

Authors' response to Executive Editor

*In particular, please note that for your paper, the following requirement has not been met in the Discussions paper:*  
• "The main paper must give the model name and version number (or other unique identifier) in the title."

This paper is merged with the accompanying paper, following the referees' comments and with approval from the Topical Editor, and we titled the new manuscript as "Global methane emission estimates for 2000-2012 from CarbonTracker Europe-CH4 v1.0".

Authors' response to anonymous referee #1

In the following, referee's comments are in *italic*, authors' responses in normal font, and references (page, line, figure, and table number) to the revised manuscript in **bold**. Please note that this paper is merged with the accompanying paper, following the referees' comments and with approval from the Topical Editor. A summary of the accompanying paper was included in the Supplementary Material of this paper.

In addition, figures and tables were revised substantially, and the following table summarizes the changes to figure and table numbers. In addition, please note that Fig. 2, Fig. 3, and Fig. 4 of the accompanying paper were moved to the Supplementary Material as Fig. S1, Fig. S2, Fig. S3.

Revised	Original	Short description
Fig. 1	Fig. 1	mTC regions + observation sites
Fig. 2	Fig. 3	Comparison with assimilated observations
Fig. 3	Fig. 4	Global mean XCH <sub>4</sub> and growth rates
Fig. 4	Fig. 5, S1	Comparison with aircraft observations
Fig. 5	Fig. 6	Model performance in Europe at assimilated sites
Fig. 6	Fig. 7	Comparison with TCCON observations
Fig. 7	Fig. 8	Comparison with GOSAT observations
Fig. 8	Fig. 11, 12, S4	Emission estimates of global and Asian temperate and tropical mTC regions.
Fig. 9	Fig. 9	Growth rates of global emission estimates
Fig. S4	Fig. 2	Land-ecosystem map
Fig. S5	Fig. S2	Comparison with TCCON observations
Fig. S6	Fig. S3	Comparison with GOSAT observations
Fig. S7	Fig. S6	Monthly mean of total emission estimates at latitudinal bands
Fig. S8	Fig. 8, 10, 13	Emission estimates of land mTC regions
Fig. S9	Fig. 14, S5	Emission estimates of ocean mTC regions.
Table 1	Table 1	Inversion set-up
Table 2	Table 2	List of observation sites used in the inversions
Table 3	Table 3	List of aircraft observation sites
Table 4	Table 5	RMSE with TCCON observations
Table 5	Table 6	RMSE with GOSAT observations
Table 6	Table 4	Global and regional emission estimates

General comments

*A lot of figures are relegated to the supplemental information, while I would prefer to see them in the main paper. I appreciate this can be a matter of preference, but I have made the case for their promotion in my comments below.*

*In most of the figures/tables showing either the model run with posterior emissions, only the posterior values are plotted. I would like to see the priors on the figures as well, in order to assess the improvement made by assimilating the data.*

We appreciate the reviewer's suggestions, and the figures and tables were revised extensively following the

specific comments below. We agree with the reviewer that the figures and tables became more complete by including the prior. The prior is added in the figures and tables in the revised manuscript. Please see below for responses and changes in more detail.

Specific comments

Pg. 5, first paragraph: I found it a bit confusing throughout the paper that the modified Transcom regions and the regions to be optimized were both referred to solely as “regions”.

We apologize for the confusion. We agree with the reviewer that the definition of 'regions' was confusing in many sentences. We have tried to make this clear in the revised manuscript by referring to modified TransCom regions as mTCs and adding the word 'optimization' for the optimization regions where the definition is unclear.

Pg. 5, line 16: What was the variance for the ice region?

The variance for the ice region was  $1 \times 10^{-8}$ . It was stated in the accompanying paper, but we agree with the reviewer that it was not clear only from this paper. In the revised manuscript, the number is explicitly added.

**Text is revised: Pg. 6, line 18.**

Pg. 6, line 30: I believe that many of these observations are not available for the whole 2000-2012 time period and you might consider giving the date range in Table 2. Does the assimilation system show any effects of observations turning “off” and “on”?

This is an excellent and very interesting point. We indeed acknowledge the sensitivity of the inversions to the set of observations assimilated, and the emission estimates could vary by turning “off” and “on” some of the observations. We surmise that the effect would be mostly in regional emission estimates, and less in continental scale estimates. This is because many “background” sites that constrain large-scale emissions have long-term observations, and the number of “new” background observations is low. The effect was especially prominent in regions that were already poorly constrained by the current background network: sites in South America had a significant effect on the increase in regional emission estimates (see Section 3.4.5). Although we did not test this further for all sites, we also discussed the effect for other observations (discrete vs discrete+continuous (see Supplementary Material), and GSN, AMY in Section 3.4.3), and we feel that this is sufficient for the paper at this point. Finally, we followed the reviewers suggestion and the date range is added to Table 2.

**The date range is added to Table 2.**

Pg. 7, line 3: In the companion paper, the minimum mdm is 7.5 ppb. Is this a difference in the approach between the two papers? I'd also suggest you include which category of observation has this value.

Firstly, we thank the reviewer for also carefully reading the accompanying paper. As pointed out by the reviewer, the minimum mdm was different in the two papers. We found that the sites with low mdm (i.e. marine boundary layer or high southern latitude sites) were particularly important to remove the effect of initial 3D mole fraction fields and to constrain background concentrations well. In addition, these sites were important to correct a bias in the latitudinal gradient of the prior emissions. In the long-term inversion we noticed that the constraints introduced by these sites was not strong enough (not shown or published), and therefore, we gave greater weight (=lower mdm) for those sites for the long-term experiment. In the revised paper, the table and description of the accompanying paper regarding mdm was omitted. The main finding of the sensitivity tests in the Supplementary Material does not change due to differences in the mdm of those sites, and therefore, we hope the reviewer agrees that the alternative description of the mdm was not necessary to be included in the Supplementary Material.

Pg. 9, second paragraph and Figure 3: I really cannot make most of these observations out on Figure 3 because of the large range of values used. In fact, I had to stare at the posterior differences for quite a long time before I was sure the three sub-figures were different. I suggest plotting the posterior panels with a different colour scale so the differences can be made out.

We apologize for difficulties that arose by the poor illustration in the figure. This is an excellent suggestion and we revised the figure by plotting posteriors in a different scale.

**Fig. 2 (original Fig. 3) is revised.**

*Pg. 9, third paragraph: A2007 and B2007 are used here (and throughout the paper), but I don't believe they are formally defined.*

We acknowledge that the words were not descriptive enough. It was referred to only once in a much earlier section. To make this clear, the text was revised in the new manuscript, where A2007 is simply written as "after 2007", and B2007 as "before 2007".

*Pg. 10, third paragraph: I think you've missed a chance to draw some conclusions about the model due to the fact that the agreement with TCCON tends to be better before the averaging kernels are applied. As noted in the paper, applying the kernels does make the comparison more valid, as the effects of the prior used in the TCCON analysis and the instruments' vertical sensitivity are taken into account. The fact that this makes the agreement worse could indicate that there's some compensating effect in the total column, which I think you can see by looking at the comparisons with IMECC in Figure 5.*

We agree with the reviewer that the point was missing in the text. This is a very interesting point in regard to the application of averaging kernels. As pointed out by the reviewer, the comparison with IMECC showed that model CH<sub>4</sub> was lower than the observations and agreement was worse at lower altitudes. The weights at lower altitudes were greater than those at higher altitudes when the averaging kernel was applied. Therefore, model XCH<sub>4</sub> decreased and agreement with retrieved XCH<sub>4</sub> became worse when the averaging kernel was applied. Although we acknowledge the effect of the averaging kernel, we decided to remove the text, as all comparisons were done with the averaging kernel applied.

*Pg. 10, line 15-16: Why not show the prior?*

We acknowledge the importance of also showing the improvements by the inversions for XCH<sub>4</sub> estimates. In the original figures, the prior was not included, because it was already clear from the comparison with the in situ observations that the prior XCH<sub>4</sub> would also be much higher than the observations. Since prior XCH<sub>4</sub> also increases significantly more than the observations and the posterior, the differences between the posterior estimates and that with the observations were less clear when prior estimates were plotted. However, in the revised figures, we included the priors.

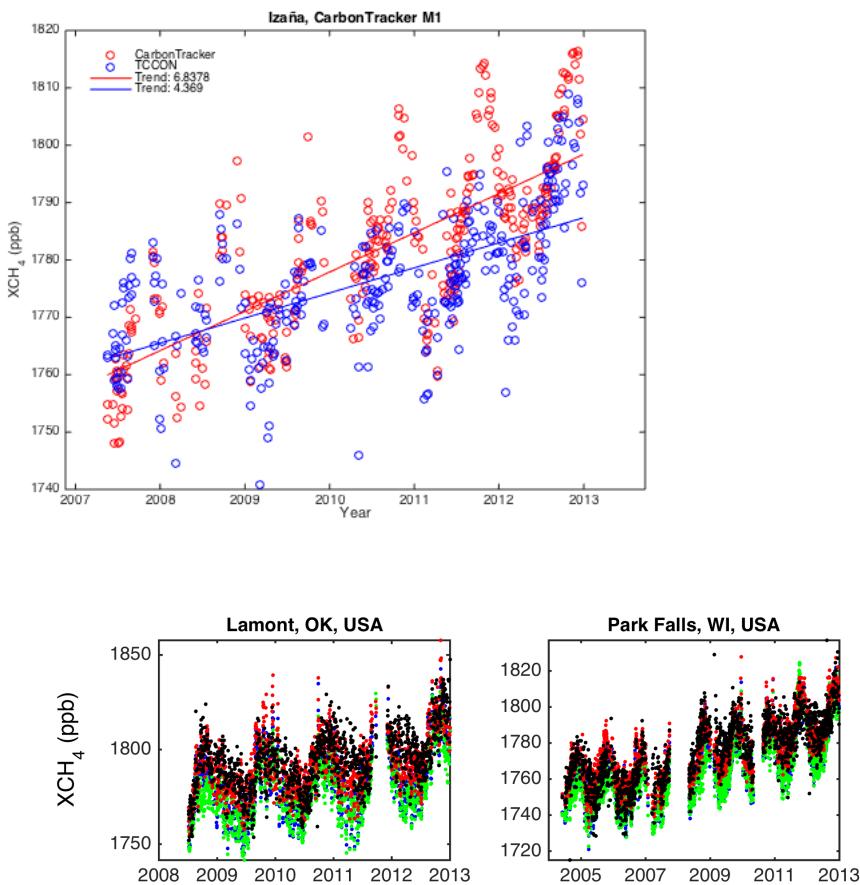
**Fig 6 (original Fig. 7) is revised.**

*Pg. 10, line 24-25: I'm not sure I understand what you're getting at here: why would missing spring observations in one year affect the trend over the study period?*

We appreciate the reviewer for noticing this inconsistency. We agree with the reviewer that the missing spring observations were not the reason for the mismatch in trends because the model trend was calculated from the same date and location of the observations, and using the averaging kernel.

We looked at the trend further, and found that the model seems to overestimate the trend (see figure below). XCH<sub>4</sub> before 2010 was lower and after 2010 was higher than the observations. A similar feature was found at other sites, e.g. at Park Falls and Lamont, such that the increasing trend for 2007-2013 was stronger in the model than in the observations. This could be caused by underestimation of emissions in northern temperate regions before 2010 and a significant increase after 2010. This may indicate that our inversion does not estimate the start of the emission increase in temperate regions correctly. The discussion is added in the revised manuscript.

**Text is revised: Pg. 12, line 9-11.**



Pg. 10, line 32: Can you not be sure that the data was not used by the inversion?

Here, we tried to address that the inversions probably did not learn much from GPA observations because mdm was set large (75 ppb). We apologize that the sentence was misleading. We are certain the data were used and assimilated in the inversions, but the high mdm limited the effect of the observations on the emissions to some extent. In theory, the effect could be checked by running inversions without the data, although such sensitivity test was not done as it would have been an expensive (CPU time) extra test. The sentence in the new manuscript is revised to address that the data were assimilated in the system.

**Text is revised: Pg 12, line 23-24.**

Pg. 11, line 1: I know observations are few and far between in Australia, but Cape Grim is in a much different environment than Wollongong! Wollongong is a hard site to model in general – see for example Fraser et al., 2011.

This is a good point. We agree with the reviewer that the XCH<sub>4</sub> cannot simply be compared because the sites are near. As pointed out by the reviewer, Cape Grim is located on an island far from cities and has little local emissions, whereas Wollongong is located in the city of Wollongong, where local emissions could have a large influence. We have added this to the discussion in the revised manuscript.

**Text is revised: Pg. 12, line 17-19.**

Pg. 11, line 6: I think this is a typo? Darwin is closer to the Equator than Wollongong.

We apologize for the incorrect sentence. The reviewer is correct. Darwin is closer to the Equator than Wollongong. The argument therefore needed to be rephrased. Considering its surrounding environment (see also the above comment by the reviewer), we are now more certain that the model bias in the Wollongong was likely due to an error in local emission estimates rather than long-range transport. We have revised the text in the new manuscript.

**Text is revised: Pg. 12, line 17-24.**

*Pg. 11, last paragraph: The text occasionally refers to Figure 8 when it should be Figure S3 and vice versa.*

*Pg. 14, line 8: Fig. Sx? I think this figure was dropped from the submitted manuscript.*

*Pg. 15, line 28: Should this be a reference to Fig. 11?*

*Pg. 16, line 22: Should this be Fig. S4?*

We apologize for the confusion that arose by inconsistencies in the figure references. We revised the references in the new manuscript.

*Pg. 12, line 31: Differences are also surely due to the different time periods studied.*

We agree with the reviewer that the differences in the increasing rate are also due to the differences in time periods compared. We added further discussion in the revised manuscript.

**Text is revised: Pg. 14, line 2-3.**

*Pg. 19, line 15: Is there something missing after i.e.? I don't understand the reference to i here.*

We apologize for the poor formulation in this sentence. The sentence was meant to say that it was not possible to conclude from this study which estimates better capture actual emissions, because the emission estimates for Africa were not well constrained by observations in either study. We revised the new manuscript accordingly.

**Text is revised: Pg. 20 line 5-6.**

*Pg. 22, line 7: I don't understand what is meant here by the extra degree of freedom.*

The sentence was meant to explain that estimates of Eurasian boreal were sensitive to estimates of other regions. Due to very sparse observations in boreal Eurasia, the scaling factors of in this mTC region acted more as a compensating effect. We revised the sentence in the new manuscript.

**Text is revised: Pg. 22 line 27-29.**

*Pg. 23, line 10: Here is an example of where I found the term "region" to be confusing, as I thought at first you were referring to the optimized regions here, where in most of your inversions only one of anthropogenic or biogenic emissions were optimized.*

Here, the first 'region' was meant to be an optimization region, and the second one to be the mTCs. Both anthropogenic and biospheric emissions were optimized in all mTC regions over land, and the system tends to change optimization regions that have large emissions near the observation sites. In many mTC regions, anthropogenic emissions were greater than biospheric emissions near the observation sites, and therefore, the inversion tends to scale anthropogenic emissions more. This then led to larger changes in the anthropogenic emission estimates for optimization regions and also in mTC regions. We revised the sentence in the new manuscript.

**Text is revised: Pg. 24 line 1-2.**

*Figure 1: Please add a reference here to Table 6 for the names of the regions.*

Following the reviewer's suggestion, the caption of Figure 1 is modified by adding a reference to Table 5 (original Table 6).

**Caption of Fig. 1 is modified.**

*Figure 2: I find the colour scheme a little difficult – very hard to make out the green regions in amongst the lighter*

blue! You could also consider blocking out the modified Transcom regions with black lines.

We follow the reviewer's suggestion and changed the colour scale of Fig. S4 (original Fig. 2), and mTC and land ecosystem borders were added.

**Fig. S4 (original Fig. 2) is revised.**

*Figure 3 (again): For the comparisons with the prior I assume it's all surface observations that are used? (As opposed to TCCON etc. . . ) Can you also plot the prior difference using only the assimilated observations? It would make it more straight-forward to assess the performance of the inversion.*

We revised the figure by plotting the prior differences between assimilated observations. In addition, the colour scale of the posterior comparison changed following the reviewer's earlier comment.

**Fig. 2 (original Fig. 3) is revised.**

*Figure 4: Do the observations appear on the top panel? Can you also plot the differences between the observations and the prior and posterior (for XCH<sub>4</sub> only)? You could consider adding a vertical line at 2007 in both panels.*

Observations were not plotted in the figure because NOAA global averages are for the surface, unlike XCH<sub>4</sub>. However, we agree with the reviewer that the figure becomes more comprehensive by adding the observations in the top panel. We followed the reviewer's suggestion and added NOAA surface global mean CH<sub>4</sub> with a second y-axis. In addition, a vertical line at 2007 is added to both panels.

We did not include the differences between the NOAA global mean and the simulated XCH<sub>4</sub>, as additional complications arise due to sampling. The top panel illustrates global mean XCH<sub>4</sub> from the model, but no satellite products have sufficient global coverage throughout the years to make a fair comparison with it. In Fig. 7 comparison with GOSAT XCH<sub>4</sub> is included, where model XCH<sub>4</sub> was sampled at the locations and times of the observations. We hope the reviewer agrees that the comparison in Fig. 7 is satisfying.

**Fig. 3 (original Fig. 4) is revised.**

*Figure 5: I would use Figure S1 here since it shows the prior concentrations too. It would also be useful to give the prior RMSEs.*

We acknowledge the importance of showing the improvements by the inversions for XCH<sub>4</sub> estimates. Following the reviewer's suggestion, we added the prior in Fig. 4 (original Fig. 5), and removed original Fig. S1.

In addition, we agree with the reviewer that it is also important to show the prior RMSE. For that, we found a table was more appropriate and therefore, both prior and posterior RMSE are summarized in Table 3 in the revised manuscript.

**Fig. 4 (original Fig. 5) and Table 3 are revised.**

*Figure 7: I would use Figure S2 here since you discuss many of the sites in that figure in the main paper. (And you don't discuss Park Falls, but it appears here.) I'd also like to see the prior concentrations included.*

*Figure 8: Again, I'd suggest using Figure S3 here and showing all the regions in the main paper. And I'd include the priors on these figures too.*

Response to the comment on Fig. 7 and Fig. 8:

We acknowledge that not only a selection of the sites and regions but all sites and regions are important to illustrate the results. We tried to make satisfactory figures with all sites and regions, but could not find a better illustration. Therefore, we decided to retain the figures, but selecting the sites and regions more carefully based on the results discussed in the text. We hope the reviewer agrees that the new figures illustrate the findings better.

**Fig. 6 (original Fig. 7) and Fig. 7 (original Fig. 8) are revised.**

*Figure 12: This figure is impossible for me to decipher. I'd replace this with the bottom panels from Figure S4. (Or even move S4 to the main paper and remove Asian temperate from Figure 11.)*

We agree with the reviewer that the figure in the Supplementary Material illustrates our results and discussion better. Following the suggestion, original Fig. S4 is moved to the main paper with modifications as Fig. 8. The original Fig. 11 is removed.

**Fig. 8 (original Fig. S4) is revised.**

*Figure 13: Since you have the space, you could include Australia in the bottom right.*

We agree with the reviewer that the figure is more complete when emission estimates for Australia are added. Following the reviewer's suggestion, the estimates are illustrated in Fig. S8 of the revised manuscript, together with other land mTC regions. We hope the reviewer agrees that the SH emission estimates could be represented in the Supplementary Material. This was done to better balance the manuscript by reducing the number of figures.

**Fig. S8 (original Fig. 13) is revised.**

*Tables 5 and 6: Please include the prior RMSEs as well.*

We agree with the reviewer that the tables are more complete when prior RMSE is added. Following the suggestion, prior RMSE is added in both tables.

**Table 4 (original Table 5) and Table 5 (original Table 6) are revised.**

Typographical comments

*Throughout the manuscript, “North” and “South” should be capitalized when referring to the continents (e.g. “North America”) but not when referring to regions (e.g. “south Africa”).*

*Pg. 1, line 31: . . . and (b) the parametrization . . .” (remove “on”).*

*Pg. 2, line 31: . . . changes in emission sources . . .*

*Pg. 7, line 4: . . . 20 times mdm because . . .*

*Pg. 9, line 4: . . . emissions and if it is actually . . .*

*Pg. 9, line 23: . . . change much in B2007 . . .*

*Pg. 9, line 2: . . . Fig. 4 are calculated . . .*

*Pg. 19, line 19: . . . optimized in M1 and M3 . . .*

*Pg. 19, line 32: . . . around the coast were not . . .*

*Pg. 22, line 12: . . . variation or lack of . . .*

*Pg. 22, line 31: . . . CTE-CH<sub>4</sub> differ by the . . .*

We appreciate the reviewer for the language correction. The manuscript is revised taking into account all the suggested changes.

In the following, referee's comments are in *italic*, authors' responses in normal font, and references (page, line, figure, and table number) to the revised manuscript in **bold**. Please note that this paper is merged with the accompanying paper, following the referees' comments and with approval from the Topical Editor. A summary of the accompanying paper was included in the Supplementary Material of this paper.

In addition, figures and tables were revised substantially, and the following table summarizes the changes to figure and table numbers. In addition, please note that Fig. 2, Fig. 3, and Fig. 4 of the accompanying paper were moved to the Supplementary Material as Fig. S1, Fig. S2, Fig. S3.

Revised	Original	Short description
Fig. 1	Fig. 1	mTC regions + observation sites
Fig. 2	Fig. 3	Comparison with assimilated observations
Fig. 3	Fig. 4	Global mean XCH <sub>4</sub> and growth rates
Fig. 4	Fig. 5, S1	Comparison with aircraft observations
Fig. 5	Fig. 6	Model performance in Europe at assimilated sites
Fig. 6	Fig. 7	Comparison with TCCON observations
Fig. 7	Fig. 8	Comparison with GOSAT observations
Fig. 8	Fig. 11, 12, S4	Emission estimates of global and Asian temperate and tropical mTC regions.
Fig. 9	Fig. 9	Growth rates of global emission estimates
Fig. S4	Fig. 2	Land-ecosystem map
Fig. S5	Fig. S2	Comparison with TCCON observations
Fig. S6	Fig. S3	Comparison with GOSAT observations
Fig. S7	Fig. S6	Monthly mean of total emission estimates at latitudinal bands
Fig. S8	Fig. 8, 10, 13	Emission estimates of land mTC regions
Fig. S9	Fig. 14, S5	Emission estimates of ocean mTC regions.
Table 1	Table 1	Inversion set-up
Table 2	Table 2	List of observation sites used in the inversions
Table 3	Table 3	List of aircraft observation sites
Table 4	Table 5	RMSE with TCCON observations
Table 5	Table 6	RMSE with GOSAT observations
Table 6	Table 4	Global and regional emission estimates

### *Overview:*

### *Scientific concerns:*

- 1. that the model assumed a fixed lifetime for CH<sub>4</sub>, even though the authors explicitly acknowledge that this assumption is unlikely to hold. It is very important that the authors qualify any reported results from this model with this assumption.*

This is an excellent point. As the reviewer points out, our results depend on the assumption of a fixed CH<sub>4</sub> lifetime, but we agree with the reviewer that the assumption is unlikely hold. Montzka et al. (2011) found an increase in OH concentrations in the beginning of the 21st century, followed by a decrease in OH concentrations after 2004-2005. Similarly, Ghosh et al. (2015) and Dalsøren et al. (2016) also obtained a decrease in the CH<sub>4</sub> lifetime in their simulations. In addition, McNorton et al. (2015) showed that although interannual variability of OH may be small, small changes in OH concentrations could lead to significant changes in CH<sub>4</sub> concentrations. We did not carry additional sensitivity test on CH<sub>4</sub> lifetime since the uncertainty in changes in OH concentrations and its relation to the CH<sub>4</sub> burden is still high, as discussed by Prather et al. (2012). We hope the reviewer agrees that further discussion added in the revised manuscript

based on suggested studies is satisfactory. We have also stated this assumption in the abstract and conclusions.

**Text is revised: see e.g. Pg. 3 line 8-15.**

*2. that the inversion violates the assumptions that form its foundation in a way that likely aliased biases in the posterior emissions estimates.*

This is a very interesting point. We acknowledge that the assumptions such as the prior emission estimates and the representativeness of the atmospheric observations affected the inversion results. As pointed out by the reviewer (see also the comments below, Detailed Comments 2), the fundamental assumption that the prior is normally distributed with mean 1 may not hold, and the bias of prior spatial distribution remains in the posterior to a certain extent. Although the posterior atmospheric CH<sub>4</sub> values in the Northern Hemisphere agreed fairly well with the observations, we find negative bias in the posterior CH<sub>4</sub> values in the Southern Hemisphere. Although we could not eliminate the bias completely nor find the exact cause, we hope the reviewer agrees that the findings are meaningful and this point is to be examined continuously in future developments. We added further discussion on this issue in the revised manuscript. Please also see the responses to Detailed Comments 2.

*3. that the prior and model-data mismatch uncertainty estimates appear to be arbitrary, and that no tests (e.g., reduced chi-squared statistic) were given to demonstrate that they accurately reflect the actual uncertainty distribution.*

This is again an excellent point. We agree with the reviewer that the prior and model-data mismatch uncertainty were indeed somewhat arbitrary. The values could not be chosen based on a theory or numerical method, because such a theory or method is not yet developed to estimate the covariance structure exactly. However, as the reviewer pointed out, the assumptions can be examined by the method presented by Michalak *et al.* (2005). We have now examined the Chi-squared statistics ( $\chi^2$ ), following the reviewer's suggestion.

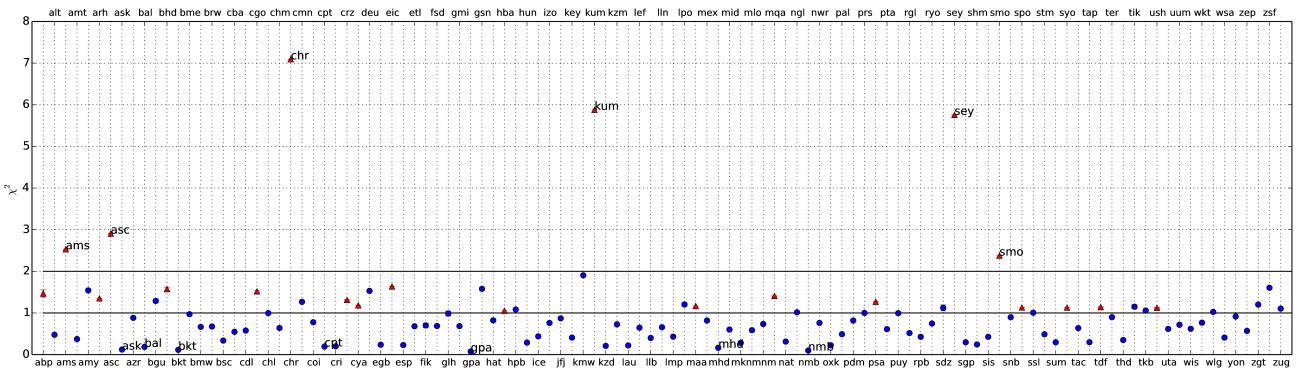
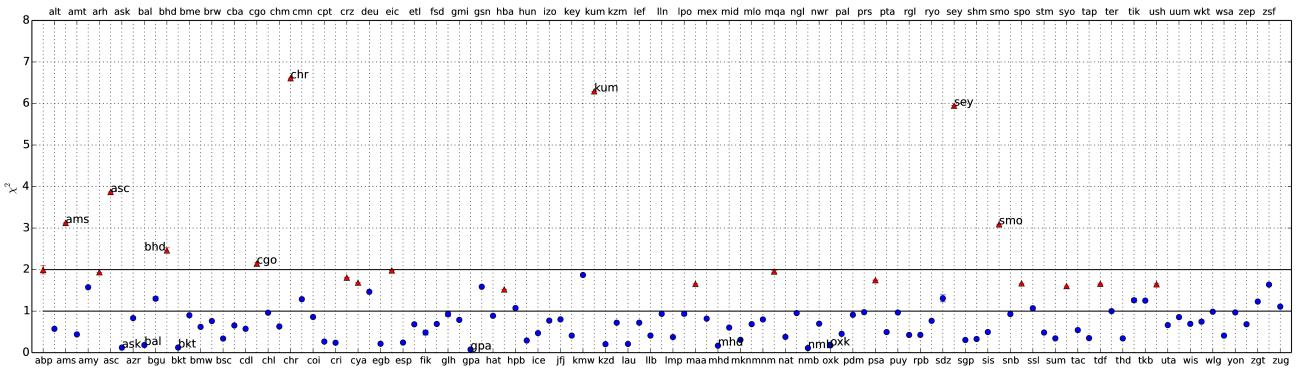
Most of  $\chi^2$  for the in situ observation sites ranged between  $0 < \chi^2 < 2$  (Fig. R1), indicating that the chosen mdm were in range of the expected value. However,  $\chi^2$  for most marine boundary layer sites are greater than one (Fig. R1), which indicates that the chosen prior mdm uncertainties were low. The  $\chi^2$  of these sites were high probably due to model errors rather than observational errors. The  $\chi^2$  for these sites were high for L<sup>62</sup>T especially because of the negative bias found around 2002, which was the most prominent among the inversions. The negative bias in L<sup>62</sup>G was not as strong as in L<sup>62</sup>T, and  $\chi^2$  was closer to one in L<sup>62</sup>G than L<sup>62</sup>T for the mbl sites (Fig. R1).

On the other hand, some sites have low  $\chi^2$ , indicating that the chosen mdm was larger than the expected uncertainty. However, mdm uncertainties for the continental sites should be assigned carefully because spatial representativity of the measurements may be low. Since the system optimizes emission estimates region-wise in this study, assigning mdm that are too small could lead to larger influence of the observations to the regional estimates than the observations would represent. The posterior ensemble distribution of  $\chi^2$  statistics followed normal distributions for all the sites based on normality tests (see Fig. R2 for an example), indicating that the normality assumption in the prior holds.

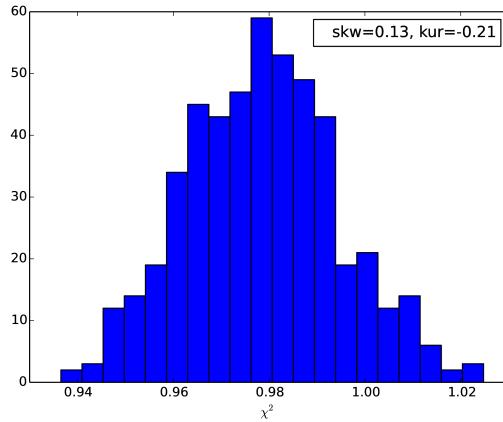
Regional  $\chi^2$  statistics were also distributed around 1 (Fig. R3). However, region mTC8 had high  $\chi^2$ , indicating that the prior uncertainty was lower than expected. This suggests that higher prior uncertainty or better prior emission estimates for the Asian temperate region was needed. On the other hand, regions such as mTC3, 5, and 6 have low  $\chi^2$ . For these regions, smaller prior uncertainties could have been used since the inversions did not retrieve much information from the observations. However, smaller prior uncertainties would lead to smaller posterior uncertainties, which may mislead the credibility of the emission estimates because having smaller posterior uncertainty does not necessarily mean the estimates are reliable. The  $\chi^2$  statistics of L<sup>62</sup>T were closer to 1 compared to L<sup>78</sup>T, whose covariance matrix was diagonal (Fig. R3). This indicates that the assumed correlations between the scaling factors were probably appropriate to a certain extent.

The values in covariance matrices could be adjusted further. However we should note that the resulting  $\chi^2$  depends on e.g. the choice of prior emission and observation data sets, and an arbitrary combination of these may or may not be better than some other, as noted by Michalak *et al.* (2005).

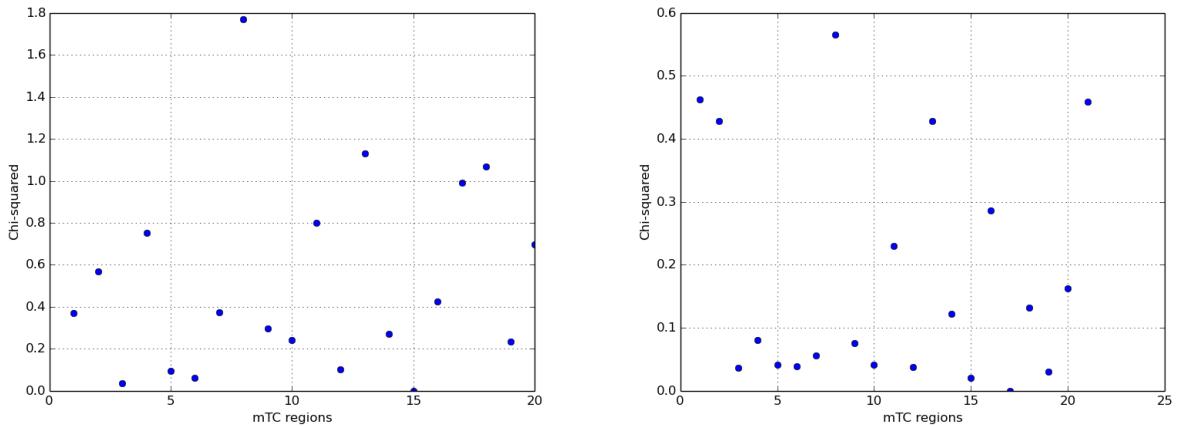
**Text is revised: see e.g. Pg.11 line 9-11, and Pg. 22, line 22-25.**



**Figure R1.** Chi-squared statistics from inversion  $L^{62}T$  (top) and  $L^{62}G$  (bottom) at the assimilated sites. The red triangles indicate marine boundary layer (mbl) sites. The sites with Chi-squared statistic larger than 2ppb or smaller than 0.2 are marked with three-letter site code.



**Figure R2.** Example distribution of Chi-squared statistic at an assimilate site (CHR: Christmas Island). Skewness (skw) and Kurtosis (kur) are shown to indicate normality of the distribution. Note that variance of the distribution was often very small.



**Figure R3.** Chi-squared statistic of regional estimates per mTC region: (left) L<sup>62</sup>T, (right) L<sup>78</sup>T.

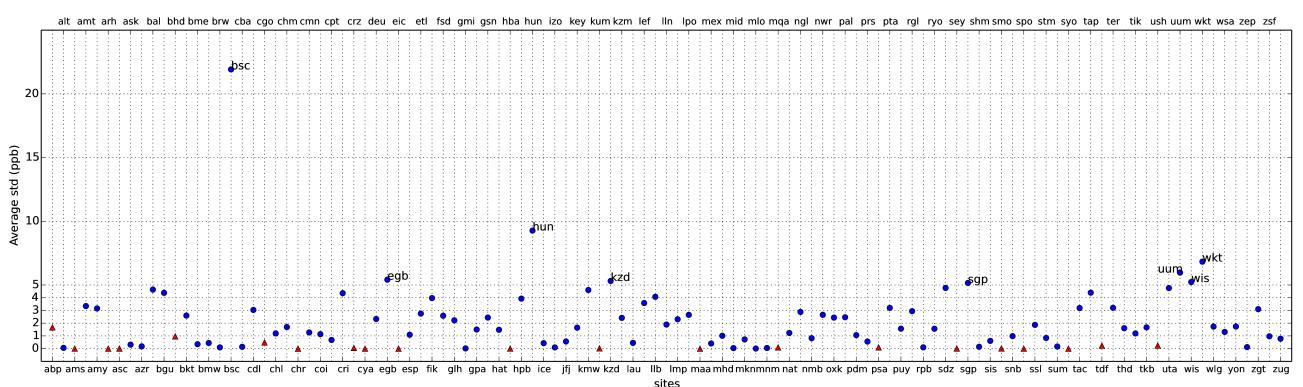
4. that the model evaluation examined only the maximum *a posteriori* estimate of the inverse model and did not give an assessment of the uncertainty estimates (similar to point 3). This paper evaluates a model that generates a statistical distribution as output – that distribution should be evaluated in its entirety.

We agree with the reviewer that the analysis based on not only the optimum (mean) posterior mole fractions, but also the ensemble distributions are important. Following the suggestion, we extended the analysis and its discussion is included in the revised manuscript.

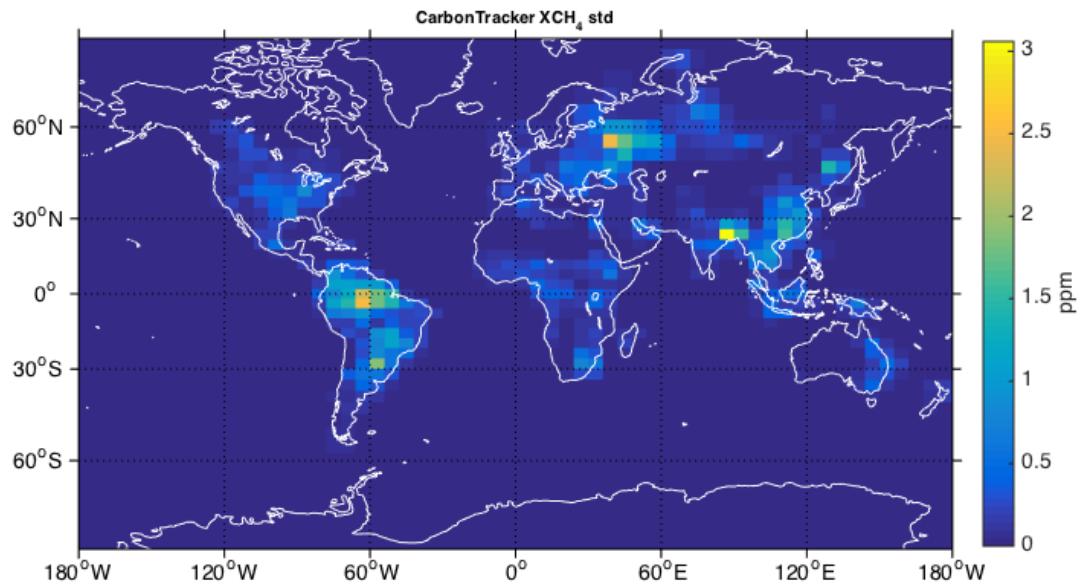
Distribution of ensemble mole fractions at assimilated in situ sites showed that the ensemble variation in CH<sub>4</sub> was small in general (<5 ppb; Fig. R4). However, Black Sea, Constanta (BSC) has the exceptionally high standard deviation (std) of the ensemble, which indicates the difficulty in the inversions to close the emission budgets nearby. The observation network around BSC was very sparse, and emission estimates around it have large uncertainty. The posterior std was also high at the sites in west and central Asia (KZD, UUM, WIS), suggesting that the emissions there were not well constrained. Small differences were found in the ensemble std at in situ sites between inversions (not shown).

Distribution of ensemble XCH<sub>4</sub> showed that the standard deviation were less than 3 ppb globally and less than 1 ppb at TCCON sites (Fig. R5, Table R1). Largest deviation was found in South American tropical region and around north west and south east Asia, again addressing the difficulty of the inversion to close budget in those regions. The results also support the finding at in situ sites. This is expected, as deviation at lower altitude affect the XCH<sub>4</sub> deviation the most.

**Text is revised: see e.g. Pg 10, line 28-30, and Pg. 12 line 33 – Pg. 13 line 1.**



**Figure R4.** Average standard deviation (std) of ensembles per site. Red triangles illustrate marine boundary layer (mbl) sites. The sites with the std higher than 5 ppb are marked with three-letter site code.



**Figure R5:** Average standard deviation (ppb) of posterior XCH<sub>4</sub> ensemble.

**Table R1:** Average standard deviation (std) of posterior XCH<sub>4</sub> ensemble per TCCON site.

Sites	std (ppb)
Ascension Island, Saint Helena, Ascension and Tristan da Cunha	0.00
Bialystok, Poland	0.37
Darwin, Australia	0.01
Eureka, Canada	0.00
Garmisch, Germany	0.17
Indianapolis, Indiana, USA	0.28
Izana, Tenerife, Spain	0.02
Saga, Japan	0.06
California Institute of Technology, Pasadena, California, USA	0.18
Karlsruhe, Germany	0.17
Lauder, New Zealand, 120HR	0.04
Lauder, New Zealand, 125HR	0.04
Lamont, Oklahoma, USA	0.53
Park Falls, Wisconsin, USA	0.21
Reunion Island, France	0.02
Sodankylä, Finland	0.08
Wollongong, Australia	0.09

Detailed Comments:

*Scientific Quality:*

For Point 1 of CH<sub>4</sub> lifetime, Point 3 of uncertainty assumption, and Point 4 of posterior distribution and analysis of posterior uncertainty, we have considered the reviewer's comments together with the comments in the Overview section, and therefore, the response is included above.

*2. The inversion setup violates one of the fundamental assumptions from which it is derived in a material way that leads me to doubt the validity of the conclusions. An inversion of this sort assumes that the error in the prior is a second-order (a.k.a. weak-sense) stationary Gaussian random process with zero mean.*

*The authors use the EDGAR 4.2 FT2010 emissions inventory as a prior anthropogenic emissions field. This is a high-resolution (0.1x0.1 degree) inventory that is known to be (and demonstrated in the paper to be) biased in its spatial distribution over a broad spectrum of scales and also biased in its temporal trend.*

*They set the prior error variance for the total emissions from any region to 0.8. The assignment is arbitrary and very likely too high. The authors effectively eliminate the bias by over-estimating the random error in the prior. Still, the posterior estimate is at the extreme of the error bounds, and so the bias in the prior affects the posterior estimate. This is visible in Fig. 3, where biases in the latitudinal distribution and seasonal cycles are visible in all posteriors.*

This is a very interesting point. It is true that the bias in the spatial distribution and temporal trend of the EDGARv4.2 FT2010 inventory has been reported, and may have violated the prior assumption. We acknowledge that the bias in the prior is one of the reasons why large prior uncertainty was needed. However, the inventory has the advantages that it provides global estimates at high resolution and long temporal coverage, which the inversions benefit from. The reported biases were not taken into account, as these are also uncertain, and it would bring uncertainty in the prior estimates in some other way, such as model bias. We assumed the inversion could correct it, but we agree with the reviewer that some bias still remained, even with the high prior uncertainty. Therefore, we would like to address that the prior estimates also need improvement, especially for regions with sparse measurement coverage. In addition, it is important to note that the bias in atmospheric CH<sub>4</sub> seen in Fig. 2 (original Fig. 3) could also be due to atmospheric transport. For example, the observed seasonal cycle was better captured using the faster vertical mixing scheme in TM5. This indicates that not only the prior emission estimates, but also the slow vertical mixing was one cause for the mismatch in the atmospheric seasonal cycle. We hope the reviewer agrees that despite the remaining bias, the finding are important, and the bias in the inventory and process-model based estimates are to be investigated in future studies.

*Additionally, the scaling factors are resolved at spatial scales of thousands of kilometers. The error in the prior varies at scales much smaller than this – producing a severe representation error. The problem is therefore likely under-parameterized (equivalent to having covariance lengths that are too long, or regions too large), and so adequate scaling factors cannot be derived that permit unbiased residuals at individual sites. As a result, the authors find strong biases in the residuals, and even throw out some sites.*

*An example of such sites are given in the paper, and the authors remove them:*

*"Strong negative bias as found in Bukit Koto Tabang, Indonesia (BKT) (-25 to -27 ppb) and Mt. Kenya, Kenya (MKN) (-18 to -23 ppb), such that the posterior mole fractions were especially low during June–October. This suggests that the measurements at those latitudes are not representative of large regions optimized in the model."*

*To solve this problem, the authors would need to perform the inversions at high resolution using covariance length scales constrained as part of an objective error characterization.*

We agree with the reviewer that the representation errors of the observations were likely high in some regions, especially where the observation network is sparse. As pointed out by the reviewer, example regions were Asian tropical regions, where BKT is located, and south Africa where MKN is located. Although those sites were assimilated in the system, the emissions in those regions were not well constrained due to lack of observations and good prior information about the emissions and their uncertainties. We acknowledge that the bias in the posterior mole fractions remained partly due to underparametrization of the system, and resolving at higher spatial resolution with carefully chosen correlation lengths would reduce such bias. However, even with a high resolution model, we would not be able to reduce the uncertainty in the regional estimates unless further information becomes available. Although we did not develop and examine the emissions further with a higher resolution optimization scheme at this point, we hope the reviewer allows us to undertake such development in

a future study as well.

Based on the reviewer's comment, further discussion on the prior assumption and representativity of the observations are added in the revised manuscript.

**Text is revised: see e.g. Pg. 19, line 1-12.**

*Presentation Quality:*

*1. Messaging*

*This work is presented as model evaluation and interpretation. The paper goes into great detail about the variations in every region, giving their trends, comparisons to other regions, and comparisons to other papers. The work needs to be boiled down to a set of key messages. My understanding is that the main messages are those described in the Overview section of this review.*

*The body of the paper needs to be focused on providing the scientific justification for the given messages.*

We appreciate the reviewer for carefully reading the manuscript and trying to understand our messages.

Following the reviewer's comment, we tried to more carefully phrase our text to better present the study and its key findings. The figures were revised substantially following the reviewers' suggestions (see also below), and the texts, including abstracts and conclusions, were revised to better highlight the key findings. Please also note that a summary of the key findings were added at the end of Summary and Conclusions.

*2. Writing*

*The paper requires extensive revision by an English language editor. Problems include: - incorrectly cased letters (e.g., "south America" should be "South America"). - inconsistent tenses and active vs. passive voice (e.g., in the abstract, line 29 "We use three configurations. . .", then line 32 "The posterior estimates were evaluated. . ."). - broken sentences (e.g., page 3, line 22 "To estimate biospheric emissions, information from an underlying ecosystem distribution map is useful, which defines the location of the sources and can help distribute larger regions over which the atmospheric signals integrate."). - truisms (e.g., page 9, line 20 "The growth rate (GR) of atmospheric methane mole fractions showed that the posterior estimates are closer to the observations than the prior, as expected"). - paragraphs that are incredibly long and rambling (e.g., page 14, line 7 to page 15, line 6; page 16, line 13 to page 17, line 5; and page 18, lines 5 - 30).*

We apologize for the inconsistencies that arose as a consequence of the weak formulation that existed in the manuscript. In this revision, we tried to more carefully phrase our text, and also had the full paper language edited by a native English speaker. Moreover, we tried to make our descriptions more clear using new labelling.

*3. Figures*

*The figures in this paper have a number of issues. Points a-c absolutely must be addressed in order for the paper to be publishable.*

*a) Figs. 9, 10, 11, 14, S4, and S5 include data and/or error bars that run off of the figure.*

We agree with the reviewer that the figures became more complete by showing the error bands fully. The y-axes of the figures were revised following the suggestion.

**Figs. 9, S8 (original 10), S9 (original 14 and S5) and 8 (original S4) are revised. Fig. 11 is removed.**

*b) Many of the figures include error bars but do not specify their meaning (1 standard deviation? 95% credible interval?)*

These were meant to be 1 standard deviation of the ensembles. We have now included the information in figure captions.

**The Figure captions (Fig. 8, S8, S9) are revised.**

*c) Figs. 4, 5, 6, 7, 8, 9, 12, 13, S1, S2, S3, and S6 are not colorblind safe.*

We apologize for the confusion that would have arose due to the choice of the colours. We tried to make the lines more distinguishable by changing the colours.

**The colours in the figures are revised.**

*d) Figs. 6 and 7 are difficult to read because of closely placed points.*

We have tried to make the points clear in the Fig. 5 (original Fig. 6) by adding a zoomed map of central Europe.

For Fig. 6 (original Fig. 7), we acknowledge that the points are close to each other, and it is difficult to distinguish each points. However, for some sites, the temporal coverage of the data was not good enough to present e.g. moving averages as lines. We tried to make points clear by changing the point sizes and shapes, but could not find a better way to present than in the revised manuscript. Therefore, we decided to retain the look and chose to illustrate using points. We hope the reviewer agrees that the intent of Fig. 6 was to give an overview of the agreement, rather than focusing on each point, and the current way of presenting is satisfactory for that.

**Fig. 5 (original Fig. 6) is revised.**

*e) Fig. 4 (top panel) should include observations.*

Observations were not plotted in the figure because NOAA global averages are for the surface, unlike XCH<sub>4</sub>. However, we agree with the reviewer that the figure becomes more comprehensive by adding the observations in the top panel. We followed the reviewer's suggestion and added NOAA surface global mean CH<sub>4</sub> mole fractions with a second y-axis.

**Fig. 3 (original Fig. 4) is revised.**