Responses to anonymous referee #2:

The study by Bousserez et al. explores the benefit of a geostationary observer with spectral coverage in the shortwave (SWIR) and/or thermal infrared (TIR) for surface flux inversion of CH4. To this end, the flux error reduction is assessed by feeding a Bayesian inversion frame work with the sampling patterns and measurement errors of several low-Earth-orbit and geostationary configurations. The geostationary SWIR+TIR configuration shows the best error reduction suggesting that inverting weekly-to-monthly fluxes on the scales of several ten kilometers is possible.

The study is of interest to the atmospheric sciences, it is generally well written. Therefore, it is suitable for publication in ACP after considering my comments:

Response:
We would like to thank the anonymous referee for their useful remarks and suggestions that helped improve the manuscript. The revised version of the paper (attached) includes significant modifications and new results that we hope address the referee's comments. Please see below our detailed responses to all the remarks and suggestions. Note that in addition to the new results produced to address the referee's comments, some errors were identified in our previous simulations and have been corrected since (in particular in the boundary condition sensitivity study). Therefore the entire manuscript has been modified accordingly and in our responses we only point to the modifications directly related to the referee's comments and suggestions.

General comments:

- In my opinion the general drawback of the approach is that model resolution is still coarse in time (weekly, monthly) and space (several ten kilometers) in comparison to the expected geostationary sampling resolution (1 hour, 4 km2 in geostationary configuration) and density. Diurnal cycle information available from the 1 h repeat cycle of the geostationary configurations, for example, is not exploited (and not discussed). Probably the diurnal cycle in the model is simply imposed. Other studies focusing on the high-resolution aspects (such as Rayner et al, AMT, 2014) should be cited.

Response:
Indeed other recent studies have focused on smaller spatiotemporal scales when analyzing geostationary observation constraints on trace gas fluxes (Rayner et al., 2014; Polonsky et al., 2014). Those works explored regional to urban size constraints, which is out of the scope of our study. Here we rather assess the relative merit of different observational configurations (SWIR, TIR, multi-spectral, and LEO vs GEO orbits) at continental to regional (50 km) scales. However, in the revised version of the manuscript we now present results for a 3-day inversion for each observational configuration, which shows in particular that the multi-spectral GEO configuration is best exploited when constraining fluxes at a time-scale of only a few days. Please see revised Section 3.1 for more details.

- Further, the model study assumes ideal measurements exhibiting purely random error characteristics. Likewise, transport model error is implemented by inflating the random errors. While these approximations might be adequate for a first assessment of sounding capabilities, I would argue that it is necessary to discuss these drawbacks and assumptions in the conclusion or discussion section.

Response:

In a real inversion framework, biases in the measurements can be estimated and removed (see, e.g., Wecht et al., 2014), therefore we rather focus on random noise in our study. However, we now mention those limitations in the conclusion of the revised paper, which has been entirely rewritten (see last paragraph in red).

- Section 2.3: What are the "observations" exactly? Is it the modelled CH4 concentration field averaging-kernel weighted as GOSAT, TES, or a SWIR+TIR instrument would deliver it? Or do you really use CH4 concentrations retrieved from GOSAT or TES? If the former, do you use a single (typical) averaging kernel or do you consider dependencies on geometry, surface temperature etc.? If the latter, how do you deal with the fact that the measured and modelled concentration fields do not match? This needs some clarification.

Response:

Yes, the "observations" are the modeled CH_4 concentration field sampled by the GOSAT, TES, or a SWIR+TIR observation operators. This is now clarified in both Section 2.2 and 2.3. We use a single averaging kernel for each instrumental configuration, as it is now explicitly stated and justified in Section 2.2 (see text in red after Eq. (8)): "A larger ensemble of averaging kernels describing a potential range of sensitivities is beyond the scope of this study given the computational cost. However, based on knowledge of thermal IR (e.g. TES) and total column (e.g. GOSAT) retrievals, use of a single averaging kernel is a reasonable approximation as our study is constrained to Northern Hemisphere summertime where the temperature and sunlight conditions provide sufficient signal for the present evaluation, and because our study looks at the relative merits of different observing approaches."

- I do not understand the role of an SVD of the posterior covariance? Why do you need it and how does it decorrelate error correlations between the layers?

Response:

Since the observational errors are correlated in the profile retrievals, it is not appropriate to apply independent perturbations at each level in our OSSE. However, in practice we can only produce independent perturbations using a random number generator. Therefore, we need to apply these independent perturbations to basis *where the errors are uncorrelated, which is provided by the SVD decomposition. This is explained in e.g., Bousserez et al. (2015) (Section 2.2., Eq. 11), or Chevallier et al. (2007) (Section 2.2).*

- Is it correct that you sample the modeled concentration field according to the GOSAT, TES, SWIR+TIR sampling patterns and then, remove all cloud-contaminated scenes based on the GEOS-CHEM cloud fraction? Please consider clarifying the text.

Response:
For each GEOS-Chem grid-cell, the GEOS-5 cloud fraction is used to remove a similar fraction of the total number of observations that fall within that grid-cell. This has been clarified in Section 2.3: " Finally, contamination by clouds is taken into account for each grid-cell by removing a fraction of the total number of observations within that cell which corresponds to the GEOS-5 cloud fraction."

- Do you consider that footprint size for a satellite observer, in particular a geostationary one, depends on distance from the subsatellite point? Are the 4 km2 geostationary resolution representative for the subsatellite point? What is it at higher latitudes?

Response:
The 4 km^2 geostationary resolution corresponds to the subsatellite point. For the sake of simplicity in our study we have neglected the impact of latitude on the satellite footprint size. Again, here we proposed an OSSE to assess the relative merit of different observational configurations, and the limitation of our setup to provide an accurate estimate of the constraints from the different observational configurations is now clearly acknowledged in the revised conclusion.

- Showing maps of exemplary "observations" could help illustrate constraint density and patterns.

Response:
We have included a map of weekly observation densities for the LEO and GEO configurations in the revised manuscript (see Figure 3).

3. Figure 3: Why do most regions show zero error reduction? Is it because the prior error covariance is defined relative (40%) wrt. to the prior fluxes which are small for large parts of the continent (figure 1)? If so, is this a reasonable setup of the inversion method? It essentially puts a hard constraint on regions with zero prior fluxes (to remain zero).

Response:
Yes, that is the reason. We have added a comment in the final paragraph of the conclusion acknowledging this shortcoming. To our knowledge, most inversions studies define the prior errors as relative to the magnitude of the fluxes. It is possible though that using an absolute error instead of a relative one for regions with small

emissions would be more appropriate.

4. Section 3.2: Would a uniform bias in the boundary conditions not be a very benign scenario? If the incoming airmasses have 2% high-biased methane and the outflow airmasses have the same 2% high-bias, the intra-domain fluxes would need little adjustments (unless there is a strong gradient between the boundaries). How would a bias in the zonal gradient between Eastern and Western boundaries affect intra-domain fluxes?

Response:
We have modified our setup in the revised manuscript. The boundary conditions are now randomly perturbed throughout the troposphere with a Gaussian noise with standard deviation 16 ppb, according to the statistics obtained from comparisons between HIPPO aircraft in situ data and the simulated methane concentrations over the Pacific ocean (representative of the west edge boundary conditions of our nested domain). Please see revised Section 3.2 for more details.

5. Technical comments

- P19020,l2: under sampling -> undersampling

Response:
Corrected.

- P19022,l6: providing -> provided

Response:
Corrected.

- P19022,l16: Calling the analysis vector xa could be misleading to many readers who are used to terminology with subscript a indicating “a priori”. But, your choice.

Response:
This is the terminology commonly used in the data assimilation/inversion literature. The subscript "a" is used for "a priori" in the retrieval literature. Here "x" denotes a flux, not a retrieval, so we think it is more appropriate to keep this notation.

- P19024,l3: inline citation: citep -> citet

Response:
Corrected.

-P19030,l18: On a weekly -> On weekly

Response:
Corrected.

- Flux figures: Units "per grid cell" are not easy to interpret since grid cell area depends on latitude. Consider replacing "per grid cell" by "per square meter" units.

Response:

The constraints on the emission scaling factors are related to the total emission in each grid-cell rather than the total emission per surface unit. Therefore we think presenting the total emission per grid-cell better help interpreting the inversion results. However, for guidance, we have now included in Table 1 a column with conversions from kgC/day/cell to kgC/day/km^2 for different latitudes that helps characterize the observational constraints in term of surface.

- Figure 6: Axes labels are small and faint.

Response:

The size of the axis labels have been increased.

- Figure 7: Consider replacing figure 7 by zooms on the relevant regions. Axes labels are too small.

Response:

We believe a map showing all regions at once offers a useful synthetic view to compare the spatial resolution of the constraints over different areas. Moreover, we have added a table (Table 2) that provides the radius of each structure shown on the maps.
The size of the axis labels have been increased.