Referee comments in bold
Responses in normal type

Major comments

• Averaging kernels have been commonly used in the modelling community for more than 10 years, in particular since the first MOPITT products. In this context, there is no point in comparing the satellite data to simple pressure weighted partial columns from a model. The corresponding figures and discussion should be removed, which would allow better focusing the whole discussion.

The importance of using averaging kernels (AKs) is clearly understood, and they are used throughout the paper to allow comparisons with independent observational data and models on a like-with-like basis, taking into account retrieval sensitivity to vertical structure and prior information. However, if the retrieval of a given vertical layer is dominated by prior information or is mostly sensitive to layers above, comparisons with AKs may show excellent agreement, despite the lack of retrieval information in the lower layer. Comparisons *without* application of averaging kernels indicate how well the retrieval observes geographical variability *within* each specified layer. We therefore consider it necessary and informative to show these comparisons as well.

• In Section 6, it is not clear whether what the authors call "Estimated Standard Deviation" (ESD) is compared to misfits using individual IASI soundings (as suggested p. 1, l. 18 and p. 9, l. 21) or to misfits using averages of IASI soundings (as suggested p. 18, l. 3). The latter would be wrong in the absence of any information about retrieval-to-retrieval error correlations, but the former would require additional information (how is the IASI sounding selected for a given HIPPO profile? – it would be wrong to re-use a given HIPPO profile several times in the statistics).

It is the former. As described, IASI soundings are selected for each HIPPO profile (all within 200km and 6 hours). These matches are the basis for computing statistics. HIPPO profiles are not reused. This is clarified further in the revised text.

• The impact of temperature uncertainty is very large (p. 10 and Figure 2) if we compare it to the natural variability of the partial column. In contrast to the ESD (as it is computed), it is very likely highly correlated in space and time. This is obviously a major limitation of the IASI information content and it should be much better highlighted and discussed. In p. 10, l. 18, the authors hope that future use of temperature retrievals from IASI will improve the situation, but (i) errors may be correlated between those retrievals and the assimilated spectral samples (at least through RTTOV) and (ii) ERA-Interim already assimilates HIRS and AMSU radiances from Metop (ok, not from IASI, but at least collocated temperature information is used).

More discussion on this point has been added to section 2 of the paper.

• Section 4 lacks explanation. It suggests that total and partial columns are compared together, which would be wrong, in contrast to what is done later with Eq. (13). The quality of the retrieved cloud parameters is also not discussed, even though these play a key role in the retrievals. The use of cloud observations from AVHRR instead of these ones suggests, maybe erroneously, that it is quite poor.

A footnote is added to explain why using AKs is not necessary for the analysis reported in section 4. We deliberately do not dwell on the physical interpretation of the cloud parameters; they are *effective* parameters intended only to account for cloud (and other) effects on the CH4 retrieval in the 7.9 micron band. They are not expected to compare well with AVHRR parameters (especially not visible optical depth). We consider it sufficient to demonstrate that they clearly do mitigate cloud-related errors on CH4 (as demonstrated in section 4).

• Section 5.2: the authors have chosen to use a GOSAT "proxy" product even though it only provides the methane total column. However, "full-physics" products (e.g., available from the same data providers and co-authors) provide methane profiles: they would be much more appropriate and would (in conjunction with proper use of the averaging kernels – Rodgers and Connor 2003) avoid the not-so-clean trick of Eq. (13).

Even though profiles may be reported, there remains only a single degree of freedom in the SWIR methane retrieval (full-physics or not), so these are no more amenable to a Rodgers and Connors approach than the proxy results. Results from the proxy method have also been demonstrated to be preferable in certain respects to those from the full-physics method, and so were selected for comparison with IASI here.

Minor comments

• There are a few typos that should be removed (e.g., p. 4 l. 13, p. 18 l.13 and l.22).

We cannot locate these typos.

• The format of the references is not unified.

We await editorial guidance on this point.

• Methane units in all figures should be ppb rather than ppm.

This is a matter of personal preference rather than convention, and we prefer not to change; ppb and ppm are both used mainly for clarity in plots. While absolute values are more compact when written in ppm (e.g. 1.8 vs 1800) differences derived from comparisons are typically small and often clearer if stated in ppb (10 ppb vs 0.01 ppm).

• P. 1, l. 18: it should be said that the column is partial.

The sentence refers to the inferred column average. We have added a clause to the sentence to make it clear that this should not be interpreted without considering the vertical sensitivity.

• p. 1, l. 25-28: the list includes both processes (e.g., fossil fuel use) and explicit sources (e.g., wetland emissions). The format should be harmonized.

We have changed the word "sources" in the sentence to "contributions".

• P. 2, l. 14: to be fair, the 2009 paper from LMD should be mentioned (Crevoisier et al. 2009). Actually, since that IASI methane product line has been well established, a few words to explain how the design of this product differs from the new one presented here would be good.

Reference to the 2009 paper has been added, together with a paragraph outlining key differences between the RAL and Crevoisier schemes.

• P. 2, l. 18: Sentinel 5P is about to be launched and could usefully be mentioned.

Mention has been added.

• P. 3, l. 24: "error" is missing before "covariance".

It is clarified later in the sentence that the covariance describes the errors of the measurements.

• P. 4, l. 1: km^2 rather than km.

This has been corrected.

• P. 4, l. 29: because of the large number of spectral samples, this way of doing does not make sense without accounting for the correlated errors (see also Stewart et al., 2014).

A footnote has been added to clarify that error correlations in the modelled RTTOV-related error are unimportant.

• P. 5, l. 28: it is not clear what the FM uses in input.

A footnote has been added to clarify that the FM takes input as for the standard retrieval scheme (except without the derived residual patterns).

• P. 6, l. 6-7: the empiricism of this 4% (70 ppb !!!!) correction in the methane amount is quite perturbing.

We have expanded slightly on this point in the text.

• P. 7, l. 11: "statistics" is missing before "are estimated".

This has been added.

• P. 20, l. 5-8: the last sentences of the paper are nearly the same as those of Crevoisier et al. (2013) that were actually a simple reformulation of the last lines of their 2009 paper. Either this idea is trivial, and it should be removed, or it would be fair to quote the previous papers.

The paragraph was drafted wholly independently of the Crevoisier paper so we were unaware of any similarity. The plans to launch Metop-C and Metop-SG are well established, however, here we make specific reference to "global, height-resolved" methane distributions which those missions are capable of providing, and which is a *new* capability with respect to the Crevoisier paper.