Anonymous Referee #2

Thank you for your effort and valuable comments on our paper. Our responses are embedded below in blue.

Received and published: 10 August 2017

The paper presents a new method for estimating mega-city emissions from satellite data in combination with a chemical transport model. It goes beyond the method presented by Pommier et al. (2013) where satellite data only were used to estimate emission trends. In general the paper is well written, and I recommend publication after the following concerns have been addressed.

General Comments:

The relatively large differences between the results presented in the manuscript using MOPITT V5 data and those in Pommier et al. (2013) should be discussed more systematically. Are those differences only due to differences in the wind direction (surface – 700 mbar averaged winds at 0.75 deg resolution vs. surface – 750 mbar averaged winds at 1 deg. resolution) as mentioned in P10 line 10? It would help to show the differences in winds to those in Pommier et al. (2013); are those larger for LA where the largest discrepancy in downwind minus upwind total column CO is found? In this context also complex topography or coastal effects should play a role, causing winds extracted from analysis files at different resolution to differ more, or even making the choice of an upwind and downwind region within the complex flow invalid.

Thank you for these remarks. We performed some extra tests to investigate the influence of the differences between our study and Pommier et al. We added a new paragraph to describe other differences between our study and Pommier et al., and the possible influence on the emission trend estimation. We agree that complex topography and coastal effects might also influence the estimation and can be somewhat different between P13 and our study due to resolution differences of the wind data. The point we want to make in this section is that the method is very sensitive to slight differences in the filtered data.

Other sources of uncertainties

Since we used a slightly different pressure level for top of the boundary layer (BL) than P13 to calculate the average wind direction, we tested the sensitivity of the relative difference calculation to the height over which the wind-direction was averaged. For this test we took the average over 12 (low BL), 15 (normal BL) or 18 (high BL) hybrid pressure layers, respectively at an average pressure of 808 hPa, 717 hPa and 613 hPa. The height of the averaging was found quite important in determining the value of the RD. For some cities, the differences were rather small, but for Moscow, Paris, Sao Paulo and Delhi, significant differences were found between the RD values for the calculations using different pressure layers. We found absolute differences of over 20%, and an opposite trend sign for Delhi, where the downwind - upwind difference between the two periods is rather small. Just as was found for the dependence on the location of the rotation point, the downwind-upwind emission estimation values are usually quite close to each other, but the difference between 2000-2003 and 2004-2008 is relatively small compared to the spread in downwind-upwind values of one period, leading to large differences in the RD values, as P13 also described in the supporting information of the paper. From this we conclude that the choice of the height over which the wind direction is averaged is important for the satellite-only technique. Since there is no objective criterion to choose the “best” height for rotating the CO column values, this introduces another systematic source of error that will affect the reliability of the results.

By extending the cloud filtering from data with less than five percent clouds, as we did by filtering on cloud diagnostic 1 or 2, to data with a maximum of zero percent clouds, as in P13, the amount of data is reduced by less than a percent. The emission estimation, however, still changes for some cities. For Paris, the downwind-upwind difference is changing by 27% for the 2004-2008 period. The absolute RD change is around 6% for most cities, although for Delhi a 21% difference was found.
We do not filter MOPITT data for retrievals containing water bodies other than rejecting water and mixed retrievals using the standard MOPITT flags. Since MOPITT is not able to measure CO in the near-infrared over areas with low albedo, such as water, this can lead to biases in the emission trend estimates in our method. For Los Angeles and Sao Paulo, which are both close to the coast, our analysis may include some scenes with fractional areas of water, while P13 filtered these out. This might explain part of the difference in RD estimation seen in Fig. 5, especially for Sao Paulo. As described in the supporting information of P13 also the averaging radius, the size of the grid cells, and the across-wind averaging distance can significantly influence the RD estimation.

As stated later also a slight change in the rotation point, e.g. related to the imperfect geolocation bias correction applied to the V5 data, causes differences; however the rotation points used in the estimate using V5 data should be identical to Pommier et al. (2013) as the same geolocation bias correction was applied to the data.
There should indeed be no difference between our study and P13 on that point because we used the same location and geolocation bias. Still we think it is important to state that a slight difference might cause a significant RD estimation difference. As we describe in Sec. 3.2.4:
This can be an important reason for the differences in emission trends found between V5 and V6. We note that the geolocation bias correction that was used in P13 and our study was slightly different from the correction done for V6 of the data by the MOPITT team (Deeter, 2012). This is a potential source of error since small location shifts can have a substantial effect on the RD estimation.

The role of the background scaling factor should be made more clear, e.g. by explicitely writing the dependence of the modelled column averages (X_mod[i]) on f_backg and f_emiss, as the model is fully linear this should be straight forward.
We added the following equation to make the role of the background more clear:
The X_{mod} is built up from data of the background simulation X_{backg} and the full simulation including emissions X_{emis} according to Eq. 5.

$$X_{mod} = X_{backg} \cdot f_{backg} + (X_{emis} - X_{backg}) \cdot f_{emis} \quad (5)$$

In this context (i.e. in section 2.3.6) also the sensitivity experiments should be introduced, where changes in “WRF’s background emissions” are applied as described in section 3.5.
We added an extra paragraph to introduce the sensitivity experiments directly afterwards:

In order to determine how sensitive our method is to different spatial averaging, different prior emissions and different filtering methods, we performed some sensitivity tests. We tested the optimization with a 10 times coarser grid, i.e., 20x20 km^2 to investigate the sensitivity to the chosen grid size and decrease the importance of patterns in the background and emission. We also used different prior emission

patterns: for 2006 we started the optimization with TNO-MACC-III emissions (Kuenen et al., 2014) for 2002 we did a test optimization starting with emissions of 2006. We also tested the sensitivity to emissions in the direct surroundings of the 200x200 km^2. Extra background simulations were performed in order to quantify this: simulations with emissions outside of the 200x200 km^2 box around Madrid, and, as the normal simulation, without emissions in the urban area where the optimization was performed.
To analyze the robustness of the method, we repeated the optimization using different data filters and investigated the effect of optimizing the absolute difference instead of the quadratic difference in Eq. 4. Four different filtering methods were tested to prevent outliers in the MOPITT data to influence the estimation: 1) Filtering out all MOPITT data that were more than three or 2) four standard deviations from the yearly 200x200 km^2 mean MOPITT CO concentration, or filtering out all MOPITT and WRF data at the same time and location that had a larger difference between them than 3) three (which is the default method) or 4) four standard deviations from the mean difference between MOPITT and WRF at the same time and location.

Appendix: The text for each appendix should include all references to figures and tables included within each appendix. The way the figures are referred to only from within the main text of the manuscript seems to suggest that the figures would be better included in the manuscript itself rather than the appendix. *We agree that some figures are more relevant in the main text, we added Fig. A1-A3 and A5 to the main text.*

Specific comments
Thank you for noting, we changed our text as suggested, except if otherwise stated:
Pg 8 Ln 16: add a period at the end of the sentence *done*

Pg 8 Ln 22: Please add the notion that the r-square value measures the explained spatial variance of the annually averaged column mole fractions (if I got this right).
Yes that is right. We added the following information:
This R^2 value quantifies the fraction of the variance in the MOPITT data that is explained by WRF. We also found a clearly visible enhancement of CO mixing ratio over the city of Madrid for this yearly period.

Pg 8 Ln 32: "both backgrounds" please explicitly state what those two different background fields are.
We changed the description of the backgrounds to make this more clear:
For each year also a background simulation was performed where the boundary and initial conditions are kept the same as in the simulations with emission but where emissions were switched off. The difference between these simulations represents the contribution of the emissions of Madrid to the simulated CO concentrations.

We added the following in the paragraph on sensitivity tests:
Extra background simulations were performed in order to quantify this: simulations with emissions outside of the 200x200 km^2 box around Madrid, and, as the normal simulation, without emissions in the urban area where the optimization was performed.

Pg 9, Ln 13: “to still maximize the available information” this is unclear; why does using column average mixing ratios maximise the information?

We removed the maximize statement and added the following explanation:

Using the column data in molec/cm^2, as done in P13, is not appropriate here, due to the effects of orography that also influence the match between the model and satellite. Instead, the column average CO mixing ratio was used. Note that we do not use the surface layer CO mixing ratio but the total column since the bias, and bias drift, of the multispectral total column product is much lower than that of one or a few layers near the surface (Deeter et al., 2014).

Pg 10 Ln 6: table A2 is referred to before table A1 *we changed the order of presenting the tables*

Pg 10 Ln 35: replace “weighing” by “weighting” *done*

Fig. 5: I suggest to separate the two time periods by colour, and the three different rotation points by symbol shape. This would make it easier to read the figure. *This is how the figure was already, therefore we did not change it.*

Pg 14 after line 20: the line numbering is incorrect, also on the following pages; I will use the indicated line numbers in the following

Pg 15 Table 1: the table needs reformatting, e.g. use shorter descriptions or labelling for the filters applied (column 4) to shorten the table *we removed the long names in column 4 to make the table smaller and clearer.*

Pg 17 Ln 39: 20x20 “optimization method” should be mentioned in the methods section under 2.3.6; why does the change from 2x2 km to 20x20 km have such impact, given the MOPITT resolution of 22 km?

The oversampling technique applied to a year of data is giving a quite detailed pattern of CO mixing ratios over Madrid, since most data are sampled at slightly different locations. Optimization on 20x20 km2 uses 100 grid cells instead of the 10000 grid cells of the 2x2 km2 grid. This leads to some grid cells in the low resolution optimization that include both the areas where emission takes place and where no emission takes place, making it better performing for the background but worse for the ‘transition zone’ between emissions and background which is why it is not surprising that the emission estimations differ.

P18 Ln 16: “changing WRF’s background emissions“ what is meant by that? Section 2.3.6 does not give any clue on what “background emissions” could mean.

We updated the description of the background simulations in section 2.3.6 and added some more explanation in line 16:

To investigate the contribution of emissions outside the optimization area on the pattern in CO in the optimization area, we performed a sensitivity test (sensitivity 1) replacing the normal background simulation, without any emissions, with a background simulation that has emissions in the area outside the 200x200 km^2 optimization area. In the ideal case these "background emissions", i.e., the emissions within the WRF domains around the optimization area, only contribute to the background of the 200x200 km^2 area around Madrid without affecting the city pattern. In this case, it is sufficient to optimize the background with only one factor.

P18 Ln 25: “replacing the normal background simulation, without emissions, with a background simulation that has emissions in the area outside the optimisation area” this seems to be in conflict with the statement in section 2.3.6 (P8 Ln 28-30) where it is

mentioned that emissions outside of the 200x200 km box around Madrid are already used in the standard case.
*It was mentioned in this paragraph that "*Most of the results in this paper are therefore based on the simplest setup for the background simulation: the one without any emissions.*", but we realize the description was not so clear. We now changed the description of the background simulation and added a paragraph to explain the sensitivity tests as explained in the answer to your comments on* Pg 8 Ln 32.

Pg 22 Ln 7: the Jacob et al. (2016) has been published as a final paper *we updated the reference*
Pg 33 29: What is specifically meant by the "oversampling method"? Does that include the rotation of the grid according to wind direction? If so, which wind was taken for the rotation of the WRF grid at each time step, WRF winds or ECMWF winds at 1 deg. as for the MOPITT observations? This needs to be clearly stated so that the reader can follow what has been done. *The oversampling method does not include the wind rotation. We explained it better now:*
For each period the oversampling method was applied to grid both WRF and MOPITT data on the 2x2km^2 grid; no wind rotation was performed.

Pg 33 line 36: "the stability of the model" may be reformulate to "a lack of spatial variability in the model" *thank you for the suggestion; we reformulated the text this way.*
Pg 33, last two sentences: those sentences are repeated from page 8 and should be removed *done*
Pg 35, Fig. A1: The observations seem to have a vary coarse resolution, as indicated by jumps with a step width of 0.1 mg/m3 (corresponding to about 90 ppb). As the background during summer months is about 80 ppb, this resolution seems a bit coarse. -> include in discussion, mention at least *We added the following sentence in the text:*
It should be noted that the resolution of the observations is 0.1 mg/m^3, especially for the background station Villa del Prado, this resolution is close to the absolute value of the measurement (0.1 mg/m3 corresponds to about 90 ppb) and could thus be considered a bit coarse for measuring background concentrations.

Pg 35, caption Fig. A2: Concentrations from only one location are shown, the text should be revised. *We revised the text accordingly, now including only Mostelos.*

Pg 40: values seem to have a second decimal point instead of a +/- *we added the ± sign in the table*