General Comments

This paper lays out the potential for current and future $^{14}CO_2$ observations to improve estimates of fossil fuel emissions in Europe. It uses two types of Observing System Simulation Experiments (OSSEs) based on either the theoretical uncertainty reduction for a well-tuned case or a more realistic case where prior uncertainties do not match differences between prior and truth. It also uses several versions of an observing network ranging from the current network to a saturated case where every grid cell in the target domain is sampled. Results are not very surprising with the current network offering useful information at the conjunction of dense networks and high emissions (and concomitant uncertainties) with the case improving as networks become more dense. Results are, however, sensitive to the proper tuning of prior covariances; a salutary result the authors are right to emphasise. The paper addresses an important problem with reasonable if not state-of-the-art tools, is clearly written and within scope.

Response:

We would like to thank the reviewer for the valuable comments and suggestions for improving our manuscript. Following the reviewer's comments, we will carefully revise our manuscript. Most of the concerns about the observation and aggregation errors raised by the reviewer were analyzed (at least partly) in Wang et al. (2017) which is cited in our manuscript. We will better remind the conclusions from this paper in the manuscript.

Please find below the point-to-point responses (in black) to all referee comments (in blue). All the pages and line numbers correspond to the original versions of text.

References:

Wang, Y., Broquet, G., Ciais, P., Chevallier, F., Vogel, F., Kadygrov, N., Wu, L., Yin, Y., Wang, R. and Tao, S.: Estimation of observation errors for large-scale atmospheric inversion of CO_2 emissions from fossil fuel combustion, Tellus B: Chemical and Physical Meteorology, 69(1), 1325723, doi:10.1080/16000889.2017.1325723, 2017.

I have two concerns about the paper, one general and one specific. the authors note the dependence of their results on the resolution of their transport model ($3.75 \times 2.5°$) but I think should do more to evaluate this. It is unlikely that anyone would use this resolution for an inversion of fossil fuel emissions targeting Europe and the guidance on network density is hard to generalise.

Response:

In this paper, our analysis focuses on the inversion of European fossil fuel emissions. However, we have worked with a global and thus coarse resolution transport model in order to: (1) properly account for the uncertainties in emissions from other continents than Europe when inverting European emissions, and (2) because we developed a system which also allows us to study the inversion of the emissions in North America and Eastern Asia.

Sect 4.2 analyses whether the uncertainty in the emissions outside Europe has an impact on the inversion of the emissions in Europe. The results indicate that this impact is in fact weak, which was not obvious to prove before doing the study. Furthermore, studies including some of the sources of uncertainties that have been ignored here could reveal, e.g. that uncertainties in the $^{14}CO2$ fluxes from oceans and land ecosystems outside Europe have a strong impact on the inversion of the emissions in Europe. A cautious account for such uncertainties could require the use of a global inverse modeling system, or of the coupling between a European scale and global scale inverse modeling systems. At our stage of investigation in this study, we thus think that the use of a global inversion system is appropriate.

The spatial resolution of LMDZ is typical for global transport models and inversion studies (Peylin et al., 2013). For example, the Transport Model 3 (TM3, 5°×4°) used for the Jena CarbonScope (Rödenbeck et al., 2006), TM5 (3°×2° without nested version) used for CarbonTracker (Peters et al., 2007), Model of Atmospheric Transport and Chemistry (MATCH, 5.6°×2.8°) and the CSIRO Conformal-Cubic Atmospheric Model (CCAM, about 220 km) used by Rayner et al. (2008), have similar spatial resolutions as LMDZv4 used here. Using a much higher resolution transport model, e.g. 1°×1°, for global simulations is computationally expensive.

In principle, we properly accounted for the representation error and its temporal and spatial correlations by using the detailed analysis of the aggregation and representation errors from Wang et al. (2017). In particular this should prevent from overestimating the effect of the spatial sampling of $FFCO_2$ and thus performance of inversions when using dense networks. In a more general way, we think that our configuration of the observation errors support our confidence in the guidance that we derived from our relatively coarse resolution inversion system regarding the impact of the network density. In our conclusions, we were cautious regarding the dependence of the results to the transport spatial resolution.

The analysis by Wang et al. (2017) provides some insights and understanding on the dependence of the results to the resolution of the transport model. However, running atmospheric inversions using higher spatial resolution model, which was out of the scope of this study, would have been the only way to assess the dependence of the results to the spatial resolution correctly, since it depends on a complex combination between the prior and observation error covariance structures together with the atmospheric transport.

We will highlight that the use of LMDZv4 aims at properly accounting for the uncertainties in $FFCO_2$ emitted over other regions outside Europe. We will also better stress the dependence of our results to the spatial resolution of the transport model but the fact this study aimed at providing some understanding of the inversion behavior and sensitivity to the network density rather than to provide a precise quantification of the uncertainty reduction that would be obtained if working with real data.

The authors can help a little here since their group has access to higher resolution models. How much do the representation and aggregation errors change with increasing model resolution. Representation error probably decreases while aggregation error increases but how much? Increased resolution makes gaps in the network inevitable, what effect will they have? this could be tested by a couple of systematic thinning experiments on the saturated network case here.

Response:

Wang et al. (2017) used the meso-scale transport model CHIMERE run with a 0.5 ° horizontal resolution to assess the statistics of the representation and aggregation errors when working with the global inversion system that is used in our study. These statistics are summarized in Sect. 2.2.2 (Page 11, lines 327-337) and Table S3 and Table S4 of this paper.

The representation error will definitely decrease with increasing spatial resolution for the transport model. Our definition (which is also that of Wang et al., 2017) of the representation error encompasses the errors associated to the representation of the emissions using a constant value within one pixel and one time step of the transport model. Therefore, our definition of the aggregation errors limits them to the errors associated with the fixed spatial distribution of the emissions within a region and month at the transport model spatial and temporal resolution. With such definitions, the aggregations errors increase when the spatial resolution of the transport model becomes finer. But such an increase is balanced, in the representation error, by the decrease of the component associated to the emission representation. Overall, the dominant pattern of the variations of the observation errors associated with the increase of the transport model spatial resolution should be the decrease of the representation error associated with the representation of the concentrations.

If following the specific framework and error definitions of Wang et al. 2017, a precise assessment of the change of the representation error in Europe as a function of the spatial resolution of the transport model would require series of European scale simulations with emission maps at different spatial resolutions (e.g. 1°, 1.5°, etc.) to feed the high-resolution transport model, and then require aggregating the output (concentrations) of the transport model at corresponding spatial resolutions (e.g. 1°, 1.5°, etc.). It would have been feasible but it was out of the focus of this previous paper. It would now be out of the scope of our paper to resume such computations, especially since properly assessing the impact of these changes of representation errors in the inversion results, would require conducting inversions with different transport model configurations (topographies, wind fields etc. are needed at different resolutions) as stated in the previous answer to the reviewer's comment.

Regarding the "gaps in the network", this phenomena is supposed to be well accounted for even when using the coarse resolution global transport model and when having two stations in each grid cell of this model thanks to a proper account for the observation errors. Representation errors indicate to the system that a station does not have a full coverage of its corresponding grid cell and so that it does not see the same information as the other station in the same grid cell (i.e. that there is already a gap between them even if using the coarse-resolution transport model of LMDZ). The physical separation of the stations in the grid of a higher spatial resolution system should not lead, in principle, to a strongly different behavior of the inversions, especially since the correlation length scale of the projection of the prior uncertainties in the concentration space is 700 km. In order to demonstrate it, we have conducted three additional experiments with different thinned networks: a) one with two sites located in the same grid cell of every two grid cells (113 sites in total); b) one site in each grid cell (117 sites in total); c) one site every two grid cell (57 sites in total). Fig. 1 shows the URs for INV-E inversions (the behavior of the results from INV-N inversions are similar but not shown). Fig. 1a and 1b show quite similar distributions and values of UR scores. The comparison between the Fig. 4g in the original manuscript with Fig. 1a and 1b here, and between Fig. 1a and 1b with Fig. 1c, show the decrease of UR across Europe due to using less sites. Since NET233 and the three thinned networks are uniformly distributed across Europe, this decrease of URs due to the gap in the networks are also nearly uniform, which confirms that the general behavior of the inversion does not significantly change due to generating gaps between the observed grid cells. We do not plan to include these results in the manuscript because they are not qualitatively very different from the ones we already showed and would not really lead to new insights or conclusions on the inversion behavior.

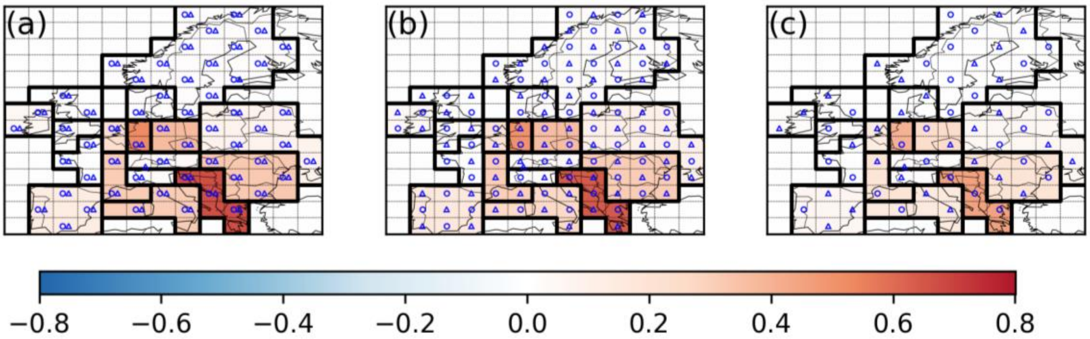

Figure 1: Average monthly uncertainty reductions in $FFCO_2$ emissions from INV-E inversions over regions delineated by solid black lines, using three networks and 2-week sampling for the inversions. The three networks are: a) two sites located in the same grid cell of every two grid cells (113 sites in total); b) one site in each grid cell (117 sites in total); c) one site every two grid cell (57 sites in total). The dots and triangles denote the locations of the observation sites where the gradients are extracted with respect to the JFJ reference site. Dots (triangles) correspond to “urban” (or “rural”) stations defined in Sect. 2.1 of the original manuscript.

We will add some discussions about how the representation and aggregation error would change when using a high-resolution transport model in Sect. 4.2, indicating that assessing properly the impact of the change these errors on the inversion results would require a large amount of work (which would be worth being investigated). We will also add cautious discussions on this general topic raised by the reviewer in the discussion section.

My other concern is for this saturated case. As I understand it, each grid cell is oversampled with two measurements. If this is the case and the transport Jacobians for the two measurements are the same then I think the two measurements can be combined into a single measurement by summing their information content. There should also be strong correlation between the two measurements in the same grid cell, accounting for large-scale errors in the transport model. In particular, I think that the relationship between the aggregation and representation errors for the two types of site is complex, interesting and perhaps important. It is quite possible that using both types of site reduces the sampling inhomogeneity necessary for aggregation errors (Trampert and Snieder, 1996; Kaminski et al., 2001).

Response:

Yes, in this case, each grid cell is sampled by two measurements at each sampling time, and the transport Jacobians for the two measurements are the same. Our modeling of R (based on the statistical estimates by Wang et al., 2017) is made such that there are full correlations between the transport errors and between the aggregation errors in the two measurements within the same grid cell. However, Wang et al. (2017) showed that the spatial correlation of the representation errors is less than 100 km, while the typical distance between the stations in NET233 network (with two sites per 3.75 °×2.5 ° grid cell) is about 200 km. Therefore we ignored the spatial correlation between the representation errors in the two measurements within the same grid cell. Wang et al. (2017) also diagnosed that the spatial correlation between representation errors for urban and rural sites is even smaller than the spatial correlation between two rural or urban sites so that it is also negligible. In addition, their analysis does not reveal any correlation between representation and aggregation errors (if following their definition of these types of errors as discussed above). Our configuration of the observation error matrix in this study exactly followed these indications.

Mathematically speaking the two measurements in each grid cell could be combined into a single measurement, but this would require the derivation of a complex observation error covariance matrix for the “combined” measurements, accounting for all the components of the observation error for individual data (measurement, transport, representation and aggregation errors) with varying standard deviations (depending on the location of the stations for the computation of transport error and on the urban or rural type of the station for the representation error) and their respective temporal and spatial correlations. In this context, such a combination would not really simplify the representation and understanding of the inversion problem of the data and of their observation errors.

We will better describe in Sect. 2.2.2 about the configuration of the **R** and associated correlations in the observation errors. We will also stress in the updated manuscript the fact that using NET233 reduces the sampling inhomogeneity and can reduce the impact of aggregation errors, as shown by the references proposed by the reviewer.