K. Schaefer

The paper describes a satellite based method to conclude about anthropogenic emissions. This is important especially in regions with strong temporal variations of emissions as China. Important is the handling of aerosol influences upon the NO2 retrieval including modelling and AOD measurements. Independent from the statements that current MAX-DOAS measurements are very limited over China it would be helpful to show a comparison between the ground-based and satellite-based remote sensing of NO2 vertical column densities and how the integration of ground-based remote sensing would improve the final results.

Response: We thank Dr. Schaefer's comments.

There are currently insufficient ground-based measurements to systematically evaluate satellite products over China. In our conclusion, we wrote:

"There are no sufficiently comprehensive independent measurements available to systematically evaluate the various NO_2 retrieval approaches. Current MAX-DOAS measurements are very limited over China, with few sites and short operation periods (Irie et al., 2012; Ma et al., 2013; Hendrick et al., 2014; Kanaya et al., 2014). In situ measurements are rare for vertical profiles of aerosols and NO_2. Our results show that the effects of aerosols and surface reflectance are highly season- and location-dependent. This clearly indicates the need for a comprehensive measurement network to validate satellite data. Nonetheless, our present study and that of Lin et al. (2014b) point the way forward for a physically more realistic NO2 retrieval by explicit inclusion of aerosol effects."

The insufficiency is because there is large spatial and temporal variability in the complex effects of aerosols and/or surface reflectance treatments, as clearly shown in our manuscript. Satellite product evaluation using limited ground-based datasets has been done in many previous studies (see the references in the above quotation), including our own work (Lin et al., 2014b). Except for those in Lin et al. (2014b), most ground-based data are not available to us currently.

In the revised abstract, we have further clarified that:

"A comprehensive independent measurement network with sufficient spatial and temporal representativeness is needed to further evaluate the different satellite retrieval approaches."

Reviewer 1

I found this to be a very important and mostly clear paper that is a wonderfully complete case study on tropospheric NO2 retrieval. It should be accepted with only a few minor changes in grammar and bookkeeping, none of particular substance. They are:

Read carefully to define abbreviations and acronyms on first use (AMFv6, OMLER, Case S_A, CRF, : : :.)

Response: AMFv6 is short for our Fortran package 'Air Mass Factor version 6'. We have decided to always just use the short name, in order not to confuse it with the actual 'air mass factor' quantity. OMLER is a widely-used satellite albedo product best understood with its short name. Case S_A etc. represent individual sensitivity cases with no full names.

12656.1 – Excluding days with high pollution is not clarified until Section 3.5. Please add a sentence to clarify here.

Response: We have updated the sentence as follows:

"The implicit aerosol treatment also tends to exclude days with high pollution, since aerosols are interpreted as effective clouds and the respective OMI pixels are often excluded by cloud screening; this is a potentially important sampling bias."

12657.5-6 Please discuss in terms of moving from implicit to explicit, since this is the direction of improvement.

Response: Updated:

"Our previous study (Lin et al., 2014b) for several locations in the North China Plain (NCP) has shown large changes in retrieved NO_2 VCDs when moving from an implicit to an explicit treatment of aerosols. In particular, NO_2 VCDs are reduced by 14% on average but are changed by (-90)–(+70)% for individual pixels when aerosol optical depth (AOD) exceeds 0.8."

12659.25 I found Figure 2 to be unnecessary. It did not clarify the retrieval procedure.

Response: Although the retrieval procedure is easily understood for advanced users, we have elected to use Fig. 2 to help other users' understanding. In particular, we clearly show in Fig. 2 that we always retrieve cloud parameters prior to the NO2 retrieval using consistent ancillary parameters. The decision is in part based on some reader feedbacks for our previous paper (Lin et al., 2014b).

12666.4-8 The wrong conclusion seems to be drawn here. If the heights are correlated would that more likely reduce errors than if they were not?

Response: The writing is indeed misleading here. We have updated the sentences to:

"Note that since the same vertical mixing and convection schemes were used to simulate aerosols and NO2, the height of aerosols relative to NO2 (above or below or mixed with NO2, relevant to our study) may be subject to smaller errors than the absolute height of aerosols. Future work is needed to better understand and constrain aerosol properties and evaluate how they affect the NO_2 retrieval."

12667.20 and later discussion – The order of rows in Figure 5 and the order of subsequent discussion in the text should be made consistent.

Response: Although we discussed the fifth row prior to the second-fourth rows, we have elected not to adjust the order of these rows, in order to be consistent with other figures, e.g., in the supplementary material.

12671.23 – "to space"

Response: Changed.

12672.7 What is the correlation for Castellanos et al?

Response: They show correlation of 0.87-0.93 for each bin of SSA. The overall correlation is reduced if all cases of SSA are considered together. The paper does not specify the overall correlation. See their Fig. 14:

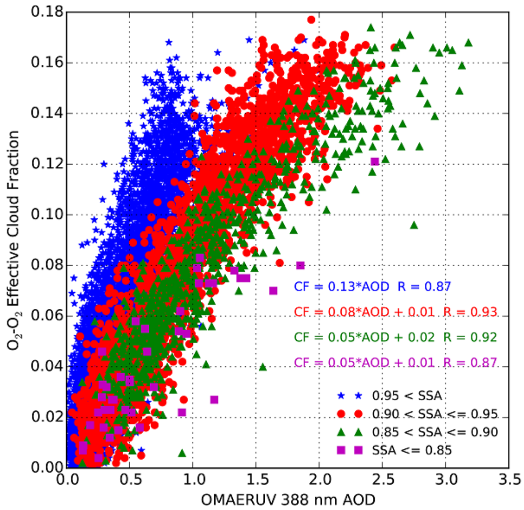

Figure 14. Comparison of the OMAERUV retrieved 388 nm AOD and observed effective cloud fraction binned by the OMAERUV retrieved SSA.

12672.26 "Several representative regions of China are considered, including: : :."

Response: Changed.

12675.15 “: : :REF from DOM. These include: : :.”

Response: Changed.

12676.12 “: : :TOA radiance is from the combination: : :.”

Response: Changed.

12676.26 “situations”

Response: Changed.

12681.10 “: : :DOMINO v2. Our cloud: : :.”

Response: Changed.

12684 – The conclusions here should certainly mention the upcoming GEMS mission and its hourly high spatial resolution measurements of China.

Response: The last sentence has been updated:

“Such retrieval efficiency enables a fast global retrieval that will be particularly important for future fine-resolution satellite instruments such as TropOMI (which is expected to have a data rate ~8 times that of OMI) and GEMS (which will be onboard a geostationary satellite with hourly measurements at a horizontal resolution of 5 x 15 km^2).”

Reviewer 2

The manuscript presents sensitivity studies to elucidate the impact of the treatment in aerosol properties and/or surface reflectance on the retrievals of NO2 and top-down estimates of NOx emissions at regional scale over China. Understanding systematic biases in the retrievals is important especially since NO2 retrievals have been used recently in several key science and policy-relevant studies (e.g., emission estimation). While this is a direct extension of their previous work (Lin et al. 2014b), the results presented in this work have potential contributions worthy of publication.

However, the reviewer has the following concerns:

1) It is important that results be validated with independent observations (e.g., groundbased measurements). While it is understandable that there are limited measurements to compare with, the results currently presented can only be interpreted qualitatively without some form of validation. A similar comment in the discussion has also been made in this regard. Systematic biases as elucidated by the sensitivity experiments can be due to biases in the inputs as well (i.e., MODIS BRDF, GEOS-Chem aerosol properties, and other prior information).

Response: we agree that comparisons with a comprehensive ground-based (and independent) measurement dataset are needed to further evaluate different retrieval approaches. And we intend to do so in the future when such comprehensive measurements are available. Current measurements are insufficient due to lack of spatial and/or temporal representativeness. We have used some of these measurements to confirm the improvements of our retrievals (a few locations, 30 days of data for a total of 127 pixels; Lin et al., 2014b). Other ground-based data are not available to us. Please see our response to Dr. K. Schaefer for more explanations.

We agree our POMINO retrieval is also subject to errors, as clearly discussed in Sect. 2.5 and the conclusion section. In the conclusion and abstract, we have also clearly stated the necessity of using comprehensive independent measurements with sufficient spatial and temporal representativeness to evaluate satellite products.

Nonetheless, as written in the end of Sect. 2.5:

"our present study, at the very least, reveals the importance of an explicit aerosol treatment for NO_2 and associated cloud-parameter retrievals at a regional scale, especially given the lack of such an explicit treatment in current satellite products. In support of our work here, Lin et al. (2014b) showed that, by explicitly accounting for aerosols with just the AOD values constrained by observations, there is excellent correlation between retrieved NO_2 VCDs and independent MAX-DOAS data (R^2 = 0.96 in day-to-day variability across the few locations being studied). Section 3.3 further shows large changes in retrieved NO_2 VCDs from an explicit to an implicit treatment of aerosols, and Sect. 4 illustrates the consequences on subsequent NO_x emission constraint. Therefore, we expect that the explicit inclusion of aerosols will improve the NO_2 retrieval, especially if more comprehensive observations become available to constrain model aerosols."

2) What are new additional important findings in this work, which were not reported in Lin et al 2014b? This distinction is not clear in the presentation. A shift in focus on these new findings would strengthen this paper.

Response:

Our previous work (Lin et al., 2014b) mainly presents the improved OMI NO2 retrieval approach for a few locations (with 30 days of data for a total of 127 pixels). This study has extended to 1) introduce a new POMINO product for the whole China domain with a highly computationally feasible retrieval method (with a OpenMP-parallelized code for pixel-specific radiative transfer calculations and no use of a look-up table), 2) revealed the large seasonal and spatial dependence of the effects of aerosol and surface reflectance treatments (which calls for a comprehensive independent measurement network for satellite product evaluation), and 3) further demonstrated the effects on emission constraint. These new works are presented and emphasized throughout the paper. Moreover, we have stated in the introduction that:

"This study extends our previous work (for a few locations; Lin et al., 2014b) to introduce an improved pixel-specific level-2 retrieval of tropospheric NO_2 VCDs over China (80 °E–130 °E, 20 °N–53 °N), Peking University OMI NO_2 (POMINO). Using a parallelized LIDORT-driven AMFv6 package (Palmer et al., 2001; Martin et al., 2003; Lin et al., 2014b), we explicitly account for aerosol optical effects, surface reflectance anisotropy, and their spatiotemporal variability. We then evaluate the individual and combined effects of an implicit aerosol treatment and changes in surface reflectance characteristics. In particular, we show large seasonal and spatial dependence of the effects of aerosol and/or surface reflectance treatments. We further illustrate the influences on subsequent NO_x emission constraints, a popular application of OMI data. Our POMINO data are available for 2004–2013 and will be updated to more recent times. Results for 2012 are presented here, by aggregating level-2 data into monthly mean values on a 0.25 °long. x 0.25 °lat. grid."

3) It is not clear whether the results of the sensitivity experiments can be interpreted in a robust manner. First, it appears (from the presentation) that the comparison between REF (POMINO) and DOM (DOMINOv2) is not a fair comparison. As mentioned by the authors, the interpretation of CRF is different between the two. The 'implicit' assumption in DOM is not entirely neglecting the aerosol contribution as it is interpreted to be the combined effect of cloud and aerosols ('effective', by way of retrieving the cloud properties). In addition, the use of 'valid pixels' for REF alone biases the comparison with DOM given that some criteria of pixels being valid are related to CRF. It would strengthen this paper if the difference between DOM and

POMINO are better described and that the implicit assumption versus explicit representation is better clarified. Can the systematic biases be quantified in DOMINO retrieval algorithm (as 'model' errors)?

Response: First, the difference between an explicit and an implicit treatment is very clearly specified in the introduction (parts of 2nd and 3rd paragraph):

"In particular, current NO_2 algorithms take an implicit approach to accounting for aerosol optical effects, with no explicit specification of aerosols in the retrievals of both NO_2 VCDs and ancillary cloud parameters. The rationales for this approach are (1) aerosols affect the retrieval of cloud parameters, so that the retrieved cloud parameters are "effective" and implicitly contain certain aerosol information, and (2) these effective cloud parameters at least partly describe the effect of aerosols on NO_2 air mass factors (Boersma et al., 2004; Boersma et al., 2011). This implicit treatment is supported by the good spatial correlation (0.66) observed between coincident MODIS aerosol optical thickness values (mostly due to scattering) and O_2-O_2 effective cloud fractions over the eastern United States (Boersma et al., 2011).

Our previous study (Lin et al., 2014b) for several locations in the North China Plain (NCP) has shown large changes in retrieved NO_2 VCDs when moving from an implicit to an explicit treatment of aerosols."

Second, comparing POMINO against DOMINO is not our main focus. Rather, we have introduced a new product POMINO, evaluate the impacts of different aerosol and surface albedo treatments by perturbing the POMINO algorithm, and conduct other analyses. Fully revealing the systematic bias of DOMINO is out of the scope of this paper.

Third, we believe we have conducted a fair comparison between POMINO and DOMINO. We have treated the pixels properly for various purposes, with clear explanations. We clearly stated in the end of Sect. 2.3:

"There are notable differences in the representation of CRF between POMINO and DOMINO. For POMINO, the CRF represents the fraction of the TOA radiance caused by clouds alone (in the context of additional contributions from the surface and aerosols). For DOMINO, however, the CRF applies to the fraction of TOA radiance caused by both clouds and aerosols, with surface reflectance represented by a geometry-independent surface albedo.

Different retrieval approaches lead to distinctive CRF values, which in turn has consequences for the selection of valid data (Lin et al., 2014b) (see discussions in Sect. 3.5). In Sects. 2 and 3, the pixels designated as "valid" by case REF are selected for analysis, regardless of their validity status in other retrievals. This choice ensures that the same set of pixels is evaluated for all retrieval methods. For the emission constraint study in Sect. 4, different sets of valid pixels specific to the individual retrieval approaches are also analyzed, in addition to the set determined by case REF."

In addition, the whole Sect. 3.5 is dedicated to discuss the implication of different retrieval approaches (not just REF and DOM but also other cases) for the choice of "valid" pixels. In particular, the section has clearly shown that an implicit aerosol treatment tends to miss pixels with high aerosols and NO2 pollution. This finding is also highlighted in the abstract and conclusion.

Furthermore, our comparison results between an implicit and an explicit aerosol treatment based on POMINO-"valid" pixels are generally consistent with the work of the KNMI team over South American (Castellanos et al., 2015).

In contrast, basing the analysis on the "valid" pixels in DOMINO would have missed lots of pixels with high aerosols and NO2 pollution and led to a low-value bias in our analysis. In addition, the overall purpose of this paper is to present and analyze POMINO, thus there is no reason to base the pixels on DOMINO.

4) Some descriptions and discussions are not clearly presented. Some terminologies and acronyms need to be described and explained, especially for readers unfamiliar with Lin et al. 2014b. Organization of Figures (numbering) is confusing. See specific comments.

Response: Please see our specific responses.

Specific Comments: 1) Abstract: Please briefly define/elaborate LIDORT AMFv6, MODIS AOD, OMLER v1, 'subsequently-constrained'.

Response: We have elected not to spell out the full names in the abstract, for conciseness and other reasons as follows. MODIS is a well-known satellite instrument. LIDORT is a well-known radiative transfer model. AMFv6 is our Fortran package for air mass factor calculation, and we have decided to always use the short name in order not to confuse it with the 'air mass factor' quantity.

We have changed the first time us of 'subsequently-constrained' as follows:

"Using POMINO to infer Chinese emissions of nitrogen oxides leads to annual anthropogenic emissions of 9.05 TgN yr^{-1}, an increase from 2006 (Lin, 2012) by about 19%."

2) Line 14 p. 12657: please briefly elaborate 'row anomaly issues'.

Response: We have added that:

"Row anomaly affects the quality of the level 1B radiance data for some viewing directions of OMI (http://www.knmi.nl/omi/research/product/rowanomaly-background.php)."

3) Line 24 p. 12659: please briefly elaborate OMCLD02 v3.

Response: Updated:

"Our cloud retrieval is focused on AMF calculations, starting with the O_2-O_2 SCDs from the official cloud product OMCLDO2 v3 (Acarreta et al., 2004)."

The OMCLDO2 product is well described in Acarreta et al. (2004).

4) Line 10-25 p. 12660: How good are the GEOS-Chem NO2 and aerosol properties, GEOS-5 profiles over China? What is the implication of using a 'relatively' coarser resolution of prior information from GEOS-Chem on the retrieval. Is the retrieval carried out at 'native' resolution of OMI or is this done after gridding to 0.25 and monthly scale? How would this impact the interpretation of your results especially in terms of consistency, variability and errors presented? What is the rationale behind using GEOS-Chem information instead of MODIS?

Response: We have conducted cloud and NO2 retrievals pixel by pixel, as clearly stated in the manuscript. All model information is collected from the grid cell covering the center of a particular pixel. Although the size of our model grid cell is larger than the size of an OMI pixel, our model grid cell size is much smaller than used in other OMI products (3 ° long. x 2 ° lat. for DOMINO [Boersma et al., 2011] and 2.5 ° long. x 2 ° lat. for OMNO2 [Bucsela et al., 2013]).

We have added several sentences in the end of this paragraph:

"As we retrieve clouds and NO_2 pixel by pixel, model information at the grid cell covering the pixel center is used. Although the size of our model grid cell is larger than the size of an OMI pixel, our model grid cell size is much smaller than used in other OMI products (3 ° long. x 2 ° lat. for DOMINO [Boersma et al., 2011] and 2.5 ° long. x 2 ° lat. for OMNO2 [Bucsela et al., 2013]). In addition, we adjust the pressure profile for each pixel based on the difference between pixel-specific surface elevation and grid cell average elevation (Zhou et al., 2009; Lin et al., 2014b)."

In the second paragraph of Sect. 2.5, we have updated that "For a given OMI pixel, aerosol data at the grid cell covering the pixel center are used during the retrieval process."

GEOS-5 is an assimilated meteorological dataset, development of which has incorporated meteorological measurements over China. There are currently no NO2 vertical profile measurements over China available to us, although the simulated NO2 profile has been validated over the U.S. (Lin and McElroy, 2010). In the end this paragraph, we have added that "The meteorological and particularly NO_2 profiles are subject to errors (Boersma et al., 2011; Lin et al., 2012, 2014b). Further research is needed to evaluate these profiles using available measurements over China."

MODIS AOD dataset is subject to missing values, especially at the daily scale. We have used MODIS AOD to constrain model AOD on a monthly basis. Other aerosol information is provided by model simulations. We have fully recognized the limitation of our retrieval, with a paragraph in the end of Sect. 2.5 to discuss the uncertainty and limitation related to aerosols:

"Several limitations constrain our ability to improve aerosol modeling. Model aerosol optical properties (AOD, SSA, phase functions) and vertical profiles are subject to errors (Drury et al., 2010; Ford and Heald, 2012; van Donkelaar et al., 2013). We used MODIS AOD data to constrain CTM-derived AOD, even though MODIS data are not free of errors (Wang et al., 2007; Wang et al., 2010; Hyer et al., 2011). No adequate observations are available to constrain other aerosol optical parameters at a regional scale with high spatial and temporal resolutions. Observation-based estimates of SSA are essentially lacking at the scale considered here, and the few results in the literature contain large uncertainties (±0.03) (Lee et al., 2007). Although the CALIOP instrument provides information of aerosol vertical profiles (Winker et al., 2009), the CALIOP profiles are limited by their spatiotemporal coverage and data quality (especially near the ground) (Ford and Heald, 2012; van Donkelaar et al., 2013). Note that since the same vertical mixing and convection schemes were used to simulate aerosols and NO_2, the height of aerosols relative to NO_2 (relevant to our study) may be subject to smaller errors than the absolute height of aerosols. Future work is needed to better understand and constrain aerosol properties and evaluate how they affect the NO_2 retrieval."

5) Line 15-16 p. 12661. Please elaborate. Is the interpretation of the results on the differences between retrieval methods affected by this?

Response: Please see our response to major comment 3.

6) Line 7 p. 12664. 'large-scale' retrieval. Please elaborate.

Response: We have revised the sentence to:

"Here we emphasize the modifications to POMINO needed to facilitate a large-scale retrieval (i.e., for a large domain in all seasons, as compared to several spot-locations investigated by Lin et al. (2014b))."

7) Order of figure discussion and introduction is confusing. Figure 3 is discussed after Figure 4 and Figure 5 for example.

Response: Figure 3 is mentioned in the 3rd paragraph of Sect. 2.4, prior to Figs. 4 and 5.

8) Line 20-21 p. 12665. Is this something the authors can compare quantitatively with MODIS data?

Response: We do not think MODIS provides a high-quality SSA dataset for such an evaluation purpose.

9) Line 8-9 p. 12666. Similar to comment 8, can this relative uncertainty in NO_2 retrievals due to aerosols be reasonably quantified by using different aerosol fields?

Response: In theory, one could use other aerosol datasets (e.g., from other model simulations) to indirectly infer related uncertainties. However, such work would still be inconclusive, as there is no accurate aerosol information at the temporal and spatial scale with such details (AOD, SSA, phase function, wavelength dependence, vertical profiles…). Such comparison work is also outside the scope of our study here. We note here that we have done various tests with aerosols for several particular locations in our previous work (Lin et al., 2014b).

10) Line 10-15 p. 12666. How is this manuscript different from Lin et al. 2014b, given that Lin et al. 2014b carried out similar sensitivity studies on explicit vs implicit assumption?

Response: Please see our response to major comment 2.

11) Line 23 p. 12667. How would this criterion (on valid pixels) bias your comparison with DOM?

Response: Please see our response to major comment 3.

12) Line 4 p. 12668. Related to comment 11, why would these differences reflect dissimilar AMF approaches given that 'invalid pixels' for DOM may actually be representing this difference?

Response: Please see our response to major comment 3. In addition, "valid" or "invalid" pixels are determined by specific AMF approaches.

13) Line 20-21 p. 12668. Was this resolution (0.05) used in the retrieval or was this regridded to 0.25?

Response: Our retrieval is done pixel by pixel, where all surface reflectance data are mapped to the particular pixel. The post-retrieval analysis is based on 0.25° gridded data, though.

14) Line 11-12 p. 12670. Again, what would be the difference between this study with Lin et al. 2014b?

Response: Please see our response to major comment 2.

15) Line 13-24 p. 12670. Should this discussion be more appropriately presented in the aerosol section or 'coupled' section?

Response: The discussion is about the effect of surface reflectance treatment alone. It should be put here.

16) Line 12-17 p. 12675. What would be the difference between this study with Lin et

al. 2014b?

Response: Please see our response to major comment 2.

17) Line 25-28 p. 12678 and Line 1-2 p. 12679. Please elaborate on the interpretation of 52% error in top-down etc. This is especially important for readers not familiar with Lin et al. 2012. Also, what is the impact of assuming the same errors for all grid cells in your emission estimates?

Response: Lin (2012) discusses the errors in detail, and is referred to for users interested in error estimate details. We have added more error information and acknowledged the uncertainty associated with the use of the same error value for all locations, as follows:

"Following Lin (2012), errors in anthropogenic emissions are taken as 60% for *a priori* and 52% for top-down (for combined errors in model simulations [~40%, Lin et al., 2012; Yan et al., 2014], satellite NO_2 retrievals [~30%, Boersma et al., 2011; Lin et al., 2014b], and emission inversion procedures [~12%, Lin, 2012]). The same errors are assigned to all grid cells, following Lin (2012). This leads to an error of 39% in the *a posteriori* emissions. Although the actual errors may be larger for individual locations, there is no such detailed information for emission constraint."

18) Line 29 p 12679. Please elaborate. Does this mean that there is no significant difference overall? What are the implications for this?

Response: On the country and annual basis, the inferred emissions are similar across the cases. The small difference is due to many compensating factors as a result of significant spatial and temporal averaging, as stated in the original manuscript and further clarified in the revised text. On the monthly and/or locational basis, emissions differ notably between individual cases, as a highlight of our study. On a daily basis, the difference is even larger (not shown). This means that previous estimates on Chinese emissions may be biased, the extent of which depends on the spatiotemporal scales being focused.

19) Line 5-6 p. 12680. What is the rationale for masking low emissions?

Response: Low emissions are masked to highlight the polluted areas. We have updated the sentence.