Response to Referee #2

We thank the reviewer very much for their constructive comments. We respond to each comment individually below, followed by changes to the manuscript.

Reviewer Comment: My major comment is that I recommend a coherent section on model and data uncertainties that may affect your analysis and conclusions. Here are some examples of what such a discussion may include:

MERRA meteorological fields: Are there any biases in precipitation or transport that may affect your results, such as through simulated wet deposition? Are there any known biases that change over time in MERRA, such as occur as new observations are brought into the assimilation system over your 20 year simulation period? These are important biases to discuss as global coverage of surface observations (e.g., wet deposition) are sparse over most of the globe.

GEOS-Chem: No model is perfect? Any known issues?

Chemistry: What are the known chemistry uncertainties in the relevant reaction mechanisms? You've answered this with your sensitivity test in Section 3.4.

Emissions: Are there biases? For instance, are the NEI NOx emissions biased?

Author Response: We agree with the referee that there could be a larger discussion on model and data uncertainties that may affect our analysis.

In response to the reviewer's comment, we have added the following discussion section to our manuscript:

"3.5 Other Considerations

A number of other uncertainties are important in an inversion of satellite NO2 columns to calculate surface NOx emissions and simulate of long-term NOy deposition. These can depend on, for example, the choice of inversion approach, errors in the satellite retrieval, and uncertainties in model processes (e.g. emissions, boundary layer mixing, chemical NOx sinks, meteorology, and dry deposition).

Cooper et al. (2017) found that the finite mass balance inversion approach used here can be improved upon by using an iterative method that performs with similar accuracy as a four-dimensional variational data assimilation. Multi-constituent data assimilation also shows considerable promise for constraints on surface NOx emissions (Myazaki et al. 2017). Satellite retrieval algorithms continue to develop with advances that will improve the accuracy of future estimates of satellite-constrained NOy deposition.

Uncertainties in model processes are also of interest. For example, uncertainties in the chemical sink of NOx alone (e.g. the rate of HNO3 formation, heterogeneous loss of N2O5 onto aerosol) can have a substantial impact on top-down emissions estimates (Stavrakou et al., 2013), suggesting more fundamental work in constraining these processes is required. Lin et al. (2010) found that top-down NOx emissions estimates over East Asia are sensitive to other model uncertainties including planetary boundary layer mixing scheme, lightning emissions, diurnal profile of emissions, and a-priori NOx, CO, and VOC emissions. Uncertainties in model meteorology are also important. For example, the MERRA precipitation fields used in our study are known to correlate weakly with observational datasets (Rienecker et al., 2011), but improvements can be expected from MERRA-2 due to the inclusion of gauge- and satellite-based precipitation corrections (Reichle et al., 2017). Finally, dry deposition schemes are also highly variable among models (Flechard et al., 2011; Hardacre et al., 2015), and future work in dry deposition evaluation should be a priority.

Nonetheless, despite these uncertainties we find a high degree of consistency between observations and our predictions in the long-term changes to deposition. Evidence continues to emerge about potential biases in bottom-up inventories (e.g. Travis et al. 2016), and our observational constraint on NOx emissions mitigates against such biases. We expect continued advancements in inversion approaches, satellite retrieval algorithms, and fundamental atmospheric chemistry processes will allow for increasingly accurate satellite-based constraints on deposition."

RC: Section 3.4: How does the model simulation of ammonia compare to observations, such as from AIRS, and the very long record of SO2, such as from the same instruments that you use for NO2

AR: The reviewer raises a good question. GEOS-Chem ammonia simulations have been explored by other groups (e.g. Schiferl et al. 2016), and suggest that the model can underestimate interannual variability and concentrations of ammonia. A new NH3 emission inventory is now available (Paulot et al., 2014), but this inventory is still only representative of a short period of time (2005-2008), thus potentially introducing similar errors in other years of our analysis anyway. Our perturbation experiment with NH3 emissions (Section 3.4) is an attempt to investigate the dependence of NOy deposition on NH3 emission uncertainties. We found that reasonable uncertainties/changes in NH3 emissions (~25%) are generally inconsequential to the spatial distribution of predicted NOy deposition over land. We did not feel the need to repeat such an experiment with SO2 emissions, since SO2 emissions tend to be less uncertain than NH3 emissions and will therefore have even less of an impact on the predicted spatial distribution of NOy deposition. We argue that an evaluation of GEOS-Chem simulated NH3 and SO2 is thus beyond the scope of our study, and would not have substantial bearing on the conclusions in our manuscript.

RC: It is no easy task to create an inter-consistent long-term data record using multiple satellite observations, so this topic deserves some discussion. What are the uncertainties and potential biases? For example, a priori vertical profiles change over time

AR: We agree with the reviewer that creating a long-term record across multiple satellite records is not necessarily a trivial task (a similar concern was raised by another reviewer). In our case, we concluded that a consistent time series can be achieved without requiring additional corrections. The reason for this is largely because the daily satellite observations were all gridded to a regular coarse grid of 2° x 2.5° latitude by longitude. Using a comparison with long-term ground-based MAX-DOAS observations, Irie et al. (2012) demonstrated that there is no inherent biases in either SCIAMACHY or GOME-2 that would preclude their combination into a single record. The work of van der A (2008) and Konovalov et al. (2010) show that a self-consistent record can be achieved by downgrading the spatial footprint of the higher resolving instruments (e.g. through smoothing or convolution) to that of the lowest resolving instrument. This is what we have achieved by gridding all the observations to 2° x 2.5°. The combination of these observations is also aided by the fact that the retrieval algorithm for obtaining tropospheric NO2 column density from all three instruments is the same (http://www.temis.nl/airpollution/no2.html), and by the fact that their overpass times are similar (between 9:30 a.m. and 10:30 a.m. local time).

We tested our approach by inspecting the timeseries from individual 2° x 2.5° pixels over selected populated regions. These are shown in Figure D1.

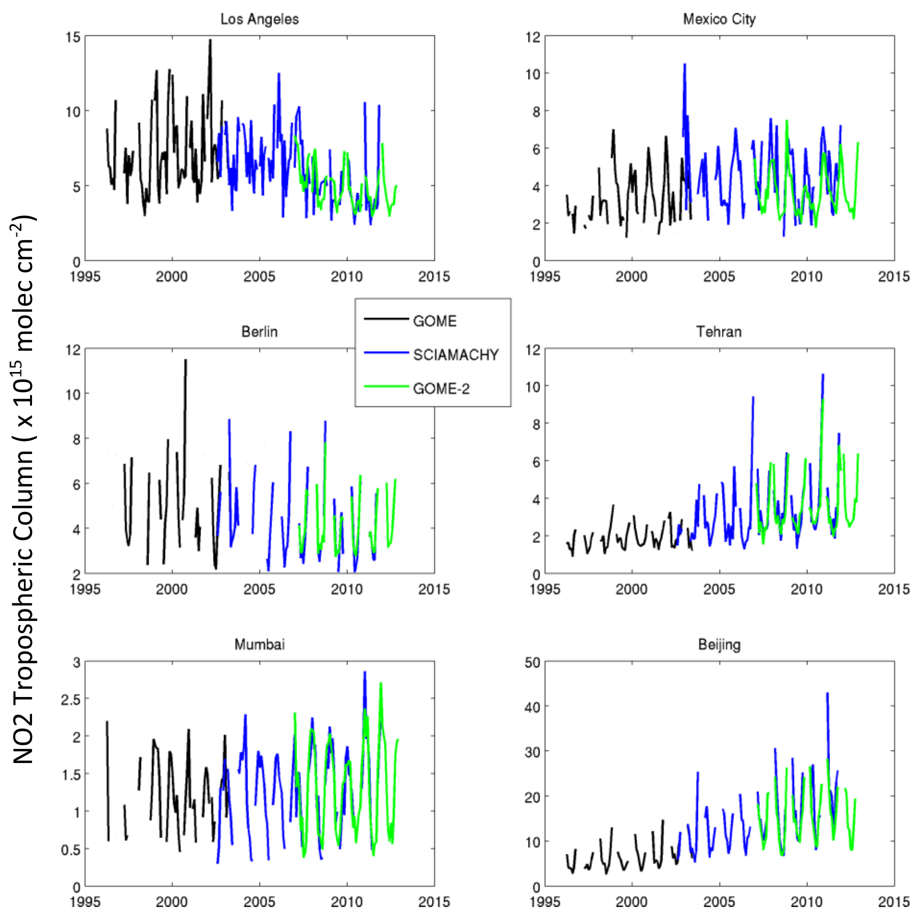

Figure D1: Monthly mean tropospheric NO2 from GOME, SCIAMACHY, and GOME-2 show consistent agreement during overlap between instruments.

Given this and the consensus in the literature, we conclude that the instrumental differences between each satellite instrument are inconsequential to our analysis.

In response to the referee's comment, we have modified our manuscript to include additional citations, and to elaborate on our reasoning for combining the satellite instrument records despite their instrumental differences:

"We calculate top-down surface NOx emissions from 1996 to 2014 using observations from GOME (1995-2003), SCIAMACHY (2002-2011) and GOME-2 (2007-). The similar overpass time of these three instruments (from 9:30 a.m. to 10:30 a.m. local time) facilitates their combination to provide consistent long-term coverage (van der A et al., 2008; Konovalov et al. 2010; Geddes et al., 2016; Hilboll et al., 2013). We achieve consistency across all three instruments despite their varying pixel sizes (320 km x 40 km, 60 km x 30 km, and 80 km x 40 km for GOME, SCIAMACHY, and GOME-2 respectively) by gridding the daily observations from each to a regular coarse grid of 2° x 2.5° latitude by longitude."

The referee also brings up the issue of changing a-priori vertical profiles over time. We agree that vertical profile shapes are an integral component of the tropospheric NO2 column retrieval, and indeed are a significant contributor to column error (at least ~10% according to Boersma et al. (2004), and up to ~100% according to Laughner et al. (2016)). However, we are limited by the current generation of available global tropospheric NO2 retrieval algorithms. Advances in satellite-derived tropospheric NO2 column retrievals will be integral to improvements in top-down emissions estimates. In response to the reviewer comment, we address this in modifications we have made to the manuscript:

"The error in individual satellite-derived tropospheric NO_2 column retrievals to be around 35-60% for polluted scenes and greater than 100% for clean regions (Boersma et al. 2004)."

"We expect continued advancements in inversion approaches, satellite retrieval algorithms, and fundamental atmospheric chemistry processes will allow for increasingly accurate satellite-based constraints on deposition."

RC: Page 4, Line 9: Since the topic of this Nowlan paper is similar and from the same group, it may be worth a sentence describing the major conclusion of this paper and how your manuscript is different/better. In fact, you may want to do briefly so the same for the other papers mentioned in this same paragraph.

AC: We thank the reviewer for this suggestion. We agree that it would be insightful to summarize the important points from Nowlan et al., although the statistical approaches taken by the other examples vary substantially from our approach and are less directly comparable. In response to this comment, we have included the following in our manuscript:

"Nowlan et al. (2014) demonstrated how satellite-inferred surface concentrations of NO2 can be combined with modeling to produce spatially continuous estimates of NO2 dry deposition fluxes. They found that dry deposition of NO_2 contributes as much as 85% of total NOy deposition in urban areas, but only represents 3% of global NOx emitted. The remaining 97% of global NOy deposition is made up of both wet and dry deposition of other reactive nitrogen oxide compounds that are not directly observed by satellite-based instruments."

RC: Figure 2. The two rows of plots look identical. Is there any way to show the differences between the two periods. If not, I'm not sure it's helpful to show both rows.

AC: We are very grateful to the reviewer for their careful attention to this figure. In fact, the two panels are identical and this was an error in the original figure (both top and bottom accidentally show results for 2000-2002). We have corrected this mistake and sincerely apologise for the error. A new version of this figure will be uploaded with the corrected manuscript, and is included here as Figure D2 for the reviewer's reference. The differences between the two time periods are now more obvious. Declines in wet nitrate deposition over this period are evident over North America and parts of Western Europe, while increases are evident over East Asia. Note that the observational coverage is also slightly different across the two periods.

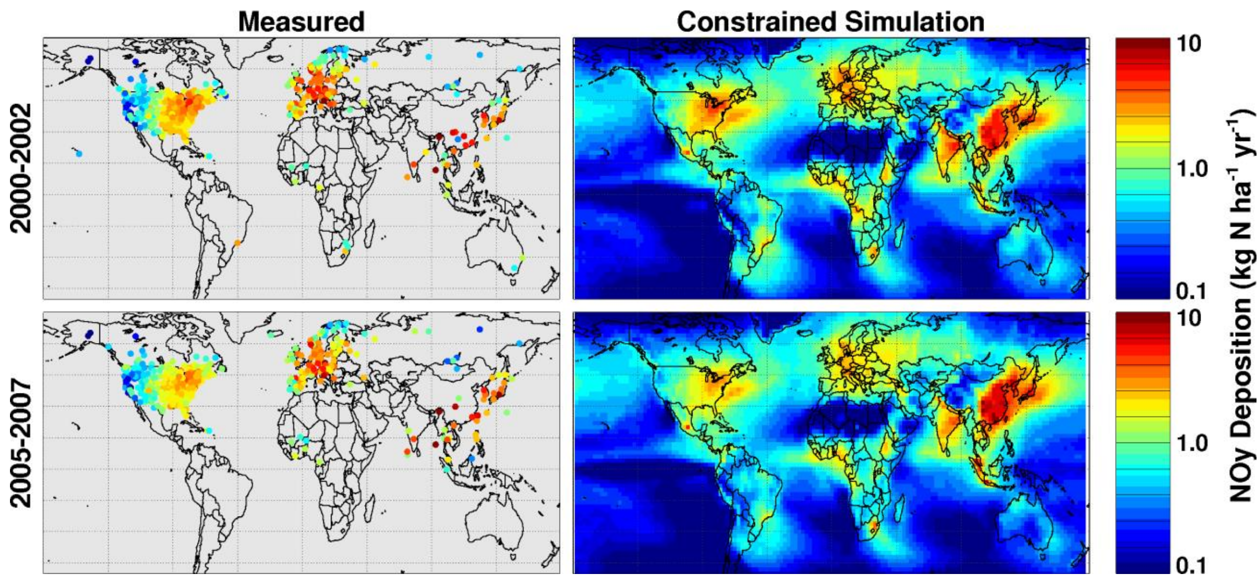

Figure D2: Corrected version of Figure 2 from the manuscript, showing results from the 2000-2002 comparison in the top panels, and results from the 2005-2007 comparison in the bottom panels.

REFERENCES:

Benkovits et al. (1996), Global gridded inventories of anthropogenic emissions of sulfur and nitrogen, *Journal of Geophysical Research-Atmospheres,* 101, 29239–29253

Irie et al. (2012), Quantitative bias estimates for tropospheric NO2 columns retrieved from SCIAMACHY, OMI, and GOME-2 using a common standard for East Asia, *Atmospheric Measurement Techniques*, 5, 2403-2411

Konovalov et al. (2010) Multi-annual changes of NOx emissions in megacity regions: nonlinear trend analysis of satellite measurement based estimates, Atmos. Chem. Phys., 10, 8481–8498

van der A et al. (2008), Trends, seasonal variability, and dominant NOx source derived from a ten year record of NO2 measured from space, *Journal of Geophysical Research-Atmospheres, 113, doi:10.1029/2007JD009021*

Zhao et al. (2017), Atmospheric nitrogen deposition to China: A model analysis on nitrogen budget and critical load exceedance, *Atmospheric Environment,* 153, 32-40

Lin et al. (2010), Constraint of anthropogenic NOx emissions in China from different sectors: a new methodology using multiple satellite retrievals, *Atmos. Chem. Phys.*, 10, 63–78

Stavrakou et al. (2013), Key chemical NOx sink uncertainties and how they influence top-down emissions of nitrogen oxides, *Atmos Chem Phys*, 13, 9057-9082

Rienecker et al. (2011), MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, *Journal of Climate*, 24, DOI: 10.1175/JCLI-D-11-00015.1

Reichle et al. (2017), Assessment of MERRA-2 land surface hydrology estimates, *Journal of Climate*, 30, DOI: 10.1175/JCLI-D-16-0720.1

Flechard et al. (2010), Dry deposition of reactive nitrogen to European ecosystems: a comparison of inferential models across the NitroEurope network, *Atmos Chem Phys*, 11, 2703-2728

Hardacre et al. (2015), An evaluation of ozone dry deposition in global scale chemistry climate models, *Atmos Chem Phys*, 15, 6419-6436

Schiferl et al. (2016), Interannual variability of ammonia concentrations over the United States: sources and implications, *Atmos Chem Phys*, 16, 12305-12328