Dear Ilse Aben,

These are reply to your comments.

1. *I asked you to respond to all remaining comments from the reviewer one-by-one. You only responded to the 2nd comment from the reviewer. The reviewer had 4 comments.*

We think we have responded to all the comments of Ref2 but in two separate files. For clarity they are modified and organized here.

As explained on page 6, biases are assessed by taking the absolute difference between model and FTS. The motivation is that biases may change sign seasonally, and therefore may not show up in annual averages when positive and negative contributions cancel out. However, whether this is a good choice or not depends on the kind of bias that is investigated. Here the focus is largely on a latitudinal bias. Suppose that there is no latitudinal bias in the annual mean, but only a latitudinally varying bias in the seasonal amplitude. By taking absolute model to FTS differences across the year you would end up with a latitudinally varying bias. In this case the choice of absolute differences was clearly not appropriate. There may not be a single solution to this problem for the biases that are investigated here, but the meaning of the numbers that are summarized in the abstract and the conclusions for stratospheric and tropospheric contribution to the bias is not clear to me. A relation with a latitudinally vaying bias is suggested, but do these numbers really reflect stratospheric and tropospheric contributions to that bias. This requires more attention, including information on how the absolute differences are calculated (on every data point like an RMS, or on monthly averages, or?).

The absolute difference between the models and FTS is only used to calculate the averaged bias over all years and all sites. The results of the absolute bias are only the numbers appearing in Page 6, line 29～30 and in the abstract. The true bias (model - measurements) is used for all other parts of the paper, including all the plots. So there are artificial latitudinal varying bias in the plots. The numbers that are summarized in the abstract and the conclusions are meant to give a impression on amplitudes of the model to the measurement difference in the troposphere and stratosphere. The model to measurements differences are calculated on every data point like an RMS.

According to the caption of Figure 3, the tropospheric and stratospheric model biases are scaled with the corresponding contributions of the troposphere and the stratosphere to the total column air mass. However, there is a danger in doing so. Suppose that the model had a latitudinally and seasonally uniform offset in the tropospheric concentration. Then the scaling with the seasonal and latitudinal varying tropopause pressure would introduce a seasonal and latitudinal variation in the bias. In that case, when you look for varying biases within the troposphere in comparison with in situ data you wouldn't find any. This is exactly what seems to be happening here. This problem is attributed to differences in the global representation of the measurements, but could also be caused by differences in the NCEP and N2O derived tropopause heights. Since CH4 shown show a sharp vertical concentration gradient just above the tropopause, the analysis may be quite sensitive to how these heights compare. The uncertainty of this needs to be assessed and discussed.

The start point of the work is that there exist a latitudinal gradient in the bias of the modelled total columns of CH4, as shown in Fig. 2. The purpose of the work is to determine whether the troposphere or the stratosphere contributes to that. The variations of CH4 total columns include the contribution of tropopause variations. To separate the total column into the tropospheric and stratospheric parts, the airmass possessed by each layer must be taken into account except for the CH4 mixing ratios. Similarly

we must take the airmass into account when separating the model bias of the total column into separate layers. The tropopause altitude influences the total and tropospheric columns of CH4 and then its inaccuracy can contribute to the corresponding model biases. However, the sensitivity test using several different definitions for the tropopause (AC2 supplement) reveals that the tropopause is not the reason to the latitudinal gradient in the tropospheric model bias.

The comparison with TES is used to investigate longitudinal variations in the bias and the global representativeness of the comparisons with HIPPO which are limited to the Pacific. Apart from the fact that it is not clear that the TES data for the troposphere are accurate enough for this purpose (sizeable offsets are seen in the troposphere, that are not due to the TM3 model), the results do not seem to support the case that is made. If anything, the latitudinal gradient in the offset is stronger in the Pacific longitude band (in red) then at other latitudes. The authors are right that the bias has a longitudinal dependence, but it works on the wrong direction. This needs to be discussed more clearly, and the message of the study should be brought in accordance with this finding.

The validation of TES measurements in the troposphere does not show a latitudinal bias (Herman and Osterman, 2014). There are some offset but latitudinal gradient in TM3 bias revealed by TES could be reliable. The results from TES actually support the conclusion that the inconsistence between the HIPPO and TCCON comparisons with the models come from the longitudinal dependence of latitudinal gradient in the tropospheric model bias. But there are writing errors in the figure caption of the Figure 6 in the manuscript. In the third panel of Figure 6, the black points correspond to HIPPO sampling area (110° W～150° E) and the red points to the region beyond it.

Looking at Figure 5, the most significant differences between the models and HIPPO seem really at the highest measured altitudes. You might debate whether they are in the troposphere or the stratosphere. I wonder how important this really is. Wouldn't it be better to conclude that the problems show up most strongly at tropopause altitudes. In that case the method of separating the troposphere from the stratosphere may actually not be so appropriate. A plausible cause could be strat-trop exchange. I don't see how the results that are presented here exclude this possibility. Yet, it is not considered as an option.

Yes, the model bias in CH4 mixing ratios indeed increases abruptly around the tropopause. However, the approach separating the troposphere from the stratosphere does not influence the latitudinal gradient in the model biases of tropospheric CH4 columns as show in the Fig. 1-5 (supplement of AC2). Only the model biases of stratospheric CH4 columns are sensitive to the separation method and appear large when the tropopause is defined as low as 1.5 PUV. But the model biases in the stratospheric CH4 columns do not present a consistent latitudinal gradient with the model biases in the total columns of CH4. The strat-trop exchange may introduce variations of CH4 mixing ratio in the troposphere and stratosphere, but our work aims at separating model biases of total CH4 columns into contributions by the troposphere and stratosphere not finding the reasons leading to such biases.

2. *Wrt. your reply on the 2nd comment your reply is still not satisfying. It is clear that scaling is meant to weigh the relative contribution from the troposphere and the stratosphere. But the point is to what extent does this influence/determine the results and thereby the conclusions ? So I can only repeat the question as posed in my previous editorial comment : 'To what extent can you exclude that the latitudinal dependent bias is not caused by the scaling approach you take (which can introduce latitudinal and seasonal variations and thus the observed bias) ?' The simulations you refered to in your reply only look at variations in trop.height. That is –I expect- a much smaller effect.*

As have been explained, the tropospheric model biases here means the measured tropospheric CH4 column minus modeled counterparts. That depend on the tropospheric CH4 mixing ratio and the tropopause pressure. The purpose of our work is to determine whether the tropospheric or stratospheric CH4 column contribute to latitudinal bias in the total CH4 column. The tropospheric CH4 columns is represented by tropospheric column-averaged CH4 mixing ratios multiplied by (1-Pt/1000) (Pt is the tropopause pressure we divide the columns by 1000 hPa just for easy presentation and does not influence results). Without the scaling the model bias will be evaluated in term of stratospheric and tropospheric column-averaged CH4 mixing ratio. But these quantities are not directly related to the total CH4 column and can not answer question proposed by the result in Figure 2 of the manuscript.

Assuming the surface emission of CH4 is prescribed the total number of CH4 molecular residing in the troposphere will be more or less a constant (fast tropospheric convection compared to slow trop-to-strat transport). However, higher tropopause results in smaller CH4 mixing ratios and lower tropopause for larger mixing ratios. Here we want to evaluate model performances in simulating total number of CH4 molecular in the troposphere. It is not correct to exclude the effects of tropopause pressure or the scaling factor.

So the question '*To what extent can you exclude that the latitudinal dependent bias is not caused by the scaling approach you take (which can introduce latitudinal and seasonal variations and thus the observed bias) ?'* is not related to the purpose of our work. The relative contribution of tropospheric and stratospheric airmass is determined by the tropopause pressure. In our simulations, we looked at variations in the tropopause height only because it is the only factor determining the scaling factor regardless of its large or small effects.

3. *The updated manuscript contains all the track changes from the previous revision. Please accept all changes made before, I only want to see the delta changes to the previous version.*

There are not further modifications relative to the the last version yet.