"**Space-borne Observation of Methane from Atmospheric Infrared Sounder Version 6: Validation and Implications for Data Analysis**"

by X. Xiong et al.

We appreciate the suggestion from O. Garciar (Referee). A substantial revise has been made after carefully considering every point as suggested by the Referee. This change makes the manuscript in a much better shape. Below is the answer to address each suggestion point-by-point.

General Comments

This paper presents the new version of the methane (CH4) retrievals, one of the most important greenhouse gases, as observed from the remote sensor Atmospheric Infrared Sounder (AIRS) aboard NASA's Aqua platform as well as a comprehensive validation study. The paper is well structured, and concise. However, the methodology used (retrieval and validation strategies) as well as all the discussion and plots shown are equivalent of an author's previous work entitled "Mid-upper tropospheric methane retrieval from IASI and its validation (http://www.atmos-meastech.net/6/2255/2013/amt-6-2255-2013.html)", but using the Infrared Atmospheric Sounding Interferometer aboard MetOp satellites. Therefore, the authors should make clearer that the approach used in the paper under discussion is a 1:1 equivalent of their prior IASI work. This clarification allows both studies can directly be compared in the current paper. In fact, including an intercomparion of the methane IASI and AIRS products would provide an added value to the paper. For example, the authors could document whether both methane products are consistent and, then, they can be combined to create a common methane database. Finally, the overall treatment of the theoretical and experimental errors is imprecise through the manuscript, so the authors should improve its description and quantification.

Answer: Following this suggestion, in the first paragraph of Section 4 we added:

"Similar to our validation to IASI CH_4 products [Xiong et al., 2013a], we first show the mean error of retrieval profiles, then compare the AIRS retrieved CH_4 with aircraft measurements in four trapezoid layers. As AIRS (20002-present) has measurements in a longer period than IASI (2008-present), more aircraft measurements are available for us to validate AIRS CH_4 production than IASI product, thus allowing us to make a closer examination of the retrieval errors in different seasons and different latitude zones in this paper than what we did for IASI validation. Similar to IASI validation (Xiong et al., 2013a), we also examine the correlation of the AIRS retrieval errors with cloud fractions and DOFs, but in more detail by checking such correlations in tropics and different seasons. An intercomparion of the CH_4 from AIRS and IASI is valuable for us to understand if both products are consistent and whether we can combine them to create a common methane database, however, such a study is out of the scope of this paper and will be addressed in another paper in the future."

Specific Comments

Title

The title of the paper suggests that the implications of the improvements of AIRS CH4 products on the data analysis will be discussed through the manuscript, but I have not identified this discussion. Do the authors refer to the "Error Analysis" discussion addressed by Section 4.2? If so, please consider revising the title of the paper or clarify this issue in the discussion.

Answer: More and more users are trying to use the AIRS and IASI retrieval CH_4 products to study CH_4 emission. However, they tend to relying on complicated mathematical methods but do not recognize these artificial variation and uncertainties imbedded in the retrieval products. For this reason we think it is important to include "implication for data analysis" in the title.

The last paragraph of Section 4 was revised as below:

From both the change of averaging kernels, or peak sensitivity layer, with atmospheric states (including thermal contrast), and the impact of the retrieval errors by DOFs and cloud factors, it is evident that the "observed" spatiotemporal variation of CH_4 by AIRS, as well as other thermal infrared sensors, not only reflects the real change of CH_4 in the atmosphere, but also includes some artificial variation due to the change of sensitivity, or DOF, and/or contamination by cloud. So, for the analysis of CH_4 distribution and/or seasonal variation using the retrieval products from AIRS (same for other thermal infrared sensors), it is important to recognize these artificial impacts. Some filtering of the data, for example, to select data based on the averaging kernels, DOFs, thermal contract, and/or cloud amounts, will help to remove these artificial impacts partially; however, it is impossible to completely remove their impact to get the real spatiotemporal variation of CH_4 accurately based on the retrieved CH_4 mixing ratios only.

Introduction

The paper basically addresses the improvement of the AIRS-Version 5 CH4 retrievals, but any information about its expected and/or observed uncertainties is given. Thereby, include this information to better quantify the real improvements done in V6.

Answer: This is not exactly right. The main purpose is to validate the AIRS-V6 CH_4 products and document it, so as to help users to better use and cite it. Comparison of the validation results with AIRS-V5 is not a major part, since such comparison needs to use the same samples. We did make substantial change in the description and discussion of uncertainties in this paper, as suggested by the Dr.O Garcias, and we think it is better to keep such discussion in later parts of this paper (see Section 2.1).

Section 2

1) To avoid being redundant in the title of the section and subsections, consider simplifying the title to "CH4 Retrieval in AIRS-V6".

Answer: Good suggestion. We did it.

2) Including a brief analysis of the impact the thermal contrast and the latitude on the sensitivity would help the discussion of the discrepancies observed in section 4.2. This can easily be done by plotting the DOFS versus latitude distinguishing the land/ocean pixels as Figure 2.b.

Answer: Good suggestion. This can be seen from the replot of Fig.3 by adding the change of DOFs with latitude.

3) The authors state that the increase of water vapour in summer enhances the total sensitivity of AIRS. Include a brief explanation supporting this statement: how does the water vapour increase the total sensitivity of methane? How is the cross-interference of water vapour taken into account in the retrieval strategy?

Answer: Thanks for pointing it out. We realized the confusion in the peak sensitivity level and sensitivity(characterized using DOF) itself, and revised the context. In actually, water vapor does not increase the sensitivity, and its impact is to move up the peak sensitive level. The uncertainty of water vapor is taken into account while solving Eq. (1), and more detail can be referred to Xiong et al. (2008).

Section 3

1) Section 3.1: The validation is carried out by using five measurement campaigns. Specify the added value of including each validation dataset.

Answer: Below is what we added as suggested:

The HIPPO campaigns provided a unique dataset for validation over a wide latitudinal range but mostly are over the ocean, and other campaigns provided measurements for validation over land (mostly over the North American) covering a longer period and in different seasons.

2) Section 3.2: The aircraftprofiles used for the validation are extended by using an Atmospheric General Circulation Model. Consequently, the smoothed aircraft profiles are a combination of the two experiments with their respectively uncertainties. Including an estimation of the expected errors in the smoothed aircraft profiles due to this extension would complete the error analysis discussion.

Answer: As suggested, before the end of Section 3.2 we add the following:

As estimated by Xiong et al. (2008), the biases in the extended part of profile would lead directly to biases in the convolved data, and these biases depend on the averaging kernels. An estimate by assuming a 50% change in the extended part of profiles shows the change in the convolved data is mostly less than 20 ppbv.

3) Section 3.3: The authors introduce an experimental quality flag based on the CH4 retrieval values and their experiences, but any postquality control implies a prior knowledge about the expected results. Have the authors thoroughly checked the criteria that defined the quality flags in the previous steps (quality flags 0 and 1)? Have the authors investigated the causes of the observed oscillation in the retrieved CH4 profiles (clouds . . .)? If so, please include this

information as well as a brief explanation about the criteria defining the quality flag equal to 0 and 1.
Answer: Yes, as the developer of AIRS CH_4 algorithm, I was responsible for the setting of quality flags.

In the end of the first paragraph of Section 2.1, we added some description as below:
"There are three types of quality flags for CH_4 retrieval products: 0 represents high quality, 1 represents good quality, and 2 represents bad quality and is not recommended to use. These flags were determined based on the retrievals of upstream products, especially the water vapor, as well as the convergence of CH_4 retrieval."

In Section 3.3, we also added
"The reasons for the oscillation are complicated, and one way to reduce the oscillation is to add constraint in the retrieval algorithm. In AIRS-V5 and V6 a static empirical damping factor is used, thus this constraint works fine for most cases but may become too weak in some specific conditions. A flexible damping factor changing with time and location is under consideration in the future versions of AIRS algorithm. Another factor is related with cloud, as the cloud-clearing radiance is used in the retrieval. We found quite a few cases with the oscillation occurred when the FOR (with 9 FOV) is not uniform."

Section 4
1) The presentation and discussion of the validation results are mixed through the two subsections. Thereby, consider merging both subsections into one. Also, part of the discussion presented in the Error Analysis could be included in the section 2.1. For example, the first paragraph of this section details the error sources affecting the CH4 retrievals, but the authors really provide a description of the error sources affecting the comparison results, i.e., the "theoretical errors" affecting the retrievals (the two first error sources cited) and errors introduced by the comparison strategy used. Therefore, consider moving the first part of this discussion to the section 2.1, where the retrieval strategy is described. Also, include an overall estimation of the expected theoretical uncertainties due to the different error sources. This information would be very useful because it indicates where a special effort should be paid to improve the precision of the CH4 products in the future.

Answer: We seriously considered this suggestion and made changes accordingly. First, we moved the first part of the "theoretical" error discussion to the section 2.1 as suggested. Instead of merging both subsections as suggested, we chose to modify the subtitles. We think the title of subsection 4.2 "error analysis" is too broad and we changed it to "Relation of the Retrieval Error with DOF and Cloud Fraction"

2) Consider moving the Figure 6.b to the section 3.1, where the different validation datasets are introduced, because it is a bit confused that the authors show the number of samples above 350

hPa when the section 3.2 states that the aircraft profiles with their ceiling beneath the 350hPa pressure level were not used in the validation.

Answer: Figure 6b was added here according to another reviewer's suggestion in last round. The caption of Figure 4 states clearly " The locations of the aircraft measurement profiles used for validation", in which the flight segments not used have not been plotted. We think it is better to keep Figure 6b figure in current way, as it helps the readers to understand the larger error above 350 hPa in Figure 6.
For clarification, in the first paragraph of Section 3.1, we added
"In each campaign, only these flight segments that have enough measurement samples covering the altitudes at least between 700 to 350 hPa are selected. Figure 4 shows the spatial locations of these aircraft profiles used".

3) The comparison between the AIRS and the smoothed aircraft profiles is done for four partial columns spanning from 777 hPa to 272 hPa when the AIRS is only sensitive to the CH4 variations in one layer (recall Figures 2 and 4). Thereby, what is the added value of comparing these 4 layers? For example, for the uppermost layer the authors are rather comparing the extended aircraft profiles (by the climate model) to the priori profiles used in the AIRS retrievals. Therefore, consider including only the comparison for the most sensitive layer, whose levels could be redefined taking into account the sensitivity studied done in previous sections. This layer could be used in the subsequent analysis.

Answer: We agree this is a good idea, but this is not a easy job and could be another study. Payne et al. (2009) (listed in the references) has tried to derive the product in the most sensitive layer. Xiong et al. (2009) also made a similar study. However, we found it is hard to define such a layer and remove the impact of varying DOF. Even we redefine such a layer, it is hard to apply the averaging kernel to smooth aircraft profiles. So, such a product is hard to use and validate, and/or compare with model.

Since most validation studies did the same way as we did here, we chose to keep the current figures and analysis. Also as we are used to measure or simulate data at different pressure levels, for the perspective of users, we also think it is better to give the validation results in these pressure layers, which will make the future users to easily understand.

Due to the limit of information content, 10 retrieval layers are used in the retrieval, and we validated the CH_4 mixing ratios in the middle 4 layers.

4) Have the authors analyzed/observed differences between the day-time and night-time AIRS overpass retrievals or between ocean/land pixels? Analysing the possible non-linearity of the AIRS retrievals would be explain part of the discrepancies found or reject these error sources in the discussion. For example, Figure 8 shows a worse correlation for comparison in summer when the AIRS sensitivity is expected to be higher. Also, the authors suggest that part of differences observed for mid-latitude region is due to the latitudinal gradient, but not include

the typical values of these gradients to justify this statement. Consider including this information.
Answer: ***We have ever checked the day/night, or land/ocean difference but have not analyzed it carefully, because such a difference of CH_4 from AIRS is not only from the variation in reality (which is impacted by surface emission, atmospheric chemistry reaction in daytime, and transport), but also has "artificial" change resulted from the change of peak sensitive layer and the information content in the retrieval. Noise in radiance measurement and cloud makes the image of the CH_4 very noisy.***
About "latitudinal gradient" for mid-latitude region, we agree with the referee. "Latitudinal gradient" for mid-latitude region may or may not be smaller than other regions, what we want to say here is the "variation range", and we changed it in the context. Thanks for pointing out this.

5) The description of the error sources is a bit imprecise (section 4.2), because the authors do not provide any estimation or error value. This makes difficult an interpretation of the relative influence/impact of the different error sources on the AIRS retrievals. An example could be when the authors detail the influence of the temporal collocation.

Answer: The subtitle of "error analysis" is inappropriate and we changed it. Also we moved the error discussion in the first half paragraph of Section 4.2 to Section 2.1. as mentioned before.

6) Regarding the discussion of Figure 10, include the description of the cloud fraction parameter (i.e., how this parameter is calculated) and clarify the discussion. The authors state that the DOFS for high northern latitude (above 60_N) is between 0.6-0.8 when the figure only distinguishes between tropics and other regions. Also, the best correlation is found in summer, when this information is not contained in the referred figure. Also, as aforementioned, including a figure in section 2 plotting the DOFS versus the latitude could help the discussion of the discrepancies observed. Finally, the authors compare independently the correlation of the retrieval error versus the cloud fraction and DOFS, suggesting the latter is the dominant factor. To analyse this I would recommend using a linear multivariate approach to take into account two variables at the same time. This analysis will allow the authors to estimate the real relative weight of each factor on the retrieval error.

Answer: in Section 4.2, we added a description of the cloud fraction parameter as below: "As the AIRS retrieval is based on one FOR, for each retrieval there is a cloud fraction parameter in the AIRS product, representing the cloud fraction in this FOR. The mean cloud fraction (same for DOF) is computed as an average of cloud fractions for all the collocated profiles with an aircraft measurement."

For clarification, we removed "DOFS for high northern latitude (above 60_N) is between 0.6-0.8" since it is not shown clearly in Figure 10; added ***(not shown)*** after "… the best correlation is found in summe.

DOFS versus the latitude has been added in Figure 3 as suggested.
Also we added a linear multivariate approach in the context, as below

Using a linear multivariate approach, i.e. a linear regression of the retrieval errors against DOFs and cloud fractions (f), in the tropics we get,

$$Bias = -79.2 + 72.4 * DOFs - 3.0 * f,$$

and the correlation coefficients with DOFs and cloud fraction of 0.74 and -0.60, respectively.

While in other regions, we have,

$$Bias = -63.6 + 57.1 * DOFs + 28.7 * f,$$

and the correlation coefficients of 0.36 and -0.04, respectively.

7) The discussion of the validation and error results do not include any comparison/reference to the previous AIRS version. Therefore, as aforementioned, the readers are not able to quantify the real magnitude of the improvements done in the new version 6.

Answer: We agree such a value will interested to some users. However, as we mentioned before, the major purpose of this paper is not for comparison with AIRS-V5. As the disk to save the old validation data and results of AIRS-V5 was damaged, we decided it is not worthy to invest time to trace back to validate V5.

Considering the validation data used in this paper is not the same as in V5, a direct comparison of the validation results and put them in a table is unfair and will be misleading, so we chose not to do so. Instead, in the end of 2nd paragraph of Section 4.2, we added: "Compared to AIRS-V5 that the retrieval bias has an apparent dependence with latitude (see the lower left panel of Figure 6b in Xiong et al., 2008), and the dependence of the retrieval errors in AIRS-6 with latitude is much smaller. In layer around 400-600 hPa, the correlation coefficient between AIRS and in situ measurements is 0.87 in AIRS-V6 while this value is 0.72 in AIRS-V5. However, we should notice the samples used in AIRS-V5 validation are not exactly the same as we used here."

8) In order to better follow the discussion of the results, consider summarizing the validation results in a Table. Also, including the uncertainties found for the previous AIRS version would help to better quantify the improvements of the new version.

Answer: See my answer in 7)

9) Revise the number of significant figures included in the Figure 7, 8 and 9.

Answer: Did as suggested.