Final response to referee comments of "Challenge of modelling GLORIA observations of UT/LMS trace gas and cloud distributions at high latitudes: a case study with state-of-the-art models"

Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2021-574, in review, 2021
F. Haenel et al.

We thank Referee 1 and Michelle Santee for their time, valuable comments and corrections to improve the manuscript.

In the following find our final responses to the comments from both referees.

Sincerely,
Florian Haenel on behalf of all co-authors

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."

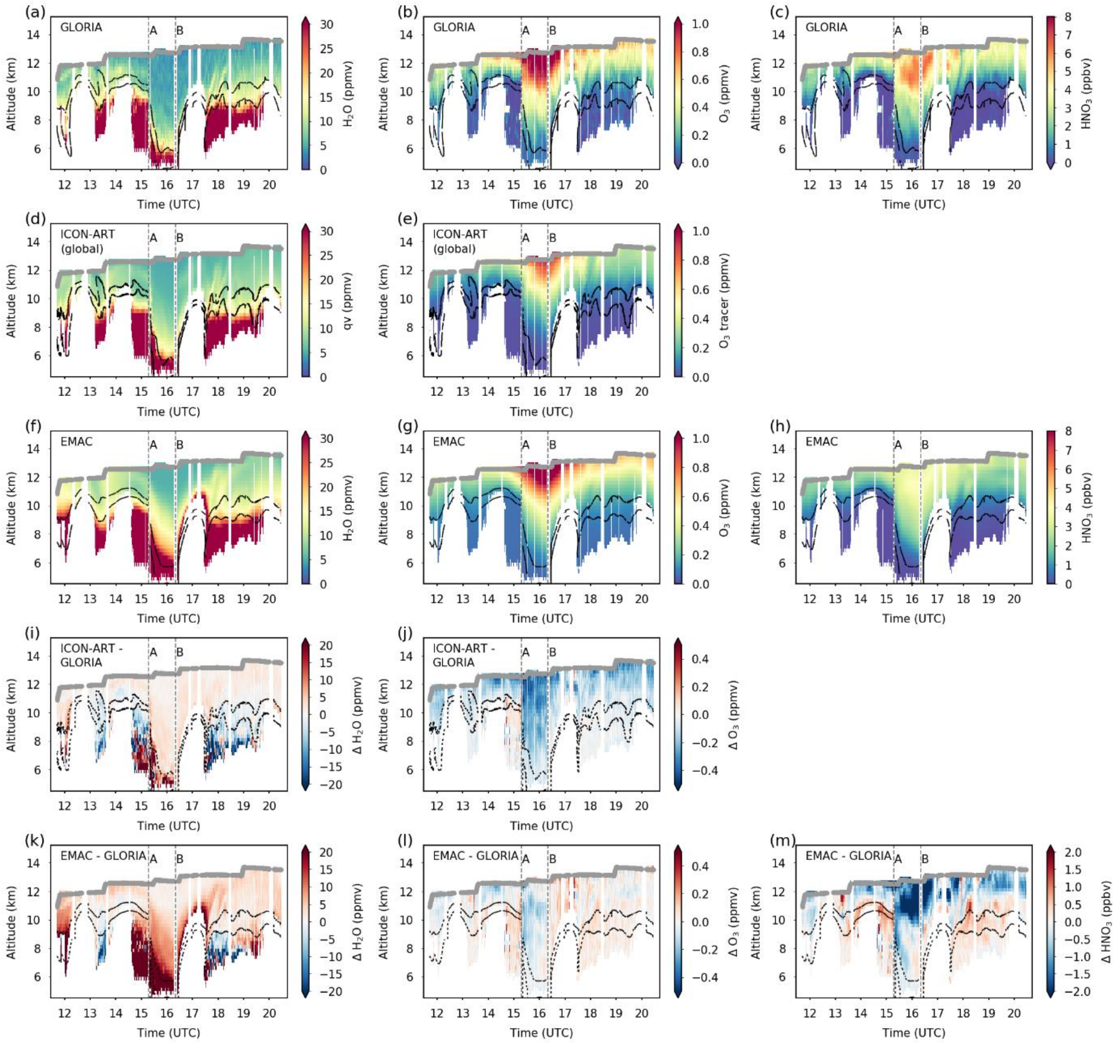

Figure 7. *Observed and modelled trace gas distributions. GLORIA observations of water vapour, ozone and nitric acid (a-c). ICON-ART (global R2B6 grid) short-term forecast of specific humidity (d) and free-running simulation of ozone using simplified ozone depletion parameterisation (e). EMAC free-running 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.*

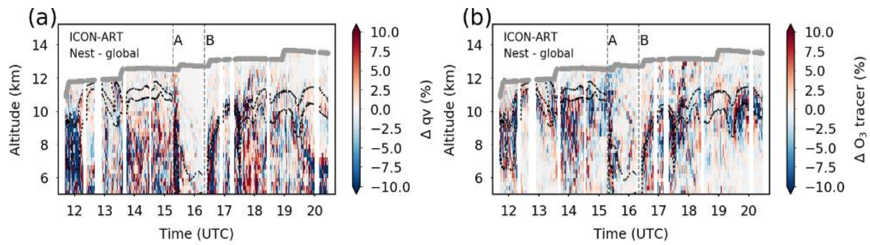

Figure 8. *Residuals between the ICON-ART nested R2B7 and global R2B6 domains of q_v (a) and O_3 (b). Black lines: 2 PVU and 4 PVU isolines as indicators for the dynamical tropopause. Grey lines: HALO flight altitude.*

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:

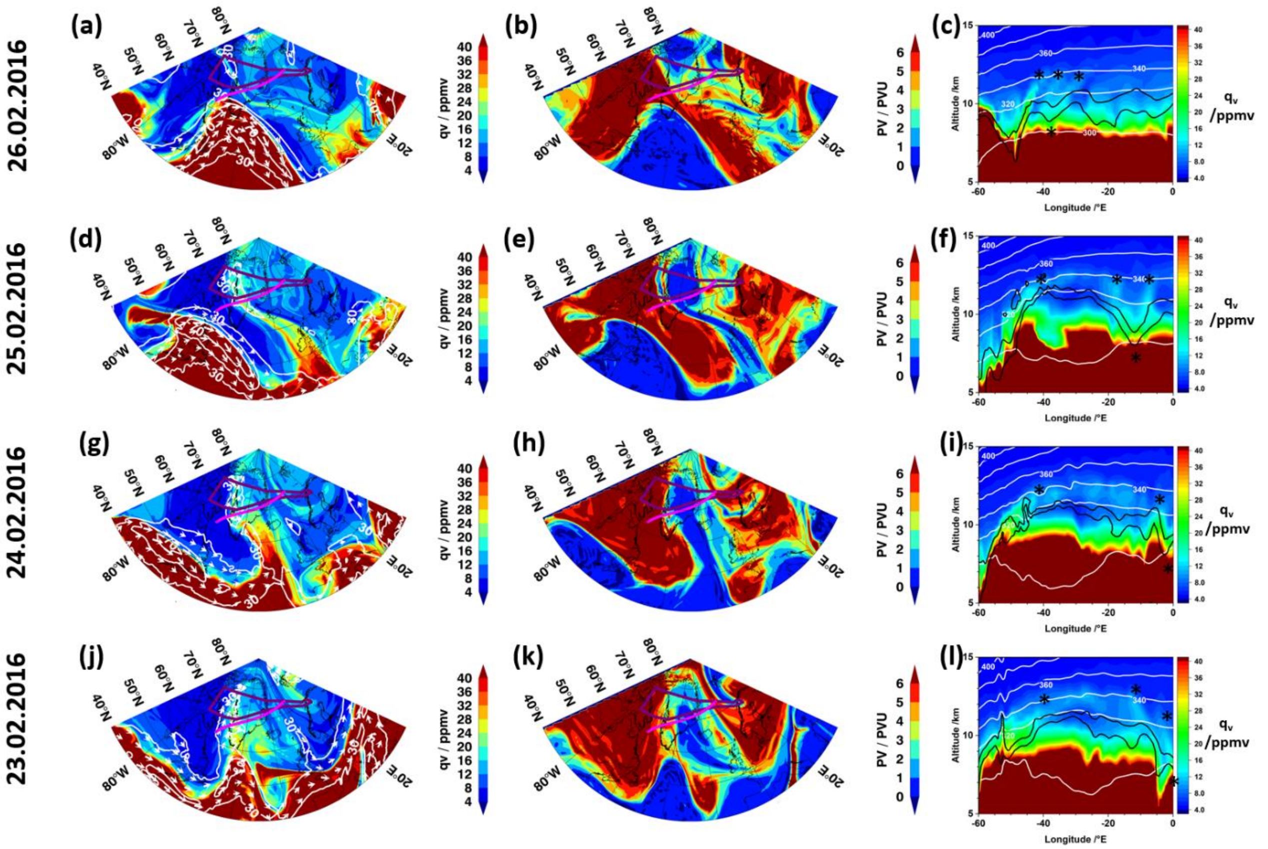

__Figure 10:__ Evolution of filaments in nested ICON-ART domain. (a,d,g,j) Horizontal distribution of q_v (coloured contour) and horizontal wind speed (white contour lines, in intervals of 20 m s^{-1}, and arrows) and (b,e,h,k) PV (coloured contour) at 10 km altitude. (c,f,i,l) Vertical distribution 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.

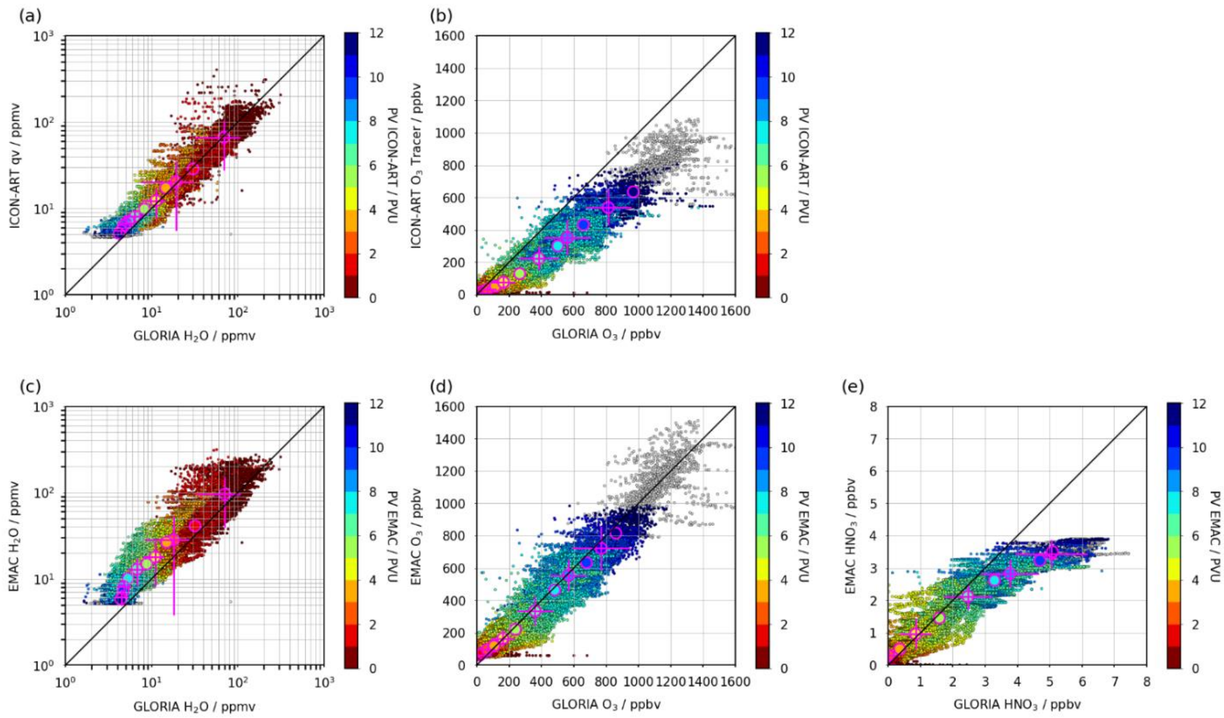

Figure 11: *Correlation of GLORIA H_2O, O_3 and HNO_3 to corresponding ICON-ART and EMAC output variables. The large data points framed in magenta are a binned representation of the small data points. Magenta bars indicate the standard deviation of the binned data points. Colour-coding: PV from corresponding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

Answer to Comment by Michelle Santee (Referee 2)

20.12.2021

We thank the Michelle Santee very much for her time, valuable comments, and detailed corrections, 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*.

Review of "Challenge of modelling GLORIA observations of UT/LMS trace gas and cloud distributions at high latitudes: a case study with state-of the-art models" by Haenel et al.

This manuscript uses GLORIA measurements taken on a PGS flight that sampled a diverse set of conditions in the UT/LMS to test the ability of two models, ICON-ART and EMAC, to simulate cloud structures and trace gas (H_2O, O_3, HNO_3) distributions. Both models are shown to reproduce the observations quite well; discrepancies between modelled and measured cloud and composition fields are quantified and their causes investigated. The paper is well organized and well written, and the figures are generally well prepared and support the discussion. I have only a few substantive issues that I would like to see addressed before the paper is accepted for publication; most of my comments are minor wording suggestions that will take very little time to act on.
Below both major substantive issues and minor points of clarification, wording suggestions, and grammar / typo corrections are listed together for each Section in sequential order.

Respectfully,
Michelle Santee

We thank Michelle Santee for the accurate summary and appreciate the positive rating.

Abstract:

- **p1L18-19: The wording "measurements taken in a challenging case study by the GLORIA" could be interpreted to imply that that flight was deliberately designed for this purpose, which I do not believe was the case. I think it would be better to simply say "measurements taken in a flight of the GLORIA" here and then add "challenging" in front of "multifaceted" in L23.**
  Agreed and done
- **p1L21: 2016, which --> 2016 that**
  Done
- **p2L3: moist-bias --> moist bias**
  Done
- **p2L7: changing of the --> changing the**

Done

- **p2L8: play only a role in case of HNO_3 --> play a role only in the case of HNO_3**
  Done
- **p2L8-9: I agree that the representativeness of these results is an important question that should be explored. However, unless I missed it, this issue is not raised anywhere in the paper other than this sentence in the abstract. It should be acknowledged elsewhere as well, at least in the Discussion and Conclusions section if not somewhere in the main text of the paper**.
  We agree that the representativeness of the results of our study should be addressed. We revisited the abstract and came to the conclusion that the corresponding statement there does not provide much helpful information. Therefore, we rephrased this part of the abstract and address the representativeness of our study now in more detail in the discussion and conclusions (Section 5):
  "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 a 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."
- **p2L10: projection --> projections**
  Done
- **p2L10: Although this study has certainly provided very useful information to characterize model biases, I am less convinced that it has really laid out sufficiently specific guidance to "define paths for further model improvements". See related final comment on the Conclusions below.**
  We agree that this wording might has overrated a bit the outcome of our study. We changed the wording to "…and provide suggestions for further model improvements." We furthermore agree that suggestions for model improvement should be summarized and discussed more clearly. We now summarize the observed model biases and provide suggestions for model improvement in the new Section 4.5 "Suggestions for model improvement" (see reply to Referee 1). Furthermore, a corresponding summary statement has been added to discussion and conclusions (Section 5).

Section 1: Introduction:

- **In a number of places in the presentation of background material (e.g., p2L13, p2L16, p2L21, p2L27, p2L28, p3L7), a few citations are given for very well-established concepts, but many other equally suitable papers could have been cited instead of or in addition to the ones listed. Obviously not all relevant papers can be referenced for these points, but "e.g." should be added in these lines to avoid giving readers the impression that the selected references are the only appropriate ones.**
  Agreed and done
  p2L17-18: spread in these trends among models while perturbating ozone and other greenhouse gas abundances --> spread among modelled trends when ozone and other greenhouse gas abundances are perturbed

Done

- **p2L18: can be --> include**
  Done
- **p2L21: knowledge on --> knowledge of**
  Done
- **p2L23: compartment --> layer**
  Done
- **p2L24: On the winter --> In the winter**
  Done
- **p3L6: sedimentation … redistribute --> sedimentation … redistributes; "eventually" is not really needed here, but if the authors want to keep it, it should come before "changes"**
  Agreed and done
- **p3L12: I do know what is meant by "(in parts)" in this sentence; if I have understood the intent, this would be better as: "in part explicitly and in part by using" or "both explicitly and by using"**
  We have changed the sentence to "in part explicitly and in part by using".
- **p3L13: such models are the models ICON --> such models include ICON**
  Done
- **p3L34: I assume that the systematic biases referred to here are in the model fields (that is, the intention is not to use the model results to validate the GLORIA data), but that should be made explicitly clear, e.g.: in the trace gas distributions --> in the modelled trace gas distributions**
  Agreed and done

Section 2: Data and diagnostics:

- **p4L10: the used model setups --> the model setups used**
  Done
- **p4L11: overview on --> overview of**
  Done
- **p4L15: aircrafts --> aircraft**
  Done
- **p4L24: combined to --> combined into**
  Done
- **p5L7: is operational --> has been operational**
  Done
- **p6L16: delete the comma after "winter"**
- *Done*
- **p6L19: air masses suitable for --> air masses whose conditions are conducive to**
  Done
- **p7L1: life time --> lifetime**
  Done
- **p7L9: does "between ~12 to 21 hours" mean "between ~12 and 21 UTC"?**

We have changed the sentence to "forecasts with lead times of ~12 to 21 hours", as we meant the lead time (i.e. the running time) of the forecast. Here, this time is identical with the time points during the flight, since the model data is interpolated in space and time to the GLORIA geolocations. For clarification, we added: "(depending on point in time during flight)"

- **p7L16: atmosphere --> atmospheric**
  Done
- **p7L20: in the --> with; T106L90MA-resolution --> T106L90MA resolution**
  Done
- **Fig. 3 caption: The corresponding T106 (T42) grid corresponds --> The T106 (T42) grid corresponds; reduces --> is reduced**
  Done
- **p7L23-25: This is a very awkwardly worded and unclear sentence. If I have understood it correctly, it would be clearer to say: "To simulate realistic synoptic conditions, surface pressure and various prognostic variables (temperature, vorticity, and divergence) are "nudged" towards the ECMWF ERA-Interim reanalysis (Dee et al., 2011) above the boundary layer and below 1 hPa using a Newtonian relaxation technique." This formulation of the sentence also introduces the term "nudged", which is used later in the manuscript but is not currently defined.**
  Agreed, thanks. We adapted the suggested wording.
- **p7L26: a comprehensive chemistry --> a comprehensive chemistry scheme**
  Done
- **p7L27-30: Because of the complexity of the punctuation in these lines, some of the commas need to be replaced with semicolons (marked in red here): "… (Sander et al., 2011); the photolysis submodel JVAL (Sander et al., 2014); the submodel MSBM, mainly responsible for the simulation of PSCs (Kirner et al., 2011); the submodel CLOUD, based on the ECHAM5 cloud scheme, simulating large scale clouds (Roeckner et al., 2006); the submodel CONVECT, calculating convection and convective clouds (Tost et al., 2006b); and …".**
  Done
- **p9L9: combined to --> combined into**
  Done
  p9L17: delete "that deviate"
  Done
- **p9L18: e.g. --> i.e.**
- *Done*

Section 3: Flight overview and meteorological analysis:

- **p10L2: concerning the decades before --> relative to preceding decades**
  Done
- **p10L6: ended by --> ended with**
  Done
- **p10L9 and L11: take off --> takeoff**

Done

- **p1011-12: headed westwards (GLORIA pointing to northward directions) --> headed westward (GLORIA pointing northward)**
  Done
- **p10L12-13: turned to a southward direction (GLORIA pointing to westward directions) --> turned southward (GLORIA pointing westward)**
  Done
- **p10L13-14: back to eastward directions and … southwards) --> back eastward and … southward)**
  Done
- **Fig. 4 and its caption:**
  - **It is very difficult to make out the overlaid white contours without greatly magnifying the plot – thicker lines and larger font for the labels would make them easier to see.**
    Done
  - **It would be convenient to have waypoints A and B marked on these maps as well as on Fig. 1.**
    Done
  - **is colour-coded in contour --> is shown by colour contours; delete "together"; occlusions --> occlusions (black overlays)**
    Done, (we furthermore added "(black and dark grey overlays)", since further fronts and another occlusion above the Atlantic are shown in panel (d) in dark grey)
- **p10L21: west- --> westward-; way-point --> waypoint**
  Done
- **p10L24: partly dissipates on --> has partly dissipated by**
  Done
- **p10L24-25: pointed subsequently towards … and into --> pointed first towards … and then into**
  Done
- **p10L29: the wording "going along with" is not completely clear. Does this mean "consistent with" or "accompanied by"?**
  We meant "accompanied by". We corrected the wording accordingly
- **p10L30-31: move "to date" to after "on record"**
  Done
- **p11L6: Why is the word "subsequently" used here? The air masses observed during the flight contained these features – they were not observed subsequent to the flight.**
  Agreed. We deleted the word "subsequently".

Section 4: Observed and modelled cloud and trace gas distributions:

- **p12L4-5: It is stated here that CI values "approaching four and higher" are indicative of cloudfree conditions. Since the color scale in Fig. 5a saturates at CI=3.0, does that imply that on this flight GLORIA never encountered air masses that can be considered cloud-free?**

Thanks for pointing out this inconsistency. The threshold value of 4 applies to spaceborne limb-sounding observations, while different threshold values were found to be suitable for airborne limb sounding. We added the following explanation:
"In the case of airborne limb observations, CI values of 2 to 4 have been found to be suitable to separate between cloud-affected and cloud-free conditions in previous studies (Johansson et al., 2018 and references therein). In the case presented here, a cloud index of ~2.5 represents the threshold between cloud-affected and cloud-free conditions."

- **p12L9: It seems a bit odd to characterize air masses affected by clouds as having an "enhanced" CI since it is actually low values of CI that indicate the presence of clouds.**
  Agreed, we changed the wording to "reduced".
- **p12L15: used cloud masks --> cloud masks used**
  Done
- **p12L18: threshold of the cloud mask for the ICON-ART- and EMAC-model at --> threshold for the ICON-ART and EMAC model cloud masks at**
  Done
- **p12L22: add a comma after "concentrations"**
  Done
- **Fig. 5b,c: I am a bit confused about why the tick marks on the cloud mask color bars are needed**
  Agreed. We removed the tick marks
- **p12L33: Discrepancies between measured and modelled clouds at lower altitudes for the system around 20 UTC are attributed to GLORIA data being affected by optically dense cloud layers above, but couldn't this explanation be applied to the mismatch for other clouds as well, such as the one between 12 and 13 UTC?**
  Yes, agreed. We now mention in the text that the same effect might explain the differences between the observed and modelled cloud systems between 12 and 13 UTC at lower altitudes.
- **p13L1: delete the comma after "GLORIA"**
  Done
- **p13L3: delete the comma after "fact"**
  Done
- **p13L8: I'm not sure what "appear more sharply in the ICON-ART simulation" means, as the cloud systems in question barely register at all in the model cloud mask.**
  Agreed. We changed the sentence to "…are barely reproduced in the ICON-ART simulation"
- **p13L10: respective --> corresponding; EMAC-standard simulation (STD) --> EMAC standard simulation (EMAC-STD); T106L90MA-resolution --> T106L90MA resolution**
  Done
- **p14L1: Recalling that --> As mentioned earlier; 2.3), however the --> 2.3); however, the**
  Done
- **p14L2: the EMAC standard simulation (STD) --> the EMAC-STD**
  Done
- **p14L11: better comparable --> more comparable**
  Done

- **p14L19: delete comma after "model"**
  Done
- **p14L21: to which degree --> the degree to which**
  Done
- **p14L25: does "~12 h to ~20 h" mean "~12 UTC to ~20 UTC"?**
  In this case we mean the lead time (i.e. running time) of the forecast, which is however identical here with times of the GLORIA geolocations. For clarification, we added "forecast lead time" prior to "between".
  Also: accumulated --> cumulative
  Done
- **p15L4-5: It is stated that "all of the observed cloud systems coincide qualitatively with a corresponding precipitation pattern at the respective geolocations in the ICON-ART-data", but the Δ H2O diagnostic does not pick up the cloud system observed by GLORIA prior to 12 UTC.**
  We agree that there are only very weak indications in the ΔH2O diagnostic prior to 12 UTC (i.e. at ~11:45 and 11:55). We added: "However, as in the case of the ICON-ART cloud mask prior to 12 UTC, only weak indications of cloud systems are found here."

- **p15L8: Although I see weak negative residuals just below the tropopause, even with the figure greatly magnified it is difficult to discern non-negligible residuals above the tropopause.**
  Agreed. We changed the range of the colour scale to -10 … 10 ppmv to make it easier to identify these patterns:

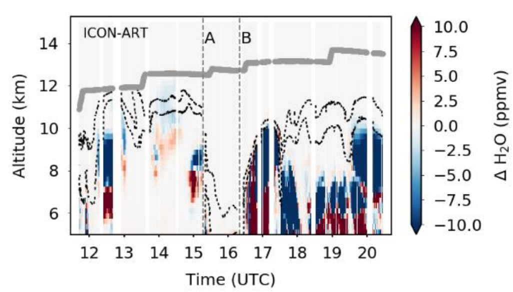

Figure 6. *Modelled short-term changes in specific humidity due to cloud processes. Residuals between nested ICON-ART domain of specific humidity and corresponding H_2O tracer without cloud microphysics. Black dashed lines: ICON-ART 2 PVU and 4 PVU isolines (lower and higher lines, respectively) as indicators for the dynamical tropopause. Grey lines: HALO flight altitude.*

- **p15L10-11: vicinity is found at 14 UTC and reaches --> vicinity at 14 UTC reaches**
  Done
- **p15L11: support that --> support the idea that**
  Done
- **p15L18: again, hints --> again hints**
  Done
- **p15L20: beside --> in addition to**
  Done

- **p16L3-4: The use of the term "precipitation" is ambiguous here – I believe that the authors mean "cirrus cloud ice particle sedimentation", but that should be clarified. I also think that it would be appropriate to add discussion putting these results about "precipitation" affecting the humidity of the LMS into the context of previous studies that have examined the impact of convection and cirrus cloud processes on moistening / dehydrating the LMS (especially in light of my previous comment that I had trouble identifying these weak signatures in Fig. 6).**
  Agreed, we meant "cirrus clouds ice particle sedimentation" and corrected the text accordingly. We rephrased p14/L19ff as following:
  "Another proxy for the characterisation of detectable cloud systems in the model is looking at the cirrus cloud ice particle sedimentation events, which include the processes of nucleation, sedimentation and subsequent evaporation of cirrus cloud ice particles. As a consequence, local irreversible dehydration is found when ice particle growth removed water from the gas phase, and hydration is found at lower altitudes where the particles sublimate."
  To provide context to previous studies, we added the following statement after p16L4:
  "Cirrus clouds under cold conditions in the LMS have been found by many observations (e.g. Lelieveld et al., 1999; Kärcher and Solomon, 1999; Spang et al., 2015) and are likely to affect LMS humidity by ice particle sedimentation (e.g. Kärcher, 2005). Furthermore, as discussed in the literature, convective hydration is known to affect the LMS and can drive air masses to saturation (Schoeberl et al., 2018; Zou et al., 2021)."
- **p16L4: affects also significantly --> also significantly affects**
  Done
- **p16L5: the major cloud systems --> the major cloud systems observed by GLORIA**
  Done
- **p16L7: the ICON-ART lacks the simulation of the --> ICON-ART fails to simulate the**
  Done
- **p16L11: add a comma after "q_v"**
  Done
- **p16L12: It should be reiterated when Fig. 7 is introduced that the presence of optically thick clouds precludes trace gas retrievals, as comparison of the patterns in Figs. 5 and 7 shows.**
  Agreed. We added: "When compared with the cloud index plot (Fig. 5a), gaps in the retrieved trace gas distributions are explained by the fact that the presence of dense clouds precludes trace gas retrievals in the affected regions. Cloud filtering is applied here prior to the trace gas retrieval."
- **p16L13-14: The tropopause is located near 10 km in all panels of Fig. 7, not just 7a. In addition, use either "around" or "~", not both (see also L16).**
  Done
- **p16L17: south-western --> southwestern; part --> flight segment**
  Done
- **p16L18-19: reach by ~2km up into --> reach as far as ~2km into**
  Done
- **p16L21: complimentary --> complementary (but "converse" is probably a better word here)**

Agreed, we have changed the word to "converse"

- **p16L25: reach up --> reach nearly up; add a comma after "altitude"**
  A closer inspection of the Fig 7b shows that the filaments in ozone reach even higher up to the flight altitude. We therefore inserted "even" instead of "nearly" and modified the end of the sentence to ", therefore deeper into the LMS than the filaments seen in the water vapour distribution." We added the comma, as suggested.
- **p16L27-29: Although I don't doubt that some nitrification at lower altitudes occurred during this winter, the morphology of the HNO_3 distribution (Fig. 7c) does not seem very different from that of O_3 (Fig. 7b) to me, and abundances of both would be expected to be higher in the LMS than in the UT. Thus I am not certain what local maxima in HNO_3 are being referred to here. The specific signatures of nitrification in this figure should be clarified.**
  Thanks for pointing out this shortcoming. We agree that the local maximum can hardly be seen with the applied range of the colour bar and the discussion is unclear. We extended the colour bar to 8 ppbv to resolve the maximum more clearly. Furthermore, there was a mistake in the time interval. Our intention is to discuss the local HNO_3 maximum below flight altitude between 15 UTC and ~17 UTC seen in the updated plot, which is not found in the O_3 distribution.
  Furthermore, Ziereis et al. (2021) discuss in their recent publication that during this flight both nitrified air masses, but also denitrified air masses that had descended from above were probed. We corrected the sentence with regard to the time interval of the local maximum. Furthermore, with reference to Ziereis et al. (2021), we now discuss that both, nitrified air masses (prior to ~15 UTC and after ~17 UTC) and denitrified air masses (between ~15 and 17 UTC) were probed at flight altitude. We discuss furthermore that the local maximum seen in the GLORIA data below flight altitude between 15 and 17 UTC is interpreted as subsided nitrified air masses which are located below denitrified air masses at flight altitude in this section of the flight.
  This interpretation is furthermore consistent with the results of the EMAC sensitivity run, which shows nitrification below flight altitude between 15 and 17 UTC (Fig. 10c, now Fig. 12c).
- **p17L8: is the comparison ("higher") with respect to ICON-ART or GLORIA? Assuming the latter: reach here higher up by 1-2 km --> reach altitudes higher than those observed by 1-2 km**
  Yes, we meant with respect to GLORIA. We modified the sentence accordingly
  p17L12: "schematically" is not quite the right word here – maybe "broadly" or "generally"?
  Agreed, we have changed the wording to "broadly"
- **p18L3: fine-structures --> fine structures**
  Done
- **p18L19: complimentary --> complementary**
  Done
- **p18L31: Thereby --> However**
  Done
- **p18L32: which is by a factor of ~5 lower than that the ICON-ART R2B7 nest --> which is lower than that of the ICON-ART R2B7 nest by about a factor of 5**
  Done
- **p19L2: delete "respective"**

Done

- **p19L3-4: It would be appropriate to acknowledge some of the previous studies that have also found substantial troposphere-to-stratosphere exchange associated with tropopause fold events; folded airmass structures reach --> airmasses in tropopause fold structures reach**
  We added references regarding tropopause folding and air mass exchange. We adapted the wording suggested by the Referee with a slight modification (i.e., "variations in the dynamical tropopause"), since a less developed tropopause fold is found here when compared with the study by Shapiro (1980). We rephrased as follows:
  "The combination of ozone and water vapour data clearly shows that air masses characterised by tropospheric moisture levels reach deeply into the LMS and are connected to variations in the dynamical tropopause. Tropopause folds and steps in the tropopause are regions where isentropic levels cross the tropopause and jet streams. They are known bidirectional exchange regions between the tropopause and stratosphere (e.g. Shapiro, 1980; Keyser and Shapiro, 1986) and to contribute to transport and mixing of tropospheric air into the LMS such as diagnosed e.g. by Werner et al. (2010), Krause et al. (2018), and Jing et al. (2018) (note however that a net exchange from the LMS to the troposphere dominates)."
  Following the suggestion by Referee 1, we added a detailed analysis of the evolution of the filaments and tropopause folds at the days before the flight (see Reply to Referee 1).
- **p19L6: shows highly --> shows a highly**
  Done
- **p19L7: I think that "broadly captures" or something like that would be better wording than "resolves in principle"**
  Done
- **p19L10: In case of --> In the case of**
  Done
- **p19L14: by both --> from both**
  Done
- **p20L5-8: I think that the flow would be improved by moving the introduction of Fig. 10 in these lines to after the end of the discussion of Fig. 9 on the following page. Also: on it) --> on them)**
  Agreed and done
- **p20L11: bias, which is known for the --> bias that is known to affect the**
  Done
- **Fig. 9 caption: EMAC and ICON-ART output --> ICON-ART and EMAC output**
  Done
- **p21L5: found and increases --> found that increases**
  Done
- **p21L8: fine-structures --> fine structures**
  Done
- **Fig. 10: It might be helpful to add an overlay outlining the zero contour, especially in Fig. 10b, since it is hard to tell where the EMAC ozone residuals change sign.**
  Done

- **p21L12: that ozone --> that the ozone**
  Done
- **p21L14: B, ozone is significantly --> B is ozone significantly**
  Done
- **p21L17: scheme by --> scheme used by**
  Done
- **p21L18-19: above the troposphere and strongly --> above the tropopause that strongly; amounts --> amounts to**
  Done
- **p21L21: while comparing --> in a comparison of**
  Done
- **p21L23: due --> due to; (de-) nitrification --> denitrification/nitrification**
  Done
- **p21L23-24: It is confusing to focus only on the evaporation of PSC particles here, as that leads to HNO_3 enhancement (renitrification). If I understand correctly, the modelled HNO_3 depletion associated with the subsided air mass encountered in the middle of the flight is being attributed to sequestration in existing PSC particles or permanent denitrification through their subsequent sedimentation. That should be clarified.**
  We agree that this sentence is confusing. As discussed by Ziereis et al. (2021), denitrified air masses are seen in the middle of the flight (as discussed by the authors, PSC particles were not detected any more in situ at flight altitude during this phase of the winter). At higher altitudes, sequestration in existing PSC particles might still have played a role here, if temperatures were cold enough. We corrected the sentence accordingly.
- **Fig. 11 caption: T106 vs T42 resolution --> T106 minus T42 resolution**
  Done
- **p22L1-5: I am not convinced of the value of including the T106 vs T42 sensitivity test shown in Fig. 11a-c, as the benefit of using the higher resolution in EMAC has already been demonstrated in the Khosrawi et al. papers mentioned here. Why was it necessary to repeat this comparison?**
  We agree that the conclusion is the same as provided by Khosrawi et al. However, we think that our results are still useful, since the study by Khosrawi et al. focused mainly on the stratosphere, while our study has a more detailed focus at the upper UT/LMS region. Furthermore, our study provides another dataset to support this interpretation. We changed the wording to: "A similar behaviour of EMAC was found in the stratosphere by Khosrawi et al. (2017), who stated …"
- **p22L2: "enhances" can have a positive connotation, hence: enhances --> exacerbates**
  Done
- **p22L4: stating --> who stated; MLS --> Microwave Limb Sounder**
  Done
- **p22L6: The findings about scavenging processes only being important for HNO_3 are presented here and later in Section 5 in a manner that suggests that they were unexpected. Did the**

authors have any expectation that scavenging processes would affect the O_3 or H_2O distributions? More background and context motivating this sensitivity test is needed.

We agree that these aspects should be addressed and added the following statement:

"Scavenging processes by cirrus cloud ice particles are capable of removing trace gases from the gas phase. Sedimentation of the ice particles is capable of removing the trapped gases from affected altitudes. While previous studies focused mainly on scavenging on liquid cloud droplets (Tost et al., 2010; Wang et al., 2010; Pierce et al., 2015; Kaiser et al., 2019), Tost et al. (2010), however, found HNO_3 values in the Northern hemisphere upper troposphere to be low due to uptake on ice particles and subsequent sedimentation. Thereby, relative changes were found to be large due to low absolute values there. In addition, the vertical redistribution of HNO_3 could induce secondary effects on other trace gases via chemical processes. In particular, altering HNO_3 could lead to changes in the budget of reactive nitrogen oxides (NO_x), which, in turn, could impact ozone (e.g. Kelly et al., 1991; Krämer et al., 2008; Schiller et al., 2008). Here, our goal is to test whether the effect of scavenging over ice on the trace gas composition is significant in the LMS in the EMAC simulation."

- **p22L7: ppbv in --> ppbv than in**
  Done
- **p22L8: Reminding --> Recalling**
  Done
- **p22L10: delete ", however,"; most parts of a region --> most of the region**
  Done
- **p22L11: delete "respective"; delete comma after "means"**
  Done
- **Fig. B1 and caption:**
  - **It seems odd to me to create an Appendix just to duplicate one figure from the main text with an additional row. It would make more sense and be easier for readers to simply add the panels showing the residuals to Fig. 7 and then refer back to that figure in this section. Some discussion of the residuals could be added where Fig. 7 is first presented as well.**
    Agreed. We added the residuals to Fig. 7 and discuss them already in this section.
  - **respective residuals between GLORIA and EMAC --> corresponding residuals (GLORIA minus EMAC)**
    Done
- **p22L12: delete comma after "region"**
  Done
- **p22L13: These findings about the impact of scavenging by high-altitude cirrus on HNO_3 in the UT/LMS should be placed in the context of other studies that have examined this issue.**
  See comment to p22/L6. We furthermore added a statement that our results are consistent with the results by Tost et al. (2010), who found a similar effect in the upper troposphere.

Section 5: Discussion and Conclusions:

- **p23L2: What does "ACM" mean? Also: during --> taken during**
  We have now spelled out "ACM" to "atmospheric chemistry model"
- **p23L13: delete "used"**
  Done
- **p23L15: by generated cloud masks from --> by cloud masks generated from**
  Done
- **p23L17: between the models are reproduced to --> between the two models are attributed to**
  Done
- **p23L18: It is not clear what "limitation of the comparison" means here.**
  Our goal was to express that the comparison of the measured quantity cloud index with cloud masks generated from the models is limited. We modified the sentence accordingly.
- **p23L19: respective --> corresponding; used for --> used as**
  Done
- **p24L6: life time --> lifetime**
  Done
- **p24L7: with comparing --> by comparing**
  Done
- **p24L9: 2019) and suggests --> 2019), which suggests**
  Done
- **p24L13: a change in --> a reduction in**
  Done
- **p24L16: show practically --> has practically**
  Done
- **p24L20: Again, "schematically" is not quite the right word here. Maybe "in a broad sense"?**
  Agreed, we have changed the wording to "broadly"
- **p24L21: simulations --> simulation**
  Done
- **p24L23-24: "continuous" is not an appropriate word here – aircraft measurements are not continuous.**
  Agreed, we have meant "continuing", done
  Also: to continuously test --> to continue to test; delete comma after "required"
  Done
- **p24L22-25: The authors "speculate" that the biases and sensitivities found in this study might help provide better forecasts and long-term projections. But it is not clear to me that they have provided "actionable" information that will really inform model development / refinement in a concrete way. It might help to add another sentence or two about how they think these results could be used to guide model improvement efforts.**
  Agreed, see comment to p2L10.

Appendix A:

- **p25L8: EMAC-model (panels g-i) between −10 --> EMAC (panels g-i) model at various times between −10**
  Done
- **p25L9: add comma after “geolocations”**
  Done
- **p26L2: and it is --> and is**
  Done
- **p26L3: the measured cloud system by --> the cloud system measured by**
  Done
- **p26L5: is dissolving --> dissolves; “supposably” is not an English word, and I cannot even guess what the authors may have meant so I am unable to offer an alternative (“supposedly” is a word but does not make sense in this context)**
  Agreed, we have changed the wording to “presumably”
- **p26L6-7: The cloud system not only appears in the model a few hours earlier than observed but it also covers a much shallower altitude domain. Is that because of the problem with “false” GLORIA cloud detections below optically dense cloud layers discussed in Section 4.1? On the other hand, EMAC also shows the cloud to have a much larger vertical extent than ICON-ART.**
  In principle, the explanation with regard to optically thick cloud tops might partly explain the discrepancy here, too. However, since some structures are seen in the GLORIA cloud index at lower altitudes here, this cloud, at least in parts, are not completely optically thick. Furthermore, it should be remembered that the comparison of the model cross sections several hours before the measurements is limited, since the atmospheric scenario changes. We added these aspects in the discussion.
- **p26L6-7: data, however --> data; however,**
  Done
- **p26L10: It is stated that the cloud “breaks apart into two pieces” at T=−6 h, but to me it seems that even at T=−10 h (Fig. A1g) there were already two connected but distinct features.**
  Agreed. We modified the discussion accordingly.
- **p26L10-11: is also dissolving --> dissolves; is also subsiding and decreasing --> subsides and decreases**
  Done
- **p26L14: Figure --> Figures**
  Done
- **p27L1: add a comma after “flight”**
  Done
- **p27L7: delete comma after “cases”**
  Done
- **p27L10: in accordance to --> in accordance with**
  Done

Recurring minor wording issues:

- **p10L19, p17L2, p21L12: it is not clear what is meant by “late” polar vortex in these lines. If I understand correctly, then “late-stage”, “late-winter”, or “aged” would be better than “late”.**
  Agreed, changed the wording to “late-stage”
- **p10L24, p11L5, p17L3: backward leg --> return leg**
  Done
- **p12L18, p14L1, p14L6, p14L8, p14L16, p15L16, p22L12, p25L8, p26L8, p27L5, p27L7, p27L9: EMAC-simulation --> EMAC simulation; EMAC-cross section --> EMAC cross section; EMACmodel --> EMAC model; EMAC-data --> EMAC data (i.e., delete hyphens)**
  Done
- **p12L18, p14L14, p15L5, p15L6, p15L13, p15L17, p15L23, p25L8, p26L2, p26L14, p27L3: ICON-ART- --> ICON-ART (i.e., delete hyphens after “ART”)**
  Done
- **p14L12, p14L14, p14L15, p14L16: GLORIA- --> GLORIA (i.e., delete hyphens)**
  Done
- **p14L27, p15L15, p15L22: at the day --> on the day**
  Done
- **p16L15, p17L1, p17L7: behind --> after**
  Done
- **p17L6, p18L3, p19L11: less details --> fewer details**
  Done
- **p18L4, p18L12, p18L13, p21L21: delete the comma after “al.”**
  Done
- **p18L10, p18L21, p19L8, p21L20: hardly --> barely, or, not well**
  Done

References

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.

Kärcher, B.: Supersaturation, dehydration, and denitrification in Arctic cirrus, Atmos. Chem. Phys., 5, 1757–1772, https://doi.org/10.5194/acp-5-1757-2005, 2005.

Kärcher, B. and Solomon, S.: On the composition and optical extinction of particles in the tropopause region, J. Geophys. Res.,104, 27 441-27 459, 1999.

Kaiser, J. C., Hendricks, J., Righi, M., Jöckel, P., Tost, H., Kandler, K., Weinzierl, B., Sauer, D., Heimerl, K., Schwarz, J. P., Perring, A. E., and Popp, T.: Global aerosol modeling with MADE3 (v3.0) in EMAC (based on v2.53): model description and evaluation, Geosci. Model Dev., 12, 541–579, https://doi.org/10.5194/gmd-12-541-2019, 2019.

Keyser, D. and Shapiro, M.: A review of the structure and dynamics of upper-level frontal zones, Mon. Weather Rev., 114, 452–499, https://doi.org/10.1175/1520-0493(1986)114<0452:AROTSA>2.0.CO;2, 1986.

Lelieveld, J., Bregman, A., Scheeren, H. A., Ström, J., Carslaw, K. S., Fischer, H., Siegmund, P. C., and Arnold, F.: Chlorine activation and ozone destruction in the northern lowermost stratosphere, J. Geophys. Res., 104, 8201–8213, 1999.

Pierce, J. R., Croft, B., Kodros, J. K., D'Andrea, S. D., and Martin, R. V.: The importance of interstitial particle scavenging by cloud droplets in shaping the remote aerosol size distribution and global aerosol-climate effects, Atmos. Chem. Phys., 15, 6147–6158, https://doi.org/10.5194/acp-15-6147-2015, 2015.

Schoeberl, M. R., Jensen, E. J., Pfister, L., Ueyama, R., Avery, M., and Dessler, A. E.: Convective hydration of the upper troposphere and lower stratosphere, J. Geophys. Res.-Atmos., 123, 4583–4593, 2018.

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.

Tost, H., Lawrence, M. G., Brühl, C., Jöckel, P., The GABRIEL Team, and The SCOUT-O3-DARWIN/ACTIVE Team: Uncertainties in atmospheric chemistry modelling due to convection parameterisations and subsequent scavenging, Atmos. Chem. Phys., 10, 1931–1951, https://doi.org/10.5194/acp-10-1931-2010, 2010.

Wang, X., Zhang, L., and Moran, M. D.: Uncertainty assessment of current size-resolved parameterizations for below-cloud particle scavenging by rain, Atmos. Chem. Phys., 10, 5685–5705, https://doi.org/10.5194/acp-10-5685-2010, 2010.

Ziereis, H., Hoor, P., Grooß, J.-U., Zahn, A., Stratmann, G., Stock, P., Lichtenstern, M., Krause, J., Afchine, A., Rolf, C., Woiwode, W., Braun, M., Ungermann, J., Marsing, A., Voigt, C., Engel, A., Sinnhuber, B.-M., and Oelhaf, H.: Redistribution of total reactive nitrogen in the lowermost Arctic stratosphere during the cold winter 2015/2016, Atmos. Chem. Phys. Discuss. [preprint], https://doi.org/10.5194/acp-2021-707, in review, 2021.

Zou, L., Hoffmann, L., Griessbach, S., Spang, R., and Wang, L.: Empirical evidence for deep convection being a major source of stratospheric ice clouds over North America, Atmos. Chem. Phys., 21, 10457–10475, https://doi.org/10.5194/acp-21-10457-2021, 2021.