shown in the appendix I find this puzzling... Is there any other diagnostic confirming this?

See previous reply: Cirrus clouds are known to occur also in the LMS. Furthermore, LMS composition is affected by troposphere-to-stratosphere exchange involving tropospheric air masses which were previously affected by cirrus clouds and HNO_3 trapping. Since HNO_3 is simulated continuously by EMAC, the cumulative effect found in Fig. 11f (now Fig. 13f) is not surprising from our point of view and does not require clouds to be present in the LMS in the entire region at the flight day.

Figure 4:

To diagnose the exchange region, add a panel showing the altitude of the PV=2pvu surface. Figure 4 currently does not provide any indication of cross tropopause exchange.

We agree that Figure 4 alone is not sufficient to explain cross tropopause exchange without diagnosing the tropopause. Following the suggestion by the Referee above, we now analyse the evolution of the filaments in Section 4.3 and include a new Figure showing the evolution of the filaments, wind and dynamical tropopause during the previous days (new Fig. 10, see above). In the caption of Figure 4, we replaced "the troposphere-to-stratosphere exchange region" by "filaments observed by GLORIA and analysed in the model data in Sections 4.2 and 4.3", since the aspect of troposphere-to-stratosphere exchange is analysed later in the manuscript.

Figure 9:

Since the overplotting of data points may mask some important details of the distributions, I strongly recommend the following: One could easily calculate the mean and standard deviation of each species in bins of e.g. 1 PVU and could overplot this on the Figures 9a)-9e).

We thank Referee 1 for the helpful suggestion. As suggested, we revised the plots and included binned data points in bins of 1 PVU in Figure 9 (now Fig. 11). We updated the discussion in the text accordingly.

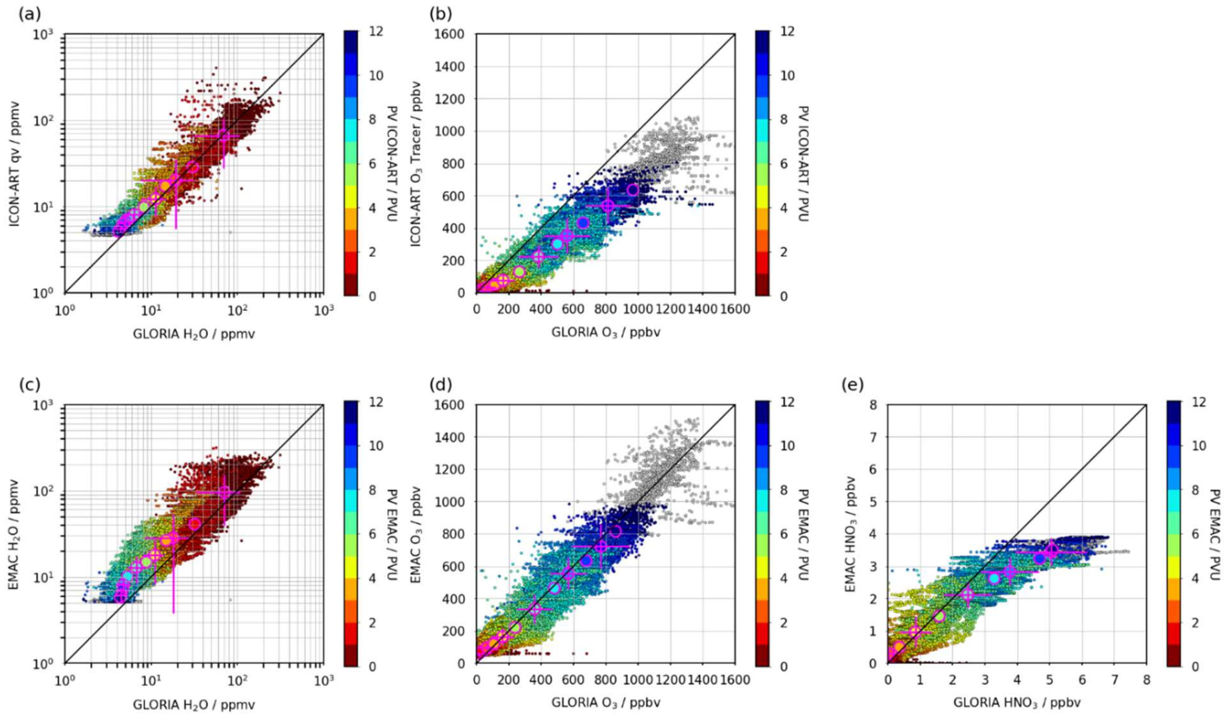

Figure 11: *Correlation of GLORIA H_2O, O_3 and HNO_3 to corresponding ICON-ART and EMAC output variables. The large data points framed in magenta are a binned representation of the small data points. Magenta bars indicate the standard deviation of the binned data points. Colour-coding: PV from corresponding model.*

Why does Fig 9.e) shows roughly a 1:1 relation for low PV-values (< 4 PVU), but a systematic difference in B1.(i)? (Eventually this discrepancy disappears after considering my previous comment to Fig.9).

See above: we clarified the representation of the residual plot (Fig. B1j, now: Fig 7m) and show EMAC minus GLORIA. The plot now shows the high bias of the model in the tropopause region with a positive sign. The high bias in the plot is consistent with the revised correlation plot in Fig. 9e, where a small positive bias is seen below 5 PVU. We thank the referee for pointing out the unclear representation.

Caption Figure 11: Please add "T106 minus T42 resolution"

Done

References

Dyroff, C., Zahn, A., Christner, E., Forbes, R., Tompkins, A. M., and van Velthoven, P. F. J.: Comparison of ECMWF analysis and forecast humidity data with CARIBIC upper troposphere and lower stratosphere observations, Q. J. Roy. Meteor. Soc., 141, 833–844, https://doi.org/10.1002/qj.2400, 2015.

Gabriel, A. and Peters, D.: A diagnostic study of different types of Rossby wave breaking events in the northern extra-tropics, J. Meteorol. Soc. Jpn., 86, 613–631, 2008.

Jing, P. and Banerjee, S.: Rossby wave breaking and isentropic stratosphere-troposphere exchange in 1981–2015 in the Northern Hemisphere, J. Geophys. Res.-Atmos., 123, 9011–9025, https://doi.org/10.1029/2018JD028997, 2018.

Nash, E. R., Newman, P. A., Rosenfield, J. E., and Schoeberl, M. R.: An objective determination of the polar vortex using Ertel's potential vorticity, J. Geophys. Res.-Atmos. (1984–2012), 101, 9471–9478, https://doi.org/10.1029/96JD00066, 1996.

Shapiro, M. A.: Turbulent mixing within tropopause folds as a mechanism for the exchange of chemical-constituents between the stratosphere and troposphere, J. Atmos. Sci., 37, 994–1004, 1980.

Tritscher, I., Pitts, M. C., Poole, L. R., Alexander, S. P., Cairo, F., Chipperfield, M. P., Grooß, J.-U., Höpfner, M., Lambert, A., Luo, B. P., Molleker, S., Orr, A., Salawitch, R., Snels, M., Spang, R., Woiwode, W., and Peter, T.: Polar Stratospheric Clouds Satellite Observations, Processes, and Role in Ozone Depletion, Rev. Geophys., 59, e2020RG000702, https://doi.org/10.1029/2020RG000702, 2021.

Ungermann, J., Blank, J., Lotz, J., Leppkes, K., Hoffmann, L., Guggenmoser, T., Kaufmann, M., Preusse, P., Naumann, U., and Riese, M.: A 3-D tomographic retrieval approach with advection compensation for the air-borne limb-imager GLORIA, Atmos. Meas. Tech., 4, 2509–2529, https://doi.org/10.5194/amt-4-2509-2011, 2011.

Wirth, V., Riemer, M., Chang, E. K. M., and Martius, O.: Rossby Wave Packets on the Midlatitude Waveguide – A Review, Mon. Weather Rev., 146, 1965–2001, https://doi.org/10.1175/MWR-D-16-0483.1, 2018.

Woiwode, W., Dörnbrack, A., Polichtchouk, I., Johansson, S., Harvey, B., Höpfner, M., Ungermann, J., and Friedl-Vallon, F.: Technical note: Lowermost-stratosphere moist bias in ECMWF IFS model diagnosed from airborne GLORIA observations during winter–spring 2016, Atmos. Chem. Phys., 20, 15379–15387, https://doi.org/10.5194/acp-20-15379-2020, 2020.