We thank the referees for their useful and knowledgeable comments that have improved our paper.

Anonymous Referee #1

Figure 3 is somewhat misleading for MERLIN or LIDAR instruments in general. While it may be adequate for "SWIR" instruments LIDAR vertical sensitivity depends on the choice of wavelengths and the number of wavelengths. Grouping LIDAR (MERLIN) with "SWIR instruments" gives the impression they have similar capabilities and limitations.

We clarified in the revised text and figure caption that SWIR in Figure 3 only applies to solar backscatter measurements.

Same applies to Figure 4 which shows the atmospheric optical depths of different gases in the SWIR at 01. nm resolution. MERLIN or a LIDAR in general can have 0.1 ppm resolution.

We clarified this in the revised text.

The quoted MERLIN precision of 1% is the "Breakthrough" precision. MERLIN has Target, Breakthrough and Threshold random and relative systematic error requirements. I suggest the authors make that clear.

The Kiemle et al. (2011, 2014) papers don't seem to make that distinction.

I wished they had gone one step further and discussed instrument biases that are often a major systematic error source. These error sources are non-stationary (have temporal variability) that are very difficult to characterize. See for example: Werle, Peter. "Accuracy and precision of laser spectrometers for trace gas sensing in the presence of optical fringes and atmospheric turbulence." Applied Physics B 102.2 (2011): 313-329 and Bomse, D. S., and D. J. Kane. "An adaptive singular value decomposition (SVD) algorithm for analysis of wavelength modulation spectra." Applied Physics B 85.2-3 (2006): 461-466. In their discussion of Rodgers the authors correctly point out that error PDFs are often assumed to be Gaussian. Instrument biases are never Gaussian and in fact they are very difficult or impossible to characterize in a real instrument prior to launch. I do not think the Adjoint or other methods (that do not assume a Gaussian PDF) offer a solution to this problem. This remains a significant metrology and calibration problem which is often overlooked. I comment the authors for their extensive list of references. For completeness I also suggest they also reference: Ohring, G., B. Wielicki, R. Spencer, B. Emery, and R. Datla, eds., Satellite Instrument Calibration for Measuring Global Climate Change, NIST Rep. NISTIR 7047, 2004.

We carefully considered this comment but would rather no change the text. The Werle and Bomse papers are specific to lasers, and the Ohring paper does not go into discussion of instrument biases except very superficially. The Kuze paper cited in the text provides a detailed discussion of GOSAT instrument errors and seems to us to provide the most useful reference.

It would be nice to add the MERLIN capabilities to Table 2

Added

I could not find where Figure 6 is referenced in the paper.

That's because it was referenced everywhere as "Fig. 6". We have replaced the first mention by "Figure 6"

Anonymous Referee #2.

This reviewer was immediately struck by the claims of GHGSat in the abstract (and elsewhere in the manuscript) despite there being little information provided in the scientific peer-review literature about the project or the data. As with all potential performances of future instruments the authors should consider toning down unsubstantiated claims.

We have toned down and qualified the GHGSat claims.

The authors provide a list of instruments (past, present and future) in Table 1. Is this consistent or an update of the CEOS list?

This is consistent with but much more detailed than the CEOS list.

Table 2 does not include MicroCarb, which is a mature mission concept led by CNES (admittedly the CNES website mentions methane as a sidenote to CO_2).

We went through the MicroCarb literature and website but the focus appears to be solely on CO_2.

Line 133. What this reviewer would find more desirable for instrument sensitivity is weighting towards the lower troposphere where variations are more sensitivity to surface fluxes. The counter argument would be that a vertical column that is uniformly sensitive to the troposphere is less sensitive to vertical mixing and therefore easier to interpret with atmospheric transport models. In any case, there appears to be two lines of thought on this topic that would be useful to at least mention in this section.

We now mention this in the paragraph and also in the conclusions.

Paragraph starting line 213. This reviewer would shamelessly argue that using the ratio directly is the more natural way to use these proxy data that are less prone to systematic errors. Mentioning the direct use of the ratio data here would help put the later discussion about inverse modelling into context.

We do now.

Statement starting on line 239. Might reflect? That sounds like supposition to this reviewer. Can the authors provide a stronger statement?

Changed to "could be due to". Hard to be stronger considering the inconsistency between the two instruments.

Lien 260. Comment: TIR measurements are useful at quantifying emissions from fires where there is a large thermal contrast between the lower troposphere and free troposphere due to intense surface heating.

We say so.

What always strikes me as an argument for geostationary concepts is the temporal sampling bias of sun-synchronous instruments. The authors have put together an informative table including local overpass time. Most times are 0930 or 1330. These times matter, although the local overpass times of the satellites are not typically defined by the methane instruments. What implications do these times have on the science objectives? Perhaps not much for the point sources and the oil/gas industry but they will have for diurnal-varying sources. Would be useful for a reader to see an acknowledgment of this bias.

We now point out that geostationary observations have unique capability to observe diurnally varying sources.

This reader thought that section 3.1 was useful but would be more appreciated in an appendix. Section 3.2 was much more relevant to a review on satellite observations of methane, but it does lean on section 3.1.

Because section 3.2 does indeed lean on section 3.1 we would prefer to leave section 3.1 where it is.

In section 3.2. What about the soil sink for methane?

Now mentioned.

Statement on line 820. This section is important for setting the scene for matching science requirements with instrument requirements. The science requirements for policy and point sources are somewhat different, but the science requirements (repeat times, time of day, accuracy/precision) for (natural) diffuse sources are different again. Wetlands, rice paddies and ruminants are significant sources of atmospheric methane but they are not addressed in this review, e.g. what kind of concept would be needed to observe them better than they are today?

We are using the Barnett Shale as example of regional source. The same analysis would apply to other large regional sources.

Thawing Arctic land surfaces are potentially an emerging source of atmospheric methane that will perhaps require an active mission but this is not discussed.

We have added this discussion to the Conclusions.

This reader is not encouraging the authors to make their review much longer (on the contrary) but there needs to be a more balanced review of the future challenges.
Line 1014. This reviewer agrees that multi-species inversions are the way forward. But the reader should be exposed to the challenges involved, even with observations collected

by the same instrument, otherwise the suggestion appears as a trivial extension to existing studies. Differences in vertical sensitivities for different gases (wavelengths) need to be carefully considered.

We now point out the difficulties.

Figure 2. Some configurations are shown. Atmospheric limb sounding has its advantages.

We now mention in conclusions the advantage of limb sounding for removing the stratosphere.

Other unsolicited comment #1:

Dear authors, Why is the detailed study for the "Geostationary Emission Explorer for Europe" (G3E), Atmospheric Measurements Techniques 8, 4719-4734, 2015, not even mentioned in your manuscript? Maybe I am wrong but I would think that the technical and also the retrieval aspects are important for your discussion, too. Best regards, J. Orphal

G3E is now in Table 1 and mentioned in the text.