We thank the reviewers for their helpful comments and recommendation, which could help us improve the paper. The missing discussion of potential vorticity and some unclear sections spring especially to mind.

In the following, we address the issues raised by the reviewers in detail except for simple typographical or technical corrections, which we simply applied.

We repeat the comments of the reviewers for convenience as indented blocks. Modified sentences of the revised paper are marked by cursive face. While we also made several small changes to figures as detailed below, we provide the new flight path/tangent point/potential vorticity plot as it is essential for some of the answers and excerpts from the revised paper.

# 1 Reply to Referee 1

## 1.1 Scientific Comments/Questions

> 1. Although some attention is given to the question of LOS pointing accuracy, it is not explained why the standard approach used for satellite infrared limb sounders, i.e, a joint pressure-temperature retrieval, is not used to circumvent the problem by retrieving tangent pressure.

We believe the ECMWF pressure to be more accurate than we could retrieve it and that the approach of deriving elevation angles instead is, in the end, equivalent.

Further, to derive the pressure/pointing from measured data, sufficient reliable features must be available in the measured spectra. There are some problems with this approach for the available CRISTA-NF data. The frequency range available to us without using multiple detectors is rather limited. Therefore, many lines typically used for pressure retrievals are therefore either unavailable to us or optically dense due to the low altitude at which CRISTA-NF operates compared to satellite instruments. As we cannot 'zoom in' on obvious spectral features, we are forced to do multi-target retrievals, which work well for rather linear problems; elevation-angle/pressure retrievals are more non-linear and experiments with synthetic data were not very encouraging to follow this approach with the given spectral range and resolution.

To clarify this matter in the paper, we added to the "A priori and model data" section: *We do not retrieve pressure as we believe the pressure data supplied by ECMWF to be of high quality. For error analysis, we assume a standard deviation of 0.1% of given pressure values, which indeed leads to a negligible error contribution.*

> 2. Horizontal resolution in the line-of-sight direction. I am surprised that Fig.9 shows figures as large as 300-400 km at low altitudes. These numbers are comparable with the resolution expected from MIPAS which has a FOV of 3-4km when projected on to the limb. One can make a crude estimate of the best horizontal resolution attainable by considering the length of

> the path within each vertical layer so how, for example, do these figures compare to that?

The given resolution figure works in a different way than in layers. Using only layers, the horizontal resolution goes to zero as the layer size is decreased towards zero. We do not believe that one can fully neglect the influence of the layer above or below, as this approach does; at least not, when one has a vertical sampling of 250 m as we do (and we strive towards 125 m).

In so far the given number is not comparable.

Using the 2-D/1-D averaging kernel matrix, it is possible to more accurately determine from where the information is actually coming from. This also includes the effect of optical thickness, which is especially important for those lower tangent altitudes. Visualising this 2-D distribution, one sees that the 1-D retrieval result of a certain altitude is the average over an airmass that is more banana shaped and follows the LOS. Using our method the part of that "banana" in the layer above is only disregarded, if its influence becomes small enough. This typically delivers larger values, but also characterises the actual horizontal smearing better. This gives a better estimate over which horizontal range, horizontal gradients will affect the retrieval result (without such gradients the point of horizontal resolution obviously becomes moot anyway).

Using this technique is very similar to how typically the vertical resolution is calculated: Collapsing the horizontal dimension of the 2-D/1-D averaging kernel matrix by summing all horizontal entries up delivers the usual 1-D averaging kernel matrix. Correspondingly collapsing the vertical dimension and summing up all vertical entries delivers a figure quite similar to what we present here. The FWHM of a sphere is a more robust measure for less well-behaving averaging-kernel matrices though (as are typical for tomographic retrievals) and therefore employed.

We added a bit of explanation to the introduction of this measure in Section "Diagnostics". We first added a reference to the von Clarmann et al. (2009) paper introducing a similar measure for MIPAS retrievals: *This approach is equivalent to performing a 2-D retrieval that enforces horizontal homogeneity as employed by von Clarmann et al. (2009).* And added further: *The resulting resolution calculated in this way is often larger than the typically used figure of FWHM within the tangent altitude only, as it also includes horizontal smoothing stemming from higher layers. This is similar to the typical vertical resolution, which always includes the horizontal component.*

3. > The importance of the demonstration of retrieving at fine vertical resolution seems overstated. Even from space, this is purely a matter of geometry and S/N, and instruments within the atmosphere itself have an obvious further advantage both in terms of integration time (S/N) and distance to the tangent point. High vertical resolution is only really useful if there is a

> commensurate increase in horizontal resolution in both directions, which cannot be said for this particular experiment (15km OK, but not 100s km in the other direction).

We agree with the reviewer that it is indeed only a function of FOV and SNR. In discussions with stake holders however, we often found different opinions with respect to potential capabilities of limb sounders. For that reason we emphasise this point with a practical demonstration of feasibility.

We do not fully agree that this high resolution is pointless unless the problem of horizontal resolution along the LOS is tackled. In case that the measurements are well aligned with atmospheric structures (as is nowadays often possible for airborne instruments using chemical forecasts), the good resolution in vertical and along-flight-track direction is quite sufficient to reproduce astonishingly clear 2-D cross-sections.

However, we also think that one should not stop here but provide a good horizontal resolution in all directions. To that aim we built the airborne GLORIA infrared limb-sounder, which is able to perform tomographic measurement patterns and should reduce the horizontal resolution below 50 km (see Ungermann et al., 2011) and we also participate in the proposal for the satellite experiment PREMIER, which also should be able to produce trace gas mixing ratios with similar horizontal resolution but global coverage (see Ungermann et al., 2010a). To that end, the given vertical resolution figures serve as a proof of concept with respect to the abilities of these instruments in the vertical direction. A detailed discussion of tomography seems however not fitting to the content of the paper.

4. It seems that most of the temperature information comes from ECMWF, which is necessarily on a relatively crude spatial grid. How is this reconciled with the claimed ability to resolve fine structures? Is it just assumed that the temperature has no fine structure, or is the retrieved temperature superimposed (but with an a priori temperature of 1K it is difficult to see how the ECMWF temperatures would be modified).

We do retrieve temperature to enhance background ECMWF temperature. Given a strong enough temperature difference to ECMWF, we would expect to retrieve these deviations from the background temperature (possibly dampened). This works well for trace gas volume mixing ratios, where we see quite large deviations from the a priori profiles by up to three sigma (see CFC-11).

Please note further that we overall reduce the regularisation strength of zero-th order by a factor of 10, which effectively corresponds to an optimal estimation regularisation of zero-th order with a variance of 10 K.

Looking at temperature in situ measurements (not contained in the paper), we also cannot see significant deviations from ECMWF background temperature beyond 1 – 2 K. Our retrieved temperature follow closely ECMWF, but deviations of about

1 K do occur regularly and are mostly consistent with in situ measurements, where available.

Further, sensitivity studies showed that temperature differences by 2–3 K changes the retrieved trace gas volume mixing ratios by an astonishingly little amount. So for the given purpose of examining dynamic structures in trace gases, the given setup fulfils its purpose.

5. (possible) horizontal inhomogeneities along the line-of-sight are cited as a potential cause of many discrepancies with the other instruments. This should be supported by a plot of potential vorticity or some other tracers (eg from MIPAS or MLS) measurements at this time which show the likelihood of such gradients.

We added contour lines of potential vorticity to the plot of flight path and tangent point locations (actually we completely revamped it). In addition we added the following description to the CFC-11 comparison: *While the in situ instrument presumably samples air from the remnant of the polar vortex, the CRISTA-NF instrument already also sees air from outside the vortex, which contains larger CFC-11 mixing ratios. This assumption is supported by the distribution of potential vorticity shown as contours in Fig. 4. At this time, the plane is located in airmasses with a potential vorticity above 26* PVU *and the instrument looks towards airmasses with lower potential vorticity. Further, the slight shift in time between detection of filaments of high CFC-11 mixing rations at 11:45* UTC *and 11:58* UTC *is also explainable by the viewing geometry. The distribution of potential vorticity suggests that the filaments are slightly slanted with respect to the viewing geometry with the aircraft entering the filament first before it comes into full view of the instrument.* In the section comparing to FOZAN measurements, we added *This might indicate that these structures stem from filaments that are not completely orthogonal to the flight path as is supported by the distribution of potential vorticity at 17* km *in Fig. 4.*

## 1.2 Minor/Technical Comments

1. Abstract: it would be helpful to state the viewing geometry (ie sideways to the flight direction)

We added *The instrument points sideways with respect to the flight direction. Therefore, the observations are also characterised by rather high horizontal sampling along the flight track that ....*

2. p6918, top: an equally important advantage of limb-sounding for trace-gas detection is that the measurements are made against the cold, uniform background of space.

We added *Due to the observation geometry of limb sounding, emissions by gases in the thermal infrared are summed up over several hundreds of kilometres of air, which* *makes this technique ideal to detect (trace) gases with small mixing ratios or weak emission lines, especially as the cold background of space allows for a high signal to noise ratio.*

3. p6918, l12: is 'passive' necessary here? (als p6942, l7). I know of no 'active' infrared limb sounders.

It is not strictly necessary, but it should help people who are not familiar with the concept of limb sounding to not confuse the technique with, e.g., LIDAR.

4. p6918, l23: unprecedented horizontal coverage? I accept that CRISTA may have provided unprecedented resolution but, from the Space Shuttle, I would expect its latitude coverage to be quite limited compared to that obtained from contemporary polar orbiters such as Nimbus-7 and UARS.

This is indeed misleading and unintended. The coverage was comparable to contemporary instruments, e.g. CRISTA-2 measured from -70 to +70 degrees enabling measuring parts of the southern polar vortex. We clarify as *CRISTA provided global limb observations of a variety of trace gases with unprecedented horizontal resolution combined with an excellent coverage during its Space Shuttle missions.*

5. p6919, l10: I assume RECONCILE is a contrived acronym so I suggest capitalising the relevant letters in the phrase in brackets to make this point (rather than it just appearing to be an explanatory phrase inserted in brackets).

In fact, RECONCILE is not an acronym. It is just the short title of the EU project, as the full title is a bit lengthy. See https://www.fp7-reconcile.eu for more information to that regard.

6. p6921, l7: what sort of S/N figures are typically obtained? In section 3 (p6925, l2) it is implied that a figure of 100 is assumed for the measurement covariance.

There are a number of factors that influence the measurements. With respect to the characteristics of the detector, the paper of Schroeder et al. (2009) gives a good overview. The relative stochastic noise of the detector is thereby in the order of 0.1%. In addition there is a constant background noise, which slightly reduces the SNR for measurements with higher tangent points depending on the channel and atmospheric situation. Added to that are further uncertainties in measuring the position of the grating, interpolation errors between spectral samples and uncertainties in elevation angle.

The factor of 1% was previously empirically determined and generates plausible chisquare values for our retrievals.

7. p6926, l7-10: This information would be better in the figure caption rather than here in the text.

We restructured this as proposed.

8. p6926, l13: Having 'temperature' as only a secondary retrieval seems surprising, since most infrared instruments would regard this as a primary retrieval, and necessary for the accurate retrieval of any other species.

Within the altitude range covered by CRISTA-NF (mostly below 20 km), the temperature product of ECMWF is quite mature and so accurate that it is actually difficult to improve upon given our means. Comparing ECMWF temperature data with TDC measurements taken onboard Geophysika during the RECONCILE campaign gave the sigma of 1K, which we took as a priori in our setups.

While it would be certainly nice to be independent of this information, it is by no means necessary for our purpose of visualising dynamic structures by means of trace gas volume mixing ratios. Further sensitivity analyses show that temperature uncertainty in the expected range of a few Kelvin are not a leading error for retrieved trace gas volume mixing ratios, i.e. the effect would be contained by the supplied error bars.

9. p6929: although the paper includes some detail on the construction of the a priori covariance matrix, this presumably has little bearing on the retrieved values if the S/N is reasonable. Are there really any advantages over a simple climatological covariance with some auto-correlation length?

The two approaches, the described one and the one proposed by the reviewer, are with properly chosen parameters equivalent in the limit. However, the presented approach poses very simple, understandable constraints on the norm of the target function and its derivative. Mathematically, it poses constraints on the continuous function $f$ representing the altitude profile sampled at the retrieval grid. First, the deviation from the a priori profile is constrained as $\alpha_0||f - f_a||$ and, second, the bias from the first order derivative is constrained $\alpha_1||f' - f'_a||$. The remaining parameters are mostly there to close the connection to the approach proposed by the reviewer and enable this scheme to deliver very similar (not identical!) results to a given climatological covariance matrix with some auto-correlation length.

The constraints posed by a covariance matrix with some auto-correlation length are per se not as easily understood, even though they are equivalent in the (typically not used) limit of infinite and zero correlation length (see Steck and von Clarmann, 2001).

The presented approach is also very easily tunable, as it is very simple to remove, e.g., the bias of absolute value by setting $\alpha_0$ to zero, which in the proposed approach would require an increase of the employed standard deviations (thereby basically discarding the "optimal estimation") and a corresponding change in correlation length to not change the imposed smoothing. Fully removing the bias of absolute value in the approach proposed by the reviewer would be rather difficult.

That said, for the given problem either approach would work and the proposed approach was indeed used by Weigel et al. (2010) and served as our starting point. The presented approach however is not only more flexible but also scalable up to 2-D or 3-D tomographic retrievals whereas climatological covariance matrices are more difficult to use as neither they nor their inverse is necessarily sparse (see Ungermann, 2011).

10. p6931, l13: some further assumption is required to construct the HNO4 covariance matrix from the Remedios climatologies which only supply the diagonal elements of this matrix.

A conservative approach might assume no correlation at all, but we typically use an auto-regressive approach with a correlation length in the order of several kilometres. Given the amount of unreliable statistics involved, this usually gives a good indication of whether this trace gas contributes significantly to the error budget of a primary target or not. If it does, one needs to do something about it.

We added in the paper: *. . . ; to assemble this matrix, it is necessary to make some assumption about vertical correlation length, which is not contained in the employed climatology.*

11. p6932, l14: doesn't this imply that your error covariance matrix is unrealistically large?

The major effect of this choice is the neglecting of potential correlations in errors. The size of the assumed error has been chosen to be certainly above the actual error, as is also asserted by acquiring typically a chi-square below 1. We wouldn't go as far as stating that it is unrealistically large. We currently assume the combined effect of stochastic errors of the instrument to be in the order of 0.6 percent. To this, the effect of pointing inaccuracy, which is also random, has to be added. All-in-all, we hope 1 percent to be a conservative assumption that is certainly in the right ballpark.

12. p6935, l5-18: in my opinion much of this explanation would be better as part of the figure caption.

We concur and moved much of this to the caption and removed redundant description.

13. p6938/39 comparisons with HAGAR. It would be useful to have the a priori CFC-11 profile plotted on Fig.10 as well. For a CFC-11 measurement to be of 'useful' accuracy it should reproduce the same deviations (at least in sign) as the HAGAR measurements from the a priori values. From the information presented here it is not possible to say whether the retrieved CFC-11 profile is actually any better than the climatology.

We added the a priori information also to the flight and ascent plots for CFC-11 and ozone. As can be seen, there is no apparent bias towards the a priori for these two gases.

14. p6939 - same as 20, applied to the ozone comparison with Fozan in Fig.12

See comment above.

15. p6941, l5: remove',' after 'We mention,'. Incidentally I think it unlikely that the differences between the two HNO3 spectroscopic databases would be large enough to contribute any noticeable difference in results.

The two databases differ by more than just the HNO3 lines. Further, our technique of using integrated windows is also more dependent of the quality of spectroscopic data of background gases. Using, e.g. the more recent HITRAN04 ClONO2 cross sections instead of the previously used HITRAN2K ones provided notable differences in ClONO2 mixing ratios and also, to a lesser extent, to O3 and HNO3. So, we would not rule out this as one of many possibilities.

16. p6941 - HNO3 comparisons with MIPAS-STR: rather than comparing single profiles I would have been more impressed if a 2D plot of the MIPAS-STR results resembled Fig 5(e).

Such a comparison is given in Fig.17 and Fig.18 of the referenced paper by Woiwode et al. (2011). We added a textual reference mentioning this fact. Both instruments see similar structures, with differences being largely explainable by the different viewing geometries: *A detailed description of the MIPAS-STR retrieval for this campaign is given by Woiwode et al. (2011); this paper also presents 2-D cross-sections of* $HNO_3$ *mixing ratios from both instruments, which underline the generally extremely good level of agreement.*

# 2 Reply to Referee 2

## 2.1 Scientific Comments/Questions

1. Page 6925, Line 16 ff.: You describe the forward model approximation using a look up table. For me it is not quite clear, what the EGA and the CGA are used for. To my knowledge, CGA is used to determine weighted means, e.g. layer mean values for discrete profiles. CGA is also applied within line-by-line models for the RT calculation. Additionally, I assume that the Look-up tables are based on line-by-line models. Unfortunately this part is not clear from Weigel et al, 2010 either. So here one or two sentences could help to make this part clearer.

The straightforward way to exploit the lookup tables would be to split the line-of-sight into individual segments and use the look-up-tables to calculate the emissivity of each individual segment. In a second step, one could follow the path from the instrument along the line-of-sight and calculate the radiance by simple multiplication and addition. However, the tabulation introduces small errors into the derived emissivities, which become enlarged by repeated multiplication necessary to determine the total transmissivity between the instrument and the local segment. The errors for this naive approach easily approaches 100 percent (we unsuccessfully tried to incorporate the radiance derived from this simple approach to improve the regression). In contrast, both the EGA and CGA method allow to determine the total transmissivity of an inhomogeneous gas cell. This is exploited in our model by directly calculating the total transmissivity between the instrument and the local gas cell and then deriving the local emissivity from the difference between the current total transmissivity and the preceding total transmissivity. This leads to a stable algorithm with astonishingly small errors given the stark inhomogeneity of the gas cells onto which EGA and CGA are applied.

We hope that this becomes clearer from the following additions: *Both methods allow to easily compute the total transmissivity between the instrument and any point on the discretised LOS, which avoids the summation of errors if only the emissivity of short segments were calculated and multiplied. The radiances derived from the two methods are in turn combined using a simple regression scheme to minimise the deviation to a more exact line-by-line model (Weigel et al., 2010). The combined method is typically subject to smaller systematic errors than either the CGA or the EGA method alone (e.g. Francis et al., 2006). Compared to conventional line-by-line calculations, the EGA and CGA methods are faster by a factor of about 1 000, since the radiative transfer is based on pre-calculated, spectrally averaged values of emissivity stored in look-up tables. The emissivity look-up tables for the forward model are prepared by means of exact line-by-line calculations utilising the RFM.*

The look-up-tables are calculated using the RFM, which is stated in the given section "The emissivity look-up tables for the forward model are prepared by means of exact line-by-line calculations utilising the RFM. "

2. Page 6926, Line 3 ff: Why can the FOV of CRISTA-NF be approximated by a Gaussian? Is this an assumption? Or from observation? If the latter is the case, provide a reference.

The FOV was already determined for the original CRISTA instrument. We provided an appropriate reference by Riese et al. (1999a): *Each measurement is affected by the FOV of the instrument, which can be approximated by a Gaussian with an FWHM (full width at half maximum) of about 3 arcmin for CRISTA-NF (identical to CRISTA; Riese et al., 1999a).*

3. Page 6927, Line 13 ff: You write that spectra from upward looking scans are used. If they are used for the retrieval, how are these spectra attributed to any altitude? There is no tangent point information available. This should be clarified.

We mention that the retrieval grid follows the typical tangent point altitude distance of 250 m and not the actual one, which implies that it is a fixed grid. We clarified this by reformulating as "*The retrieval grid has a fixed spacing of 250 m below 20 km and thereby follows roughly the typical tangent point altitude distance.*" However, it is difficult to attribute upward looking spectra to some fixed altitude, a situation quite similar to nadir measurements. The sensitivity of these measurements almost always peaks around the instrument altitude (with some exceptions for trace gases with volume mixing ratios peaking above the aircraft such as ozone). However, the kernels of these upward looking measurements exhibit different shapes, so some information about trace gas distributions above flight level can be derived. In fact, we do not even plot trace gas mixing ratios derived from above the aircraft as these are of a rather poor quality. We found however that employing these measurements in our retrieval lead to an increased quality at flight altitude as the influence of (wrong) a priori information above flight level is diminished by having better estimates about the top column and some information about distribution therein available.

Even for downward looking spectra it is difficult to attribute them to a single altitude as most employed channels are not fully optically thin in the troposphere and therefore often peak in sensitivity before the tangent altitude. This poses serious issues especially for altitudes below 10 km and is partly responsible for decreased retrieval quality there. So we do not in fact attribute individual measurements to retrieval altitudes except for some rough measure of where we can expect sensible results. The retrieval algorithm receives as input parameters only location of the instrument, viewing elevation angle, and 1-D representation of the atmosphere on the other side.

4. Page 6930 Line 17 ff and 6937 Line 20 ff: I agree with reviewer number 1 that the description of the horizontal resolution along the line of sight is unclear and misleading. To which extend does the horizontal resolution agree with the length of the partial column (lets call it actual footprint) through the layer at the tangent point? Is it better or worse than the actual footprint? Especially for the lowest tangent altitudes this discussion is not possible, as the path along the line of sight before and after the tangent layer is not covering the tangent layer anymore but the layers above. So you cannot get any information on the field at this altitude outside the footprint as this is observing other altitudes. Here you should rethink the use of horizontal resolution and use other argumentation.

It is difficult to compare the employed measure with the footprint of a single measurement, as usually several measurements that are sensitive to a given altitude somehow contribute to a given retrieved value. In addition horizontal smoothing can cause the retrieval to basically extrapolate from different altitudes, which happens often for the lowest values of the retrieval grid and also, as the reviewer points out, for the altitude(s) of the lowest measurement(s).

However usually the 2-D/1-D averaging kernel for a 1-D retrieval resembles a "banana" with largest values between the instrument and the tangent point (e.g. see Fig. 16(b) by Ungermann et al. (2010)) . For optically thin conditions, it is rather symmetric around the tangent point. The algorithm will then almost always deliver a larger horizontal resolution than the "footprint" (or FWHM within the tangent altitude layer as we refer to it in the paper) in the tangent point layer, as the contribution of the vertical layers above stem from points horizontally further away from the tangent point. As these points do enter the final value, it is correct to take them into account for the calculation of the horizontal resolution, especially as this is already standard practise for determining the vertical resolution (at least for the measure of FWHM of the 1-D AVK matrix).

For the lowest point, the horizontal resolution as given in the paper is indeed potentially meaningless, but even more so is the vertical resolution! Here, it is valuable to look in addition at the distance between the vertical location of the maximum of the 1-D AVK matrix row and the altitude this row belongs to. While being 0 (or close to 0) for altitudes several kilometres below the instrument, this number will increase from a certain point on and linearly increase with distance as soon as the retrieval altitude drops below the lowest measurement.

As mentioned in our reply to reviewer 1, we added the following explaining section: *The resulting resolution calculated in this way is often larger than the typically used figure of FWHM within the tangent altitude layer only, as it also includes horizontal smoothing stemming from higher layers. This is similar to the typical vertical resolution, which always includes the horizontal component.*

5. Page 6931, Line 19 to Page 6932, Line 6: The description of the pointing error assessment is difficult to follow. It could be improved.

We tried to clarify this by elaborating the concept a bit more:

*The effect of random elevation angle errors of individual spectra is covered by the error budget of 1 % per measurement. In addition, a potential systematic elevation angle offset error consistent over all spectra of one profile is treated separately. The absolute error in retrieved trace gas mixing ratios induced by such an constant elevation angle offset is especially large in the presence of strong vertical gradients, but not very meaningful. An offset in elevation angle places the real structures in trace gas mixing ratios at slightly shifted altitudes. The actual mixing ratios in the structures are thereby not largely affected (at least for the offsets potentially present). In a conventional error estimate, a slight shift of a peak in mixing ratio thus results in a very large absolute error in the vicinity of this structure.*

*To capture the influence of this error in a more practical manner, we perform retrievals with an elevation angle offset increased and decreased by 0.02°. These derive structures with a slightly different altitude structure. We then map the shifted structures back to their original location by removing the geometrical effect of the elevation angle offset from the retrievals. By calculating the difference in trace gas mixing ratios between these retrievals and the original one, we separate the effect of the elevation offset angle error on absolute values of trace gas mixing ratios from its effect on the* *vertical placement of structures. The effect on absolute values of trace gas mixing ratios is then incorporated into our error estimates while the effect on vertical trace gas structure placement is captured in Tab. 4.*

6. Page 6933, Line 11/12: What do you mean by quiet time. Is this the plane standing before and after the flight? If yes this should be mentioned. The use of the expression quiet time is unlucky.

We replaced “quiet time” with the more accurate *a time span with few aircraft movements and vibrations.*

7. Page 6934, Line 27: What do you mean by a fit. This is the first and only time you mention this. Please clarify this.

We indeed did not give the statistical background of the employed retrieval scheme. We modified the paper by describing the selection criterion with terminology we did introduce: *To filter out profiles with obvious defects, e.g. due to stronger movements by the aircraft than the attitude system can compensate for, only profiles where the term describing the fit to the measurements in the cost function could be reduced to less than 0.65 are used (i.e.* $(\vec{F}(\vec{x}_\mathrm{f})-\vec{y})^\mathrm{T}\mathbf{S}_\epsilon^{-1}(\vec{F}(\vec{x}_\mathrm{f})-\vec{y}) < 0.65$*). This value was chosen in an ad-hoc manner to consistently filter out profiles with obvious defects.*

8. Page 6935, Line 14 ff: The description of the observed distributions of the various trace gases could be improved. When reading this section I had difficulties to distinguish with statements related to individual species and statements applicable to all gases:
   - Discussing CFC-11: How do you define the location of air masses inside or outside the polar vortex?
   - CFC-11 and ClONO2: What are typical distributions/profiles and why do you expect the two species to be anti correlated?
   - The color scale is consistent top the error bars: Is this valid for water vapour only or for all species described?

We updated the description of the trace gases and hope that it is now more accessible.

The very low mixing ratios of CFC-11 suggest vortex air that was brought down from higher altitudes. This is supported by increased potential vorticity values. In addition, we lightened our tone, stating that this is only a possibility.

The use of “correlated” is probably wrong. CFC-11 usually increases with altitude while ClONO2 decreases in this altitude range. In so far one would expect an airmass to have either high CFC-11 or high ClONO2 mixing ratios. In so far these trace gases can function as tracer up to a point, as especially ClONO2 is subject to a lot of chemistry in the polar vortex. Typical climatological profiles for these gases can be found in the Figures at the end of the paper, where the a priori profiles are shown in addition to CRISTA-NF and MIPAS-STR profiles.

The error for water vapour is especially large, which in combination with the stark increase of volume mixing ratio at lower altitudes makes it difficult to settle on a linear colour scale. For all other trace gases, we picked a scale for the error plots that is one tenth of the color scale for the mixing ratio. Please note also, that the error bars of the water vapour profile almost always include the a priori profile. We added the sentence with respect to water vapour's colour scale to explain why we did not choose a colour scale showing more detail than the one employed.

This is the new section, where we dedicated one paragraph for each primary target: *Panel (a) shows CFC-11 mixing ratios. CFC-11 mixing ratios typically fall with altitude and low volume mixing ratios indicate usually stratospheric air. The very low CFC-11 volume mixing ratios to the upper left above 15* km *show probably the extent of the polar vortex (compare also with potential vorticity distributions Fig. 4) whereas the low volume mixing ratios to the lower right are likely remnants of polar vortex air. The air mass with low volume mixing ratios on the upper right could be of a different origin as the ozone volume mixing ratios in this filament seem to be too high to be a recent remnant of vortex air. Please note the thin filament of increased CFC-11 mixing ratio at ≈ 16 km on the right, probably originating from mid-latitudes. This thin filament has a vertical extent of 0.5 to 1 km and its signature can be found in all primary target species except water vapour.*

*The distribution of* $CCl_4$ *in panel (b) is very similar to the distribution of CFC-11, but fewer filaments are visible due to the decreased vertical resolution compared to CFC-11.*

$ClONO_2$ *is a chlorine reservoir species found in abundance in the stratosphere. The* $ClONO_2$ *mixing ratios depicted in panel (c) show largely high mixing ratios where low CFC-11 mixing ratios are given and vice versa. This is expected from their climatological mixing ratios, which decrease in altitude for CFC-11 and increase for* $ClONO_2$ *in the presented altitude range.*

*Panel (d) shows that the water vapour content in this polar atmosphere is roughly 5* ppmv*, which is consistent with climatological profiles. The colour scale for the water vapour graph was chosen to be consistent with our error bar, implying that we cannot significantly resolve more details than depicted.*

*In panel (e), the mixing ratios of the stratospheric trace gas* $HNO_3$ *is shown. Please note the fine structures in* $HNO_3$ *mixing ratios, especially the horizontal filaments of increased mixing ratio of about 6* ppb *at 12 km and 13.5 km in the middle part of the figure as well as the structure at 12 km to the right, where air with increased* $HNO_3$ *mixing ratio apparently surrounds a filament with decreased mixing ratio.*

*The last panel (f) depicts the mixing ratios for* $O_3$*. Its distribution is similar to the other stratospheric trace gases. The comparatively low volume mixing ratios to the upper left and lower right indicate air with depleted ozone most likely stemming from the polar vortex.*

*An in-depth discussion about the atmospheric situation is given by Kalicinsky et al. (2012).*

9. Page 6936, Line 3 ff, Figure 6: Where does the horizontal structure in the error distributions for ClONO2 and H2O come from? It appears to be related to the a priori profiles?

We do not see a horizontal structure in the error for $H_2O$ retrievals, but the vertical structure depends largely on the existing trace gas volume mixing ratios as our errors are often proportional to the trace gas volume mixing ratio at hand.

The horizontal structure in error comes largely from three factors. First, the horizontal distribution of trace gases themselves, as especially the noise errors are proportional to the given mixing ratios. Second, we tried to compensate for the effect of an angle elevation offset, but some effect of different vertical location of trace gas structures remain; this explains the large errors around 16km after 12:00 UTC. Third, our regression scheme necessarily uses different correction parameters for different altitudes of the instrument. As some of our predictors depend on trace gas concentrations along the line-of-sight (see Francis et al., 2006), a change in regression parameters will increase or lower the influence of a certain gas on the calculated radiances and thereby the retrieval result. This obviously affects the gain matrix and thereby the error calculations. We tried to keep this effect small by optimising the regression with respect to which gases to include in the regression, but a remainder of this effect can be seen in $CCl_4$ and $ClONO_2$ where the error structure changes at 11:25 and 12:10 coinciding with aircraft altitude changes.

The a priori information is most likely not responsible, as only the a priori information for water vapour and temperature vary from profile to profile.

10. Page 6936, Line 18 ff: The discussion of the measurement distributions is unclear: Why do we get values above 1 for mixing ratios below the detection limit? This should be clarified. "It can be removed... What can be removed? The description of possible cut-off in the retrieval altitudes is unclear and could be improved.

We were concerned with the increased measurement contribution of 1.2 in water vapour and $ClONO_2$, even though this feature seems to occur quite frequently in the transition region between good and no sensitivity in plots of measurement contribution (e.g. even in Fig. 3.5 in the classic retrieval book of Rodgers (2000) ). Most authors do not discuss this feature, though. We tried to explore why this feature is there, tried to remove it and finally found only an unsatisfiable solution: do not retrieve the relevant quantity in an altitude range without sufficient measurement information. We found this unsatisfying as it somehow beats the purpose of adding a priori information to the retrieval and as it had some side effects on other diagnostics quantities. As the retrieved mixing ratios in the affected altitude range are very stable and consistent with other non-affected trace gases, we decided to accept this artifact.

We improved the description in the following way: *We hypothesise the increase above 1 to be an artefact of the low trace gas mixing ratio, which is below the detection limit of the IMWs employed here. The bump above 1 can be removed, e.g. by not retrieving* $ClONO_2$ *in the lowest 10 km of the profile but assuming the a priori values instead; this delivers nearly identical* $ClONO_2$ *mixing ratios above 10 km with the measurement contribution staying close to 1. However, the abrupt change between retrieved* $ClONO_2$ *mixing ratios and values fixed to the a priori in the lower part of the profile causes other undesirable artefacts in retrieved mixing ratios and diagnostics, so it was not applied to the presented results. The increased measurement contribution in water vapour could be treated similarly by not retrieving water vapour mixing ratios above 12* km*, as this removes the oversensitivity. However, we do not see the increased measurement contribution as a problem as the retrieved mixing ratios in the affected altitude range are close to the detection limit and therefore of low quality to begin with.*

11. Page 6939, Line 22-23: ... might be horizontal gradients...: Here you could use additional information (e.g. maps of pot. vorticity) to show that the vortex boundary really was affecting the LOS. The argumentation is becoming less speculative.

This is indeed a sensible suggestions and we consequently added such a map to the paper (see below) and added a description of the relation between observed discrepancies and potential vorticity structures. We obviously looked at the meteorological situation but wrongly decided to leave this out of this more retrieval oriented paper. See last Scientific Comment/Question of Reviewer 1 for details.

12. Page 6940, Line 9 ff: The offset discrepancy between the ascent-profile comparison (Fig 13) and the flight altitude comparison (Fig 12) is not sufficiently treated. The statement on the horizontal averaging should be supported by some (at least rough) estimates of the influence of the horizontal inhomogeneities.

As stated in the last Scientific Comment/Question of Reviewer 1, we modify the flight path plot to also contain contour lines for potential vorticity, which shows the likely orientation of filaments 17km. Further, we modified the paragraph in question: *The agreement between FOZAN and CRISTA-NF is much better for the ascent profile than for the comparison along the flight track. The air measured by FOZAN during the ascent of the M55-Geophysica lies well within the region from which most radiation is received by CRISTA-NF, so that a better fit is expected. Please note also, that the largest horizontal gradients in ozone are expected in the altitude region above roughly 16* km *in the vicinity of the polar vortex due to ozone depletion in vortex airmasses. Further, the agreements on flight level is better for the northward part of the flight,* *where the potential vorticity suggests more similar air masses along the LOS. This suggests that the discrepancy in ozone mixing ratios at the observer position evident in Fig. 12 is at least partially an artefact of the spatial averaging characteristics of the remote sensing measurement method.*

13. Page 6927, Line 28 ff: What do you mean by aggressive standard deviation?

We wanted to express that we choose a small standard deviation and thereby basically accept an offset bias of our retrieval towards ECMWF temperature data.

We replaced the adjective "aggressive" with *comparably small*, referring to the larger temperature standard deviations given by the Remedios climatology.

14. Page 6928, Line 13 ff: Refer $\mathbf{S_a^{-1}}$ to section 3 : E.g "... assemble the a priori covariance matrix $\mathbf{S_a^{-1}}$".

We implemented the suggestion.

15. Page 6930: You could flip Equ 4 and 5.

We follow the suggestion and offer the definition of A after the definition of G.

16. Page 6930, Line 14/15: What do you mean by the inverse of the diagonal entries of A being a measure for vertical resolution? How is this related to the FWHM approach, which is actually used?

The inverse of the diagonal is one potential (dimensionless) mesure of resolution. By multiplying it with the retrieval grid sampling, one makes it comparable to the FWHM measure. As the paper states, we employ the FWHM of the averaging kernel matrix row. The inverse of the diagonal is certainly a very useful measure, but does not allow to differentiate between resolution in various dimensions.

We try to remove confusion by not mentioning the simple inverse of diagonal/reciprocal data density measure and provide a reference to Rodgers instead: *This paper employs the FWHM of the averaging kernel matrix rows as measure of vertical resolution (e.g. Rodgers, 2000, p.61f).*

17. Page 6932, Line 11 ff: You describe the set up of $\mathbf{S}_\epsilon$ and give reasons for the use of a diagonal matrix. this is redundant or connected to Page 6925, line 1ff. You could make a reference and mention that this concerns the set up of $\mathbf{S}_\epsilon$, explicitly.

We added the symbol $\mathbf{S}_\epsilon$ and a couple of words to make the connection explicit: *Thus, as stated above, we assume an uncorrelated error covariance matrix $\mathbf{S}_\epsilon$ with an assumed error of 1 %, as was already used by Weigel et al. (2010).*

18. Page 6936, Line 10/11: You refer to two scans (taken at 11:07 UTC and 12:30 UTC) as profiles. As you work with profiles of trace species, Temperature etc, I recommend to use the expression scan rather than profile

when referring to these two scans. This is applicable to all places in the paper.

We follow the suggestion and relate to such a set of spectra as a "vertical scan".

19. Page 6936, Line 14-18: The discussion about the profiles of water vapour and ClONO2. The direction should be consistent for both profiles, going from the surface up or down to the surface, respectively.

We agree and modified the section as follows: *Going down from top to bottom, the measurement contribution for water vapour is first very small as the signal of these low water vapour mixing ratios is below the detection limit in the spectral range used for retrieval. It then increases above 1 and then drops back to 1 and stays there down to the lowest valid data points. Similarly but in opposite direction, going again from top to bottom the measurement contribution of* $ClONO_2$ *is first close to one, then increases and finally drops towards the surface due to a lack of signal.*

20. Page 6937, Line 5-13 ff: These results show... to 250 to 300m Put this section to the end of the paragraph. Currently the discussions of the resolution are interrupted. Why are the vertical resolution profiles of water vapour and ClONO2 so strange? This could be mentioned.

We follow the first suggestion. The increase in resolution of ClONO2 is handled together with the less affected CCl4. We clarify this portion and added *The resolution of water vapour similarly increases towards higher altitudes as the signal to noise ratio of the employed IMWs is insufficient for the present water vapour volume mixing ratios. Please note the strong connection between the drop in measurement contribution of* $ClONO_2$ *and water vapour and the worsening vertical resolution.*

21. Page 6941, Line 3: The MIPAS-STR retrieval grid is generally finer ... than what? Than the measurement grid? Here you could also mention the estimated vertical resolution of MIPAS-STR.

We clarified *The MIPAS-STR retrieval grid is mainly finer than its FOV...* and *Its resulting vertical resolution is largely 1 to 2* km*, which is typically about twice as much as the corresponding CRISTA-NF resolution.*

22. Figure 14: I think you could remove the a priori profiles and the error bars. You dont really discuss them in the text and so they are not necessary. Especially due to the large error bars the a priori dominates most of the figures and distract from the two profiles to be compared. Especially for the H2O comparison, you should reduce the x-axis to the values covered by both instruments. You can easily reduce the range to 15 to 20 ppmv and show the good agreement.

We prefer to leave the a priori profiles in there as demonstration of the fact that the presented trace gas volume mixing ratios (with the potential exception of water vapour) are not obviously biased towards the a priori profiles. We already tried to mitigate the visual clutter by using a dotted error bar. We follow the suggestions of the reduced value range and also reduced the number of dots in the a priori standard deviation error bars to further reduce the clutter.

# References

von Clarmann, T., De Clercq, C., Ridolfi, M., Hölpfner, M., and Lambert, J.-C.: The horizontal resolution of MIPAS, Atmos. Meas. Tech., 2, 47–54, doi:10.5194/amt-2-47-2009, 2009.

Francis, G. L., Edwards, D. P., Lambert, A., Halvorson, C. M., Lee-Taylor, J. M., and Gille, J. C.: Forward modeling and radiative transfer for the NASA EOS-Aura High Resolution Dynamics Limb Sounder (HIRDLS) instrument, J. Geophys. Res., 111, doi:10.1029/2005JD006270, 2006.

Kalicinsky, C., Grooss, J.-U., Guenther, G., Ungermann, J., Olschewski, F., Knieling, P., Hoefer, S., Blank, J., Stroh, F., Riese, M.: CRISTA-NF high resolution retrieval results in the vicinity of the polar vortex, in preparation.

R. J. Purser and H.-L. Huang. Estimating effective data density in a satellite retrieval or an objective analysis. *J. Appl. Met.*, 32:1092–1107, June 1993. doi:10.1175/1520-0450(1993)032⟨1092:EEDDIA⟩2.0.CO;2.

Riese, M., Spang, R., Preusse, P., Ern, M., Jarisch, M., Offermann, D., and Grossmann, K. U.: Cryogenic Infrared Spectrometers and Telescopes for the Atmosphere (CRISTA) data processing and atmospheric temperature and trace gas retrieval, J. Geophys. Res., 104, 16 349–16 367, doi:10.1016/S0273-1177(97)00172-5, 1999a.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific, 2000.

Schroeder, S., Kullman, A., Preusse, P., Stroh, F., Weigel, K., Ern, M., Knieling, P., Olschewski, F., Spang, R., and Riese, M.: Radiance calibration of CRISTA-NF, Adv. Space Res., 43(12), 1910–1917, doi:10.1016/j.asr.2009.03.009, 2009.

Steck, T. and von Clarmann, T.: Constrained profile retrieval applied to the observation mode of the Michelson Interferometer for Passive Atmospheric Sounding, Appl. Optics, 40, 3559–3571, doi:10.1364/AO.40.003559, 2001.

Ungermann, J.: Tomographic reconstruction of atmospheric volumes from infrared limb-imager measurements, Ph.D. thesis, Wuppertal University, 2011.

Ungermann, J., Hoffmann, L., Preusse, P., Kaufmann, M., and Riese, M.: Tomographic retrieval approach for mesoscale gravity wave observations by the PREMIER infrared limb-sounder, Atmos. Meas. Tech., 3, 339–354, doi:10.5194/amt-3-339-2010, 2010a.

Ungermann, J., Kaufmann, M., Hoffmann, L., Preusse, P., Oelhaf, H., Friedl-Vallon, F., and Riese, M.: Towards a 3-D tomographic retrieval for the air-borne limb-imager GLORIA, Atmos. Meas. Tech., 3, 1647–1665, doi:10.5194/amt-3-1647-2010, 2010.

Ungermann, J., Blank, J., Lotz, J., Guggenmoser, T., Kaufmann, M., , Preusse, P., and Riese, M.: A 3-D Tomographic Trajectory Retrieval for the Air-borne Limb-imager GLORIA, Atmos. Meas. Tech. Discuss., 4, 3805–3859, doi:10.5194/amtd-4-3805-2011, 2011.

Weigel, K., Riese, M., Hoffmann, L., Hoefer, S., Kalicinsky, C., Knieling, P., Olschewski, F., Preusse, P., Stroh, F., Spang, R., and Volk, C. M.: CRISTA-NF measurements during the AMMA-SCOUT-$O_3$ aircraft campaign, Atmos. Meas. Tech., 3, 1437–1455, doi:10.5194/amt-3-1437-2010, 2010.

Woiwode, W., Oelhaf, H., Gulde, T., Piesch, C., Maucher, G., Ebersoldt, A., Keim, C., Höpfner, M., Khaykin, S., Ravegnani, F., Ulanovsky, A., Volk, C. M., Hösen, E., Dörnbrack, A., Ungermann, J., Kalicinsky, C., and Orphal, J.: MIPAS-STR measurements in the arctic UTLS in winter/spring 2010: Instrument characterization, retrieval and validation, Atmos. Meas. Tech. Discuss., submitted, 2011.

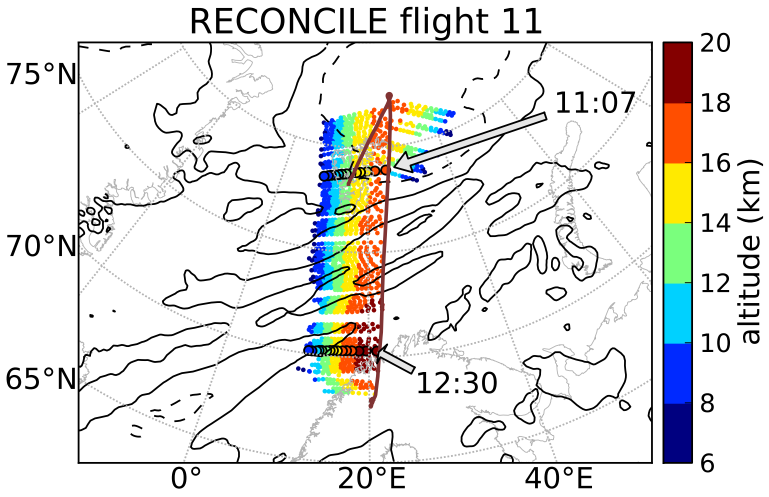

Figure 4: Plot of tangent point locations of CRISTA-NF measurements taken during RECONCILE flight 11, the second flight of 2 March 2011. The flight path is drawn in dark red. The Locations of every fourth tangent point are drawn as circles with colour code according to tangent point altitude. Thick black contour lines indicate a potential vorticity of 20 PVU at 17 km at 12:00 UTC (indicating the position and orientation of filaments across the flight path and dashed black contour lines show 26 PVU at the same altitude and time (roughly indicating the location of the polar vortex core).