Reply

To the Comment on “Comparisons of CH4 satellite GOSAT and ground-based FTIR measurements near Saint-Petersburg (59.9°N, 29.8°E) by N.M. Gavrilov et al., Anonymous referee #3

Dear Anonymous referee #3,
Thank you very much for your useful comments, which help to improve our paper. Our replies are given below in bold font:

In the paper " Comparisons of CH_4 satellite GOSAT and ground-based FTIR measurements near Saint-Petersburg (59.9°N, 29.8°E)” by Gavrilov et al., submitted for publication in AMT, the authors present an analysis of the comparisons between V01.xx and V02.xx GOSAT data with ground-based remote sensing measurements at the St. Petersburg site. It yields some insight into the V01.xx and V02.xx differences, as well as the properties of the ground-based CH_4 measurements at St. Petersburg. Due to the limited number of data and the
methodology used, results are very preliminary in nature, as is acknowledged by the authors. Nevertheless it is useful to those who wish to further explore the ground-based measurements from this site.

General comments

(1) My first general comment deals primarily with Chapter 3. As the authors themselves state,the amount of overlapping data between the ground-based and satellite measurements is small. However the implications of this statement are not made concrete, certainly when discussing the results in tables 1 and 2. For instance, table 1 yields the average δCH_4 as -13.0 ppb, and its standard deviation as 25.8 ppb. This implies that, using 95% confidence bands (standard deviations are poor estimators of the uncertainty on the bias, better to use the 2 sigma standard error (2.std/sqrt(N)) or (certainly for small samples) the actual 95% confidence bands) and a sample size of 9, the average δCH_4 ranges between -32.8 and 6.8 ppb. Therefore the following statement “which shows that GOSAT CH_4 data version V01.xx are about 0.7% lower than the ground-based FTIR values” holds little significance as the uncertainty is so large (between ~1.8% lower and ~0.4% higher). It cannot even be asserted that the V01.xx data has a negative bias. A similar analysis can be made of table 2. Here the 95% confidence interval on δCH_4 ranges between -3.8 and 11.2 (or -0.2% and 0.63%). Thus in this analysis one cannot state with confidence that for both versions there is a definitive positive or negative bias, nor that the versions themselves yield different results as both confidence intervals show significant overlap.The robustness of the second method (table 3) is better (V01.xx average δCH_4 between -20.6 and -9.4 ppb, V02.xx average δCH_4 between -3.2 and 4.4 ppb). Therefore I would focus my analysis entirely on this second method, merely stating that using 1 day overlap yielded similar results.

There are two answers. The first answer: if one decrease confidence from 95% to 68%, he will obtain 1-sigma confidence intervals for average δCH_4 of (-22.9 to -3.2) for Table 1 and (0 to 7.5) for Table 2.There is no overlap and one can see different signs of biases, but with confidence less than 95%. The second answer: we agree that confidence of our comparisons is not very high. However, please keep in mind that to increase substantially the numbers in Tables 1 and 2 we need several years of measurements, provided only several sunny days per year in our place. We are going to continue our measurements and hope to make comparisons that will be more confident in the future. We made some corrections in the text (see below).

(2) The authors use the CH_4 retrieval approach detailed in Sussmann et al., 2012. However when comparing his NDACC retrieval results with those from TCCON (Sussmann et al.2013), Sussmann takes into account the NDACC-TCCON a priori differences. Since the authors, in their discussion, compare their analysis with several GOSAT-TCCON validation results, such a correction should have been made. At the very least the impact of any NDACC vs. TCCON bias and the lack of any a priori correction, must be discussed in this context.

We had no intentions to copy the studies by Sussmann et al. (2012). We used generally the same approach, but some differences still exist. We just mentioned results of Sussmann et al. (2012) in the discussion. Detailed studies of the influence of a priory correction require special research. We added this statement into our discussion.

(3) The authors compare their results with the biases found in other studies. They typically list said studies average bias and standard deviation. For instance the average V02.xx δCH_4 in Yoshida et al. (2013) is -5.9 Ŕ} 12.6 ppb, with a 723 sample size. However to answer if the obtained bias for St. Petersburg is compatible with the one obtained by Yoshida et al., we again have to look at the confidence interval on their bias [-6.8 to -5.0 ppb, assuming a normal distribution]. This is very narrow due to the large sample size, and does not overlap with the 95% confidence range obtained from the St-Petersburg analysis [-3.2 to 4.4 ppb], indicating that there is a significant difference between TCCON and St. Petersburg FTIR, contrary to what the authors claim.However looking at Yoshida et al., the bias ranges between -12.3 and 4 ppb depending on the station (ignoring stations with N<20). Therefore the overall average bias (drawn from all individual pairs) is a poor indicator of the potential bias at one particular site. This applies to all studies cited.

In principle, we agree that overall average biases give not enough information about the potential bias at one particular site. Nevertheless, these characteristics are widely used in the literature, and we mentioned these studies in our discussion and compared them with our results. We added respective statement into the discussion.

(4) Given a potential NDACC-TCCON bias, the range of biases found in the TCCON-GOSAT comparisons, and the uncertainty on the St. Petersburg measurements due to the limited amount of data, the statement that "measurements of Saint-Petersburg could provide reasonable agreement with satellite data" could be applied to a fairly wide range of potential St. Petersburg-GOSAT validation results. The authors need to assess this range, in order for their results to hold any real value.

We changed the statement in the revised copy of the paper.

Specific comments

Page 7042

- line 17: Here you list NDACC and TCCON acronyms, without specifying what they stand for (something you did do for GAW and for TCCON on the next page).- **Corrected.**
- line 20: Here you mention SCIAMACHY, AIRS, IASI and TAS studies. Do you imply validation studies or development studies of the algorithms? I ask because I find the listed references to be a mixture of both and I would have selected different examples depending on the focus. In any case the listed references are a set of examples thus add e.g. as in (e.g. Xiong et al., …) – **We mentioned all kind of studies, without specific focuses, e. g. is added.**

Page 7043

- line 8: … gases in the infrared range – **Corrected.**
- line 12: infrared **- Corrected**
- line 13-14: …are also carried out by the international… **- Corrected**
- line 29-Page 7044, line 1: comparing them with FTIR measurements from the TCCON network – **Corrected.**

Page 7044

- line 2: Later, comparisons between XCO2 and XCH4 obtained with other GOSAT retrievals and TCCON… **- Corrected**
- line 11: performed at ~60°N using modified NDACC retrieval algorithms.- **Corrected.**
- line 19-21: Move "for the optical path differences of 180 cm" from the end to the beginning of the sentence – **Corrected.**
- line 22: PROFITT – **Corrected.**

Page 7045

- line 10: for retrieving greenhouse gases – **Corrected.**
- line 11: the nearest upper air soundings site Voejkovo – **Corrected.**

Page 7046

- line 24: in most cases. Replace “most” by the number as in “in 6 out of 9 cases”- **Corrected.**
- line 1: To compare ground-based and GOSAT XCH4 near Saint-Petersburg, we found… - **Corrected.**
- line 15: Does Figure 1, correspond with the data found in Table 3? If so mention this or move to relevant paragraph.- **No, Fig. 1 shows all available data without filtering.**
- line 25: already detailed in my general comment – **We removed part of this phrase.**
- line 27: shows slightly larger standard deviation. I would not use those words for a difference between 17 and 16.9 ppb, certainly regarding the large errors and the fact that in table 2, 17 becomes 16.9! The message should be that they are very similar – **Corrected.**

Page 7047

- line 5: smaller than the deviations – **Corrected.**

Page 7048

- line 20: Several other algorithms for XCH4 retrieval from GOSAT data were compared… Even better would be to list their acronyms (RemoTeC, UoL, etc. and references) – **References are added.**

Page 7050

- line 1: delete “over”, - **Corrected**
- line 12: delete “from the Earth’s surface”- **Corrected.**

Page 7059

- Table 3: “априорный” ? – **Corrected.**
- Table 3: V02.xx instead of V01.xx in the second line of the table – **Corrected.**

Page 7060

- Fig 1: …near Saint-Petersburg with FTIR(SPb) and GOSAT (data version… - **Corrected**

Page 7061

- Fig 2: line 1: St. Petersburg (try to be consistent Saint Petersburg, Saint-Petersburg and St. Petersburg are found in the text) – **Corrected.**
- Fig 2: line 3: Change “The line 2 corresponds” to “Line 2 corresponds “ -**Corrected**
- Fig 2: the order in which the data is plotted is confusing. 3=V01, while 1=V02, line 4 is shifted V02 and 5 shifted V01.?? I would make V01=1, V02=2, shifted V01=3,shifted V02=4 and the SPD=GOS line = - **Corrected**

Yours sincerely.
Nikolai M. Gavrilov, Maria V. Makarova, Anatoly V. Poberovskii, Yury M. Timofeev.