Referee (#2)

The authors would like to thank Referee #2 for his/her thoughtful and helpful comments and suggestions. Below are the comments by Referee #2 in blue and answers in black. Any modification made to the text has been highlighted within a green box. The line numbers correspond to the version of the manuscript available for online discussion.

Comment (1): Since the structural error is just the 1-sigma variation among the retrievals, I wonder if the fact that so many of them use the Kleipool albedo and fairly coarse models is underestimating the uncertainty due to albedo and profile shape. Thorne et al. (2005) specified that structural uncertainty should be "aggregated over many independent, plausibly constructed datasets..." The paper should address whether the uncertainty due to parameters shared among a large percentage of the retrievals used in the study may bias the interpretation to an underestimate of the uncertainty.

The principal point raised by the reviewer is a good one. However, only 3 out of 7 groups are using the exact same albedo values in their retrievals. It could still be argued that most of the albedo datasets (5 out of 7) come from the Kleipool et al. (2008) database. Therefore the estimation of the structural uncertainty could indeed be biased by the surface albedo originating from the same database (though represented differently) for most of the retrieval groups.
In order to test this hypothesis, we have estimated the structural uncertainty using 3 groups that use distinctly different surface albedo values (BIRA-IASB (MODIS BSA + OMI min LER for ocean and gap filling), KNMI/WUR (minimum LER from Kleipool et al. (2008)) and University of Leicester (mode LER from Kleipool et al. (2008)) and two groups that use exactly the same albedo (KNMI/WUR and NASA, both minimum LER from Kleipool et al. (2008)). The table below shows the structural uncertainty estimated for these two different ensembles and the 1-sigma relative uncertainty of the albedo values from the different datasets:

	Institutes	**1sigma of albedo datasets**	**Structural AMF uncertainty**
Identical albedo datasets	KNMI/WUR, NASA	0	19.7%
Different albedo datasets	BIRA-IASB, KNMI/WUR, U. Leicester	30%	17.7%

We conclude that the estimation of the structural uncertainty is of the same order for the two different retrieval ensembles, so the fact that the surface albedos values come from the same database does not appear to be a clear driver of the overall structural uncertainty calculation. Nevertheless, the structural uncertainty using only two or three retrievals is smaller than the overall structural uncertainty calculated with the 7 different groups. This indicates that two retrievals only are insufficient to represent all the structural differences (use of BRDF for the surface reflectivity, different surface pressure values, aerosol corrections, cloud corrections,…) that are represented by the ensemble of 7 retrieval groups used in the manuscript.

We have added some discussion on this topic in line 436 (p. 19):

"Most of the surface albedo values used in the retrievals come from the Kleipool et al. (2008) database, which is based on OMI surface reflectance climatology. However, due to the different representations of surface reflectance within this database, only three retrieval groups use the exact same albedo values. We investigated if this could bias the estimation of the AMF structural uncertainty, and we concluded that that is not a clear driver of the overall structural uncertainty calculation."

Comment (2): Regarding the discussion in the final paragraph of p. 21, the point that validation of the a priori profiles is important is well taken, and I agree that estimating the effect of only the spatial (or temporal) resolution of the chemical transport model on the retrieval would require a very specific study. However our understanding of structural uncertainty is that comparing the AMFs calculated by a variety of retrievals allows a characterization of the total error independent of the parametric uncertainty calculations.
Given that the highest resolution a priori profiles used here were the 0.5 x 0.667 degree profiles in the POMINO retrieval, and that Valin et al. (2011), Heckel et al. (2011), and Yamaji et al (2014) all show that model resolution < 20 km is necessary to capture the nonlinearity of NOx chemistry; that Russell et al. (2011), McLinden et al. (2014), and Kuhlmann et al. (2015) show profiles at <= 15 km resolution significantly change the AMF, and Vinken et al. (2014) used a sub-grid plume parameterization in their retrieval with 0.5 x 0.667 degree profile resolution to a similar effect, my concern is that the overall uncertainty in the AMF derived from a structural uncertainty that does not include any retrievals using profiles with < 20 km resolution misrepresents the true uncertainty and bias. If adding at least one retrieval with < 20 km resolution a priori profile is impractical at this point, then at a minimum an extended discussion of the likelihood that the AMF uncertainty derived here is underestimated should be developed.

Unfortunately there is not yet a global retrieval that uses high-resolution a priori NO_2 profiles on a global scale; the specific retrievals are available only for particular regions (such as at city or regional scale, oil sands, shipping lanes) and particular studies. Ideally one would have to create a global AMF dataset using high resolution a priori profiles. This was not the main goal of this study and because of time constraints we will add some discussion on the topic, which we believe is very relevant both for this study and for satellite retrievals for current and future missions.

The table below shows the main quantitative findings from different studies (Kuhlmann et al. (2015), Mclinden et al. (2014), Heckel et al. (2011)) on the effect that using a high resolution a priori NO_2 profile have on specifically on AMF values:

Study	AMF coarse	AMF high-res	Notes
Kuhlmann et al. (2015)	1.19 GEOS-Chem	0.82 CMAQ	Profiles not validated 40% smaller AMF_{HR}
McLinden et al. (2014)	AMF coarse	AMF coarse / 1.9	DOMINO:EC AMF ratio = 1.9 around emission sources LUT have very few points
Heckel et al. (2011)	=	=	50% underestimation over land with coarse CTMs.

Based on this simple literature survey, we created different high-resolution AMF databases in which the AMFs are 50% smaller over polluted areas to simulate the effect that using a high resolution a priori profile has on the AMFs. To create these simulated high-resolution AMFs we applied a 50% reduction to one of the members that participated in the comparison. Then we included this "new" AMF member in the comparison and performed the complete analysis as done in the manuscript. In this way we obtain a new AMF structural uncertainty estimate. By including or excluding the "new" members we may obtain an estimate of how much our original structural uncertainty may be biased because of the lack of a high-resolution AMF dataset in the ensemble. We did the following three experiments:

Exp. 1: Applied 50% reduction to NASA AMFs to create the new database (HR1)
Exp. 2: Applied 50% reduction to U. of Leicester AMFs to create the new database (HR2)
Exp. 3: Applied 50% reduction to U. of Leicester AMFs and WUR AMFs to create two databases (HR31, HR32).

The statistics of the comparison for each of the experiments is summarized in the tables and figures below (in line with Table S6 and Fig. 9 in the manuscript):

Experiment 1						
	FEBRUARY			AUGUST		
	Mean	Median	σ	Mean	Median	σ
BIRA	-23	-24	16	-18	-21	16
IUP-UB	-5	-5	21	-15	-14	16
Leicester Uni.	-4	-4	16	-6	-5	11
MPIC	-16	0	43	-5	3	34
NASA	-9	-9	11	-9	-9	11
WUR	12	12	14	9	7	11
HR1	45	45	6	45	45	6

Table 1: Statistical parameters of the comparison with the model mean in experiment 1 ($((\overline{AMF} - AMF_x)/\overline{AMF})*100$, in %) of total tropospheric AMFs over the globe for polluted pixels (>1•10^{15} molec/cm^2)

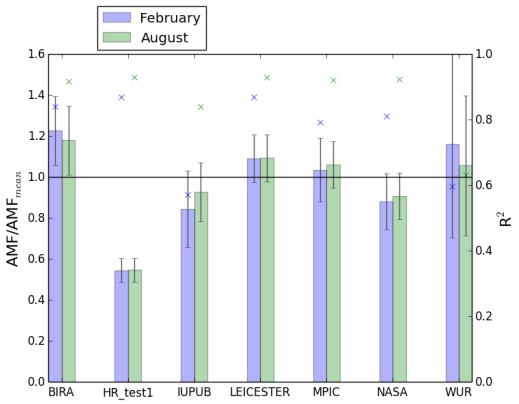

Figure 1: Ratio of tropospheric NO_2 AMFs by each group to the ensemble mean (left axis, bars) and the correlation coefficient (right axis, cross) for experiment 1.

	Experiment 2					
	FEBRUARY			AUGUST		
	Mean	Median	σ	Mean	Median	σ
BIRA	-23	25	16	-18	-21	17
IUP-UB	-5	-5	21	-15	-14	16
Leicester Uni.	-4	-4	14	-6	-5	10
MPIC	-16	-1	45	-5	3	34
NASA	-9	-9	12	-9	-9	12
WUR	11	11	13	9	7	11
HR2	48	48	7	46	47	5

Table 2: As Table 1 but for experiment 2

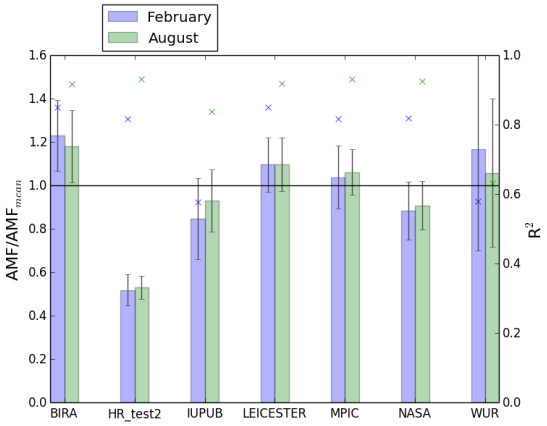

Figure 2: As Fig. 1 for experiment 2.

	Experiment 3					
	FEBRUARY			AUGUST		
	Mean	Median	σ	Mean	Median	σ
BIRA	-32	-34	17	-27	-30	18
IUP-UB	-13	-13	23	-24	-22	18
Leicester Uni.	-11	-11	15	-13	-14	11
MPIC	-25	-7	42	-13	-3	37
NASA	-17	-17	13	-17	-17	13
WUR	5	5	13	3	1	11
HR2	44	44	8	43	43	6

Table 3: As Table 1 but for experiment 3

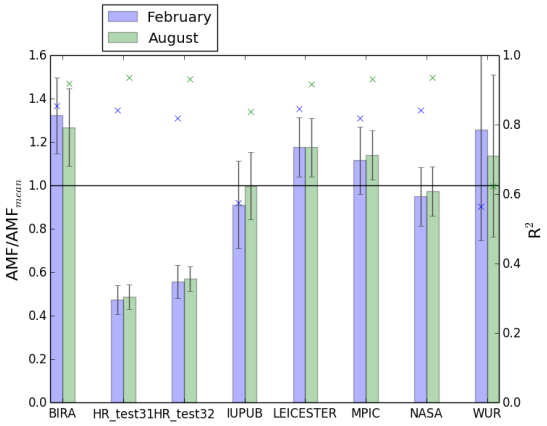

Figure 3: As Fig. 1 for experiment 3.

We compared the standard deviations in these tables to Table S6 in the supplement. For the individual comparisons, the standard deviations with respect to the model mean do not change considerably, the order of magnitude stays within a couple of percent points. In terms of the estimation of the AMF structural uncertainty we conclude that:

1. AMF structural uncertainty over polluted areas (>1*10^{15} molec/cm^2) increases by 1% in February and 3% in August.

2. AMF structural uncertainty over polluted areas (>1*10^{15} molec/cm^2) increases by 3% in February and 3% in August.

3. AMF structural uncertainty over polluted areas (>1*10^{15} molec/cm^2) does not increase in February and increases 6% in August.

These results indicate that the effect of lacking a hi-res AMF member on our AMF structural uncertainty is likely not very strong. The effect of course is notable in the AMF values as showed in the mentioned studies and also visible in the different figures of the different experiments. However, the original ensemble of 7 retrievals used in the comparison accounts for most of the possible structural differences in the AMF calculation.

We have extended the discussion on the effect of the a priori profiles in page 21: (two first paragraphs are already in the original manuscript, for context purpose)

Selecting a specific chemistry transport model thus influences the AMF structural uncertainty via differences in the profile shape. These differences in the profile shape depend on the different characteristics of the models (e.g. spatial and temporal resolution and parameterization of different processes in the atmosphere). Previous studies analysed how using different CTMs influences the NO_2 retrievals due to the change in the profile shapes used to calculate the AMF values. Heckel et al. (2011) compared retrievals using fine and coarse resolution models and concluded that using one AMF value for a large heterogeneous scene can lead to 50% bias in the retrieved NO_2 columns. Vinken et al. (2014) reported much smaller average differences of 10% in retrieved NO_2 columns mainly due to different emission inventories used in TM4 (3° x 2°) and GEOS-Chem (0.5° x 0.67°). According to Laughner et al. (2016), different temporal resolution also influences a priori profile shapes; they found differences in the retrieved NO_2 column for individual days up to 40% that were mostly explained by day-to-day wind direction variations that were not captured in the monthly averages.

All these aspects influence the estimation of retrieval (and AMF) theoretical uncertainties. In order to quantitatively estimate the effect of one model characteristic alone (e.g. the spatial resolution) on the AMF structural uncertainty it would be necessary to compare AMF calculated with the same approach but with just that specific characteristic being different in the profile shapes generated by the CTM. Such a specific sensitivity analysis has not been done in this study but should be considered in future AMF comparisons.

To test the robustness of our structural uncertainty estimate, we did some experiments by creating new AMF databases to simulate the effect of high resolution a priori profiles on AMF values. Kulhmann et al. (2015), McLinden et al. (2014) and Heckel et

al. (2011) reported that AMFs calculated using coarse resolution a priori profiles are overestimated over polluted areas by approximately 50%. When including synthetic AMF databases emulating the use of high resolution a priori profiles, the estimated AMF structural uncertainty is not strongly affected (increases by 3-6%). This indicates that with the ensemble of retrievals used in our comparison the estimate of the structural uncertainty in the AMF calculation may be considered a robust estimate.

Comment (3): The final statement on p. 21: "It is worth to note that using averaging kernels will reduce the effect of the a priori trace gas profile chosen in the retrieval scheme." Requires additional discussion. My understanding of the use of averaging kernels is that they are useful in two ways:
1) When comparing satellite retrieved VCDs against a model, applying the AKs to the model effectively "retrieves" the model trace gas profile, thus the dependence on the a priori profile in the retrieval is the same for both the observed and modeled column, and cancels out (Eskes and Boersma, 2003).
2) Alternately, one could use AKs to implement one's own a priori profiles in the retrieval.
Only in the first case would I say that the dependence on the trace gas profile is reduced, and that only applies when comparing to a model. Work using the satellite columns directly (e.g. Duncan et al. 2010, Beirle et al. 2011, Valin et al. 2013, Mebust and Cohen 2014, Lu et al. 2015, Liu et al. 2016, etc.) would not be able to use AKs in this way.

When averaging kernels are being applied, for instance when comparing retrieved NO_2 columns with modelled NO_2 distributions or with observed NO_2 profiles (aircraft, balloon), the comparison will become self-consistent in terms of using a priori information. Using the averaging kernel reduces systematic and random differences between modelled and satellite-observed columns because the representativeness of the modelled state for the observed state improves (e.g. Boersma et al., 2016). We agree that the retrieval of NO_2 columns will stay sensitive to the choice of the a priori profile, but using the averaging kernel provides a data user with the means to improve the consistency associated with the a priori profiles in interpreting the satellite data.

As both reviewers have raised their concern in this particular statement, we have tried to make it clearer:

It is worth to note that using averaging kernels in satellite applications (e.g. when comparing retrieved NO_2 columns with modelled NO_2 distributions or observed NO_2 profiles) will reduce the representativeness errors in the comparisons associated with the a priori trace gas profile used in the retrieval scheme (e.g. Boersma et al., 2016).

Technical corrections:
p. 21 l. 476 - PRevious (the R should be lowercase) - corrected
p. 21 l. 479 - NO2 (the 2 should be subscript) - corrected
p. 21 l. 480 - (Laughner et al. 2016) > Laughner et al. (2016) - modified

References

Beirle et al.: Megacity Emissions and Lifetimes of Nitrogen Oxides Probed from Space, Science, 333, 1737-1739, 2011.

Duncan et al.: Application of OMI observations to a space-based indicator of NO_x and VOC controls on surface ozone formation, Atmos. Environ., 44, 2213-2223, doi: 10.1016/j.atmosenv.2010.03.010, 2010.

Eskes, H. J. and Boersma, K. F.: Averaging kernels for DOAS total-column satellite retrievals, Atmos. Chem. Phys., 3, 1285-1291, doi:10.5194/acp-3-1285-2003, 2003.

Heckel et al.: Influence of low spatial resolution a priori data on tropospheric NO_2 satellite retrievals, Atmos. Meas. Tech., 4, 1805-1820, doi: 10.5194/amt-4-1805-2011, 2011.

Kleipool, et al.: Earth surface reflectance climatology from 3 years of OMI data, J. Geophys. Res.: Atmospheres, 113, doi:10.1029/2008JD010290, d18308, 2008.

Kuhlmann et al.: Development of a custom OMI NO_2 data product for evaluating biases in a regional chemistry transport model, Atmos. Chem. Phys., 15, 5627-5644, doi:10.5194/acp-15-5627-2015, 2015.

Laughner, J. L., Zare, A., and Cohen, R. C.: Effects of daily meteorology on the interpretation of space-based remote sensing of NO_2, Atmos. Chem. Phys. Discuss., 2016, 1–27, doi:10.5194/acp-2016-536, 2016.

Liu et al.: NO_x lifetimes and emissions of cities and power plants in polluted background estimated by satellite observations, Atmos. Chem. Phys., 16, 5283–5298, doi:10.5194/acp-16-5283-2016, 2016.

Lu et al., Emissions of nitrogen oxides from US urban areas: estimation from Ozone Monitoring Instrument retrievals for 2005–2014, Atmos. Chem. Phys., 15, 10367-10383, doi: 10.5194/acp-15-10367-2015, 2015.

McLinden et al., Improved satellite retrievals of NO_2 and SO_2 over the Canadian oil sands and comparisons with surface measurements, Atmos. Chem. Phys., 14, 3637-3656,doi:10.5194/acp-14-3637-2014, 2014

Mebust and Cohen, Space-based observations of fire NOx emissions coefficients: a global biome-scale comparison, Atmos. Chem. Phys., 14, 2509-2524, doi:10.5194/acp-14-2509-2014, 2014.

Russell et al., A high spatial resolution retrieval of NO_2 columns densities from OMI: method and evaluation, Atmos. Chem. Phys., 11, 8543-8554, doi:10.5194/acp-11-8543-2011, 2011.

Thorne, P. W., Parker, D. E., Christy, J. R., and Mears, C. A.: Uncertainties in climate trends: Lessons from Upper-Air Temperature Records, Bull. Am. Meteorol. Soc., 86, 1437–1442, doi:10.1175/BAMS-86-10-1437, 2005.

Valin et al., Effects of model resolution on the interpretation of satellite NO_2 observations, Atmos. Chem. Phys., 11, 11647-11655, doi:10.5194/acp-11-11647-2011, 2011.

Valin et al., Variations of OH radical in an urban plume inferred from NO_2 column measurements, Geophys. Res. Lett., 40, 1856–1860, doi:10.1002/grl.50267, 2013.

Vinken et al., Constraints on ship NO_x emissions in Europe using GEOS-Chem and OMI satellite NO_2 observations, Atmos. Chem. Phys., 14, 1353–1369, 2014.

Yamaji et al., Influence of model-grid resolution on NO_2 vertical column densities over East Asia, J. Air. Waste. Manage., 64, 436-444, doi:10.1080/10962247.2013.827603, 2014.