Authors' response to anonymous referee #2

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 $XCH_4$ 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 $CH_4$ 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 $CH_4$ 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 $CH_4$ 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 $CH_4$ concentrations. We did not carry additional sensitivity test on $CH_4$ lifetime since the uncertainty in changes in OH concentrations and its relation to the $CH_4$ 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 distribute 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 $CH_4$ values in the Northern Hemisphere agreed fairly well with the observations, we find negative bias in the posterior $CH_4$ 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 $L^{62}T$ especially because of the negative bias found around 2002, which was the most prominent among the inversions. The negative bias in $L^{62}G$ was not as strong as in $L^{62}T$, and $\chi^2$ was closer to one in $L^{62}G$ than $L^{62}T$ 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 necessary mean the estimates are reliable. The $\chi^2$ statistics of $L^{62}T$ were closer to 1 compared to $L^{78}T$, 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.**

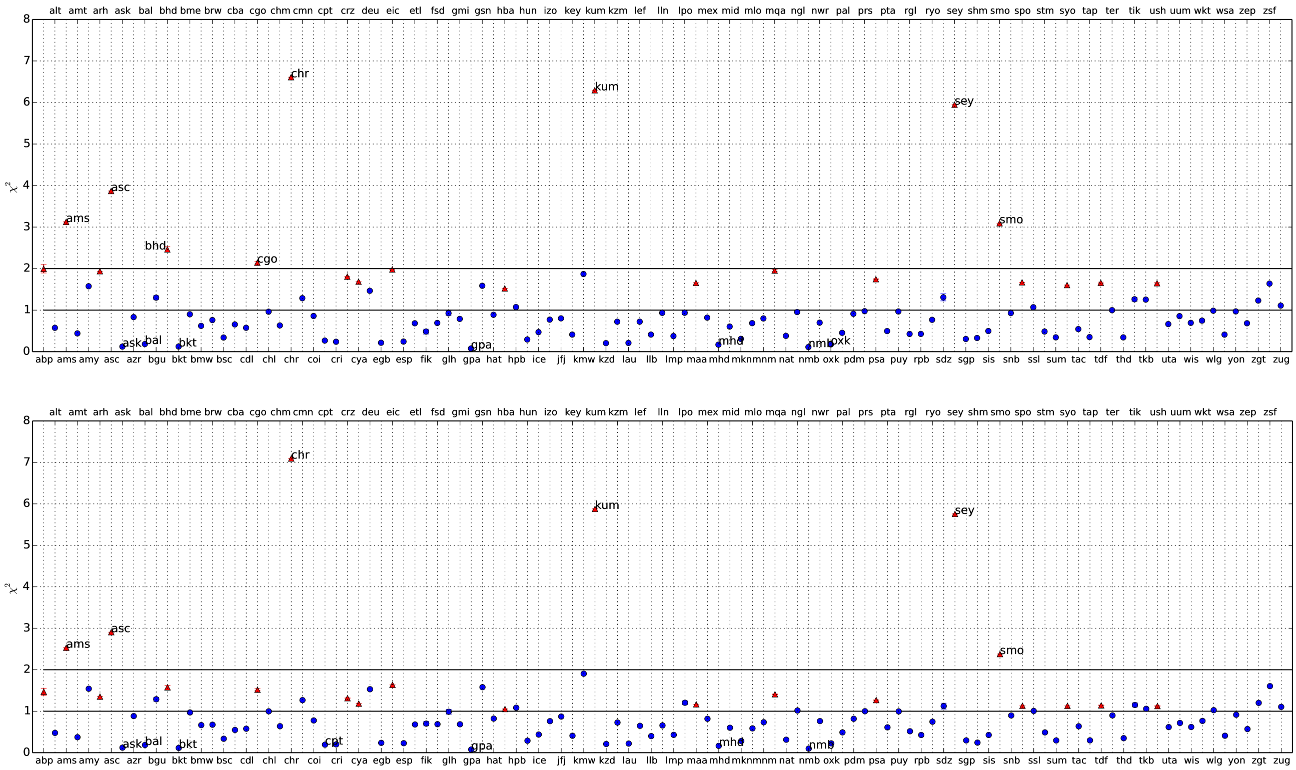


**Figure R1.** Chi-squared statistics from inversion $L^{62}T$ (top) and $L^{62}G$ (bottom) at the assimilated sites. The red triangles indicate marine boundary layer (mbl) sites. The sites with Chi-squared statistic larger than 2ppb or smaller than 0.2 are marked with three-letter site code.

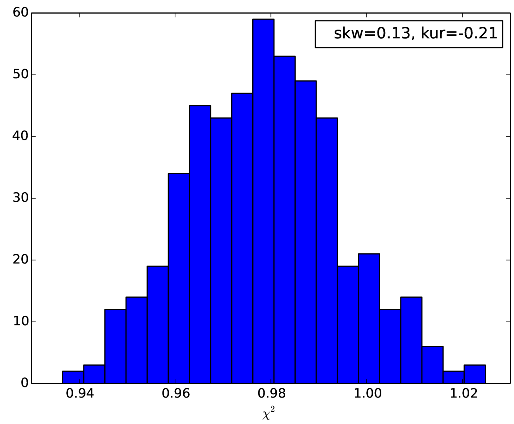


**Figure R2.** Example distribution of Chi-squared statistic at an assimilate site (CHR: Christmas Island). Skewness (skw) and Kurtosis (kur) are shown to indicate normality of the distribution. Note that variance of the distribution was often very small.

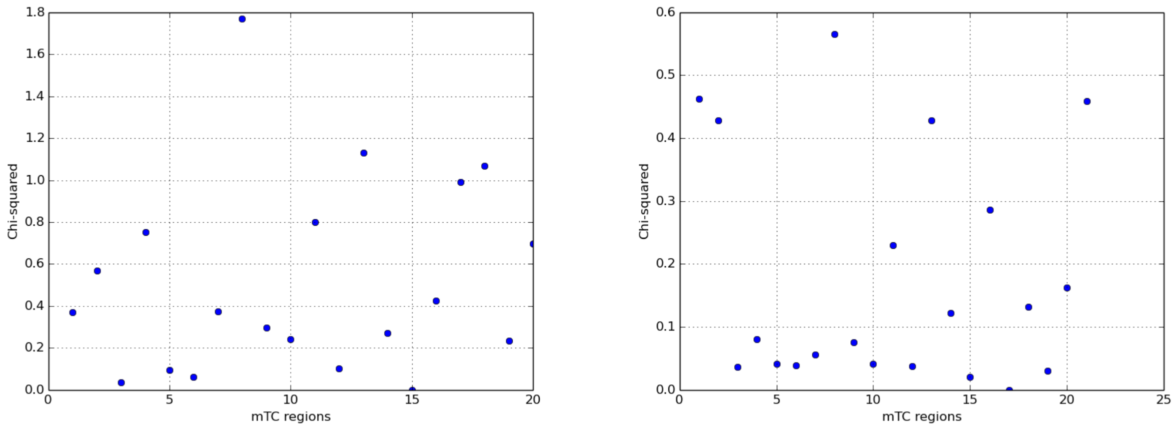


**Figure R3.** Chi-squared statistic of regional estimates per mTC region: (left) $L^{62}T$, (right) $L^{78}T$.

*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 $CH_4$ was small in general (<5ppb; 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 $XCH_4$ 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 $XCH_4$ deviation the most.

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

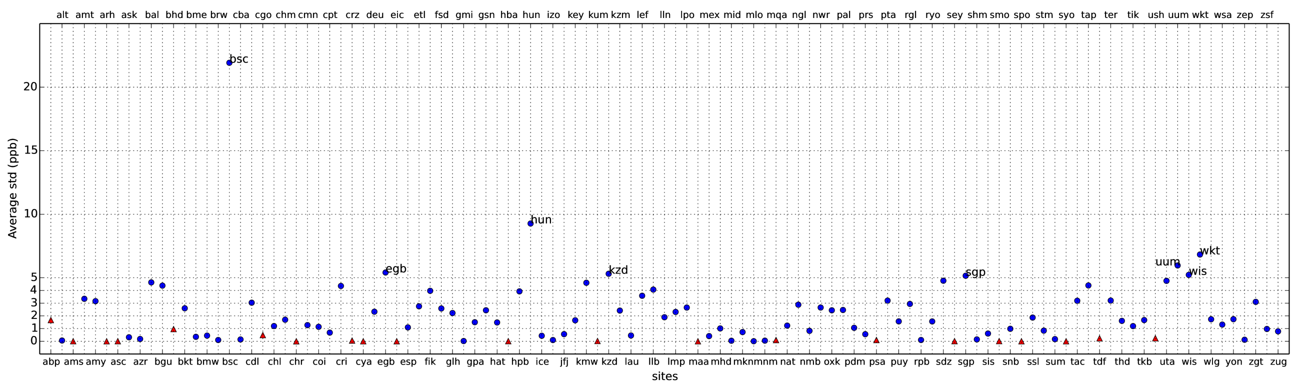


**Figure R4.** Average standard deviation (std) of ensembles per site. Red triangles illustrate marine boundary layer (mbl) sites. The sites with the std higher than 5 ppb are marked with three-letter site code.

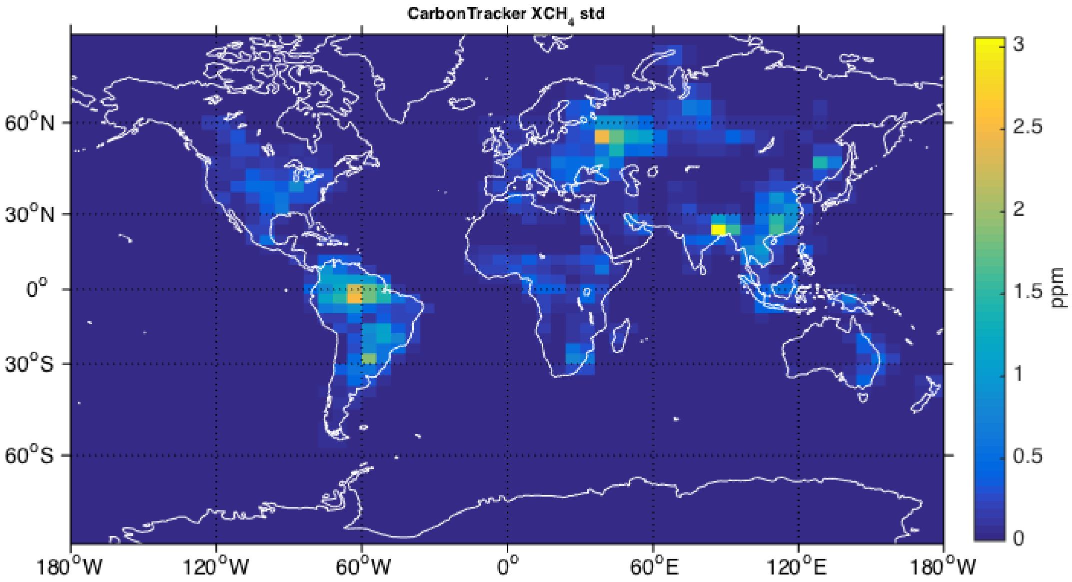


**Figure R5:** Average standard deviation (ppb) of posterior $XCH_4$ ensemble.

**Table R1:** Average standard deviation (std) of posterior $XCH_4$ 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 $CH_4$ lifetime, Point 3 of uncertainty assumption, and Point 4 of posterior distribution and analysis of posterior uncertainty, we have considered the reveiwer'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 $CH_4$ 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 luck 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 $XCH_4$. 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 $CH_4$ mole fractions with a second y-axis.

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