Reply

To the interactive 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 #1 Received and published: 8 October 2013

First of all, we greatly appreciate the referee’s helpful comments and suggestions. We thank the referee for the time to improve our manuscript. We have carefully considered the referee’s comments and will correct our manuscript as much as possible in the revised version. The following are our responses to the referee’s comments (given below in bold font):

Gavrilov et al. describe ground-based FTS measurements of column-averaged methane at St Petersburg and a comparison with GOSAT data.

General remarks

Previous GOSAT validation studies have been based upon the use of TCCON data. It is nice to see that also FTS measurements of the NDACC network are now used for validation of GOSAT methane as it had been successfully performed in the previous case of ENVISAT validation. Complementary use of methane data from both NDACC and TCCON networks has become possible after a complex process elaborating nontrivial recipes on how to achieve agreement between NDACC and TCCON data of methane, as documented in a series of papers by Sussmann et al. Previous GOSAT validation studies have been based upon (TCCON) sites below 55 N and it is certainly interesting to look at a station located at 60 N as done in this paper.

Therefore, we encourage publication in AMT. However, the paper needs several major revisions prior to final publication.

Major points

1) I perfectly agree with the comments of “anonymous referee #3” (dated 23 Sep 2013, unfortunately entitled “referee #1” within the review-text). I would request to address all comments of referee #3 properly before the paper can be finally accepted. In particular I confirm, the use of the standard deviation as a proxy for the statistical error of a mean (bias) is not correct. The (95 % confidence) standard error of the mean (2 sigma / sqrt(n)) should be used instead. As stated by referee #3 this implies that two major conclusions of the paper will be reversed: The bias derived by Gavrilov et al. does not agree with the bias derived by Yoshida et al., and in fact is significantly different considering 95 % confidence intervals. Furthermore, it can only be concluded that the limited data set of Gavrilov et al. does not allow to infer statistically significant differences between the biases for GOSAT V01 and V02.

Probably, the referee #1 knows a joke about “lie”, “big lie” and “statistics”. In fact, statistics itself cannot prove or reject anything. It gives just probabilities. We may find many funny results with statistical confidence 99.99%. To make conclusions in all cases we should take into account physical reasons. From these physical reasons, we cannot say that biases for GOSAT V01 should be the same as for V02. Results in Tables 1 and 2 also show smaller biases for GOSAT V02 compared to V01. The number of data in Table 1 is not much smaller than those in previous comparisons between GOSAT V01 and aircraft and ground-based northern FTIR stations. This number cannot be increased, because GOSAT stops processing and distribution of V01 data. As far as Table 1 is already exists, we think it may be useful for history of such comparisons.

To get 95% confidence we need several time increases in the data number in Table 2. We need about 10 - 20 years of measurements for that. Should we wait until then for the comparisons? No, the history of science knows many examples when experimentalists compared a few first data points and confirm their results later using better statistics. In fact, we do not validate the GOSAT data. We verify the data of our measurements

using satellite data. Our results, showing that our measurements near Saint-Petersburg at least not contradict to satellite and other ground-based FTIR data, are important for us and for other users of our data. Nevertheless, we understand that such comparisons are very preliminary and we will work on further validation of our measurements.

Correct your wording throughout the text, to make clear whether you are talking about i) “standard deviation” (of one data set), “standard deviation of differences” (between two coincident data sets), “standard error of the mean (bias)” (see above).

We consider different quantities and calculate different standard deviations. We corrected wording to make it clear.

2) The authors follow the NDACC retrieval strategy suggested by Sussmann et al. (2011), but important site-specific details are not mentioned. What are the mean degrees of freedom for signal, what is the mean water vapor column at St. Petersburg? Why do you restrict solar zenith angles to +/- 3 h around local noon? Did you look at a possible airmass dependency of you retrievals? Also the station altitude is missing.

We added listed information to the paper. We restrict solar zenith angles to +/- 3 h around local noon because the GOSAT data for Saint-Petersburg is usually available for around the local noon. We were trying to analyze airmass dependencies using the ideology of N. M. Deutscher et al. Atmos. Meas. Tech., 3, 947–958, 2010, but our analysis shows that we probably have not enough data to make reliable conclusions about such dependencies. This may be because our measurements are performed in conditions far from background in presence of relatively large natural and anthropogenetic CH4 sources. Hence, we need more data for definite conclusions.

3) I do not agree with the statement (p 7043, l 10): “For validation of satellite observations of greenhouse gases, the special monitoring network TCCON (…) was set up. While satellite validation is an important goal of TCCON it is certainly not the only goal. The authors are invited to read Wunch et al. (2011) and amend other TCCON goals.

In fact, goals of TCCON are not very important for the present paper. We moved and changed the mentioned phrase.

4) Clarify on p 7049: What is limiting the statistical error of your bias compared to Yoshida et al.: the limited number of measurements at St. Petersburg or the high variability at St. Petersburg?

We think both reasons are important: high variability and limited statistics. Both require increasing number of measurements.

Technical corrections

1) You use various non-standard wordings for XCH4: “methane column-mean mole fractions (abstract)”, “average CH4 mole fractions (abstract)”, “column average mole fractions of methane (p 7043)”, “column methane mole factions (p 7050)”. Use either “column-averaged dry-air mole fractions of methane” or simply “column-averaged methane”. – **Corrected.**

2) p 7046, l 24-26: Is 16.9 larger than 17.0? - **The text is modified.**

3) p 7046, l 25: -13 > -13 ppb; ppb is also missing at several other places in the text. – **Corrected.**

4) p 7047, l 1-3: “When several ground-based or satellite XCH4 values were registered during a day, we used respective daily means in Tables 1 and 2.” This explanation should be placed earlier in relation to the first time where Table 1 is addressed. Also clarify: Does Table 3 also show number derived from daily means? Add this information to the caption of Tab. 3. – **No, Table 3 uses individual spectra. Text is corrected.**

5) p 7047, l 11: 17 pp > 16.9 ppb – **The text is corrected.**
6) p 7049, l 7: I cannot see the range 0.01 – 1.8 % from Tables 2 or 3. – **Yes minimum value in Table 2 is 0.04%. Corrected.**
7) p 7049, l 8: 0.2 +/- 0.8% > 0.2 % with a standard deviation of the differences of 0.85
%. – **Corrected.**
8) Use 1 decimal for ppb0s (example 16.9 ppb) throughout the text, never use roundedvalues (example 17 ppb). – **We think, this suggestion is too much optimistic taking into account small numbers of analyzed measurements. It would be more appropriate to use rounded values for our data in ppb everywhere in the text. We keep 1 decimal for ppb in Tables.**
9) Tables 1-3: Medians are given in the Tables. The Discussion only deals with the mean values. Therefore you could remove the medians from the Tables.
For random Gaussian processes Average=Median. In all tables Averages are very close to Medians. We consider this as an indication that our statistical estimates are not as bad as one may assume. We added this statement into discussion.

Yours sincerely.

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