**Response to anonymous referee #2**

*This paper uses a multi-inversion hierarchy to derive top-down constraints on N2O emissions for 2011. The goal is to make a detailed evaluation of the 3 different methods and their impacts on inversion results. All methods are based on the adjoint of the GEOS-Chem chemical transport model, where 4D Var is considered the "standard" approach, as well as two alternative ways for aggregating the results, given that the existing observational network is insufficient to fully constrain $N_2O$ emissions at the gridscale level. The first approach uses the 4DVar method, but aggregated to the traditional 6 continents and 3 oceans. The more novel approach tested is the new SVD-based technique based on the "prior- preconditioned Hessian of the 4D-Var cost function." An additional goal is to address the impact of initial condition uncertainties using 6 different approaches. This analysis is performed first and an optimal approach is selected for use in the evaluation of the 3 different inversion methods.*

*The paper is well written and logically organized. While some of the mathematics, particularly the SVD approach, are beyond my ability to evaluate, I found the results and discussion interesting and insightful. My main criticisms are, first, there seems to be a predisposition to claim the SVD results as the "best estimate of the true global flux." This conclusion is not clearly based on objective criteria. Other interpretations that might be more critical of SVD are not discussed, including the odd, spiky SVD results (e.g., in South America, Africa and the Tropical Oceans in Figure 7). Second, there is an unwarranted emphasis on the results of Chen et al. 2016, which are often presented as though they were primary results of the current study (see further comments below). However, these are minor criticisms of what is overall an impressive and interesting body of work. I recommend publication with some relatively minor revisions detailed below.*

**We wish to thank the reviewer for their positive evaluation of our manuscript. Please find our responses to specific comments below, where the comment is in italics and our response is in bold.**

*Abstract L31-32 "the inversions reveal a major emission underestimate in the US Corn Belt (which may extend to other regions), likely from underrepresentation of indirect N2O emissions from leaching and runoff. Please clarify an underestimate relative to what? Also, the last part of this sentence is supported only on p12L30 with a reference to Chen et al. 2016. It is not supported by the current study and does not really belong in the abstract as a primary new finding.*

**We have deleted the reference to leaching and runoff here, and clarified that the underestimate is in the prior bottom-up inventory used.**

*As an aside, I will make a few comments about Chen et al. 2016, which is referenced multiple times (e.g., again on P17L27) as the source of the conclusion that the underestimate of indirect emissions is responsible for the underestimate of agricultural emissions in prior inventories. Realistically, I don't think the Chen et al. methodology is able to separate indirect and direct* *emissions. Their prior direct agricultural source is based on EDGAR, which is at least somewhat reliable since it is computed using gridded N inputs from fertilizer, etc. multiplied by emission coefficients. In contrast, the indirect source is based on the CLM45-BGC nitrate leaching and runoff flux, which is unreliable and almost certainly wrong (see, e.g., Houlton et al., Nature Climate Change, 5, 398, 2015). The Chen methodology then assumes those 2 prior sources accurately represent the spatial and temporal distribution of direct and indirect N2O emissions, respectively. That methodology is fraught with uncertainty. Moreover, the fact that (as stated on p16L28) indirect emissions peak earlier than direct emissions is a red flag that something is wrong. This result doesn't make sense, given that indirect emissions, by IPCC definition, occur later and downstream/downwind of direct emissions.*

**We would like to clarify a few points in the above comments on the Chen et al. (2016) paper. First, the Houlton et al. (2015) paper cited by the reviewer used the CLM-CN coupled model, which has a different solution for nitrate leaching and runoff than the CLM45-BGC model used in the Chen et al. (2016) paper. Second, the Chen et al. (2016) methodology does not assume the temporal distribution of a priori emissions is correct, as they solve for monthly fluxes. Third, a more recent paper (Griffis et al., PNAS, 2017), obtains very similar seasonality using different a priori emissions, which supports the Chen findings. However, Chen et al. (2016) is not being reviewed here and so we focus the rest of our response on issues relevant to our manuscript.**

**Based on the IPCC definition: "Indirect pathways involve nitrogen that is removed from agricultural soils and animal waste management systems via volatilization, leaching, runoff, or harvest of crop biomass", so there is not an indication of seasonality here. Indirect emissions in the US Corn Belt are high in April-June when tile drainage and stream discharge peak. Additionally, many farmers in the US Corn Belt apply fertilizer in the fall, which would serve as a source of nitrogen to be released in the spring. As such, we have added the following sentence to the end of this section: "Fall fertilizer application is also common in the US Corn Belt—more than one third of corn farmers in Minnesota do their main N application during this time (Beirman et al., 2012)—which could explain the October peak in the SVD-based results, and provide a source of nitrogen that would be released in the early spring thaw and subsequent runoff period." We also added the following sentences in Section 4.3.1 to indicate that different processes (beyond just leaching and runoff) could be driving the overall underestimate of emissions in this region: "However, other processes could also contribute, such as freeze-thaw emissions or direct emissions after spring fertilizer application. The timing of these processes, and that of peak stream flow, correspond to the dominant modes of ambient $N_2O$ variability observed in this region (Griffis et al., 2017)."**

*P2L24 Crutzen et al., 2008; Davidson, 2009 are not really bottom-up emissions. They are based more on a top-down approach (in a global box model sense) of comparing the observed atmospheric N2O increase to the rate of external N inputs and anthropogenic N fixation.*

**Thank you for the clarification. We have deleted the term "Bottom-up" from the beginning of this sentence.**

*P3L10-12 It seems somewhat overcritical to say previous aggregation has been informal and ad hoc. It's been based largely on geographical and political boundaries, i.e. North vs. South America, Pacific vs. Atlantic Ocean, etc., which are logical regions of interest.*

**Good point. We have replaced the phrase "in an informal ad-hoc way" to "based on physical or political boundaries".**

*P3L10-18 Exact totals are given for the ocean, GFED and EDGAR non-agricultural sources, but the Saikawa non-agricultural source and the EDGAR agricultural source are not specifically stated, yet these are the largest component sources. Please report them too. Two additional points are that the Saikawa source was based on a global model without cropland, such that it included a "non-agricultural" soil source from land such as the US Midwest where crops are grown. Also, the EDGARv4.2 total is about 1.7 TgN/yr from industry, waste water and energy. To bring up the reported 2.3 TgN/yr, I wonder if the authors have included the EDGAR savanna, forest, grass, and agricultural fire fluxes (of 0.84 Tg N/yr), which might be redundant with the GFED source? (Note: my numbers are from 2005 and thus may be slightly different from 2008.)*

**Thank you for catching this. We did not include the EDGAR fire sources in our 2.3 Tg N $yr^{-1}$, but accidentally included indirect emission from $NO_x$ and $NH_3$ deposition (~0.4 Tg N) and manure management (~0.2 Tg N) in this total rather than in the reported agricultural source total. Thus, the EDGARv4.2 total for 2008 is about 1.7 Tg N $yr^{-1}$ as you noted. We now report specific totals for the Saikawa source (7.5 Tg N) and EDGAR agricultural soil direct+indirect (3.5 Tg N $yr^{-1}$) and manure management sources separately.**

*P4L29 Please state the time resolution of the inversion somewhere around here.*

**We have added "monthly" to the first line of this paragraph.**

*P5L19 Negative emission scaling factors may be appropriate for some oceanic regions, especially during seasonal cooling in regions where the biological source is small and thermal solubility-driven uptake may dominate the air-sea flux.*

**We do have negative fluxes where uptake dominates the air-sea flux in our a priori oceanic emissions. Our inversion approach does not require positive fluxes, but it does assume that the sign of the a priori flux is correct in each grid square. We now mention this explicitly in the text here.**

*P5L27-29 Can we infer from this that the total observational uncertainty (which is also referred to as model-data mismatch uncertainty) is typically about 0.45 ppb? It would be useful to state this. It is interesting and unexpected that the observational uncertainty dominates the model representation uncertainty. At only 0.2 ppb, the model representation error seems substantially underestimated. Also, considering that the grid resolution is 4x5 degrees, how many grid boxes actually "surround" any given observation and what kind of heterogeneity is missed inside the actual box that contains the measurement?*

**Correct, this would correspond to a mean observational error of ~0.45 ppb, which we now mention in the text. We also note that values extend up to ~4 ppb. Given the coarse**

**horizontal resolution, we could be underestimating the representation error for near-source observations. However, we have since run a test standard inversion with tripled observational error and get very similar results (global flux of 17.8 Tg N).**

*P7 The SVD-method is complex to the point of being unfathomable for many readers (including me!), so we must take it on faith that the calculation is accurate. Given the lengthy form of equation 5, I am concerned that it would be easy for human errors to slip into the calculation. What assurances do we have that such errors will be detected?*

**Three of the co-authors have rechecked the equations for accuracy, and no errors have been detected.**

*P9-10 I found this section difficult to follow and did not emerge with a clear understanding of why certain initialization methods are better than others. I'm not sure what to suggest to help clarify, but one step might be to include some columns for the NH-SH gradient in Table 2 in addition to (or perhaps instead of) the separate NH and SH bias columns. Those are not really referred to in the text, while the "overly strong interhemispheric gradient" is mentioned on P9L26 but is not obvious in Table 2.*

**We have tried to clarify here that an overly strong interhemispheric gradient is indicated by the fact that the model has a high bias in the Northern Hemisphere and a low bias in the Southern Hemisphere. We have also removed the subsequent reference to the interhemispheric gradient and replaced it with the following sentence: "The interpolation methods without subsequent spinup (AprZonal, AprKriging) perform better in terms of initial model:measurement bias – in the global mean and in each individual hemisphere."**

*P11L18 The statement that SVD "appears to provide the best estimate of the true global flux" seems based on fairly limited and/or subjective criteria. Furthermore, it is not obviously true that SVD agrees best with HIPPO. In fact, it seems to agree worst from 30S-30N. (This is attributed on P11 to the fact that "the spatial distribution (in the tropics) is particularly difficult to resolve," but this is not necessarily a satisfactory explanation.) Is the comparison to HIPPO based on subjective visual inspection or some more quantitative measure? Also, are we sure the HIPPO calibration scale is not systematically biased from the data used in the inversion, especially given the adjustments described in section 2.3?*

**We have now deleted the claim that the SVD-based inversion provides the best estimate of the true global flux. We also outline more specifically where/when the agreement with HIPPO is improved. The sentence now reads: "It also gives a better comparison to HIPPO IV and V measurements in the southern extratropics and to HIPPO V in the northern extratropics (see below)." We have also edited the last two sentences of Section 4.1 to read "The lower global flux obtained with the SVD-based approach (Fig. 3 and Table 3) is thus the reason for this correction, implying that the global annual a priori flux (from all sources combined) may be too high. We note that a slight low bias does emerge in the tropics in the SVD-based approach, where observational constraints are low." We have also edited the language in the conclusions to be consistent with this. As for the calibration,**

**we have adjusted the HIPPO QCLS data based on concurrent flask observations, which are on the NOAA scale. We now mention this at the end of Section 2.3.**

*P13 section 4.3.3 The results for Europe indicate a fairly dramatic reduction from the prior. Please state the total non-agricultural prior source in EDGARv4.2. How much of the total 1.70 Tg N prior does it comprise?*

**We had already included the contribution of EDGARv4.2 non-agricultural sources to the total in Europe (~40%) in Section 4.3.3. However, since we accidentally lumped manure management and indirect emissions from $NO_x$ and $NH_3$ deposition in that total, we have revised this number. The total European non-agricultural source in EDGARv4.2 is about 0.5 Tg N, which is about 30% of the total prior emissions here. The resulting relative a posteriori adjustments for soil and non-agricultural sources, when integrated over Europe, are thus comparable in magnitude, and we have edited the text to reflect that.**

*P14L31 The results of 3.35-3.48 are above the range found by Buitenhuis (2.4 +/- 0.8).*

**Correct, we already note at the end of this sentence that our optimized oceanic fluxes are higher than that found by Buitenhuis. However, they are closer to that estimate than the results of Thompson et al. (2014), which is what we meant by "more consistent with". We have now clarified this in the text.**

*P17L20 It's perhaps notable here that the EDGAR industrial source has dropped by about a factor of 2 in version subsequent to v4.2 used here.*

**Thank you for mentioning this. We have added a note mentioning this at the end of this bullet point in the text.**

*P17L23-24 I think the main issue here is that the seasonality in the existing inventories used here is governed by natural soil emissions from a model without crops. The EDGAR agricultural source with no seasonality is then added. However, the hotspot of emission is in agricultural areas where the seasonality is influenced by spring fertilizer input. Thus the seasonality of existing inventories is predictably wrong from the outset.*

**The reviewer is correct that we should expect some degree of seasonal bias given that annual EDGAR fluxes were used a priori. We have now deleted the phrase "than our current inventories suggest" from this sentence, and emphasize that the optimized seasonality is consistent with other studies.**

*P31 Figure 3 caption. Do the bars show the median (as currently stated) or the mean of the 3 inversions? How meaningful is the median of just 3 values? Would it be better to just show all 3 results + prior, i.e. 4 bars per region?*

**The bars do show the median of the a posteriori values as stated—the median was chosen so it is easy to infer all three values (min, median, and max) from the figure. We previously tried plotting 4 bars per region but found the plot too busy, so prefer to keep the two bars per region, but we have thickened the a posteriori error bars to more easily see the range.**

*P34. I'm not sure Figure 6 adds much value to the paper. Furthermore, why is the KCMP measurement of primary interest to the current study?*

**KCMP is of primary interest to our study given that i) it is the only site with a near-persistent model underestimate, ii) it is in an agricultural region comprised of drained lands (now mentioned at the end of Section 4.4.1), and iii) the emission processes for this ecosystem type are not well represented in current emission inventories, and have been linked to underestimated indirect $N_2O$ emissions associated with leaching and runoff. As for the value of Fig. 6, we find it helpful to show the seasonal model biases in $N_2O$ mixing ratio that exist before showing the a priori and a posteriori seasonal fluxes in Fig. 7.**