Reply to comments raised by Referee 1

The original comments are written in plain text, our replies are given in italic

1) General comments

This study compares the accuracy of two algorithms in determining methane (CH4) concentrations from IASI sensor data. The retrieval results with the two algorithms are compared with each other and with independent observations and model data using various methods.

It is crucial to validate IASI CH4 data comprehensively for scientific use, and this paper's scientific significance is high. Therefore, I believe this paper's content deserves to be published in AMT.

However, the structure and description of this paper are somewhat confusing. In particular, the intercomparison part in Section 4 needs a more organized structure. The "short summary" subsection should be included in each intercomparison subsection, with tables, to aid readers in comprehending the validation results.

Also, regarding the gap in the data for 2013, it is important to include clear communication in the data section that this issue exists. Then, a comparative validation distinguishing between before and after data, or at least an intercomparison between before and after RAL and LMD, should be performed to show the presence or absence of an impact.

In addition, several figures need more legends or are unreadable. Please improve visibility.

Considering the above, I recommend that this paper should be published after examining the issues and making necessary revisions.

I would like to thank the reviewer for the useful comments. A common comment of both reviewers pertains to the structure of the paper and upon reflection, I have to concur with their observations.

Particularly, the comparisons with independent reference data are impacted by the various temporal instabilities that are present in both datasets be it gradual or sudden. In the current structure, these temporal issues are discussed in depth after the in situ comparisons (in part because the temporal hiatus was discovered during the project). We therefore suggest a change in the structure of the paper in which we first focus on the LMD and RAL side-by-side comparisons using CAMS as an intermediate, including a figure that shows the LMD-RAL global bias distribution (as the bottom row of current figure 4) as a function of time (see figure below as an example). Then, after a discussion on the various temporal dependencies that are at play, we *will turn to the independent reference data, where we break up the RAL data into two segments due to its bias shift in 2023.*

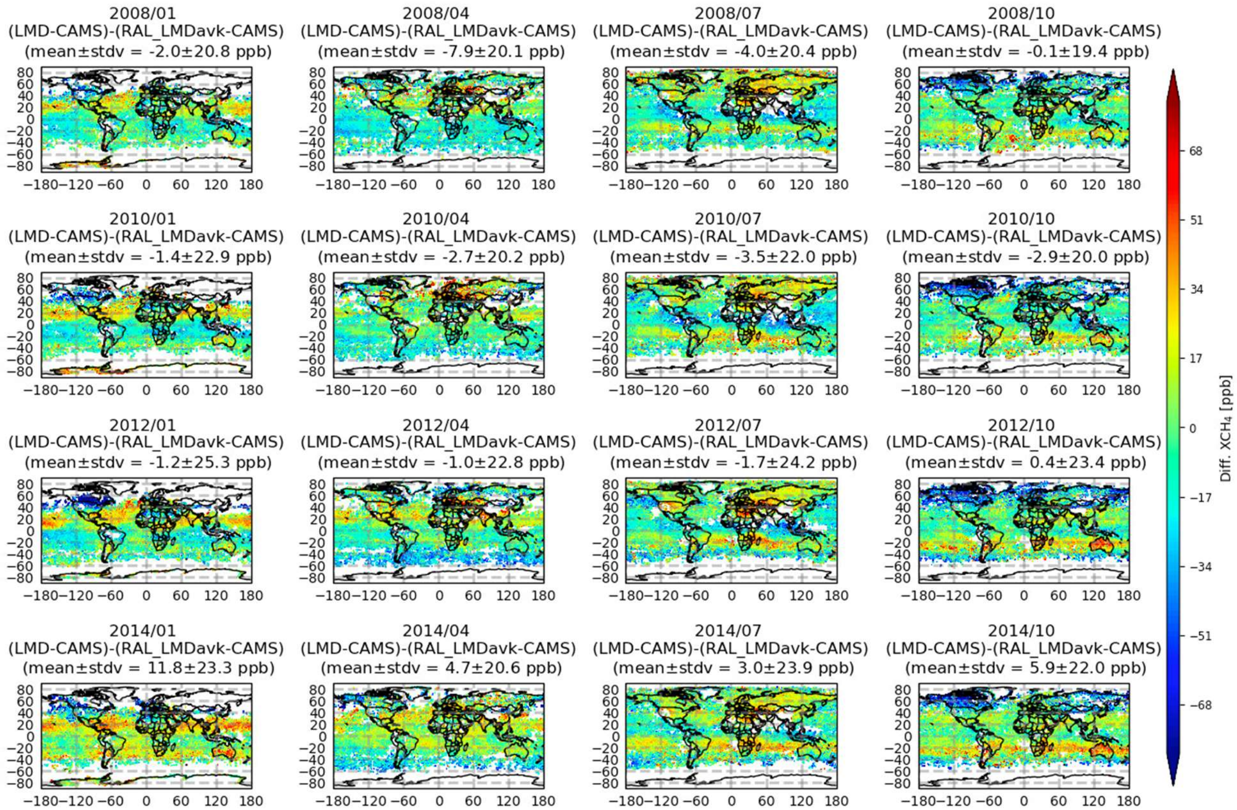

Figure 1: The difference in LMD and RAL_LMDavk biases relative to CAMS for different months (columns) and years (rows)

2) Specific comments

・Page3 Line20 (P3L20): about IFOV of RAL measurements

Why use the IFOV with the highest brightness temperature among the four IFOVs instead of the average of the four?

The IFOV with highest BT is assumed to be the one with least amount of potential cloud contamination. The RAL team therefore saw no obvious benefit in averaging over the four IFOVs.

・P8L19: about vertical sensitivity of LMD and RAL_LMDavk

You note that there is a difference in altitude concerning the sensitivity of the two, but how much difference would this be in mtCH4? Can mtCH4 be evaluated using, for example, climatological values?

As can be seen in figure 2, the altitude range to which both products show significant sensitivity differs substantially with RAL having peak sensitivity at ~600 to 400 hPa and LMD between 400 and 200 hPa. Furthermore LMD's near surface sensitivity is practically 0. Smoothing RAL with the LMD averaging kernel does *align them better but there still is a vertical shift in the peak sensitivity region (with LMD's peak at lower pressures). Of importance to note here is that therefore LMD is more sensitive to methane residing in the so-called UTLS (Upper Troposphere-Lower stratosphere) and it is just in this region that a steep decline in the CH4 concentrations takes place as one moves from the troposphere into the stratosphere. The difference between RAL and LMD mtCH4 (due to smoothing errors alone!) is thus highly dependent on the true atmospheric state and the location of the tropopause height. A significant hint as to the magnitude of this vertical sensitivity bias are the two bottom rows of Figure 4, where a direct comparison between the 2 mtCH4 products (3rd row) show a bias around -12 to -15 ppb, while if we compare them individually against CAMS their bias ranges between -1.7 to 0.4 pbb (4th row). Thus we can state with some confidence that the differences in vertical sensitivity between LMD and RAL_LMDavk still amounts to a ~10 ppb bias, but locally they can be much higher (as can be observed when comparing specific regions between rows 3 and 4 of Figure 4)*

As to whether mtCH4 can be evaluated using climatological values, we can state that these can indeed be used for a qualitative evaluation to some extent (for instance to verify trends and large-scale spatial distributions), but we felt that in the context of this work a chemical transport model is a much more suitable dataset.

- P8L23: about grid averaging

Why did you use a 1-degree grid for intercomparison? This corresponds to a 100 km grid on a spatial scale, which might be too large, considering that the source of CH4 is more localized than CO2.

It is true that CH_4 sources are more localized and that, if we were to focus on issues such as emission hot-spot localization etc., a 1-by-1 degree resolution might be too large. However given that IASI's near surface sensitivity is relatively low, a fair degree of mixing/dilution has already taken place once air masses reach the altitude range where the satellite's sensitivity is at its optimum. Furthermore, in this article we focus on global larger scale phenomena (regional biases, long-term trends and seasonality) which do not require a high resolution approach.

- P9L4 and P9L10-L17: about selection criteria

The time frame for simultaneous observation with IASI is currently at ±6 hours. However, I believe this interval may be too long, considering the horizontal and vertical transport of CH4. CH4 distribution can vary significantly over such a period, so it may not be accurate to consider the observation as genuinely simultaneous.

Additionally, comparing ground-based FTIRs to the satellite dataset is unfair due to the differences in time and longitude ranges for each comparison. Can the conditions be as consistent as possible?

As with the previous answer, IASI's vertical sensitivity range dampens the effect of localized sources and a more relaxed collocation approach can thus be justified. That said, finding the right collocation criteria is always a compromise between having a large enough data sample and minimizing the introduction of additional biases into the comparison. Using our fairly relaxed collocation criteria, we indeed cannot guarantee that on a station to station basis, no biases are introduced into the system. It is therefore important to look at the network dataset as a whole, certainly when focusing on larger scale phenomena.

- P9L5: about the mean of the satellite values

How many satellite data points are usually averaged? And how large is the variability (stdv) of these data?

The number of individual RAL satellite data points that are typically averaged range between 1 and 24, with a mean of 8.1. The stdv of the RAL co-located CH4 is about 12.4 ppb. For LMD the number of averaged data ranges between 1 and 15, with a mean of 3.9. The stdv of the LMD co-located CH4 is about 9.1 ppb.

- P10L21: about eq.(7)

Please tell me how this equation was derived. Also, please explain what $C'_{r,R}$ on the left side represents.

$C'_{r,R}$ is the retrieved RAL XCH4 where the impact of its own a priori has been replaced by the TCCON a priori. The equation stems from Rodgers [Rodgers, C.D. (2000) Inverse Methods for Atmospheric Sounding: Theory and Practice. World Scientific, River Edge.] where the retrieved quantity $C_r = C_{ap} + A(x - x_{ap})$, with x being the true state profile and x_{ap} the a priori profile and C_r and C_{ap} the retrieved and a priori mole fractions respectively. The equation directly follows from when you were to calculate the difference between two C_r using 2 different a priori profiles. We will clarify this in the text.

- P11L9, L24, and L28: about equations (9), (11), (12)

What does the right-hand side of equations (9), (11), and (12) represent, respectively? Please add an explanation.

Equation 9 is as equation 7 (replacing the influence of the a priori). Equations 11 and 12 pertain to the LMD product which does not include averaging kernel information. Therefore it contains no a priori information and no transformations regarding an a priori (as with equations 9 and 7) is required. Instead, it employs a sensitivity profile with a weighting function wf (see equation 1).

However in the equations there looks to be a space between w and f where there should be none as it is one parameter, which leads to confusion. This will be fixed and information will be added in the text.

- P13L5-L6: about internal consistency

Why was the internal consistency only checked in October 2014? Could there be differences depending on the season, especially summer and winter? Also, is there any impact of the 2013 gap?

The internal consistency has been checked only for October 2014 because the main objective here was to analyze the possible impact of slight instrumental inter-pixel calibration defects in the L2 data. There is no need to perform seasonal analysis to investigate this kind of systematic effects. The number of analyzed pixels (~90000) on a global scale is large enough to assess the systematic instrumental biases.

The 16th of May 2013 gap is a result of a change in the ground configuration for the spectral calibration introducing a band dependent PSF (Point Spread Function). This modification improved, among others, the interpixel spectral calibration especially in Band 2 as shown in the following figures. However, we do not expect these near negligeable interpixel spectral departures to have a significant impact on Level 2 retrieval since these departures especially in Band 2 are much weaker than the specification for IASI instruments ($\Delta v/v = 2.10^{-6}$).

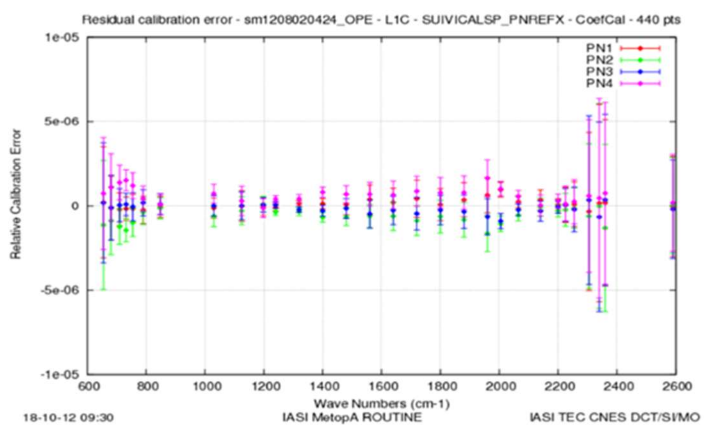

Figure 2 : Residual inter-pixel relative spectral calibration ($\Delta v/v$) for the orbit of 2012/08/02 04h24 (Jacquette et al., 2013)

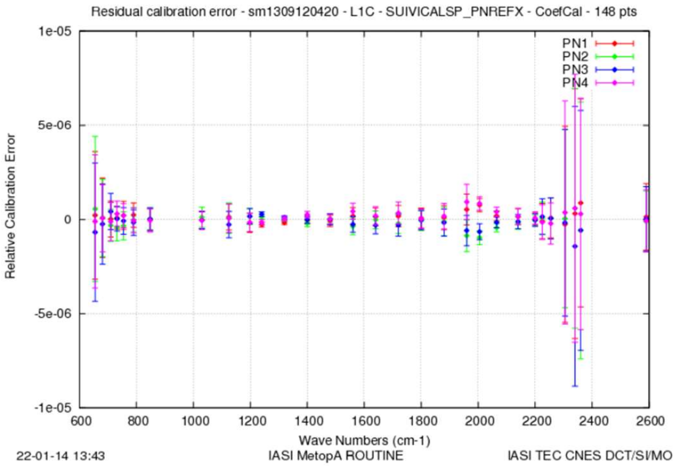

Figure 3 : Residual inter-pixel relative spectral calibration (Δν/ν) for the orbit of 2013/09/12 04h21 (IASI quarterly performance report from 2013/09/01 to 2013/11/30)

E. Jacquette, B. Tournier , E. Péquignot , J. Donnadille , D. Jouglet , V. Lonjou , J. Chinaud , C. Baque , L. Buffet : IASI spectral calibration monitoring on MetOp-A and MetOp-B, 3rd IASI conference, 4-8 February 2013, Hyères, France

- P13L30: IFOV selection

The authors claim that "RAL selects the best IFOV among 4 of them.", but in the previous section they mention "the one with the highest brightness temperature". Why is this the best?

The IFOV with highest BT is assumed to be the one with least amount of potential cloud contamination. We will mention this in the revised paper.

- P14L16: about SAT

What is “SAT”? Does SAT mean RAL and/or LMD measurements?

SAT pertains to satellite and hence either RAL or LMD (smoothed and unsmoothed). This will be clarified in the text.

- P15L13-L14: about the correlation between HIPPO and RAL measurements

What is the correlation between the two values obtained from the best fit?

I'm not sure what you are implying with 'best fit'. Lines 13-14 only mention the linear fit in the correlation plot. If this is the fit you allude to, then by definition R=1.

- P15L15-L18: comparison with IAGOS

Why is the correlation coefficient with IAGOS lower than other comparisons? Is such a low correlation coefficient due to the spatial bias of IAGOS measurements?

There are many factors at play here. One is indeed that the IAGOS data is more geographically dispersed compared to HIPPO and Aircore and its descend and ascend profiles are typically located at or near urban centers, which could imply that local biases have a stronger effect on the IAGOS data. Another important aspect is that the vertical range covered by IAGOS is more restricted compared to HIPPO, and even much more so compared to Aircore. This entails that it relies more on the extrapolation of its data, yet another source of uncertainty. We will add this to the discussion of the results.

- P17L24-P19L9: about subsection 4.7

This subsection repeats previous statements, making it redundant. It would be more useful to list each table in a related subsection.

While we do think that a short summary is of use after each section, you are correct that the information presented under 4.7 goes beyond the scope of a summary and much of its content (and tables) should be presented under the related subsections. This will be applied to the restructured paper.

- P19L10-P19L15: about discontinuity in RAL L2 data in mid-2013

Does the discontinuity in mid-2013 also affect the intercomparisons made in Section 4? For example, the comparison with AirCore and IAGOS was made using all data before and after the discontinuity. Would there be a difference in bias before and after?

Also, on page 23, the authors only compared partial columns between RAL and CAMS in 2012. Is there any difference in partial column bias before and after 2013?

This is indeed a valid point. In the current structuring of the paper, the discussion of the temporal effects come after the comparisons with reference data. This will be changed and comparisons will then be evaluated prior and post May 2013.

The partial column comparisons between RAL and CAMS were carried out for all years (for instance, shown below is the upper-lower RAL bias for January 2014). While there might be *small differences in the absolute values, the observations and conclusions drawn from these comparisons remain the same. We will add a line to convey this message.*

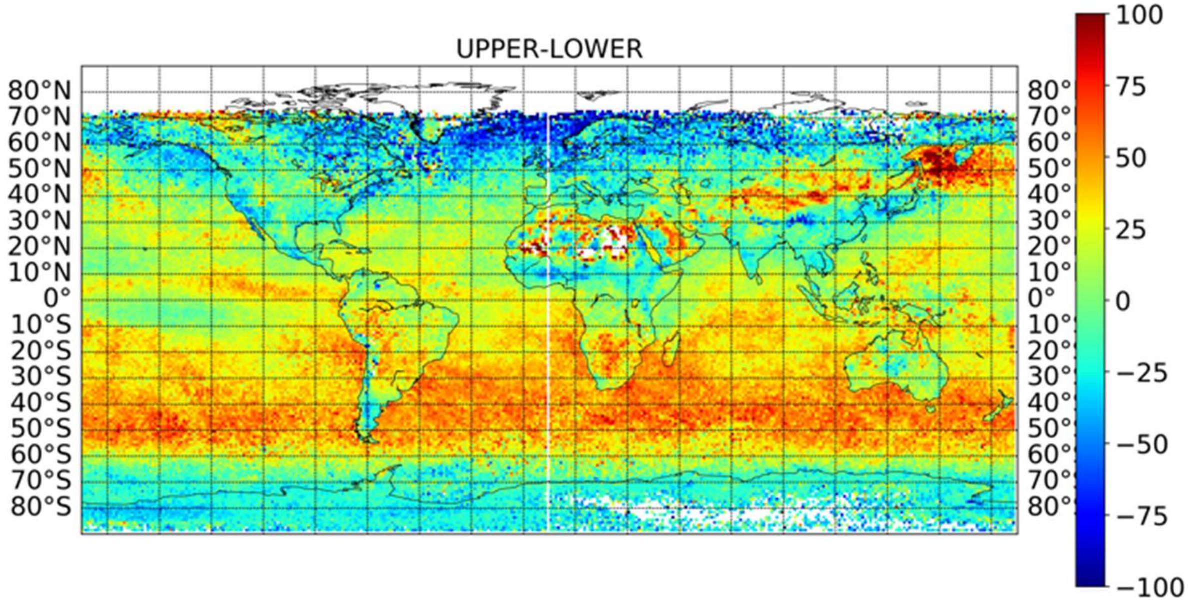

Figure 4: upper-lower RAL qCH4 bias for January 2014

- Figure 5, 6, 7, 12 and 13:

The legend should be relocated or resized to avoid overlapping the plots. In Figure 13, please add a legend.

This will be done

- Figure 5, 6 and 7:

Could you please explain the meaning of the dotted lines in the correlation diagram?

This is a simple linear fit (without forcing the fit to go through (0,0)). We will make this more clear in the figure text.

3) Technical corrections

P2L9: andh => and

P4L23: (about 30 km; (Karion et al., 2010)) => (about 30 km) (Karion et al., 2010).

P5L3: CAMS model) as => CAMS model as

P7L9: in (Massart et al., 2014) => in Massart et al. (2014)

P9L20: According to (Rodgers and Connor, 2003) => According to Rodgers and Connor (2003),

P10L4: “$\boldsymbol{c_i}$” in bold should be plain.

P10L7: respectively; => respectively,

P10L7: XCH4 => XCH4. (need period)

P15L23, L28 and L29: XCH4 => qCH4

Table 1: ParkFalls => Park Falls?

Figure 1: Rikubetsu should be “TCCON (yellow)”, not “NDACC (green)”.

Figure 4: “The last three rows show …” maybe “The last two rows show …”.

Figure 6: In the x- and y-axis labels, “XCH4” should b

e “qCH4”.

Thank you for finding these errors. The suggested technical corrections will all be implemented