We thank Mark Wenig for his positive review and valuable comments. We respond to each specific comment below. The comments and questions from the referee are in italic font and blue color.

It is not clear to me how you can derive the uncertainty of 20% for the monthly grid cell emissions just by comparing 2 years. Van der A et al. (2016) showed that the average NOx emission over Eastern China stays more or less constant in those two years, but also shows that the different provinces have their peak NOx emissions in different years, so the assumption of constant emissions might only hold on average but not for individual locations. What data are you comparing exactly, daily values or monthly? What starting conditions are you varying? Since with this approach you can only determine the precision but not the accuracy of your results, have you tried to determine systematic errors as well?

The uncertainty we present here is based on the statistics. We first calculated the difference in monthly emissions of each grid cell by comparing the emissions of the two years. The domain includes 15609 grid cells, which means we have 15609 x 12 samples. The average difference is about 20%, which includes the uncertainty and the trend (which is small). We compared monthly emissions of each grid cell in 2012 by running DECSO v5 with different initial emission inventories and starting years. The difference is less than 20% as well. In the same way, we calculated the difference of each province from two different runs (30 x 12 samples), which is less than 2%. We only calculated the precision in this study. But we plan to analyze the accuracy in more detail by comparing different emission inventories in a follow-up study.
We change the text on line 27 page 9:
"... However, we can roughly estimate the precision of the emission based on the year-to-year variability in the derived monthly emissions per grid cell in 2012 and 2013, since there is no significant trend in these two years (van der A et al., 2016). We calculate the average difference in monthly emissions between 2012 and 2013 over all grid cells (15609 grid cells), which is about 20%. We verify the result by comparing the derived monthly emissions for 2012 from DECSO with a run with a different initial emission inventory and starting year. We conclude that the precision is about 20% for each grid cell. We do the same calculations on a provincial level and find that the provincial monthly emissions have a much better precision of less than 2%. "

Would it be possible to use model data to test your algorithm? You could integrate the model output over height to simulate the satellite measurement, add some noise and then apply your retrieval technique. Of course you cannot determine how the model uncertainties affect your emission estimates, but at least you could compare improvement efforts to the algorithm.

We agree that this is a very good strategy to monitor the improvements. This has been done in section 5 of Mijling and van der A (2012). However, we use a different approach for the new version. We tested our improvement by using one grid cell box model with artificial observations. We haven't mentioned this in the paper. Earlier results of this box model were also mentioned in Mijling and van der A (2012).

You mention the precision of monthly emissions, but since you have the word 'daily' in your algorithm name, you might want to refer to the daily emission estimates.

The DECSO algorithm derives daily emissions because we use satellite observations on a daily basis. The response in emission updates, however, depends on the number of daily

observations (after filtering), and the retrieval error. Current retrieval products from instruments like OMI and GOME-2 lack the spatial/temporal resolution nor the accuracy to capture strong day-to-day changes in emission. Monthly emissions have an almost full coverage of the whole domain and are our final product. That's why we analyze the precision of monthly emissions.
To emphasize this, we add the following sentence on line 9 page 7:
"…over East Asia. The final results are monthly emissions for this period. As we showed …"

In Sec. 2 you describe how you filter the observations, so it might be interesting to get to know how much data is left after that.

After the filtering, we have about 2000 observations per day over the domain of East Asia. Note that this has a strong seasonal cycle due to cloud climatology and snow (which for example lowers the response time in winter and rainy seasons.)

We add this information on line 20 page 3:
"… the retrieval product. After this filtering, we typically have about 2000 observations per day for each domain."

It might be helpful to better describe what the variables in Eq. 1 depend on. e^f(t) e.g. looks like it only depends on the current day, but it depends mainly on the previous day, right? Unfortunately I couldn't find the time to read all the referenced papers with the more detailed algorithm description, so maybe more details are given there, but a more detailed in this paper might help. Does the observed NO2 column concentrations vector y only contains observations from the same day or also includes previous days?

Yes, $e^f(t)$ depends on the previous day. We assume a persistent emission model, which means the forecasted emission of the current day is equal to the analysis of the emission from the previous day. The NO2 column concentration vector only contains the observations from the current day. To make it more clear, we change the text below at line 30 page 3:
"… the observed NO_2 column concentrations **y** at time t and the forecasted …"

We also change the text at line 34 page 3:
"….satellite footprint. $\mathbf{e}^f(t)$ is, equal to the analysis of the emissions from the previous day, following a persistent emission model. The Kalman…"

We add more description about DECSO in Section 2 (page 3 line 5):
"The essential part of DECSO is the calculation of the sensitivity of the NO_2 column concentrations (on a footprint of the satellite) to the gridded NO_x emissions, in which the transport of NO_2 over the model domain is taking into account. The transport is calculated using an ensemble of 150 isotopic 2-D trajectories for each grid cell. For the trajectory, we use the operational meteorological forecast of the European Centre for Medium-Range Weather Forecasts (ECWMF) interpolated into half an hour time steps. The inversion method used in DECSO is based on an extended Kalman filter. The emissions and their error covariance derived from DECSO are independent from the a prior emission inventory after a spin-up time of about 3 months."

The sentence starting in line 12 on page 5 is a little confusing. You might want to add 'forecasted' and 'measured' to the sentence ". . .the [] emissions of one day are equal to the [] emissions of the previous day" to be consistent with the equation. Please clarify if the equation and the sentence are only true on average.

The equation is based on an assumption that emissions are the same from day to day within a certain error margin. We could refine our persistency model with more information on biogenic emissions, but this would mean that we add a priori information based on land use, temperature and soil moisture. This would add additional complications and we still would miss for example the building of a new power plant. Therefore, we prefer using a very simple model without a priori information that is able to follow changes on a time scale of days or longer.

We change the sentence into:
"In the DECSO algorithm, we use a persistent emission model, which assumes that the <u>forecasted</u> emissions of <u>the current</u> day are equal to <u>the analysis of</u> the emission of the previous day. "

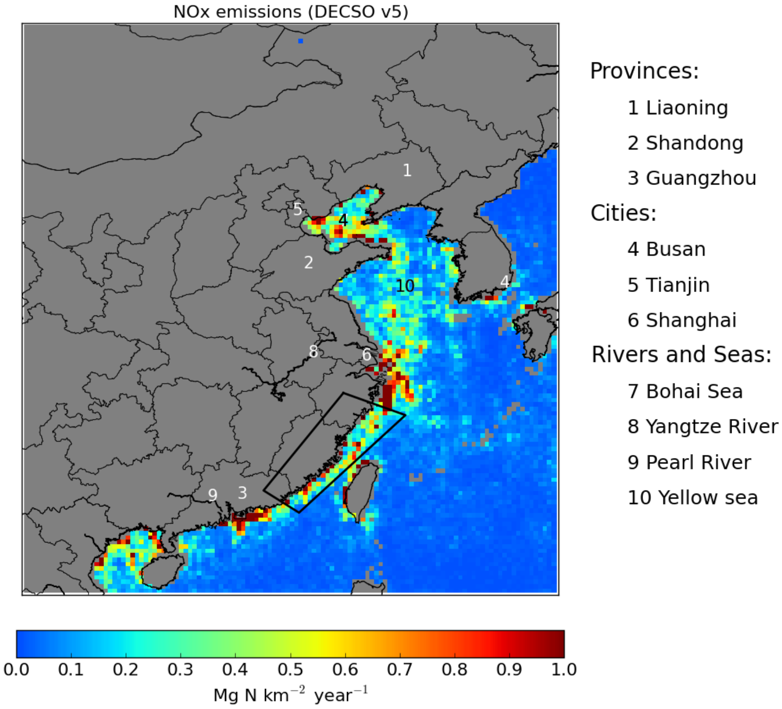
In addition to the correlation coefficients provided in lines 18ff on page 7, you could also mention slope and offset of a linear fit between modeled and measured NO2 columns, because a consistent over- or underestimation is equally important and doesn't show in the correlation coefficient.
We calculate the correlation coefficients with satellite observations on the spatial distribution to show locations of emissions of DECSO v5 are more precise than of the old version. The slopes with and without fixed offset of a linear fit between modeled and measured NO2 columns are both improved in DECSO v5. However, we decide not to present the slope and offset in the paper because the datasets of two versions are not comparable, since we calculate the coefficient following the two steps below:

1. We used the daily modeled columns which are projected in to the footprint of satellite observations and applied with Average Kernel. This dataset is from the assimilation process.
2. We regridded both daily modeled and observed NO2 columns used in the assimilation process on the footprint into the grid cells and calculate the monthly, seasonal and yearly average to calculate the correlation coefficient of spatial distribution. Note that the observation data used here are the data used in the DECSO algorithm, which means the observation data are not exactly the same in two versions. DECSO v3b uses an OmF satellite data filter and DECSO v5 doesn't have the filter but an emission update constraint. So all the outliers of modeled and satellite columns are filtered out if the OmF are large in DECSO v3b.

The further validation of emissions with both in-situ and satellite observations is future work.

Adding the locations of the cities mentioned in the text to Fig. 7 would make it easier for those not familiar with this area of the world to follow the description in Sec. 4.2.

We add the locations of the cities and provinces in Fig 7.
We change Figure 7 in the paper.



Can you provide a correlation coefficient of the emission data from Fig. 7 with the ship location density from Fig. 8? It seems to be quite good for the quadrangle marked in Fig. 7 but not so much in the Yellow Sea where you detect some emissions as well, any idea why that is?

Unfortunately, we don't have the underlying data of figure 8. Therefore we cannot calculate the correlation coefficient of the emission data from Fig 7 with the ship location. We filter the grid cells, which include any part of the land. The ships showing in figure 8 in the yellow sea are mainly at the coast and inland water. We add the description of the filtering on line 24 page 8:
"... derived with DECSO v5. We filter out the grid cells including any part of the land because we cannot distinguish shipping emissions from land-based emissions. ..."

We add the following sentence on line 32 page 8:
"...by DECSO v5 shown in figure 7. Many of the ship locations in figure 8 are close to the coast or on inland water and therefore are not visible in figure 7."

You write that changing the threshold value of the sensitivity matrix H reduced the errors and I'm wondering how you determined the optimal threshold.

We have tested several threshold values of the sensitivity matrix H by comparing the results over some isolated hot spots. When the threshold is too high, we are running into numerical problems. 0.1 is the optimal choice based on these tests.
On line 33 page 6, we add:
"..... In this study we set the threshold value to 0.1 hour based on several tests using different threshold values"

References

Mijling, B., and van der A, R. J.: Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale, Journal of Geophysical Research: Atmospheres, 117, 10.1029/2012JD017817, 2012.

van der A, R. J., Mijling, B., Ding, J., Koukouli, M. E., Liu, F., Li, Q., Mao, H., and Theys, N.: Cleaning up the air: Effectiveness of air quality policy for SO_2 and NO_x emissions in China, Atmos. Chem. Phys. Discuss., 2016, 1-18, 10.5194/acp-2016-445, 2016.