Reply to Referee 2

We thank the reviewer for his/her very useful and constructive review and respond to each of the points raised below.

This manuscript presents a comparison of five models, meteorology, and inversion frameworks in estimating nitrous oxide emissions. The objectives are to quantify N2O emissions globally and regionally and to identify the potential causes of discrepancies between models/inversion frameworks. The paper is well-written and reaches important conclusions, for example, about the representation of stratosphere-troposphere exchange in models and the shift in emissions from the tropics to mid-latitudes, and should be published in ACP.

1. General comments

1. I do not understand some of the results, particularly from the MOZART model. The model mole fractions simulated by the posterior fluxes do not appear to match of observations (as shown in Figures 2,3, A1, A2, A3). If the inversion is doing what it should, incorrect fluxes may be derived but the simulated mole fractions would come close to the observations (at individual stations as well as in the growth rate). Is there a problem with the inversion? The authors state that there is a problem with having too high initial conditions, but this should be "absorbed" into the derived emissions. While the degree to which the emissions are tied to the prior is governed by the prior emissions and prior uncertainty, the results still indicate that the inversion did not reach meaningful conclusions (i.e., poor validation as there is not a good fit to the observations). If the derived fluxes are then incorrect, this would also significantly affect the posterior simulation of spatial distributions and seasonal cycles – therefore, it is not clear to me why the model can be excluded for emissions estimation but used for other analyses, which are still based on the posterior fluxes. In any case, would it not be straightforward for the inversion to also solve for the initial condition as part of the inversion to minimise these issues?

Concerning the MOZART4 model, we agree that this model has an offset with respect to the observations, of approximately +1 ppb, which is due to too high initial conditions (as is also stated in the paper and is the reason why we do not include the derived fluxes from MOZART4 in the median flux estimate). While what the reviewer states is true, i.e. that the too high initial conditions will be "absorbed" into the derived fluxes, the time taken for the modeled mixing ratios to match the observed ones will be governed by the lifetime of N2O (which is long, circa 120 years), the prior uncertainty assigned to the fluxes and the uncertainty assigned to the observations. Therefore, it is not necessarily the case, that within the period of inversion that the modeled mixing ratios will exactly match the observations. It is also true, that it would be relatively straight-forward to solve for the initial conditions as part of the inversion, however, this was not done in the MOZART4 inversion. We emphasize also that the purpose of the study was to examine current inversion frameworks that estimate N2O and identify their problems and understand the differences between them, and this is exactly one problem that was identified.

2. The authors do mention that each CTM could be used with any inversion framework but it would be good to stress in the conclusions that the results do not *necessarily indicate issues with the model specifically but the combination of the two as well as other parameters (e.g., type of observation assimilated – monthly, weekly, etc, as well as the choices in prior and observational uncertainties).*

We agree that the differences between models are also sensitive to the resolution of assimilated observations (monthly, weekly, daily etc.) as well as to choices for the prior and observation uncertainties. However, the differences in the spatial distribution between the posterior fluxes are likely to be due to a significant extent also to transport differences, albeit that the magnitude of the differences also depends on the aforementioned inversion parameters. For example, TM5-I, has the highest emission estimate for the SH tropics, which can be understood in that TM5 has the longest inter-hemispheric exchange time of all the CTMs, while LMDZ4-I has the lower than prior estimated emissions in the SH tropics but highest emissions in the NH tropics and sub-tropics and correspondingly, a short inter-hemispheric exchange rate (see section 3.2.1).

We feel that the summary and conclusions already reflect the dependence on the combination of the inversion framework and CTM as in the following sentence (from the Summary and Conclusions):
"Moreover, we have identified emission patterns that are robust, that is, common to all inversion frameworks as well as those that depend strongly on the modelled transport and/or inversion set-up."
(This sentence is followed by a list of the salient results).

3. Does the general uncertainty analysis only include the spread in the models or does it account for uncertainties from the inversions (for example, on a regional level, the uncertainties derived in each inversion could be large but the spread small)? Please make this more clear in the text.

We have reported the median and MAD values for the emissions from all models (except MOZART). These values, therefore, do not reflect the uncertainty for each individual model. Not all inversions are currently able to report uncertainties, e.g. some of the 4D-var models. In LMDZ4-I the uncertainties are calculated using a Monte Carlo approach. The uncertainties for LMDZ4-I are, in general, higher than the derived MAD values for each region indicating that the uncertainty in the inversions (if the LMDZ4-I uncertainties are representative, which we consider them to be) is larger than the spread.

We have now added the following sentences to section 3.2.2 and added the LMDZ4-I regional uncertainties to Table 7:
"Figure 9 shows the annual mean total emissions for 7 sub-continental and 3 ocean regions from each of the inversions and the prior, in addition, the corresponding range, median, and MAD of the emissions, as well as the uncertainty calculated from a single inversion model (LMDZ4-I), are given in Table 7. The calculated uncertainties per region are larger than the corresponding MAD values indicating that the spread of posterior emissions is smaller than the uncertainty calculated for a single inversion."

4. All figures and text should be made clear that simulated values for each model come from the posterior fluxes.

We state this in section 3.1.1:
"Figure 2 shows the observed annual mean meridional gradient (2006 to 2009) compared with simulations by each CTM integrated with the corresponding posterior fluxes."
and section 3.1.2:
"Figure 4 shows the annual mean (2006 to 2009) seasonal cycles from the posterior model simulations and observations at 6 key background sites."

We have also now included this in the captions of Fig. 2, 3 and 4.

5. It would be useful in the figures showing comparisons to observations to also show the observational uncertainty (e.g., Figs 3, A1, A2, A3). For example, in Fig A3, the LMDZ growth rate is also a bit low but it might be within the observational uncertainty so that would be helpful to see.

We have now included the observational uncertainty in Figs. 3, A2 and A3. Note that we do not show this in A1 as with the number of points it would not be possible to see the error bars.

Specific comments:

1. Page 5273 Line 24 The citation of Forster et al. 2007 for N2O being the third most potent greenhouse gas is not correct (it has it as the fourth). There is updated work showing N2O surpassing CFC-12 (e.g., NOAA Annual Greenhouse Gas Index).

We have updated this to the Fifth Assessment Report, 2013.

2. Page 5277 Line 26 How does each model/inversion choose the prior and measurement uncertainties?

We have added a description of how each model calculates the prior flux uncertainty to a new section, section 2.2.3, and changed the former section 2.2.3 to 2.2.4. For the measurement uncertainties, each model uses the estimate given by the data providers (see section 2.3). In addition to the measurement uncertainty, each model estimates a model-representation uncertainty also in the observation space (described in section 2.2.3).

3. Page 5280 Line 16 Even measurements from the same station could have systematic errors when calibration gases are changed (step changes in time series).

Before using the observations, all timeseries' were checked for step-changes and outliers. If step changes were found, then the timeseries was not used. The data from the in-situ sites (where this could potentially be a problem) are all of a very high quality and the calibration is regularly and systematically checked and the data corrected if necessary. Therefore, we do not think that such site-internal calibration errors are a problem in this study.

4. Page 5280 Line 22 Why do only certain models/inversion frameworks have the capability to solve for inter-calibration offsets? Couldn't all models/inversion frameworks solve for these offsets as additional parameters?

Yes, such an offset optimization could in theory be included in all inversion frameworks. We did not insist that all models optimize the offsets, but all models without this optimization did use observations that had been corrected using the prior estimates for the offsets. The reasons we did not insist on this are 1) TransCom is a voluntary community that receives no funding, therefore, TransCom model inter-comparisons rely on to a large extent on people's free time and 2) we wanted to present an inter-comparison of the existing models as they currently are in order to determine the spread of the models as well as to understand what causes this spread.

5. Page 5281 Line 11 Be more clear about where this R^2 value was calculated from (is this comparing the mean of 2006-2008, all of the models, etc.)?

We mean that the R-squared was equal to or greater than 0.92 for all models, in other words, none of the models had R-squared less than 0.92 for the correlation of the modeled versus the observed gradient. We have tried to make this clearer.

6. Page 5281 Line 11 If the initial conditions are incorrect, this will also affect the gradient (see general comment above).

It is true that a north-south bias in the initial conditions (if these are not optimized in the inversion) could affect the north-south gradient of the posterior simulated mixing ratios. In MOZART4-I the bias in the initial conditions appears to be fairly constant across latitude and so the simulated gradient appears not to be greatly affected by this problem.

7. Page 5283 Line 11 The notation of STT is suddenly used, I think, instead of STE. Is STT meant to be something different (if so, it is not mentioned)?

STT is Stratosphere to Troposphere Transport and is not equivalent to STE (Stratosphere-Troposphere Exchange). We have now added the definition of STT.

8. Page 5284 Paragraph 1 To be more clear, the discussion of the global totals and the connection with how the observations were assimilated should be expanded on. It is not inherently to do with the model or the inversion methodology but how the data was assimilated. The three models (MOZART-I, ACTMt42I67-I and TM3-I) could assimilate at the sampled time, in principal. It should be mentioned that the low totals that result from using monthly means, for example, is largely due the fact that NOAA flasks are filled for "background" conditions so averaging the model's simulation into a monthly mean will tend to be higher than a background value.
Second, the statement of low MOZART emissions being a result of high initial conditions should be moved before the discussion of the temporal resolution of the data assimilation. It seems to me, from the way it reads now, that the primary reason for the low global emissions is because of assimilating monthly means but the high initial conditions are likely a major reason (again pointing to the general comment above). Additionally, there are some typos in the ACTMt42I67-I naming. On this page, it is labeled ACTMt32I67.

We have now changed the first paragraph of 3.2.1 to mention the problem of the MOZART4-I initial conditions before discussing the temporal resolution of the data assimilation. We have also corrected the typo with respect to ACTMt42l67-I.
The reviewer is quite right that the NOAA flasks are generally sampled for "background" air, therefore, the monthly average of the simulated mixing ratio will usually be an overestimate. We have included this fact in the discussion in section 3.2.1.

9. Page 5285 Paragraph 2 This is a nice discussion of the effect of inter-hemispheric mixing time. Could you expand on how that would affect the "other hemisphere"? For example, TM5-I over-predicts the SH emissions because of a too slow mixing time. This should cause the NH emissions to be under-predicted. Similarly, the reverse should be true for LMDZ-I. Also a diffusive PBL in this model should result in higher concentrations simulated in the upper troposphere. How does this compare with other models and HIPPO?

This is an interesting point. In fact the accumulated emissions of TM5-I still exceed those of e.g. LMDZ until circa 30°N. The reason for this cannot be determined from these results alone but it may be at least in part also owing to transport errors in LMDZ, e.g. too diffusive PBL.

Concerning the diffusive PBL in LMDZ, we compare all models to the total columns from HIPPO in January and November 2009 (see Fig. 3): LMDZ underestimates the total column mixing ratios compared to HIPPO. However, I am not certain that HIPPO is the best dataset to determine this as HIPPO is over the Pacific Ocean and remote from the main sources. In Part I of this study (published in ACP in 2014) the prior modelled mixing ratios were compared to repeated aircraft profiles at 4 sites, and at the northern mid-latitude continental site, ULB, LMDZ did overestimate the mixing ratio above 3000 m compared to the observations (see Fig. 4).

10. Page 5287 Line 17-18 Please remind the readers about what the first and second criterion mean. Is this the definition of "significant" starting on the previous page?

We now repeat the criteria here.

11. Page 5288 Paragraph 2 South Asia experiences a double maximum in April and September. The authors argue the peak in April could partially be an artifact due to too strong STT. What is causing the September peak?

There is also a peak in the prior emissions around August-September, thus, the peak found a posteriori, may also be influenced by the prior. In any case, a September peak may not be unexpected since this corresponds with the end of the summer monsoon period (around April to September) and may result from increased emissions due to the warm and moist soil conditions.

12. Page 5289 Paragraph 1 For South and Tropical America, what could cause the maximum in September?

We think that this is likely to be only an artifact of the transport, as the models cannot capture the correct seasonality a priori of the atmospheric mixing ratios resulting in the April minimum in the fluxes, and to compensate, a maximum around September. This is explained on p5289, L8-13.

13. Page 5289 Paragraph 2 LMDZ predicts a much larger amplitude than any other model for the 90-30S ocean. What could cause this?

Again, it is most likely to be due to transport error. LMDZ (and also the other models but to a lesser extent) does not capture the seasonality of the observed atmospheric mixing ratios a priori. Thus in order to match the observed mixing ratios, lower emissions are estimated in the inversion in April (and compensatingly high emissions in September). We have now added this to p5289, paragraph 2.

14. Page 5290 Line 11-13 How was the 16.1-18.7 TgN/yr global emissions calculated? Where does this uncertainty of 0.7 TgN/yr come from? Does it include the spread in models as well as the uncertainties from the inversions? Is it an average of all of the years? If this uncertainty is based on the range in the inverse methods/models, then it cannot really be considered a "true" estimate of the uncertainty because the models do not form an ensemble of independent estimates (see general comment).

The uncertainty (0.7 TgN/y) is the median absolute deviation (MAD) of the inversion results average for 2006 – 2008. This number does not include the uncertainty from the inversion. As explained above, not all the inversions provide estimates of the posterior uncertainty but this number is slightly higher than the posterior uncertainty of LMDZ4-I (0.5 TgN/y for the global total). While it is true that the inversion results are not independent estimates, we think that this number 0.7 TgN/y is the best estimate of the uncertainty that we can provide. Alternatively, we can just state the range and give no uncertainty estimate.

15. Page 5300 Table 2 What does the '11 regions' under the 'scale length in B' mean for the ACTMt42l67 model?

This was just to indicate that ACTMt42l67 solves the fluxes regionally (11 land and 11 ocean regions) thus the fluxes within each region have a correlation of 1. However, we realize that this is a bit confusing and hence we have changed removed this now.

16. Page 5308 Fig 2 The figure is not plotting the gradient but the zonally averaged mole fraction. Also, the sentence 'The grey shaded area shows the range of values for the model using the prior fluxes' is difficult to understand. Is it the range from each model aggregated together?

We agree that "gradient" is a bit misleading as it is the meridional mole fraction. It is, however, not the zonally averaged mole fraction, as it is the mole fraction sampled at varying sites from north to south. We have now corrected this in the caption. The grey shaded area shows the range of prior values for all models.