Reply to the Interactive comment of reviewer #1 on "Influence of under-sampled a priori data on tropospheric NO 2 satellite retrievals" by A. Heckel et al.

We would like to thank the reviewer for the important and helpful comments which lead to some additional calculations and results. We will address the reviewers concerns in detail below.

This paper addresses the important issue of spatial undersampling of a priori information needed by NO2 retrievals from space. Using low resolution a priori information for high resolution retrievals means that sharp spatial gradients in a priori NO2 profiles, surface albedos and terrain height are not properly taken into account, resulting in systematic errors in the retrievals. For clouds, this problem is less relevant as cloud information is usually retrieved from the same instrument at the appropriate resolution.

The authors deserve considerable credit for taking up this issue that has been identified before, but was never quantified. The set-up of the experiment –comparing retrievals with high resolution a priori information to retrievals with spatially smoothed a priori information– makes sense, and the outcome is in line with expectations based on retrieval theory. The paper has been written well, and the authors put their results in the perspective of what we know from literature. I also like the fact that the authors have quantified the error reduction when improving to 90 x 90 km2 a priori NO2 profile data. That resolution comes close to the global chemistry transport model capability of 1 x 1, which is expected to be the next feasible resolution generating a priori profiles that will be used in standard retrievals in the foreseeable future.

All in all, this paper makes an important contribution in pointing the way forward for improving satellite retrievals of minor trace gases with the DOAS approach. I absolutely support the main conclusion that state-of-science retrievals should focus on using a priori information with the appropriate spatial resolution. In case of sensors such as OMI and GOME-2, this means that especially the spatial resolution of the a priori profiles needs to be improved.

The difficulty I have with this paper is that the final error estimates to my opinion exaggerate the problem at hand. The abstract states that the relative uncertainties can be more than a factor of 2, but the paper does not make sufficiently clear that this factor of 2 uncertainty does not hold for current standard retrievals. The error presented by the authors is representative for a theoretical retrieval approach for a non-existing instrument with 15 x 15 km2 spatial resolution that has to rely solely on 3 x 3 a priori information for both NO2 profiles, albedos, and aerosols. But standard retrievals all use a priori information at significantly better resolution than the 3 x 3 used here, especially for albedo and terrain height, but also for a priori NO2 profiles. Thus, the authors should either make clear that their error estimates are valid for such a theoretical retrieval based on 3 x 3 information, or use current, more realistic spatial resolutions for pixel sizes, NO2 profiles, albedo, and aerosols, and come up with a better (and likely smaller) error estimate that would hold for standard retrievals.

We agree with the reviewer that our study is dealing with an extreme case in the sense that it evaluates the effect of using very coarse input data to measurements taken at 15 x 15 km resolution. The study does intentionally not aim at evaluating the uncertainties in specific retrievals as the results would not have been general in that case. The idea was to separate effects by changing resolution on different aspects of the a priori data and evaluate the effects on a simplified scenario. The resolutions applied are somewhat arbitrary but not unreasonable – most CTMs used for the a priori data have spatial resolutions of between 2 and 3 degrees, and instrumentation currently being planned such as TROPOMI will have ground pixels of up to 7 x 7 km2.

However, as the reviewer points out, the resulting uncertainties will probably be larger than those of current retrievals, and in this sense, they may be misleading to readers not fully aware of the details. We therefore have performed additional simulations to evaluate situations with more intermediate resolutions. From these simulations we present the impact of improved resolution of the albedo, the aerosol and the combined effect of improving albedo and aerosol. Finally we show that a considerable reduction of uncertainties is already achieved by using the NO2 profiles at 90 km resolution in combination with 15km resolution albedo and aerosol input. The additional results are discussed in detail in section 3. Furthermore we specified our conclusions more carefully to avoid misinterpretation.

To underline this:

** The authors use a single a priori NO2 profile on 3 x 3 resolution as a reference. The implicit assumption in the paper is that the air mass factors calculated for this 3 x 3 profile represent common practice in standard retrievals. But it's not. Standard retrievals use a priori profiles with resolutions of 2 x 2.5 (Dalhousie, NASA), 2 x 3 (KNMI), so the 3 x 3 profile is not so 'typical' as the authors claim: it is a factor of 1.5-1.8 too coarse. Therefore, the conclusion that single AMF errors lead to errors of 50-100% is too strong, and only holds for retrievals that use a 3x3 degree a priori NO2 profile, and to my knowledge such retrievals do not exist.*

Agreed, but we believe that the difference between 2 x 3 and 3 x 3 degrees resolution is less significant considering the theoretical nature of our study and the fact that the spatial resolution of most models also depends on latitude and therefore is not fixed relative to the satellite resolution.

** The authors conclude that the 'spread of AMF values', not captured by the single AMF value, leads to errors up to 100% . But even if the 15-km profile happened to be identical everywhere within the 3x3 domain, one would still observe a distribution of AMFs. Such a distribution reflects the variability of other a priori parameters (e.g. albedo, terrain height). Most standard retrievals do take high-resolution variability in surface albedo, terrain height, etc. into account (e.g. NASA, KNMI, EMPA to name a few), and their AMFs therefore capture at least part of the spread shown in Fig. 3(b). The authors should make a distinction between spread in AMFs caused by hi-res a priori profile variability not captured by current retrievals and spread caused by hi-res surface albedo, terrain height variability that is accounted for by current retrievals, and not report just the combined number as this may mislead readers into believing that the 50-100% is in fact the profile-shape undersampling error.*

We agree with the reviewer that the effects of the different a priori data need to be discussed separately. This was already done in the original manuscript to some extend but in response to the reviewer's comment we have now evaluated situations with different combinations of low and high resolution inputs.

With respect to the point that many retrievals already use some a priori data sets at higher spatial resolution we again agree but as stated above have tried to separate this theoretical study from specific settings used in current retrievals. However, in the revised manuscript, we now have made the point clear that many retrievals already have some a priori data at higher resolution.

** The WRF-Chem a priori profiles at 15 x 15 km2 are too high-resolution to represent the spatial variability needed for current sensors such as OMI (24 x 13 km2 at best) or SCIAMACHY (60 x 30 km2), and too coarse to be representative for future missions such as TROPOMI (7 x 7 km2). So the spread in Fig. 3(b) is too strong to be representative for OMI/SCIAMACHY resolutions, and inappropriate as a basis for error estimates for these instruments. Using WRF-Chem profiles on 30 x 15 km2 would be much more representative for OMI (as 60 x 30 would be for SCIAMACHY), and also lead to smaller* *differences and smaller error estimates between the distribution and the domain-average, single AMF values. This is in line with the 15 vs. 90 km study reported on in 3.4.*

> Again, we do not try to make this a study on the uncertainties in a specific retrieval on data of a specific instrument. We could co-add profiles to come to the resolution of 30 x 15 km, but in that case questions such as the orientation of the OMI pixels and their exact location relative to the centre of the model grid cell will become relevant. Investigating such effects is interesting but clearly out of the scope of this simple sensitivity study which tries to keep independent from sensor details.

** Section 3.2: to my knowledge, none of the standard retrievals is still using 'single AMF' values based on 3x3 albedo maps. Instead most account now for high-resolution spatial variability in the surface albedo. The authors might consider comparing the use of MODIS-based pixel-size albedo estimates to the 0.5 x 0.5 or 1x1 albedos now used (NASA, KNMI, EMPA, Dalhousie, Bremen), and update their error estimate to relevancy.*

> In response to the reviewer comments, we have extended our analysis to include more different resolutions. We also tried to give more credit to improvements already implemented in current retrievals.

The application to other seasons (section 4.2) is interesting, but it is not clear to me whether also seasonal changes in NO2 profile shape have been accounted for. From the description on P1911, it seems that only the solar zenith angle has changed to reflect lower sun in wintertime, and all other variables have been kept constant. Keeping the surface albedo constant is perhaps justifiable, but we know that the vertical and spatial distribution of the NO2 profiles differs significantly between seasons (as the authors also acknowledge on lines 9-12). So to properly evaluate seasonal differences in the errors, I would encourage the authors to use an appropriate WRF-Chem simulation for a wintertime day and evaluate the combined effects of low sun and wintertime NO2 profile on the retrievals.

> We agree that this section does only deal with one aspect of the seasonality and have rephrased this section accordingly. Unfortunately, no winter model run from the same model is available so that we cannot provide the full seasonal test.

Some aspects of the influence of a priori profiles on retrievals, irrespective of their spatial resolution, have not been addressed at all. Previous work (e.g. Hild et al., 2002; Martin et al. 2006, Beirle-papers) has clearly shown that NO2 in the upper troposphere for instance from lightning leads to increased values for the air mass factors. Because the authors do not state whether lightning NOx production is included in WRF-Chem in the first place, and neither whether free and upper tropospheric NO2 contributed significantly to the NO2 burden on 29 August 2005, we can only guess whether the results presented here are truly representative. Suppose that WRF-Chem does not include the lightning NOx source, then the conclusions presented here are too strong. On the other hand, if we suppose that WRF-Chem simulated too much NO2 aloft, then the conclusions might even be at the cautious side. The authors should inform us to what extent NO2 in the free troposphere is taken into account in WRF-Chem, and also to what extent the simulation of 29 August 2005 can be regarded as typical.

> The model simulations did not take recent elevated levels of NOx due to lightning into account. It is certainly correct that the presence of lightning NOx is an additional source of uncertainty of the NO2 profile and subsequently the AMF. However, this theoretical study does not compare to actual measurements but compares the impact of the different resolutions of a priori data relative to the high resolution model data set. In that way the

results can be regarded as representative for regions or air masses which have not recently experienced lightning activity.

On page 1897, the authors state that the 'uncertainty in the a priori information used' (in the calculation of the air mass factor – KFB) is a factor that is 'systematic', and contrast it to measurement noise, which they characterize as of 'random nature'. I presume the authors actually refer to the error contributions from both noise and a priori here (instead of the 'factors'), and imply that a priori information leads to systematic errors in the retrieval. But a priori information error contributions are not strictly systematic in the sense that they represent a fixed and unchangeable bias. To illustrate this: the assumed surface albedo might be too low for one particular pixel on a particular day (due to e.g. vegetation growth), and too high for the same location one day later (e.g. soil darkening due to precipitation). So I do not think that a strict distinction between fitting errors as random errors on the one hand and air mass factors as systematic errors on the other hand does justice to the complexity of the issue. Certainly, air mass factors have significant systematic components, but sometimes can be regarded –at least partially- as consisting of random contributions as well.

We agree the error budget is more complex than the simplified example we give in that sentence. However, the essential fact is that the uncertainty in the AMF due to the a priori data sets is not purely random but can introduce significant systematic biases. Hence it is not possible to simply remove this uncertainty by temporal averaging. We have rewritten this sentence accordingly to highlight the complexity of this issue.

Minor issues

P1895, L20: the papers cited are concerned with changes in emissions or NO2 columns over periods of more than a decade, so I would suggest not to call this 'short trends'.

Done

P1895, L29: 'quantitative analysis with high accuracy'. Any analysis obviously needs to be of the highest possible accuracy, but I think the authors are actually referring here to the need for accurate absolute retrieved quantities. I suggest they rephrase.

We agree and rephrase the respective sentence to: "These applications rely on quantitative analysis with high absolute accuracy on single measurements to derive reasonable emissions estimates."

P1896, below equation (2), delta z should also be defined for completeness.

Done

P1896, L18-22: I suggest to point out here that bAMF_z depends on the assumed albedo, aerosols, clouds, and terrain pressure to make clear where 'This step' actually relies on the appropriate a priori data.

We agree and modified the paragraph accordingly.

P1896, L24: I suggest the authors specify what 'on average' here means.

The use of 'on average' in this context is misleading. The sentence has been corrected.

P1897, L9: I dispute that 'the resolution of the a priori dataset . . . has often not been improved from that used for GOME-1'. Improvements in the terrain height resolution were the topic of papers by Schaub et al. [2007], and Zhou et al. [2009]. OMI NO2 products switched to higher-resolution albedo datasets as of 2009.

We agree that we should acknowledge the respective improvements and have rewritten accordingly.

P1898, L21: does ‘atmospheric profile’ here refer to one a priori NO2 profile?

We meant one set of a priori parameters to calculate the AMF for the domain. Clearly the use of ‘atmospheric profile’ is not quite to the point. We have rewritten the sentence.

P1899, L16-17: I think the statement that ‘this kind of spatial variability’ (in the NO2 distribution - KFB) is not represented in standard retrievals is too strong. I think it would be more appropriate to state that the spatial variability is not sufficiently represented in the a priori information used for standard retrievals (and perhaps also state which retrievals you refer to as standard retrievals: are these are the Dalhousie, Bremen, NASA, and KNMI retrievals?). The KNMI algorithm for instance calculates a more representative a priori profile based on the closest 4 (2 x 3 native resolution) grid calls to the centre of the pixel. Such a smoothing step will not completely resolve the gradients discussed in this study, but it results in a better representation of the NO2 spatial distribution compared to the single large model grid cell case studied here as reference case.

We rephrased the paragraph strengthening the theoretical nature if this study and avoiding direct comparisons as these need too much detail to be justified.

P1901, L19-20: it was not clear to me why all data from August 2005 were averaged. Previously the authors stated that the authors intend to evaluate the effect of under-sampling a priori data for a single day, so why now use a monthly mean here? Is MODIS AOT for 29 August 2005 not covering the whole domain?

While the MODIS combined product for land and ocean at 550nm reports AOT values for the whole domain, the respective land only product at 465nm shows repeatedly gaps close to the coast line of the San Francisco bay. Hence the choice of the monthly average AOT at 465nm instead of the AOT of actual day.

P1902, L223: typo ‘obtaine’.

Corrected.

P1905, L26-27: I think the authors should clearly state here that aerosols enhance the photon path length in the boundary layer over shielding, provided that the aerosol and NO2 vertical profiles are identical, as they assumed in this calculation.

We agree and write this point more clearly.

P1906, end of 3.3: it would be interesting to still put a number on the aerosol effect. This can then be directly compared to the error estimates for the hi-res profile and surface albedo effects.

Done.

P1915, L13: the uncertainties range here from -5 1015 to +5 1014 molec. cm-2. Is it really 1014? Later on I see that this can be read off from Figure 10, which has not been presented before. Perhaps it is good to remind us that the -5 1015 effect mainly illustrates the AMF decreases because of improved sampling over polluted land scenes, whereas the +5 1014 molec. cm-2 illustrates the increases in AMF because of better spatial sampling over oceans.

Following the additional simulations Figure 10 has been replaced and discussed in more detail. We also highlight the differences between polluted and less polluted pixels. The latter will be dominated by the ocean pixels but more rural land pixels are included as well.