Response to referee 2:
We thank the reviewer for his/her comments. Below are our responses in blue.

In the course of responding to the reviewers, two main changes occurred: (1) a trend discussion was added and (2) we emphasize the lack of MLS penetration below the upper troposphere throughout the paper.

Trend discussion

Before the conclusions the following paragraphs will be added:

As an example, Figure 8 shows the H_2O and O_3 trends in the tropics computed using monthly zonal mean deseasonalized anomalies of the raw model fields, as well as using all the available satellite-sampled measurement locations and only those passing the screening criteria in the tropics. As shown, when all available measurement locations are used, the MIPAS and MLS sampling allows accurate derivation of trends, with values matching those calculated from the raw model fields almost exactly. However, when only those measurements passing the screening criteria are used, both instruments have limitations: MIPAS trends are impacted because of the large percentage of measurements screened out below 100 hPa, which introduces non-negligible artifacts (up to 80% change for H_2O and up to 20% change for O_3); MLS trends are impacted because of the reduced vertical resolution, which limits its usefulness to the upper troposphere and above. Note that the impact of quality screening on MIPAS trends can be mitigated by using a regression model similar to the ones used by Bodecker et al. (2013) and Damadeo et al. (2014). These models have been shown to mitigate the effects of the non-uniform temporal, spatial and diurnal sampling of solar occultation satellite measurements. Furthermore, MIPAS trend analysis can be restricted to regions less affected by deep convection (for example, the mid tropical Pacific) to minimize the quality screening effects.

The estimated number of years required to definitively detect these trends is also shown in Figure 8. These estimates were computed assuming a trend model similar to the one described by Tiao et al. (1990), Weatherhead et al. (1998), and Millán et al. (2016), with a seasonal mean component represented by the monthly climatological means. As shown, with the MIPAS screened fields additional years are required for robust trend detection (up to ~150 years for H_2O and up to ~40 years for O_3 versus 50 years and 20 years, respectively, when all available measurements are used).

Similar analyses were performed for other latitude bands. Although the magnitude of the trends derived when using the MIPAS screened measurement locations was also impacted in these cases, no significant difference was found in the number of years required to detect such trends. In addition, no significant artifacts were found for HNO_3, CO or temperature for either the trend magnitude or the number of years required to detect such trends. Note that, when using real data, the effect of instrument noise upon trends will be negligible due to the vast number of MIPAS or MLS measurements associated with each monthly latitude bin. Drifts and long-term stability issues on these datasets [i.e., Eckert et al., 2014; Hubert et al., 2016; Hurst et al., 2016] will have to be corrected.

In the conclusion section, the trend discussion will be changed to:
These biases affect trends derived from these measurements using a simple regression upon monthly zonal mean data substantially affected by clouds. Further, the number of years required to detect such trends may increase due to the extra noise added to the time series by screening out measurements.

The following figure will be added (as figure 8 of the revised paper):

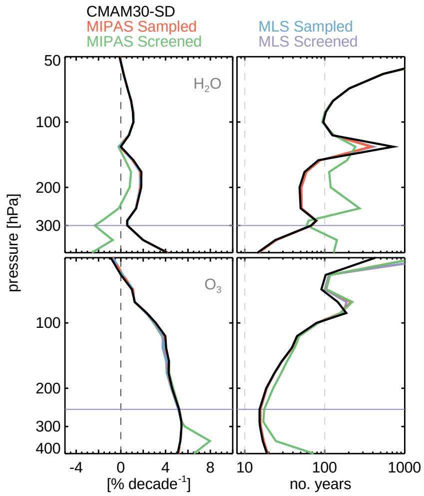


Figure 8: (left) H_2O and O_3 trends computed based on monthly zonal mean deseasonalized anomalies for the tropics (20S to 20N) using the raw model fields, all the available satellite measurements (MIPAS or MLS sampled) and only those measurements passing the screening criteria (MIPAS or MLS screened). Note that for O_3, we only use data starting from 2000 to capture the expected period of O_3 recovery. A purple line indicates the bottom (largest pressure) of the recommended range of the MLS retrievals. (right) Number of years required to detect such trends.

References:

Bodecker et al 2013: 10.5194/essd-5-31-2013

Damadeo et al 2014: 10.5194/acp-14-13455-2014

Tiao et al. (1990): 10.1029/JD095iD12p20507

Weatherhead et al. (1998): 10.1029/98JD00995

Millán et al (2016): 10.5194/acp-16-11521-2016

Lack of MLS penetration

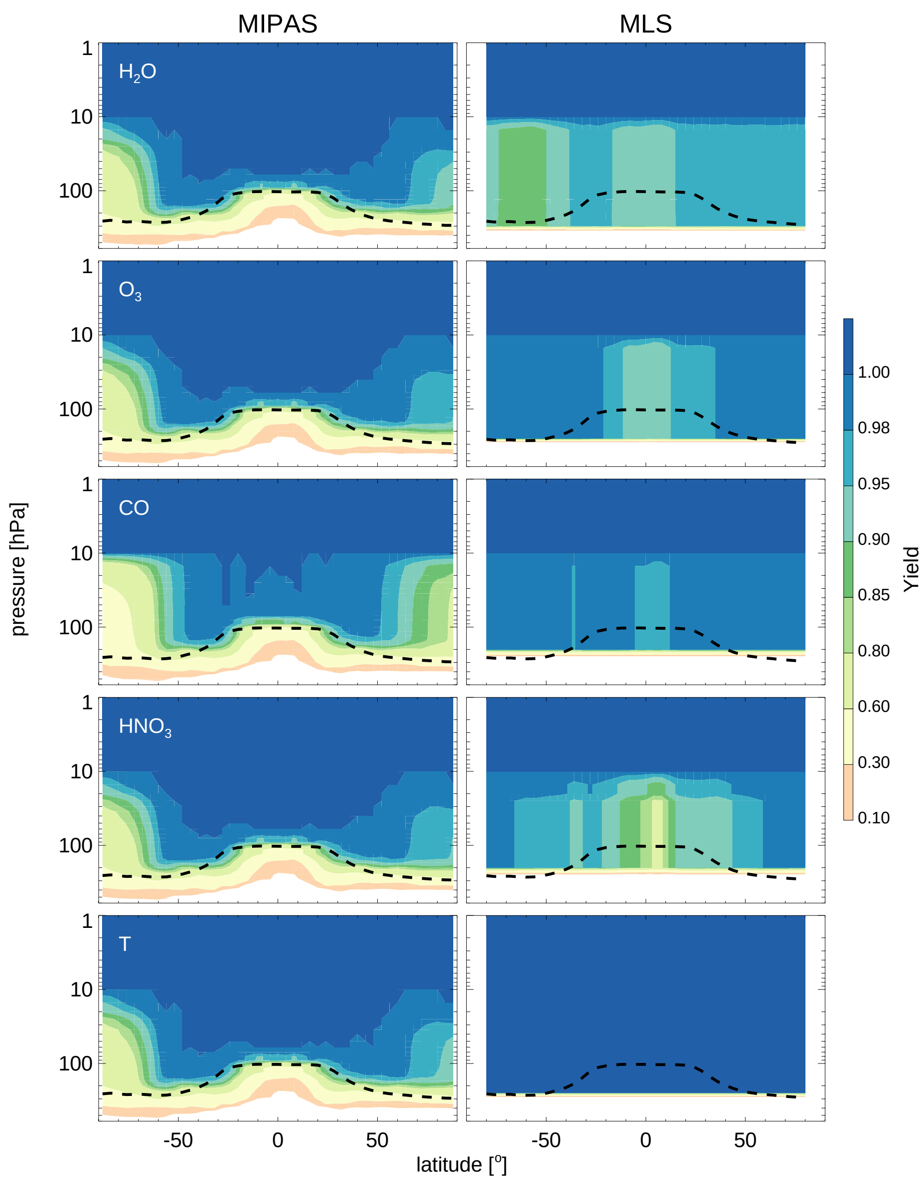
To emphasize more the MLS caveat, the sentence in P1L16 (in the abstract) will be changed to: In contrast, MLS data quality screening removes sufficiently few points that no additional bias is introduced, although its penetration is limited to the upper troposphere while MIPAS may cover well into the mid troposphere in cloud-free scenarios.

In P5 L8 this sentence will be changed: In contrast, in general MLS yield values are better than 90%, although the measurements do not extend below the upper troposphere.

In a similar manner P7 L18 will be changed to: However, continuum absorption in the microwave suppresses signals from the mid and lower troposphere in a limb viewing geometry, limiting the MLS vertical range to the upper troposphere and above while MIPAS may cover well into the mid troposphere in cloud free scenes.

And in the new paragraph about trends we included: However, when only those measurements passing the screening criteria are used, both instruments have limitations: MIPAS trends are impacted because of the large percentage of measurements screened out below 100 hPa, which introduces non-negligible artifacts (up to 80% change for H_2O and up to 20% change for O_3); MLS trends are impacted because of the reduced vertical resolution, which limits its usefulness to the upper troposphere and above.

Furthermore, we noticed that the Figures were not displaying the correct MLS pressure cut off. The revised versions showcase much better the lack of MLS penetration (as an example, the updated Figure 2 is shown below). Also, we superimpose a mean thermal tropopause derived from MERRA2 to Figure 2, 3, and 4.



Updated Figure 2

Reviewer comments

The paper is dedicated to the characterization of sampling biases in infrared and microwave limb sounding instruments, with MIPAS and MLS taken as examples. The paper is a continuation of a series of publications on characterization of sampling biases. The new aspect is analyzing the influence of quality screening on data representativeness.

MAJOR COMMENTS

1) It is stated in the abstract that " analysis of long-term time series reveals that these additional quality screening biases may affect the ability to accurately detect upper tropospheric long-term changes using such data" (similar statements are on page 6 and in conclusions) However, the performed analyses are insufficient for such statement. It is rather expected that the screening of cloudy conditions results in biased estimates, and that the variability might not be represented properly. However, biased estimates, not perfect correlation coefficient with the full time series and R2 do not necessary imply that the long-term trends are inaccurate. Furthermore, if the sampling patterns do not change over time, a large part of sampling uncertainty can be removed in the trend analysis by consideration of deseasonalized anomalies. In order to make such statement on ability of accurate trend detection, the authors should perform trend analysis using the full and sub-sampled datasets and support their statements by quantitative estimates. Another, a simpler solution, is to remove these abovementioned statements on ability to accurately detect trends from the manuscript.

See discussion on trends above.

If the authors will decide to extend the analyses, it would be also interesting to investigate the influence of sampling patterns on ability to reproduce the natural cycles.

We decided not to expand the manuscript upon the ability to reproduce natural cycles because we believe that is outside the scope of the current paper.

2) The value of the paper will be increased significantly, if the presented analyses of sampling biases using the modelled data are enhanced with comparison of real experimental data from MIPAS and MLS. Such analyses would illustrate whether the observed biases are explained by sampling patterns.

After careful consideration, we decided not to include a comparison of the real data, because such comparisons will suffer from the fact that we do not know the truth and because such comparisons can be found in several validation papers. The fact that extensive validation of these data sets has been documented in previous validation papers is now noted in the revised manuscript.

MINOR COMMENTS

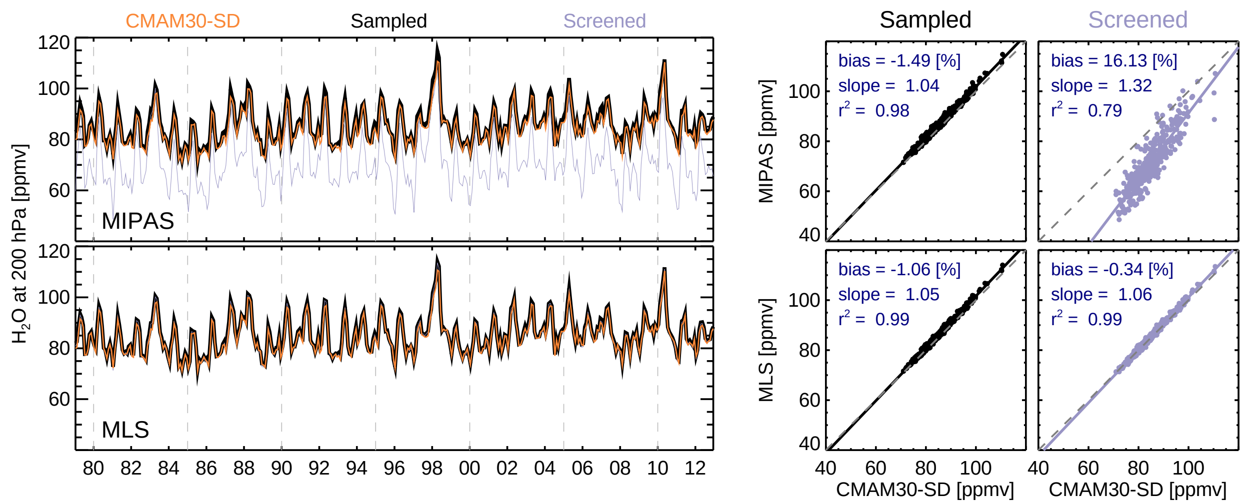
1) P.4, L.5 : Please write the version of the IMK/IAA processor. We will add: in particular version 5.

2) P.2 L.11: The sampling uncertainty has been also discussed in (Sofieva et al., 2014). In this paper, the authors analysed the sampling biases for 6 satellite instruments and proposed a parameterization of sampling uncertainty in monthly zonal mean data.

Sofieva, V. F., Kalakoski, N., Päivärinta, S.-M., Tamminen, J., Laine, M. and Froidevaux, L.: On sampling uncertainty of satellite ozone profile measurements, Atmos. Meas. Tech., 7(6), 1891–1900, doi:10.5194/amt-7-1891-2014, 2014. 3)

We will modify that section as follows: For the limb sounding technique, Sofieva et al., (2014) estimated the sampling biases in zonal mean ozone profiles from six limb-viewing satellite instruments and proposed a simple parameterization to estimate them. Toohey et al., (2013) characterized the sampling bias for H_2O and O_3 ...

Figure 5: Please use more distinct colors in scatter plots.
The colors were changed, see below:



The caption will be changed accordingly: The dashed gray lines are the 1:1 line, and the solid lines are the linear best fits, whose slopes are given.