Authors' response to anonymous referee #1

In the following, referee's comments are in *italic*, authors' responses in normal font, and references (page, line, figure, and table number) to the revised manuscript in **bold**. Please note that this paper was merged with the accompanying paper, following the referees' comments and with approval from the Topical Editor. The summary of this paper was included in the Supplementary Material of the accompanying paper.

Unfortunately, the manuscript is not well written and would improve from English language editing. There are also inconsistencies and misunderstandings especially in the Introduction section; for example page 1, lines 31/32: which models are based on assimilation techniques? It is rather that in atmospheric transport inversions assimilation techniques are often employed to invert the transport. Another example just follows on page 2, line 1: the inversion method does not provide information on the emissions; it is rather the observations that provide information. I suggest the authors (some of the co-authors are native or close-to-native speakers and some are world-leading experts in inverse modelling) carefully go through the manuscript to improve the language and correct these misunderstandings of how transport inversions work.

We apologize for the misunderstandings and inconsistencies that arose as a consequence of the weak formulation that existed in the manuscript. In this revision, we have phrased our text more carefully, and also had the full paper language edited by a native English speaker. Moreover, we tried to make our descriptions more clear using new labeling.

The main problem of this manuscript, however, is its aim, as the authors state: 'The aim is to introduce the set-up and choices for an inversion system, which will be used in long term studies and presented in an accompanying paper.' Is this really sufficient for a publication in its own or could that have been merged into the accompanying paper that looks at long-term methane emissions and trends. And how much do we really learn from a sensitivity study covering a very short time period (5 months) with no interannual nor seasonal variability for an inversion aiming to analyse long-term emission fields and trends? On top of that the CTE-CH_4 system has essentially already been published elsewhere (see Tsuruta et al., 2015).

After reflecting on the comments from the reviewer, and following the suggestions, this paper was merged with the accompanying paper. Although experiments were carried only for a short time period, the findings were important and could hold also for experiments on longer time periods. Since the previously published Tsuruta *et al.* (2015), the system was further developed, and this paper extends the analysis with the new version of the model.

- Why did the authors finally choose S1 and S5 for the long-term experiment in the accompanying paper? Based on the evaluation in this manuscript here there are not any indications to rule out any of the tested configurations. At least the authors do not provide any objective reasons.

We agree with the reviewer that the choice of the simulations to be extended over the longer time period was based on expert judgment, and not so much on derived metrics. In the end, we did not find specific reasons to choose other set-ups based on uncertainty estimates and agreement with the NOAA in situ observations. Therefore, in the new manuscript, we decided to present each estimate as an equal realization of the surface fluxes. We hope the reviewer agrees that this is a more balanced representation of our results.

- Is there anything the inverse modelling community can learn from the experiments here in general? Maybe the results of some of the experiments can be analysed in more depth such that the findings can be generalised?

We acknowledge that some of the experiments were specifically meant to test the configuration of this system, following a custom of designing and presenting a new inverse modelling framework. In addition to the real flux estimates that are interesting in themselves, we also believe that the sensitivity to vertical transport, and the attempt to separately estimate biosphere and anthropogenic fluxes are useful to other, similar attempts, given that the sparsity of systems that perform inverse modeling of CH_4 fluxes. In the revised manuscript, sensitivities to those were analysed in depth based on multi-year simulations and validation with model independent observations. Moreover, the extensive evaluation presented here sets a target for future studies that we also find important.

- It would be illustrative to run some longer experiments to account for seasonal and interannual variability for some selected set-ups.

This is an excellent suggestion, and we agree that our system would profit from further tests of robustness and sensitivity. As we mention in the revised manuscript, testing different prior emission patterns as well as varying the OH-sinks over time would be the first things to try. However, to benchmark our system and describe it for future reference, we feel that the revised manuscript already is quite extensive, which is merged with the accompanying paper. In the revised paper we present results from longer-term experiments to examine the robustness and sensitivity of the estimates on seasonal and interannual variabilities to the number of parameters, the vertical mixing schemes in the transport model, and briefly to the set of observations. We plan to continue using our system at FMI for the near future, and we hope the reviewer allows us to undertake these follow-up tests in a future study as well.

- Although the title suggests that the inversion system focuses on Europe the analysis of the results in the manuscript does not. What is the effect of the zoom over Europe? And although you zoom over Europe you only distinguish between four regions.

The reviewer correctly remarks that our paper presents results for the full globe, rather than just focusing on Europe. We applied the zoom over Europe, since the observation network in Europe is most dense and we have a primary interest to study European CH_4 fluxes with our system in the future. With this higher resolution, CH_4 abundance at closely located sites can be resolved separately, taking into account meteorological parameters and transport of each grid cell. However, the regional inversion over Europe is only trustworthy if we can realistically constrain the inflow and outflow of CH_4 across regional boundaries, and for that we need to take into account the fluxes in the rest of the world. The current paper thus lies the foundation for future regional inversions. Note that the identifier “Europe” in our title refers to the origin of the CarbonTracker branch we use (to distinguish it from NOAA's CarbonTracker) and does not necessarily identify the focus of the study.

- In Fig 4 plotting the relative differences in the uncertainty estimates would be much more illustrative.

We thank the reviewer for this suggestion. The figure was modified accordingly, and moved to supplementary material of the accompanying paper.

See Supplementary Material of accompanying paper, Fig. S3.

- In Fig 5 the fit against the observations when transporting the posterior fluxes is not very impressive in this kind of plot. Maybe time series for some selected stations would illustrate the improvement much better.

Here we aimed to illustrate the large scale and seasonal patters, and less focus on the agreement at individual stations. The format also corresponds to the presentation of residuals in other CarbonTracker applications, where the format is generally well appreciated. The requested time series are now part of the revised Figure 2, although we decided to retain the look.