

Response to comments on “Potential of European ¹⁴CO₂ observation network to estimate the fossil fuel CO₂ emissions via atmospheric inversions” by Y. Wang et al.

We thank the three referees for their very detailed reviews. Their comments have allowed us to improve the manuscript by better emphasizing its strength. In this final response, we keep all the answers to the reviewers in previously response, but also add changes in the revised manuscript, highlighted with dark red. All the pages and line numbers correspond to the original versions of text.

General Comments

This paper lays out the potential for current and future ¹⁴CO₂ 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., Kadyrov, N., Wu, L., Yin, Y., Wang, R. and Tao, S.: Estimation of observation errors for large-scale atmospheric inversion of CO₂ 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 (3.75×2.5 9

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 $^{14}\text{CO}_2$ 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, $5^\circ \times 4^\circ$) used for the Jena CarbonScope (Rödenbeck et al., 2006), TM5 ($3^\circ \times 2^\circ$ without nested version) used for CarbonTracker (Peters et al., 2007), Model of Atmospheric Transport and Chemistry (MATCH, $5.6^\circ \times 2.8^\circ$) 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. $1^\circ \times 1^\circ$ 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 highlight that the use of LMDZv4 aims at properly accounting for the uncertainties in FFCO₂ emitted over other regions outside Europe, by adding in line 112: “Although the results are presented only over Europe, we use a global inversion system and the global transport model LMDZv4 to ensure that uncertainties in FFCO₂ emitted over other regions of the globe are properly accounted for and to study their impact on the inversion of the FFCO₂ emission in Europe. LMDZv4 has a 3.75 °×2.5 ° longitude×latitude horizontal resolution and 19 layers in the vertical between the surface and the top of the atmosphere. This spatial resolution is comparable to that of transport models used in state-of-the-art global inversions (Peylin et al., 2013). We assess the potential...”

We 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 by:

- Adding in line 110: “... at ICOS-like stations. The study primarily aims at providing a typical quantification of the inversion performances and at understanding qualitatively how the inversion behaves depending on the level of FFCO₂ emissions, on the knowledge on these emissions and on the network density. ...”
- Adding in line 521: “This study provides understanding of the inversion behavior and sensitivity to network density, but the precise quantification of the performance of the inversion is largely dependent on the spatial resolution of the transport model. Wang et al. (2017) showed that the representation error contributes the most to the observation errors, followed by the transport and measurement errors. ...”

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.

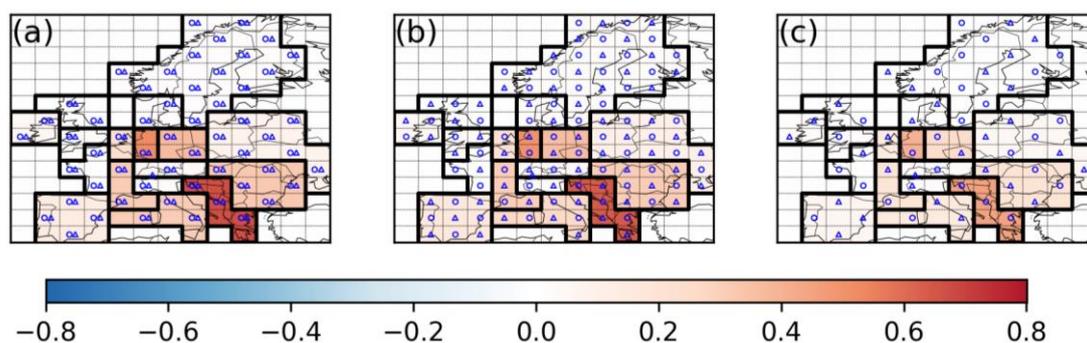


Figure 1: Average monthly uncertainty reductions in FFCO₂ emissions from INV-E inversions over regions delineated by solid black lines, using three networks and 2-week sampling for the inversions. The three networks are: a) two sites located in the same grid cell of every two grid cells (113 sites in 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 add in line 521: “followed by the transport and measurement errors. Following the definition of the observation errors in Wang et al. (2017) and in this study, the representation and the transport error are highly dependent...” We highlight the fact 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) by adding in line 527: “... to improve the results from atmospheric inversion of FFCO₂ emissions at regional scale. A proper quantification of the change of representation and transport error as a function of spatial resolution, and of the impact of this change on the performance of the inversion system would require a series of transport models and inversions at varying spatial resolution which are out of the scope of this study but which would be worth being investigated in the future.”.

We also add cautious discussions on this general topic raised by the reviewer in the discussion section by:

- Rewriting the paragraph from lines 93-107: “In this study, we study the potential of an atmospheric inversion system to quantify FFCO₂ emissions at regional scales (i.e. the size of a medium-sized country in Europe like France or Germany) over the European continent based on continental-scale networks of atmospheric CO₂ and ¹⁴CO₂ measurements. Special attention is paid to the representation and aggregation errors induced by the use of a coarse grid transport model. Wang et al. (2017) derived the statistics of these errors for the inversion system that we apply here, which is based on the Laboratoire de Météorologie Dynamique LMDZv4 global transport model (Hourdin et al., 2006) and our study strongly relies on their results. They highlighted that both the representation and aggregation errors have large magnitudes, and could thus strongly reduce the ability of the inversion to

filter the information on the uncertainties in regional FFCO₂ emissions. They also stressed the fact that the spatial scales of the correlations in the representation and aggregation errors are smaller than that of the projection in the atmospheric observation space of the typical uncertainties in the prior estimates of regional emissions (called “prior FFCO₂ errors” hereafter). More precisely, with their modelling configuration they obtained values smaller than 200 km and larger than 700 km respectively for these spatial scales. Therefore, if the observation networks are dense enough to provide information at finer spatial scale (typically with distances from a given station to the closest ones being systematically smaller than 700 km), the impact of aggregation and representation errors on the inversion of the regional budgets of FFCO₂ emissions could be small (Wang et al. 2017). In this study, we account for the aggregation and representation errors using their detailed and quantitative characterization and check whether using dense networks could overcome the limitations brought by coarse resolution transport models and by the uncertainties in the distribution of the emissions at high resolution when retrieving regional emission budgets. Using the error estimates from Wang et al. (2017) ensures that our inverse modelling system does not overestimate the potential of measurement networks that are dense compared to our coarse transport model resolution but whose distances between the sites are larger than the spatial scales of local atmospheric signals from the anthropogenic emissions.”

- Adding in line 151: “... assumed to be one urban and one rural distant by more than 200 km in order to combine data for the structures of representation errors are different (i.e. which have a different view in terms of the scale of FFCO₂ emissions). Any of the transport model pixels provides such locations since having areas of nearly 10⁵ km² (Wang et al. 2017).”
- Adding in the Sect. 4.2, line 520: “... 2) the observation errors bear complex temporal and spatial correlations which are close to the prior FFCO₂ errors (Wang et al., 2017). Such a result illustrates the need for using a suitable observation error characterization (here based on the results from Wang et al., 2017) to prevent the stations having a full coverage of information on the emissions in the model framework shown here even when the observation network is as dense as NET233. A proper account for the observation errors and their temporal and spatial correlations avoid overestimating the potential of the atmospheric inversion in OSSES when using a coarse resolution transport model.”

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 \mathbf{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 $3.75^\circ \times 2.5^\circ$ 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 rewrite the paragraph in lines 323-337 to better describe the configuration of the \mathbf{R} and associated correlations in the observation errors:

“In this study, we use the estimates of the standard deviations and of the correlation functions for these different types of observation errors from Wang et al. (2017) to set up the \mathbf{R} matrix. Wang et al. (2017) sampled representation and aggregation errors by using simulations with a mesoscale (with higher resolution than LMDZv4) regional transport model and by degrading the spatial and temporal resolution of the emission maps in the input of this model and in the output FFCO₂. Based on these samples, the standard deviation of ϵ_r was characterized by a function of season and on whether a station is “urban” or “rural” (see Sect. 2.1). For ϵ_a , the standard deviation for spring/summer and autumn/winter were derived. **The standard deviation of the transport error at a given site is set-up proportional to the temporal standard deviation of the 1-year long time-series of the high-frequency variability of**

the detrended and deseasonalized simulated daily mean afternoon mixing ratios in the grid cell of the transport model, at which the sites are located. Such an estimation of transport error which relies on some results from Peylin et al. (2011) aims at representing the typical value for global transport models, not that of the specific transport model used in this study. The temporal auto-correlations in the representation and aggregation errors were characterized by Wang et al. (2017) using the sum of a long-term component and a short-term component: $r(\Delta t) = a \times e^{-\Delta t/b} + (1-a) \times e^{-\Delta t/c}$ where Δt is the timelag (in days) and a, b, c are parameters optimized by regressions against the samples of the errors. Furthermore, we do not include temporal auto-correlations in the transport error for simulated daily to 2-week mean afternoon FFCO₂ gradients, since previous studies of the auto-correlations of the transport errors have not evidenced that they should be significant at daily scale (Lin and Gerbig, 2005; Lauvaux, 2009; Broquet et al., 2011). This choice follows the corresponding discussion by Wang et al. (2017) and implicitly ignores that transport model errors likely bear long-term components (often referred to as “biases”, Miller et al., 2015) even when being dominated by components on short timescales. The corresponding values of the standard deviation and the modelling of temporal autocorrelation of the observation errors for 2-week/daily mean afternoon FFCO₂ gradients are listed in Table S3 and Table S4.

A simpler account of the spatial correlations in the observation errors is derived from the diagnostics of Wang et al. (2017). We do not account for the spatial correlation in the representation error, as the scale of the spatial correlation according to Wang et al. (2017), i.e. 55-89 km, is much smaller than the size of the grid cells of the global transport model $\mathbf{H}_{\text{transp}}^{\text{LMDZ}}$ used for the inversion. When there are more than two sites are located in the same grid cell of the transport model, we consider that the aggregation errors and the transport errors are fully correlated between these sites, according to the definition by Wang et al. (2017). We do not account for spatial correlations between aggregation errors for measurements made at sites in different grid cells, because the scale of the spatial correlation is 171 km and is smaller than the size of the grid cell, according to Wang et al. (2017). Finally, we do not account for spatial correlations between transport errors or measurements made at sites in different grid cells.”

We 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 by adding in line 141: “...of the LMDZv4 transport model (Fig. 1c). The NET233 network is denser than NET17 and NET43 in the high emitting regions, e.g. Germany, and also covers the region that is not well sampled by NET17 and NET43. However, the location of its 233 sites is not intended to be optimal since the emissions have a very heterogeneous spatial distribution. Their homogeneous spreads allow us to reduce the impact of representation and aggregation errors (Tramper and Snieder, 1996; Kaminski et al., 2001) and to assess the impact of having a dense network for all control regions.”

Response to comments on “Potential of European $^{14}\text{CO}_2$ observation network to estimate the fossil fuel CO_2 emissions via atmospheric inversions” by Y. Wang et al.

We thank the three referees for their very detailed reviews. Their comments have allowed us to improve the manuscript by better emphasizing its strength. In this final response, we keep all the answers to the reviewers in previously response, but also add changes in the revised manuscript, highlighted with dark red. All the pages and line numbers correspond to the original versions of text.

The authors present an OSSE study of the capability of ICOS $^{14}\text{CO}_2$ observations to constrain European fossil fuel CO_2 fluxes and their trends. The study is well structured and should be published. I have a few comments which I'd like the authors to address before publication.

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. 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.

Major comments

1. Line 112: The assumption that $^{14}\text{CO}_2$ measurements can be accurately translated into FF CO_2 , i.e., there are no spatial patterns introduced due to the other terms (especially disequilibrium and nuclear plants in the European context), is a big one. Those terms will not only affect annual emission estimates, but also the ability to detect trends, as countries change their nuclear power generation capacity and switch to wood-fired domestic heating (e.g., Germany). I understand that modeling the full $^{14}\text{CO}_2$ budget is beyond the scope of the authors' framework, but it should be possible to estimate the impact, by e.g. modelling just the nuclear or disequilibrium contribution as a tracer in a transport model and looking at the change in $\Delta^{14}\text{C}$. Have the authors done that? Unless that concern is addressed, the actual numbers from the manuscript are hard to trust.

Response:

Indeed, the modeling of the full $^{14}\text{CO}_2$ budget is beyond the scope of this study. It would be necessary to study the impact of uncertainties in other $^{14}\text{CO}_2$ fluxes on the $^{14}\text{CO}_2$ gradients, to precisely quantify the corresponding errors when converting the gradients of atmospheric $^{14}\text{CO}_2$ measurements into FFCO₂ gradients. Graven and Gruber (2011) estimated the sources of ^{14}C from nuclear power generation and spent fuel reprocessing and used the global TM3 transport model at $1.8^\circ \times 1.8^\circ$ resolution, which is slightly higher than LMDZv4 used in our study, to simulate their

continental-scale influences on $\Delta^{14}\text{C}$. Their results showed that nuclear enrichment may cause an impact of -0.9 [-0.6, -1.4] ppm in FFCO₂ for Orleans, France (48.8 N, 2.5 E) and an impact of -0.7 [-0.4, -1.3] ppm for Heidelberg if nuclear ¹⁴C enrichment was not accounted for. These two sites are representative of European continental sites that are close to nuclear power plants. Turnbull et al. (2009) estimated the impact in large-scale FFCO₂ signals caused by ignoring other ¹⁴C fluxes including cosmogenic production, ¹⁴C disequilibrium between atmosphere and biosphere and between atmosphere and ocean, and ¹⁴C source from nuclear power plant (the estimate of ¹⁴C sources from nuclear power plants is not as accurate as Graven and Gruber, 2011), using the same transport model LMDZv4 as in this study. Turnbull et al. (2009) showed that impact caused by other ¹⁴CO₂ sources in translating atmospheric measurements of ¹⁴CO₂ into FFCO₂ is mainly from terrestrial biosphere, whereas the contributions from ocean CO₂ exchange and cosmogenic production of ¹⁴C contribute are weak. According to Turnbull et al. (2009), neglecting the influences from biosphere leads to an error typically between 0.2 and 0.8 ppm. Miller et al. (2012) estimated the impact of biospheric disequilibrium ¹⁴C fluxes in North America and get a similar value as Turnbull et al. (2009), ranging from less than 0.2 ppm to 1.4 ppm.

We work with OSSEs so that what matters in our system is the correct representation of the uncertainties in the different components of the model, not a correct representation of these components. In our OSSE framework, by ignoring the fluxes of ¹⁴CO₂ other than the dilution of ¹⁴C in CO₂ by fossil fuel emissions, we implicitly assume that the uncertainties in these fluxes has a weak impact, not that these fluxes themselves are ignored. The results mentioned above show that the uncertainties in the signals of these ¹⁴CO₂ fluxes may cause some impact on the interpretation of FFCO₂ using atmospheric ¹⁴CO₂ samples indeed, but also that this impact is below 1 ppm and thus smaller than the components of observation errors, e.g. measurement error (1 ppm), representation error (0.17-2.56 ppm) and transport error (0.52-4.15 ppm) in the paper. In this context, we assume that the influence of the uncertainties in ¹⁴CO₂ fluxes other than the dilution of ¹⁴C in CO₂ by fossil fuel emissions on the inversion of fossil fuel emission should be relatively weak. Our estimate of URs and MRs could be slightly over-estimated due to the ignorance of other sources of uncertainties from other ¹⁴CO₂ fluxes. But this study aims at understanding how the inversion system behaves when dealing with uncertainties in the FFCO₂ emissions versus observation errors, about how the variation of the UR as a function of regions, of level of emissions, of the density of networks, rather than about providing absolute values of UR that would apply when conducting real-data applications. Further investigations accounting for uncertainties in other ¹⁴CO₂ fluxes will be needed to refine those numbers.

We explain the assumptions underlying the conversion of ¹⁴CO₂ into FFCO₂ gradients with a 1 ppm uncertainty given that we work with an OSSE framework (i.e. it does not mean that the nuclear plant, cosmogenic and biosphere fluxes themselves have been ignored) by rewriting the paragraph between lines 108-118: “Our inversion system solves for monthly FFCO₂ emissions in different regions of Europe over a

period of one year by assimilating synthetic observations of atmospheric gradients of FFCO₂ mixing ratios obtained from co-located CO₂ and ¹⁴CO₂ measurements at ICOS-like stations. The study primarily aims at providing a typical quantification of the inversion performances and at understanding qualitatively how the inversion behaves depending on the level of FFCO₂ emissions, on the knowledge on these emissions and on the network density. Furthermore, we assume here that the uncertainties in the signals from ¹⁴CO₂ fluxes other than the FFCO₂ emissions, such as that from terrestrial biosphere, oceans, nuclear power plants and cosmogenic production, should have a moderate impact on the order of magnitude of the inversion performances that are analysed in this study. This leads us to ignore these uncertainties and consider that the only uncertainties in the FFCO₂ mixing ratios data are related to the instrumental precision of CO₂ and ¹⁴CO₂ measurements. In practice, in the frame of this study, which focuses on the propagation of uncertainties, this is mathematically equivalent to assuming that ¹⁴CO₂ is a perfect tracer of FFCO₂. However, this does not imply that the signal from natural fluxes and nuclear power plants could be ignored when processing real data.”

To recall the fact that the uncertainties associated with the signals from ¹⁴CO₂ fluxes other than the dilution of ¹⁴C in CO₂ by fossil fuel emissions are expected to be relatively small compared to the uncertainties in the atmospheric ¹⁴CO₂ caused by fossil fuel emissions and by other types of observation errors and to stress that future studies will be needed to quantify the impact from uncertainties in the nuclear plant emissions, cosmogenic production, biogenic fluxes, etc., we modify the sentences in Sect. 4.2, between lines 502-507: “... This 1-ppm standard deviation approximately corresponds to the errors in the atmospheric measurements and ignores uncertainties in the conversion of ¹⁴CO₂ and CO₂ measurements into FFCO₂. Uncertainties in various fluxes that influence atmospheric ¹⁴CO₂, such as those from cosmogenic production, ocean, biosphere and nuclear facilities, bring errors to the conversion of ¹⁴C measurements into FFCO₂ (Lehman et al., 2013; Vogel et al., 2013). Over land regions, heterotrophic respiration is expected to be one of the main contributors to the large-scale signals of atmospheric ¹⁴CO₂ (Turnbull et al., 2009). Over some areas of Europe, ¹⁴C emissions from nuclear facilities may have even larger influences than plant and heterotrophic respiration (Graven and Gruber, 2011). The level of uncertainties in these fluxes and how much their influences on the FFCO₂ gradients will introduce additional errors remains to be quantified. According to the simulations by Graven and Grubber (2011), Turnbull et al. (2009) and Miller et al. (2012), one can expect that the impact of signals from the uncertainties associated in the estimate of these fluxes, on the conversion of atmospheric ¹⁴CO₂ measurements to FFCO₂, are typically below 1 ppm, i.e. much smaller than the observation errors that have been accounted for in this study, justifying that we have ignored these fluxes. However these signals may have complex spatial and temporal patterns leading to significant impact on the quantification of the inversion performances. Uncertainties in the trends of these fluxes could also impact that in the fossil fuel trend detection. Therefore, in future studies, especially if working with real data, the impacts from uncertainties in the ¹⁴CO₂ fluxes other than the anthropogenic fossil fuel emissions need to be

investigated and accounted for by modelling all these $^{14}\text{CO}_2$ fluxes, their atmospheric $^{14}\text{CO}_2$ signals and associated uncertainties.”

References:

- Graven, H. D. and Gruber, N.: Continental-scale enrichment of atmospheric $^{14}\text{CO}_2$ from the nuclear power industry: potential impact on the estimation of fossil fuel-derived CO_2 , *Atmos. Chem. Phys.*, 11(23), 12339–12349, doi:10.5194/acp-11-12339-2011, 2011.
- Turnbull, J., Rayner, P., Miller, J., Naegler, T., Ciais, P. and Cozic, A.: On the use of $^{14}\text{CO}_2$ as a tracer for fossil fuel CO_2 : Quantifying uncertainties using an atmospheric transport model, *J. Geophys. Res.*, 114(D22), doi:10.1029/2009jd012308, 2009.
- Miller, J. B., Lehman, S. J., Montzka, S. A., Sweeney, C., Miller, B. R., Karion, A., Wolak, C., Dlugokencky, E. J., Southon, J., Turnbull, J. C. and Tans, P. P.: Linking emissions of fossil fuel CO_2 and other anthropogenic trace gases using atmospheric $^{14}\text{CO}_2$, *J. Geophys. Res.*, 117(D8), doi:10.1029/2011jd017048, 2012.

2. Line 145: The authors say that the inversion interpret the gradient between JFJ and other sites. I do not understand how that is implemented. Is it that JFJ is the only background site in the network, and hence the inversion implicitly interprets gradients w.r.t. JFJ (much as a global CO_2 inversion might interpret everything w.r.t. MLO and SPO)? Or is it that the pseudo-obs are fed in after explicitly subtracting the JFJ time series, in which case the model’s observation operator looks like “site – JFJ” at each individual site? Basically, the authors say in words that they interpret the gradient w.r.t. JFJ, but I do not understand how that is implemented in practice.

Response:

Our implementation corresponds to the second option explained by the reviewer, i.e. the pseudo-observation are differences between the data at other sites and at JFJ for a given time, and the observation operator relates the fluxes to “site minus JFJ” gradients. To clarify this, we revise the sentence in lines 159-160: “This correction is based on (i) a set of gradients of FFCO_2 mixing ratios **between the different measurement sites and JFJ** sampled during the afternoon (see Sect. 2.2.2) across Europe, called hereafter the “observations” \mathbf{y}_o , (ii)...”, in lines 237-238: “In this study, we first consider 2-week integrated afternoon data. **More precisely, we first consider 2-week averages of afternoon FFCO_2 gradients with respect to JFJ.** In addition, we present tests with daily afternoon **gradients**, for which **the corresponding** sampling scheme would be more costly.”, and in line 249: “... $\mathbf{H}_{\text{transp}}$ is the atmospheric transport model, and \mathbf{H}_{samp} samples the FFCO_2 gradients **with respect to JFJ** corresponding to the observation vector from the transport model outputs (Wang et al. 2017).”

3. Line 201: I’m having trouble deciphering the meaning of “mismatch reduction”, and its bounds and limits. Instead of describing it in words after equation (4), could the authors please write down the mathematical expressions for ϵ_a and ϵ_b ? Since I did not know what those ϵ ’s were, I also could not interpret maps of MR (e.g., Figure 4). In particular, I did not understand what negative vs positive MR meant.

Response:

The “misfits” (i.e. the “mismatch” in the reviewer’s comment) that are considered for the “misfit reduction” are the differences between the prior or posterior and “true”

estimates of the emission budgets. For a given region i and month m in the control vector \mathbf{x} , the prior misfit $\mathbf{x}_{i,m}^b - \mathbf{x}_{i,m}^t$ is denoted $\boldsymbol{\varepsilon}_{i,m}^b$ and the posterior misfit $\mathbf{x}_{i,m}^a - \mathbf{x}_{i,m}^t$ is denoted $\boldsymbol{\varepsilon}_{i,m}^a$. We compute a misfit reduction for each region-month emission budget as the relative difference between the prior and posterior misfits:

$$\text{MR}_{i,m} = 1 - \boldsymbol{\varepsilon}_{i,m}^a / \boldsymbol{\varepsilon}_{i,m}^b$$

We only have one practical realization for \mathbf{x}^b , \mathbf{y}_o and \mathbf{x}^a and thus a single realization of the misfits for each month and region. However, we want to have a statistical assessment of the performance of the inversion system based on the misfits that could be compared to the scores of uncertainty reduction. Therefore, we also consider the typical MR at the 1-month scale for a given region i which is the relative difference between the quadratic mean of the monthly prior misfits and the quadratic mean of the monthly posterior misfits.

$$\text{MR}_i = 1 - \frac{\sqrt{\frac{1}{12} \sum_{m=1}^{12} (\boldsymbol{\varepsilon}_{i,m}^a)^2}}{\sqrt{\frac{1}{12} \sum_{m=1}^{12} (\boldsymbol{\varepsilon}_{i,m}^b)^2}}$$

In all cases MR values could theoretically range between $-\infty$ and 1. When, on average, the posterior emission estimates are closer to the synthetic truth than the prior estimates, the MR is positive (the inversion reduces the misfits). Conversely, when the posterior emission estimates are further from the synthetic truth than the prior estimates, the MR is negative (the inversion increases the misfits). MR is null when the posterior misfits are as large as the prior misfits, i.e. the inversion do not decrease or increase the misfits.

We revise the paragraphs in lines 202-210: “true values for the corresponding emission budgets. MR range from negative values (when the inversion deteriorates the precision of the estimation) to 1 (or “100%”; when the inversion provides a perfect estimate of the emissions).

We focus on uncertainties and misfits at both monthly and annual scales. However, we can have only one practical realization for \mathbf{x}^b , \mathbf{y}_o and \mathbf{x}^a following the protocol of that is presented in Sect. 2.3. Therefore, the assessment of the performance of the inversion for a given region-month using the corresponding score of MR may be over- or under-estimated due to the lack of sampling of the prior and observation errors. Consequently, at monthly scale, in order to strengthen the evaluation of the theoretical uncertainties based on these single realizations of the prior and posterior misfits, we compare, for a given region, the quadratic mean of the twelve monthly misfits (called “monthly misfits” without mention of a specific month in Sect. 3) to the quadratic mean of the standard deviations of the twelve monthly uncertainties (called “monthly uncertainties” without mention of a specific month in Sect. 3), which characterizes the average monthly uncertainties over the year. This computation implicitly assumes that the twelve monthly misfits through a year follow the same statistical distribution, and represent twelve independent realization of this distribution. In such a situation, the comparison between the averages of the prior and posterior monthly misfits give a good indications of the error reduction that should not be highly skewed by sampling errors. In the result section, for a given region i ,

UR and MR scores derived at the “monthly” scale without mention to a specific month will correspond to the relative difference between the prior and posterior values of these average monthly uncertainties and misfits from a whole year of inversion:

$$UR_i = 1 - \frac{\sqrt{\frac{1}{12} \sum_{m=1}^{12} (\sigma_{i,m}^a)^2}}{\sqrt{\frac{1}{12} \sum_{m=1}^{12} (\sigma_{i,m}^b)^2}} \quad (5)$$

$$MR_i = 1 - \frac{\sqrt{\frac{1}{12} \sum_{m=1}^{12} (\varepsilon_{i,m}^a)^2}}{\sqrt{\frac{1}{12} \sum_{m=1}^{12} (\varepsilon_{i,m}^b)^2}} \quad (6)$$

At the annual scale, ...”

Minor comments

1. Line 151: For NET233, each grid box is supposed to have one urban and one rural site. I’m not sure that’s a good strategy. Wouldn’t it be better to designate urban/rural depending on the nearest NET43 site? I mean, there could easily be grid boxes where it was more realistic to put two rural or two urban sites.

Response:

The design of NET233 was not intended to be optimal. We wanted to conduct a test with many sites whose distribution would be homogeneous and cover all the control regions. In this case, the variations of UR from one region to the other one are a direct consequence of variations in the emission uncertainties, and not as a consequence of the variations in the network density. Putting most of the sites near the areas with the highest emissions could have larger URs than spreading them homogeneously across Europe, but will not distinguish the role of network density and the role of emissions uncertainties themselves in the variations of the UR from one region to the other one.

Even though the distribution of the emissions is highly heterogeneous in Europe, the grid cells of the transport model used in this study have a 3.75 °x2.5 ° resolution (i.e. they cover areas that are much larger than megacities like London or Paris), so that we can assume that nearly all pixels have both rural and urban locations.

We clarify the rationale for the NET233 network in Sect. 2.1 by adding in line 141: “... in which two sites are placed in each European land pixel of the LMDZv4 transport model (Fig. 1c). The NET233 network is denser than NET17 and NET43 in the high emitting regions, e.g. Germany, and also covers the region that is not well sampled by NET17 and NET43. However, the location of its 233 sites is not intended to be optimal since the emissions have a very heterogeneous spatial distribution. Their homogeneous spreads allow us to reduce the impact of representation and aggregation errors (Tramper and Snieder, 1996; Kaminski et al., 2001) and to assess the impact of having a dense network for all control regions.”

2. Line 233: In the ICOS protocol, are the two-week samples going to be filled continuously, or are they only going to integrate mid-afternoon (or nighttime) air?

That would very much change the sensitivity of the observations to FF CO₂, and the impact of transport errors.

Response:

The present protocol for almost all of the ICOS ¹⁴CO₂ sites is to fill continuously two-week samples. And some sites fill continuously daily/weekly atmospheric ¹⁴CO₂. However, the option of intermittent filling of air samples is feasible in practice and has been used (Levin et al., 2008; Turnbull et al., 2016). On the other hand, the state-of-the-art transport models used for atmospheric inversion studies have still difficulties in simulating the vertical mixing during night-time and in the morning. So we prefer to consider this option of two-week afternoon sampling in the OSSE. We keep in mind the fact that using samples filled continuously over two weeks instead of during the afternoon only would definitely change the transport condition and sensitivity of the observation to FFCO₂, and thus may give different values of UR and MR. But we assume that it should change the result in a quantitative but not qualitative way.

We add a sentence in line 232 to mention that common practice of ¹⁴CO₂ sampling is continuously over the course two weeks: “Current atmospheric ¹⁴CO₂ samples in Europe are usually filled continuously over the course of two weeks (Vogel et al., 2013; Levin et al., 2013). However, state-of-art inversion systems generally make use of data during the afternoon only, ...” We already mentioned in the manuscript in the same paragraph that intermittent filling of air samples is practically feasible so that our definition of the observations in this study can be seen as a compromise between the current requirement of state-of-the-art inversion systems and current measurement practices.

References:

- Turnbull, J. C., Keller, E. D., Norris, M. W. and Wiltshire, R. M.: Independent evaluation of point source fossil fuel CO₂ emissions to better than 10%, Proc. Natl. Acad. Sci. U. S. A., 113(37), 10287–10291, doi:10.1073/pnas.1602824113, 2016.
- Levin, I., Hammer, S., Kromer, B. and Meinhardt, F.: Radiocarbon observations in atmospheric CO₂: determining fossil fuel CO₂ over Europe using Jungfraujoeh observations as background, Sci. Total Environ., 391(2–3), 211–6, doi:10.1016/j.scitotenv.2007.10.019, 2008.

3. Line 237: Are the authors assuming that two week average $\Delta^{14}\text{CO}_2$ will translate into two week average FF CO₂? What the two week average $\Delta^{14}\text{CO}_2$ represents depends on the method of collection; an open tray will fix CO₂ proportional to the partial pressure of CO₂, while a bubbled trap will fix all the CO₂ in the ingested air. The former represents average FF CO₂ weighted by the total CO₂ mole fraction, while the latter represents average FF CO₂ over two weeks. Which one applies for the ICOS protocol?

Response:

The sampling technology (https://www.icos-cal.eu/crl/radiocarbon_samples) that ICOS utilize for two-week integrated samples follows the method developed in the 1970s (Levin et al., 1980). The sampling system is equipped with a small aquarium pump, which actively collects about 15 m³ of air during the two week period. CO₂ is collected by chemical absorption in CO₂-free NaOH solution in the so-called

Raschig-tube samplers. In this context, the FFCO₂ represents the average FFCO₂ over two weeks (the second way that the reviewer mentioned).

We clarify the fact that our definition of FFCO₂ corresponds to the average afternoon FFCO₂ over the course of two weeks here in line 237: “In this study, we first consider 2-week integrated afternoon data. **More precisely, we first consider 2-week averages of afternoon FFCO₂ gradients with respect to JFJ.** In addition, ...”

References:

Levin, I., Munnich, K. O., Weiss, W.: The effect of anthropogenic CO₂ and ¹⁴C sources on the distribution of 14C in the atmosphere, Radiocarbon, 22, 379–391, 1980.

4. Line 253: Did the authors model a diurnal cycle in FF CO₂ emissions? According to Nassar et al. (2013), the diurnal cycle can be fairly large over populated areas. Along with the selective mid-afternoon sampling used by the authors, the impact could be sizeable.

Response:

We fully account for the diurnal cycle of emissions in our computations. The synthetic truth of the OSSEs and the synthetic FFCO₂ observations are modelled using the emission map IER-EDG, which is an hourly emission products and has a clear diurnal cycle in the emissions. However, the emission map used to compute the observation operator is PKU-CO₂ version 1 (Wang et al., 2013), which is an annual product and which does not have temporal profiles, so that there is no diurnal/seasonal cycle in $\mathbf{H}_{\text{distr}}^{\text{PKU}}$. The mismatch between the temporal profiles of synthetic true emission and $\mathbf{H}_{\text{distr}}^{\text{PKU}}$ contribute to the so-called aggregation error (Wang et al., 2017). As shown in Table S3, the typical aggregation error is only 0.17-0.30 ppm, indicating that such difference only has a small impact on the simulated FFCO₂ signals. Furthermore, the inversion results indicate that the inversion can significantly reduce the uncertainties and misfits of the estimate of monthly emission budgets for large regions, even if using flat temporal profiles for the emissions while the truth is modeled with diurnal, weekly and seasonal temporal profiles for the emissions. One explanation is that the two-week mean afternoon samplings integrate the signal from both daytime and nighttime emissions across Europe due to the atmospheric transport.

We add a discussion about this in Sect 4.2, line 496: “...yet mainly over high emitting regions. **In particular, Sect. 3 indicates that the inversion can significantly reduce the uncertainties and misfits in the estimate of monthly emission budgets for large or high emitting regions, even though the observation operator used by the inversion assumes flat temporal profiles for the emissions while the true emissions have diurnal, weekly and seasonal temporal profiles. This confirms that the two-week mean afternoon ¹⁴CO₂ samplings integrate the atmospheric signal transported from both daytime and nighttime emissions across Europe which can be filtered from the signal from local emissions to provide large-scale information on the emissions.**”

References:

Wang, R., Tao, S., Ciais, P., Shen, H. Z., Huang, Y., Chen, H., Shen, G. F., Wang, B., Li, W., Zhang, Y. Y., Lu, Y., Zhu, D., Chen, Y. C., Liu, X. P., Wang, W. T., Wang, X. L., Liu, W. X., Li, B. G. and Piao, S. L.: High-resolution mapping of combustion processes and implications for CO₂

emissions, Atmos. Chem. Phys., 13(10), 5189–5203, doi:10.5194/acp-13-5189-2013, 2013.
Wang, Y., Broquet, G., Ciais, P., Chevallier, F., Vogel, F., Kadyrov, N., Wu, L., Yin, Y., Wang, R. and Tao, S.: Estimation of observation errors for large-scale atmospheric inversion of CO₂ emissions from fossil fuel combustion, Tellus B: Chemical and Physical Meteorology, 69(1), 1325723, doi:10.1080/16000889.2017.1325723, 2017.

5. Line 264: Does “practical” refer to the operator used to generate pseudo-observations from the “true” fluxes?

Response:

No, \mathbf{H}^{prac} means this operator is a practical representation of the flux distribution and atmospheric transport that is used in the inversion system, while \mathbf{H}^{OSSE} is the theoretical “true” operator used to generate pseudo-observations for OSSEs. We explained in line 264 the $\mathbf{H}^{\text{prac}} = \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{PKU}}$ and we also wrote in line 364 “The synthetic observations are generated using $\mathbf{x}^{\text{IER-EDG}}$ and the operator $\mathbf{H}^{\text{OSSE}} = \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{IER-EDG}}$, which relies on the same $\mathbf{H}_{\text{samp}}^{\text{coloc}}$ and $\mathbf{H}_{\text{transp}}$ operators as the \mathbf{H}^{prac} observation operator used in the inversion system.” So the difference between the practical operator \mathbf{H}^{prac} and \mathbf{H}^{OSSE} is their last sub-operators $\mathbf{H}_{\text{distr}}^{\text{PKU}}$ and $\mathbf{H}_{\text{distr}}^{\text{IER-EDG}}$, respectively. To clarify such a use of the “practical” term to name \mathbf{H}^{prac} , we revise the sentence in Sect. 2.2.2, line 264: “In sum, the observation operator used in the practical configuration of the inversion system is defined by $\mathbf{H}^{\text{prac}} = \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{PKU}}$.”

6. Line 296: Is the covariance model global, or is this done only over Europe?

Response:

This covariance model (which ignores the spatial correlations) is applied for the estimate of the prior uncertainty variance and temporal correlations for each of the control regions over the globe. We revise sentences in lines 295-296 to emphasize the fact that it is applied to all regions over the globe, not only to regions in Europe: “..., except for the Balkans where they reach up to 44%. We assume that there is no spatial correlation of the prior uncertainty between different control regions. For each control region of the globe, the statistics of the difference between the monthly emission budgets from the two maps are fitted by a covariance model that combines four different covariance matrices” and we delete the last sentence (lines 300-301) of this paragraph.

7. Line 363: Can the authors explain how they obtained the IER hourly inventory? I tried to download it from their website, but given that each month had to be separately downloaded, that was very inconvenient. An email to the contact person listed on the website bounced, so that was a dead end too.

Response:

We downloaded the “global fossil fuel emissions” under the “product” tab month by month. These files are hourly emissions for each month and for three emission heights. We summed the emissions at different emission heights together (our simulation of FFCO₂ assumes that all the emissions are emitted at surface). In this case, we obtained twelve files of global hourly emission fields for one year and each

file corresponds to one month.

8. Line 366: This will only get at the random error in transport, not any systematic error in transport modeling. Have the authors tried quantifying the impact of systematic errors in LMDZ, say by using ^{222}Rn or SF_6 ?

Response:

We prefer to consider that errors can have long spatial and temporal scales rather than a “systematic” component, since what the people usually call biases or systematic errors vary in time and can be difficult to predict and diagnose. The inversion system integrate such temporal and spatial consistency of the errors through the temporal and spatial correlations in **R**.

The assessment of the model errors over long temporal scales, e.g. by using real tracer data was out of the scope of this study and of that of Wang et al. (2017). It has been the topic of numerous model inter-comparison studies in the past including LMDZ such as Peylin et al. (2011) and Patra et al. (2011). Locatelli (2015) also studied in details the transport errors by LMDZ. These studies show that the skills of LMDZ are in line with state-of-the-art transport models and there is no evidence that LMDZ has a systematic error compared to other transport models. Furthermore, our aim here is not to assign transport errors for the specific transport model we use, but rather for a typical global transport model in order to produce general results.

The structure of the transport error is hard to assess for a specific transport models. Some studies have made some attempts to characterize it, but have come to different conclusions. For example, Lin and Gerbig (2005) determined the correlation timescale to be 2-3 hours for U-/V- winds. Broquet et al. (2011) analyzed the distribution of differences between the simulated and measured atmospheric mole fractions of ^{222}Rn (Radon). Temporal auto-correlations with 3 to 6-hour timescales are found in Broquet et al. (2011) based on such an analysis. Lauvaux et al. (2009) estimated the potential of the temporal correlations for transport error based on an ensemble of perturbations of the simulation of atmospheric transport. They found there are negative correlation between the atmospheric CO_2 mole fractions in the day and the night and the correlations (of the errors) with a lag time of 1 day are close to 0 for midafternoon data (Fig. 4a in Lauvaux et al., 2009). Miller et al. (2015) investigated the magnitude of temporally covarying atmospheric transport errors and found that removing temporal covariances in the transport would underestimate the transport error at the monthly scale, but their study did not try to use any function to describe the temporal auto-correlations so that their results are difficult to generalize.

Most of these studies found temporal error correlations that are generally smaller than 1 day. If considering daily to monthly concentration averages, such a correlation should increase (if the transport error is a combination of error components at high and low frequencies). However, these various studies prove that the correlations in the transport error are challenging to estimate and the majority of existing inversion studies do not account for such potential correlations in atmospheric transport modeling (Peylin et al., 2013; Chevallier et al., 2010; Kadygrov et al., 2015; Peters et al., 2007; Gurney et al., 2008; Niwa et al., 2012). So, we keep the traditional

assumption.

We revise the paragraphs about the configuration of transport errors, as well as representation and aggregation errors in lines 328-337: "... For ϵ_a , the standard deviation for spring/summer and autumn/winter were derived. The standard deviation of the transport error at a given site is set-up proportional to the temporal standard deviation of the 1-year long time-series of the high-frequency variability of the detrended and deseasonalized simulated daily mean afternoon mixing ratios in the grid cell of the transport model, at which the sites are located. Such an estimation of transport error which relies on some results from Peylin et al. (2011) aims at representing the typical value for global transport models, not that of the specific transport model used in this study. The temporal auto-correlations in the representation and aggregation errors were characterized by Wang et al. (2017) using the sum of a long-term component and a short-term component: $r(\Delta t) = a \times e^{-\Delta t/b} + (1-a) \times e^{-\Delta t/c}$ where Δt is the timelag (in days) and a, b, c are parameters optimized by regressions against the samples of the errors. Furthermore, we do not include temporal auto-correlations in the transport error for simulated daily to 2-week mean afternoon FFCO₂ gradients, since previous studies of the auto-correlations of the transport errors have not evidenced that they should be significant at daily scale (Lin and Gerbig, 2005; Lauvaux, 2009; Broquet et al., 2011). This choice follows the corresponding discussion by Wang et al. (2017) and implicitly ignores that transport model errors likely bear long-term components (often referred to as "biases", Miller et al., 2015) even when being dominated by components on short timescales. The corresponding values of the standard deviation and the modelling of temporal autocorrelation of the observation errors for 2-week/daily mean afternoon FFCO₂ gradients are listed in Table S3 and Table S4.

A simpler account of the spatial correlations in the observation errors is derived from the diagnostics of Wang et al. (2017). We do not account for the spatial correlation in the representation error, as the scale of the spatial correlation according to Wang et al. (2017), i.e. 55-89 km, is much smaller than the size of the grid cells of the global transport model $\mathbf{H}_{\text{transp}}^{\text{LMDZ}}$ used for the inversion. When there are more than two sites are located in the same grid cell of the transport model, we consider that the aggregation errors and the transport errors are fully correlated between these sites, according to the definition by Wang et al. (2017). We do not account for spatial correlations between aggregation errors for measurements made at sites in different grid cells, because the scale of the spatial correlation is 171 km and is smaller than the size of the grid cell, according to Wang et al. (2017). Finally, we do not account for spatial correlations between transport errors or measurements made at sites in different grid cells."

References:

- Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T. and Ciais, P.: A European summertime CO₂ biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, *J. Geophys. Res.*, 116(D23), doi:10.1029/2011jd016202, 2011.
- Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krümmel, P.

- B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morgu í J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S. and Worthy, D.: CO₂ surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *J. Geophys. Res.*, 115(D21), doi:10.1029/2010jd013887, 2010.
- Gurney, K. R., Baker, D., Rayner, P. and Denning, S.: Interannual variations in continental-scale net carbon exchange and sensitivity to observing networks estimated from atmospheric CO₂ inversions for the period 1980 to 2005, *Global Biogeochem. Cycles*, 22(3), GB3025, doi:10.1029/2007GB003082, 2008.
- Kadygrov, N., Broquet, G., Chevallier, F., Rivier, L., Gerbig, C. and Ciais, P.: On the potential of the ICOS atmospheric CO₂ measurement network for estimating the biogenic CO₂ budget of Europe, *Atmos. Chem. Phys.*, 15(22), 12765–12787, doi:10.5194/acp-15-12765-2015, 2015.
- Lauvaux, T., Pannekoucke, O., Sarrat, C., Chevallier, F., Ciais, P., Noilhan, J. and Rayner, P. J.: Structure of the transport uncertainty in mesoscale inversions of CO₂ sources and sinks using ensemble model simulations, *Biogeosciences*, 6(6), 1089–1102, 2009.
- Locatelli, R., Bousquet, P., Hourdin, F., Saunois, M., Cozic, A., Couvreur, F., Grandpeix, J.-Y., Lefebvre, M.-P., Rio, C., Bergamaschi, P., Chambers, S. D., Karstens, U., Kazan, V., van der Laan, S., Meijer, H. A. J., Moncrieff, J., Ramonet, M., Scheeren, H. A., Schlosser, C., Schmidt, M., Vermeulen, A. and Williams, A. G.: Atmospheric transport and chemistry of trace gases in LMDz5B: evaluation and implications for inverse modelling, *Geosci. Model Dev.*, 8(2), 129–150, doi:10.5194/gmd-8-129-2015, 2015.
- Lin, J. C. and Gerbig, C.: Accounting for the effect of transport errors on tracer inversions, *Geophys. Res. Lett.*, 32(1), L01802, doi:10.1029/2004GL021127, 2005.
- Miller, S. M., Hayek, M. N., Andrews, A. E., Fung, I. and Liu, J.: Biases in atmospheric CO₂ estimates from correlated meteorology modeling errors, *Atmos. Chem. Phys.*, 15(5), 2903–2914, 2015.
- Niwa, Y., Machida, T., Sawa, Y., Matsueda, H., Schuck, T. J., Brenninkmeijer, C. A. M., Imasu, R. and Satoh, M.: Imposing strong constraints on tropical terrestrial CO₂ fluxes using passenger aircraft based measurements, *J. Geophys. Res.*, 117(D11), D11303, doi:10.1029/2012JD017474, 2012.
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M., Petron, G., Hirsch, A. I., Worthy, D. E., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C. and Tans, P. P.: An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proc. Natl. Acad. Sci. U. S. A.*, 104(48), 18925–30, doi:10.1073/pnas.0708986104, 2007.
- Peylin, P., Houweling, S., Krol, M. C., Karstens, U., Rödenbeck, C., Geels, C., Vermeulen, A., Badawy, B., Aulagnier, C., Pregger, T., Delage, F., Pieterse, G., Ciais, P. and Heimann, M.: Importance of fossil fuel emission uncertainties over Europe for CO₂ modeling: model intercomparison, *Atmos. Chem. Phys.*, 11(13), 6607–6622, doi:10.5194/acp-11-6607-2011, 2011.
- Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R. and co-authors 2013. Global atmospheric carbon budget: results from an ensemble of atmospheric CO₂ inversions. *Biogeosciences* **10**, 6699–6720.
- Wang, R., Tao, S., Ciais, P., Shen, H. Z., Huang, Y., Chen, H., Shen, G. F., Wang, B., Li, W., Zhang, Y. Y., Lu, Y., Zhu, D., Chen, Y. C., Liu, X. P., Wang, W. T., Wang, X. L., Liu, W. X., Li, B. G. and Piao, S. L.: High-resolution mapping of combustion processes and implications for CO₂ emissions, *Atmospheric Chemistry and Physics*, 13(10), 5189–5203, doi:10.5194/acp-13-5189-2013, 2013.

9. Line 367: I know it is usual practice in the OSSE world to perturb the measurements according to the error statistics of R, but I have never understood why, unless it is done in an ensemble of multiple realizations of the measurements. In an ensemble of inversions with different measurements from the same network, it makes perfect sense to produce those measurements using perturbations according to R, since the resulting spread in the flux estimates then gives the uncertainty due to R. However, for a single inversion, perturbing the measurements according to R only ensures that the posterior will be different from the “true” flux, without any way to infer the significance of that difference. As in, how do the authors know that the MR’s they estimate are not because in the one realization of the measurements they used,

some of them just happened to be skewed in one direction? This is especially a concern for the NET17 network, since there are so few measurements, with scant opportunity to average over the perturbations.

Response:

We carefully analyzed and discussed the UR and MR results keeping in mind and explicitly recalling the problems of working with a single realization of \mathbf{y}_o and \mathbf{x}^b . The risk of under- or over-estimating the posterior error in the emission budgets when using one realization only for the prior and measurement errors is the reason why we compute the “statistics of misfits” at the monthly scale and the “MR for monthly emissions” (see our answer to major question 3). By assuming that the monthly misfits all follow the same distribution, we roughly consider that we have an ensemble of twelve realizations of monthly inversions for which the average-based MR give a reliable indicator or the error reduction that should not be significantly skewed by sampling errors.

Of note is also that even when considering NET17 and a 2-week sampling strategy, we have 416 data over the year and thus 416 realizations of perturbations to individual observations. Finally, the scores of UR should definitely be considered as the reference indicators of the inversion behavior, and the MR are mainly analyzed to provide confidence in these UR, as explained in Sect. 2.2.1.

The perturbation of the observation according to \mathbf{R} even when a single realization of the observation and prior errors is considered is a common practice in inversion because, we believe, this is what still makes most sense if having to assess the uncertainty reduction using a single inversion. In other words, we do not have a simple idea of perturbation that would make the single realization more adapted for assessing the typical impact of the errors.

We revise the paragraphs in Sect. 2.2.1, lines 203-204 to clarify the use of one realization of observations: “We focus on uncertainties and misfits at both monthly and annual scales. However, we **can** have **only** one practical realization for \mathbf{x}^b , \mathbf{y}_o and \mathbf{x}^a following the protocol of that is presented in Sect. 2.3. Therefore, the assessment of the performance of the inversion for a given region-month using the corresponding score of MR may be over- or under-estimated due to the lack of sampling of the prior and observation errors. Consequently, ...”

References:

Basu, S., Miller, J. B. and Lehman, S.: Separation of biospheric and fossil fuel fluxes of CO₂ by atmospheric inversion of CO₂ and ¹⁴CO₂ measurements: Observation System Simulations, Atmos. Chem. Phys., 16(9), 5665–5683, doi:10.5194/acp-16-5665-2016, 2016.

10. Line 374 and Figure 4: The authors solve for monthly emissions over a year, but report a single UR/MR map of monthly emissions. Is this the RMS of UR/MR values over 12 months, or the UR/MR calculated from the RMS of the posterior errors, or...? As in, can the authors give a mathematical expression of what is being shown in Figure 4 as the “monthly” UR/MR, in terms of their control vector and/or covariance matrix?

Response:

This is related to the major comment #3. We computed the quadratic mean of the twelve values of monthly uncertainties and misfits to derive the so-called “average monthly URs and MRs”. As stated in the response to the major comment #3, we add two equations to show that we compute the UR and MR “at monthly scale” based on the quadratic mean of twelve monthly values across one year. Please see the changes correspond to the major comment #3.

11. Line 381: I think the reference to Figure 4(d) should actually be to 4(e). Likewise, in line 384, the references should be to 4(e) and 4(g).

Response:

The reviewer is right. In line 381, the reference to figure is Fig. 4e and on line 384, the reference to figure is Fig. 4e and 4g. We are sorry about these mistakes. We went through the paper again carefully and feel that there are not such mistakes anymore.

12. Line 409: “... the posterior misfits are even larger than the prior misfits.” Why does the inversion allow this? For stations within the blue regions, is this obvious from looking at the atmospheric FF CO₂ time series, that post-optimization the time series is further away from the pseudo-data than pre-optimization? I suspect the perturbed measurements are to blame (see earlier comment).

Response:

The primary reason for getting posterior misfits in the emissions that are larger than the prior misfits in the emissions is connected to the fact that the prior uncertainty matrix **B** of the inversion system does not match the statistics of the actual errors in the prior estimate of emissions:

1) As demonstrated by Wang et al. (2017), the signature of the errors in the prior estimates in the FFCO₂ emissions has a smaller amplitude than the observation errors and the ability to filter this information for a proper correction of the emissions strongly relies on the knowledge of the prior uncertainty covariance. If **B** misses the amplitude, temporal and spatial correlations of the actual errors, the system can easily translate observation errors into its corrections to the emissions. While decreasing the differences between the simulated FFCO₂ concentration time series and the atmospheric data, it would increase the misfits to the true emissions.

2) Some of the region-months are poorly constrained by the observations due to the network distribution and to the meteorological conditions (even when using NET233), and the corrections to emissions from a poorly constrained region-months is imposed by the extrapolation of the corrections to other region-months following the patterns of **B**. If those patterns are wrong (typically, if spatial correlations in **B** for the errors of two neighbor regions are negative while the spatial correlations between the actual errors of these regions are positive), the system could apply corrections with a wrong sign or amplitude to the poorly observed region-months, which can easily lead to an average increase of the errors for such regions. The problem is similar when the network can constrain the sum of the budgets for several region-month but not these region month individually (due to being too coarse). In

this case, if the structure of **B** is wrong, the repartition of the constraints from the observations between these different region-months could be erroneous.

The fact that the aggregation error has a non-Gaussian and biased distribution while the inversion system believe it is Gaussian and unbiased can also feed the generation of negative MR as well as problems of sampling errors when computing scores of MR as suggested by the reviewer. However we believe that the impact of these additional factors is relatively small because when the setup of **B** matches well (even though not perfectly) with that of the actual prior errors, we have very few negative MRs.

We add in lines 546 to discuss the negative scores of MR: "... This degradation occurs even when using daily measurements or the network NET233. A first explanation is that the signature of the errors in the prior emission estimates in the FFCO₂ fields has a smaller amplitude than the observation errors and thus the ability to filter this information for a proper correction of the emissions strongly relies on the knowledge of the prior uncertainty covariance. If **B** misses the amplitude and the temporal and spatial correlations of the actual errors, the system can translate observation errors into corrections to the emissions. Furthermore, some of the region-months are poorly constrained by the observations (due to the meteorological conditions and/or to the observation network spatial distribution), and the corrections to such region months is imposed by the extrapolation of the corrections to other region-months following the uncertainty structures characterized by **B**. If those structures do not represent the actual errors correctly, the system could apply corrections with a wrong sign or amplitude to the poorly observed region-months. A similar problem occurs when the network can constrain the sum of the budgets for several region-months but not the individual budgets of these region months (due to being too coarse). If the structure of **B** is wrong, the repartition of the constraint from the observations between these different region-months can be erroneous. All these analysis reveal the difficulty to capture the signatures of uncertainties in the prior emission estimate from the assimilated prior-model data misfits in our specific inverse modeling problem and thus to derive good corrections when the prior uncertainty covariance matrix is not configured properly."

References:

Wang, Y., Broquet, G., Ciais, P., Chevallier, F., Vogel, F., Kadyrov, N., Wu, L., Yin, Y., Wang, R. and Tao, S.: Estimation of observation errors for large-scale atmospheric inversion of CO₂ emissions from fossil fuel combustion, *Tellus B: Chemical and Physical Meteorology*, 69(1), 1325723, doi:10.1080/16000889.2017.1325723, 2017.

13. Line 417: The correlation is not between uncertainties, but between corrections from the prior emission.

Response:

We do not agree with the reviewer. Fig. 6 shows the correlations in the **B** and **A** covariance matrices. Since **B** and **A** characterize the uncertainties in the prior and posterior estimates, the correlations in both **B** and **A** are indeed between uncertainties rather than corrections.

Following Eq. (2) in the main text, the correction for \mathbf{x} is actually $\mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{y}_o-\mathbf{H}\mathbf{x}^b)$, so that the covariance matrix for the corrections is $\mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{R})\mathbf{R}^{-1}\mathbf{H}\mathbf{A}$, and the correlations correspond to this matrix are between corrections.

14. Line 455: I'm surprised at the low UR for the NET233 network. Why are there so many white areas (low UR) still?

Response:

This is related to the fact that the signals from the uncertainties in the prior estimate of emissions are well below the observation errors in these regions and that the observation errors have similar error structures (temporal and spatial correlations) to the signals of uncertainties in the prior estimate of emissions, as discussed in Sect. 4.2 (line 516-520). This demonstrates that the proper account for representation and aggregation errors avoid overestimating the potential of the atmospheric inversion in OSSES when using a coarse resolution transport model.

We emphasize this by adding discussions in line 520: “which are close to the prior FFCO₂ errors (Wang et al., 2017). Such a result illustrates the need for using a suitable observation error characterization (here based on the results from Wang et al., 2017) to prevent the stations having a full coverage of information on the emissions in the model framework shown here even when the observation network is as dense as NET233. A proper account for the observation errors and their temporal and spatial correlations avoid overestimating the potential of the atmospheric inversion in OSSES when using a coarse resolution transport model.”

15. Line 484: In real trend detection situations, the transport will vary year by year, as will the disequilibrium and nuclear fluxes of ¹⁴C. Can the authors estimate how big an impact this will have on the trend detection?

Response:

There is not a strong reason to think that there could be some major change in the uncertainties in the estimate of nuclear and disequilibrium fluxes of ¹⁴C from year to year, which is what really matters on the inversion results (rather than changes in the absolute value for these fluxes), as discussed in the response to the first major comments. This response also indicates that we assume that such uncertainties should have a relatively weak impact on the inversion results.

The transport conditions (and thus \mathbf{H} and the uncertainties in the transport modeling in \mathbf{R}) may vary significantly from year to year which could lead to variations of the sensitivity of the observations to the emissions and variations of the scores of uncertainty reduction. We could even observe trends in the transport conditions over several years (e.g. the shallowing of the PBL between the 1990s and the 2000s by Aulagnier et al., 2009; Ramonet et al., 2010). However, the impact of the resulting changes in atmospheric transport condition on the UR and posterior uncertainties, and thus on our computation of the trend detectability is difficult to anticipate. We assume that this should not result in a behavior that is different from the one described by our computation with constant posterior uncertainties from year

to year (it should change the result in a quantitative but not qualitative way). Conducting tests with varying posterior uncertainties would thus fall out of the scope of our assessment of the typical uncertainties in trends for a typical uncertainty in annual emissions, and we thus think that it is not worth investigating it.

We delete the simple statement in line 479: “...fully independent ~~and ignoring the changes in transport on decadal scales (Aulagnier et al. 2009; Ramonet et al., 2010)~~, we calculate the uncertainty in relative trends ...” and add a paragraph to discuss that the changing transport may have an impact on the trend detection but quantifying this impact would require detailed studies and further investigation after line 490: “Our assumption that the posterior uncertainties in annual emissions have the same amplitude from year to year should not strongly drive the results, so the results here give a good indication of the level of uncertainty in the trend detection for a typical level of uncertainty at the annual scale. However, changes of the transport from year to year or on decadal scales (Aulagnier et al. 2009; Ramonet et al., 2010) may change the level of the sensitivity of the observations to the emissions, i.e., the level of the atmospheric constraint of the inversions which leads to uncertainty reduction, and thus the level of posterior uncertainties on the same timescales. A more complex model accounting for varying levels of annual posterior uncertainties may thus be useful to refine the quantification of the uncertainty in the trends. Of note is that the level of uncertainties in the trends could be increased if the modeling framework accounts for the trends in the transport or in the sources of ¹⁴CO₂ other than the fossil fuel emissions. Such trends in the modeling errors may have to be considered for applications with real data.”

References:

- Aulagnier, C., Rayner, P., Ciais, P., Vautard, R., Rivier, L. and Ramonet, M.: Is the recent build-up of atmospheric CO₂ over Europe reproduced by models. Part 2: an overview with the atmospheric mesoscale transport model CHIMERE, *Tellus B*, 62, 1, 2009.
- Ramonet, M., Ciais, P., Aalto, T., Aulagnier, C., Chevallier, F., Cipriano, D., Conway, T. J., Haszpra, L., Kazan, V., Meinhardt, F., Paris, J.-D., Schmidt, M., Simmonds, P., Xueref-Rény, I. and Necki, J. N.: A recent build-up of atmospheric CO₂ over Europe. Part 1: observed signals and possible explanations, *Tellus B*, 62(1), 1–13, doi:10.1111/j.1600-0889.2009.00442.x, 2010.

16. Line 499: As far as I can tell, Basu et al. (2016) did not estimate UR.

Response:

Basu et al. (2016) did not estimate UR, but they analyzed the misfits between the inversion-estimated fluxes and true fluxes such as when we analyze scores of MR. The text assumed that the “error reduction” can be characterized by either UR or MR. We revise the sentence accordingly to avoid the confusion: “We paid attention (as compared to previous OSSEs published for the USA) to account for aggregation and representation errors, which is the reason why our inversions do not provide as impressive error reductions (**uncertainty and misfit**) as **the misfit reduction** of Ray et al. (2014) and Basu et al. (2016). However, ...”.

17. Line 531: The authors seem to suggest that the boundary condition – a bane of most regional inversions – does not affect their flux and uncertainty estimates. Is this

because everything is referenced to JFJ?

Response:

Our simulation indicated that the reason why, in this study, the uncertainties in the emissions remote from Europe (i.e. what would feed the boundary conditions in a regional model) does not significantly impact the FFCO₂ inversion in Europe is the atmospheric diffusion of the FFCO₂ signal associated with these uncertainties. The high emitting regions outside Europe (USA, China) are far from the European continent. The diffused signals from these regions does not yield large gradients between the European sites which are mainly due to the emissions from European continent. Therefore, these signals should not impact the inversion of fluxes within Europe. Assimilating gradients to JFJ definitely helps the system understand that a large scale signal over Europe should not be connected to European emissions, but inversions system assimilating FFCO₂ data at individual site could also naturally avoid to correct for emissions between the measurements sites based on such a large scale signal.

The boundary conditions for traditional regional inversion of natural CO₂ fluxes assimilating total CO₂ concentrations which only target at natural CO₂ fluxes can be more critical, since there are large fluxes from Atlantic Ocean and other adjacent continents like Asian Russia which can cause significant spatial patterns between European sites. But still, some studies, such as that of Lauvaux et al. (2008) showed that the influence of the boundary conditions is not significant in their regional inversion of the natural CO₂ fluxes in the South West of France.

At last, we keep in mind the fact that if the inversion would account for uncertainties in the biogenic and ocean ¹⁴CO₂ fluxes, the situation could be different and could pose problems for the boundary conditions of the regional scale systems. There could be large ¹⁴CO₂ fluxes from Atlantic Ocean and other adjacent continents like Asian Russia which can cause significant spatial patterns of ¹⁴CO₂ within Europe.

In the manuscript, we already discussed the fact that “the inversion system mainly exploits the signals of the gradients between the European sites to constrain the European emissions” (lines 533-535). We add a sentence in Sect. 4.2, line 539 to discuss the last point we discussed above: “... by controlling the regional transport model boundary conditions. **However, such a conclusion may need to be re-evaluated when processing real data and accounting for uncertainties in other types of ¹⁴CO₂ fluxes, since, e.g., parts of the Atlantic ocean fluxes may have a significant signature on the European ¹⁴CO₂ gradients.**”

References:

Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., Davis, K. J., Ciais, P., Denning, A. S. and Rayner, P. J.: Mesoscale inversion: first results from the CERES campaign with synthetic data, *Atmos. Chem. Phys.*, 8, 3459–3471, 2008.

18. Eqs. C-3 and C-4: I believe there are errors in these two formulae. If \vec{p} is obtained by minimizing the cost function J

$$J = \frac{1}{2}(\vec{y} - \vec{\tilde{y}})^T R^{-1}(\vec{y} - \vec{\tilde{y}}) \quad (1)$$

where $R = \text{cov}(\vec{y})$ (R in their case contains the posterior error estimates on fluxes),

then the optimal estimate of \vec{p} and the corresponding covariance are

$$\vec{p} = (X^T R^{-1} X)^{-1} X^T R^{-1} \vec{y}$$

$$\text{cov}(\vec{p}) = (X^T R^{-1} X)^{-1} X^T R^{-1} X (X^T R^{-1} X)^{-1}$$

Response:

The reviewer is right if $\text{cov}(\mathbf{Y})$ (in the review's equations, it's \mathbf{R}) is not diagonal or if the diagonal terms (variances of the variables in \mathbf{y} vector) are not equal. In fact, the last equation proposed by the reviewer can be further simplified to:

$$\text{cov}(\vec{p}) = (X^T R^{-1} X)^{-1}$$

The equation we wrote as Eqs. C-3 and C-4 only applies when $\text{cov}(\mathbf{Y})$ is diagonal and the terms in \mathbf{y} all have equal variance, that is $\text{cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}$ where σ^2 is the variance for all variables and \mathbf{I} is an identity matrix. The equivalence between our equations and those proposed by the reviewer can be proved:

$$\begin{aligned} \text{cov}(\mathbf{p}) &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{cov}(\mathbf{Y}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \mathbf{I} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} = (\mathbf{X}^T \sigma^{-2} \mathbf{I} \mathbf{X})^{-1} = [\mathbf{X}^T \text{cov}(\mathbf{Y})^{-1} \mathbf{X}]^{-1} \end{aligned}$$

In this sense, the equations proposed by the reviewer is more generalized. In this paper, we indeed used the $\text{cov}(\mathbf{Y})$ in the equal variance case, so that the computation and results of the paper are correct.

In order to avoid any confusion, we revise the sentence in lines 478-479: “Assuming that the **absolute values of the standard deviations of the** uncertainties in annual emissions of different years (in Tg/year) are **identical and that these uncertainties are** fully independent ...” and we change the equations C-3 and C4 as proposed by the reviewer:

$$\mathbf{p} = (\mathbf{X}^T \text{cov}^{-1}(\mathbf{Y}) \mathbf{X})^{-1} \mathbf{X}^T \text{cov}^{-1}(\mathbf{Y}) \mathbf{y} \quad (C-3)$$

$$\text{cov}(\mathbf{p}) = (\mathbf{X}^T \text{cov}^{-1}(\mathbf{Y}) \mathbf{X})^{-1} \quad (C-4)$$

19. Table 2: In columns 2 and 3, I believe rows 3 and 4 have been flipped.

Response:

The reviewer is right. We correct this mistake accordingly.

20. Figure 2(a): Am I supposed to see 56 colors in the world map? I don't. I think the problem is that the country and state boundaries overlap with the region boundaries. I would suggest, at least in the world map, only showing the region boundaries from the control vector and eliminating the country and state boundaries.

Response:

Here, we do not use 56 colors in the map actually because it is hard to find 56

colors that are visibly differentiable. In fact, we repeatedly use 12 colors for non-adjacent regions. For example, the Northern Europe, Middle East, one region in the USA and one region in China are all red. But because they are in different continents, so that they represent four different regions. We change Figure 2 by removing the lines country and state boundaries, as the reviewer suggested, and mention the fact we repeatedly use 12 colors for non-adjacent regions in the caption:

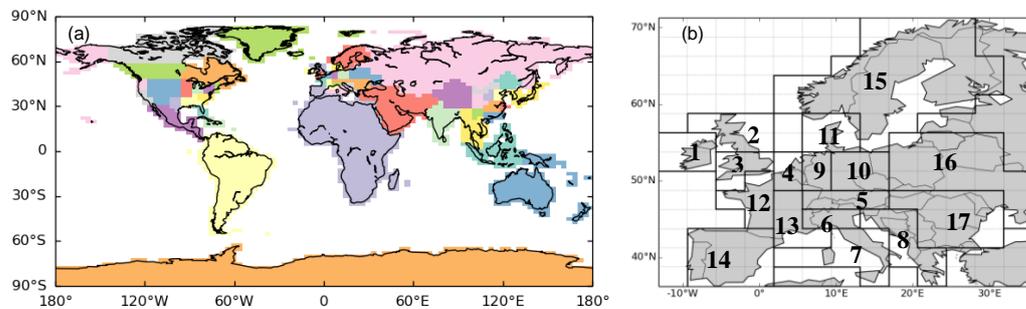


Figure 2: (a) Map of the 56 regions whose monthly emission budgets are controlled by the inversion; (b) zoom over the 17 control regions in Europe. **In (a), we repeatedly use twelve colors for non-adjacent regions. For example, the Northern Europe, Middle East, one region in the USA and one region in China are all red. But since they are in different continents, they represent four different regions.**

5 **Potential of European $^{14}\text{CO}_2$ observation network to estimate the fossil fuel CO_2 emissions via atmospheric inversions**

Yilong Wang^{1,*}, Grégoire Broquet¹, Philippe Ciais¹, Frédéric Chevallier¹, Felix Vogel¹, Lin Wu¹, Yi Yin¹, Rong Wang¹, Shu Tao²

10 ¹Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ- Université Paris Saclay, 91191, Gif-sur-Yvette CEDEX, France

²Laboratory for Earth Surface Processes, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China

*Correspondence to: Yilong Wang (yilong.wang@lscce.ipsl.fr)

Abstract. Combining measurements of atmospheric CO_2 and its radiocarbon ($^{14}\text{CO}_2$) fraction and transport modeling in
15 atmospheric inversions offers a way to derive improved estimates of CO_2 emitted from fossil fuel (FFCO₂). In this study, we solve for the monthly FFCO₂ emission budgets at regional scale (i.e. the size of a medium-sized country in Europe) and investigate the performance of different observation networks and sampling strategies across Europe. The inversion system is built on the LMDZv4 global transport model at $3.75^\circ \times 2.5^\circ$ resolution. We conduct Observing System Simulation Experiments (OSSE) and use two types of diagnostics to assess the potential of the observation and inverse modeling frameworks. The first
20 one relies on the theoretical computation of the uncertainty in the estimate of emissions from the inversion, known as “posterior uncertainty”, and on the uncertainty reduction compared to the uncertainty in the inventories of these emissions which are used as a prior knowledge by the inversion (called “prior uncertainty”). The second one is based on comparisons of prior and posterior estimates of the emission to synthetic “true” emissions when these true emissions are used beforehand to generate the synthetic fossil fuel CO_2 mixing ratio measurements that are assimilated in the inversion. With 17 stations currently
25 measuring $^{14}\text{CO}_2$ across Europe using 2-week integrated sampling, the uncertainty reduction for monthly FFCO₂ emissions in a country where the network is rather dense like Germany, is larger than 30%. With the 43 $^{14}\text{CO}_2$ measurement stations planned in Europe, the uncertainty reduction for monthly FFCO₂ emissions is increased for UK, France, Italy, Eastern Europe and the Balkans, depending on the configuration of prior uncertainty. Further increasing the number of stations or the sampling frequency improve the uncertainty reduction (up to 40% to 70%) in high emitting regions, but the performance of the inversion
30 remains limited over low-emitting regions, even assuming a dense observation network covering the whole of Europe. This study also shows that both the theoretical uncertainty reduction (and resulting posterior uncertainty) from the inversion and the posterior estimate of emissions itself, for a given prior and “true” estimate of the emissions, are highly sensitive to the choice between two configurations of the prior uncertainty derived from the general estimate by inventory compilers or computations on existing inventories. In particular, when the configuration of the prior uncertainty statistics in the inversion
35 system does not match the difference between these prior and true estimates, the posterior estimate of emissions deviate significantly from the truth. This highlights the difficulty to filter the targeted signal in the model-data misfit for this specific

inversion framework, the need to strongly rely on the prior uncertainty characterization for this, and, consequently the need for improved estimates of the uncertainties in current emission inventories for real applications with actual data. We apply the posterior uncertainty in annual emissions to the problem of detecting a trend of FFCO₂, showing that increasing the monitoring period (e.g. more than 20 years) is more efficient than reducing uncertainty in annual emissions by adding stations. The coarse spatial resolution of the atmospheric transport model used in this OSSE (typical of models used for global inversions of natural CO₂ fluxes) leads to large representation errors (related to the inability of the transport model to capture the spatial variability of the actual fluxes and mixing ratios at sub-grid scales), which is a key limitation of our OSSE setup to improve the accuracy of the monitoring of FFCO₂ emissions in European regions. Using a high-resolution transport model should improve the potential to retrieve FFCO₂ emissions, and this needs to be investigated.

1 Introduction

CO₂ emitted from fossil fuels is the major contributor to the increase of atmospheric CO₂ (Ballantyne et al., 2015). Knowledge of FFCO₂ emissions and their trends is essential to understand the drivers of their variations and assess the effectiveness of agreed upon emission reduction policies over time (Pacala et al., 2010). At national scale, FFCO₂ emission inventories are derived based on energy and fuel use statistics, combustion efficiencies and emission factors. These inventories have low uncertainties in OECD countries, and large uncertainties in developing countries due to uncertain energy data and fuel-specific emission factors (Liu et al., 2015; Ballantyne et al., 2015; Andres et al., 2014; Ciais et al., 2010). At sub-national and intra-annual scales, the uncertainties in the estimates of FFCO₂ emissions are higher than at national and annual scale (Ciais et al., 2010; Wang et al., 2013) because subnational intra-annual estimates require either the top-down disaggregation of national annual emissions relying on uncertain socio-economic proxies (Wang et al., 2013; Pregger et al., 2007; Oda and Maksyutov, 2011; Andres et al., 2012), or a detailed knowledge of local activity data for a bottom up-scaling of emissions (Gurney et al., 2009). The comparison of different emission maps of that kind also suggests large uncertainties due to treatment of administrative or land/water borders, the use of different proxies, and different spatial resolutions of the maps (Andres et al., 2016), etc. In consequence, national budgets obtained by aggregation of emission maps may have larger uncertainties than those based on national energy use and fuel accounting systems.

Atmospheric inversions exploit the observed variability in atmospheric mixing ratios of CO₂ to quantify CO₂ fluxes. Inversions have been applied for natural CO₂ sources and sinks based on CO₂ observations (Broquet et al., 2011; Chevallier et al., 2010; Peylin et al., 2013). Recent attempts to quantify FFCO₂ emissions with inversions based on atmospheric CO₂ measurements have stressed the importance to measure mixing ratio gradients very close to the emitting source, such as a city (Staufer et al., 2016; Cambaliza et al., 2014; Lindenmaier et al., 2014) or a power plant (Turnbull et al., 2016). Away from the emitting source, the atmospheric signals of FFCO₂ emissions mixes with those of natural fluxes, so that FFCO₂ emissions can hardly be monitored by atmospheric CO₂ measurements only (Shiga et al., 2013). Because of this, monitoring FFCO₂ emissions

at national scales, using continental networks of stations located outside the vicinity of the largest sources, is only possible when measuring an additional tracer specially sensitive to the signal of FFCO₂ emissions (Miller and Michalak, 2017; Basu et al., 2016). Radiocarbon in CO₂ is arguably the best tracer (Levin et al., 2003; Turnbull et al., 2006). Pacala et al. (2010) proposed to estimate national fossil fuel emissions of the US with an inversion based on measurements of radiocarbon in CO₂. Assuming 10,000 atmospheric ¹⁴CO₂ observations at 84 sites per year and a transport model of 5°×5° horizontal resolution, they suggested that the inversion could reduce the relative uncertainty in monthly emissions of the US from 100% (prior) to less than 10% (posterior). Ray et al. (2014) assumed virtual FFCO₂ observations are sampled every 3 h from a network of 35 measurement towers, and their inversion at 1°×1° resolution could reduce errors on 8 days country-level fossil-fuel emissions from about 15% (prior) down to 7% (posterior). Basu et al. (2016) developed an inversion system at 1°×1° resolution to account for the fact that ¹⁴CO₂ is not a perfectly accurate tracer of FFCO₂ alone and that its mixing ratio is also affected by natural fluxes. They showed that given the coverage of ¹⁴CO₂ measurements available in 2010 over North America (969 measurements per year), the US national total fossil fuel emissions can be constrained with a relative precision of 1% for the annual mean, and less than 5% for most months.

In all these pioneer studies, the actual spatial scale of the areas emitting FFCO₂ is smaller than the grid sizes of the transport models (from 100 to 500 km). The misfits between the spatial scales controlled or modeled within the inversion system and those of actual emissions, or those of the FFCO₂ patterns in the atmosphere generate errors known as *aggregation* and *representation* errors (see Sect. 2.2.2), which strongly affect the inversion of FFCO₂ emissions (Wang et al., 2017). Those errors were not formally accounted for in previous FFCO₂ inversion studies.

In recent years, as part of the ICOS project, a rather dense network of standardized, long-term and high precision atmospheric measurements of CO₂ has been set up in Europe. Some of the ICOS sites also measure ¹⁴CO₂ and this type of measurement will be extended in the near term with the aim of determining gradients of FFCO₂ mixing ratios across the European continent. The ICOS atmospheric network is expected to sample 2-week integrated ¹⁴CO₂ at about 40 stations (1,000 analyses per year; ICOS Stakeholder handbook 2013 at <http://www.icos-uk.org/uk-icos/sites/uk-icos/files/documents/Stakeholders%20Handbook%202013.pdf>). In this context, network assessment studies are needed to understand how much this ¹⁴CO₂ network will improve the knowledge on FFCO₂ emissions.

In this study, we study the potential of an atmospheric inversion system to quantify FFCO₂ emissions at regional scales (i.e. the size of a medium-sized country in Europe like France or Germany) over the European continent based on continental-scale networks of atmospheric CO₂ and ¹⁴CO₂ measurements. Special attention is paid to the representation and aggregation errors induced by the use of a coarse grid transport model. Wang et al. (2017) ~~already evaluated derived~~ the statistics of these errors for the inversion system that we apply here, which is based on the Laboratoire de Météorologie Dynamique LMDZv4 global transport model (Hourdin et al., 2006) and our study strongly relies on their results. ~~Their results~~ They highlighted that both the representation and aggregation errors have large magnitudes, and could thus strongly reduce the ability of the inversion to filter the information on the uncertainties in regional FFCO₂ emissions. They also stressed the fact that the spatial scales of

the correlations in the representation and aggregation errors are smaller than that of the projection in the atmospheric observation space of the typical uncertainties in the prior estimates of regional emissions (called “prior FFCO₂ errors” hereafter). More precisely, with their modelling configuration they obtained values smaller than 200 km and larger than 700 km respectively for these spatial scales. Therefore, if the observation networks are dense enough to provide information at finer spatial scale (typically with distances from a given station to the closest ones being systematically smaller than 700 km), the impact of aggregation and representation errors on the inversion of the regional budgets of FFCO₂ emissions could be small (Wang et al. 2017). In this study, we account for the aggregation and representation errors using their detailed and quantitative characterization ~~use the results from Wang et al. (2017) to take into account representation and aggregation errors~~ and we check whether using dense networks could overcome the limitations brought by coarse resolution transport models and by the uncertainties in the distribution of the emissions at high resolution when retrieving regional emission budgets. Using the error estimates from Wang et al. (2017) ensures that our inverse modelling system does not overestimate the potential of measurement networks that are dense compared to our coarse transport model resolution but whose distances between the sites are larger than the spatial scales of local atmospheric signals from the anthropogenic emissions.

Our inversion system solves for monthly FFCO₂ emissions in different regions of Europe over a period of one year by assimilating synthetic observations of atmospheric gradients of FFCO₂ mixing ratios obtained from co-located CO₂ and ¹⁴CO₂ measurements at ICOS-like stations. The study primarily aims at providing a typical quantification of the inversion performances and at understanding qualitatively how the inversion behaves depending on the level of FFCO₂ emissions, on the knowledge on these emissions and on the network density. Furthermore, ~~we~~ we assume here that the uncertainties in the signals from ¹⁴CO₂ fluxes other than the FFCO₂ emissions, such as that from terrestrial biosphere, oceans, nuclear power plants and cosmogenic production, should have a moderate impact on the order of magnitude of the inversion performances that are analysed in this study. This leads us to ignore these uncertainties and consider ¹⁴CO₂ is a perfect tracer of FFCO₂ so that the only uncertainties in the FFCO₂ mixing ratios data are related to the instrumental precision of CO₂ and ¹⁴CO₂ measurements. In practice, in the frame of this study, which focuses on the propagation of uncertainties, this is mathematically equivalent to assuming that ¹⁴CO₂ is a perfect tracer of FFCO₂. However, this does not imply that the signal from natural fluxes and nuclear power plants could be ignored when processing real data. ~~In particular, we ignore the signals of ¹⁴CO₂ fluxes (natural fluxes, nuclear power plants) other than those of FFCO₂ emissions.~~

Although the results are presented only over Europe, ~~we use a global~~ we use a ~~global~~ inversion system and the global transport model LMDZv4 is global to ensure that uncertainties in FFCO₂ emitted over other regions of the globe are properly accounted for and to study their impact on the inversion of the FFCO₂ emission in Europe. LMDZv4 has a 3.75 °×2.5 ° longitude×latitude horizontal resolution and 19 layers in the vertical between the surface and the top of the atmosphere. This spatial resolution is comparable to that of transport models used in state-of-the-art global inversions (Peylin et al., 2013). We assess the potential of this inversion to improve the estimates of regional fossil fuel emissions based 1) on the statistics of the theoretical prior and posterior uncertainties provided by a Bayesian statistical framework, and 2) on the statistics of the misfits between the prior

and posterior estimates of emissions against the assumed “truth” generated by the choice of another emission inventory independent from the one used as prior (see Sect. 2.3). The second type of assessment is used to test the impact of error structures that can hardly be accounted for by the representation of the prior and model uncertainties in the theoretical framework of the atmospheric inversion.

The presentation of the results first focuses on regional FFCO₂ emission budgets over one year. It also explores the monitoring of the decadal changes of FFCO₂ emissions, compared to a baseline year, which is also of importance since it corresponds to climate mitigation targets set for the Kyoto Protocol and the Intended Nationally Determined Contribution. The trends of FFCO₂ emissions over multiple years can be computed using simple regression of series of annual emissions estimates from inventories or atmospheric inversions. The relative uncertainties in decadal trends (e.g. the relative uncertainties in regression slopes) tend to be lower than that in the emission budget of a given year (Pacala et al., 2010), implying that changes can be monitored more accurately than annual budgets. Here, we provide a quantitative analysis of how accurate the trends of national annual FFCO₂ emission can be monitored using measurements of FFCO₂ mixing ratios.

The paper is organized as follows. Section 2 gives a full description of the inversion and OSSE framework. Section 3 analyzes the statistics of the posterior uncertainties and misfits from inversions using different observation networks. Section 4 evaluates the potential of atmospheric inversion for the monitoring of decadal changes and discusses the relevance of using a coarse-resolution transport model in the inversion system to quantify regional FFCO₂ emissions. Conclusions are drawn in Sect. 5.

2. Methodology

2.1 The configurations of the observation network

We consider three different observation networks, in which the number of the stations ranges from 17 to 233. The minimum network (NET17) includes 17 sites, based on existing European ICOS ¹⁴CO₂ stations in 2016. Using these sites and possible future additional ¹⁴CO₂ stations listed in the 2013 ICOS Stakeholder handbook (available at <http://www.icos-uk.org/uk-icos/sites/uk-icos/files/documents/Stakeholders%20Handbook%202013.pdf>), we also consider an intermediate ¹⁴CO₂ network of 43 sites (NET43). The NET17 and NET43 networks have high densities in France, Germany, UK and Switzerland, but remain sparse in Eastern Europe (Fig. 1). The corresponding site locations are given in Table S2. We also test a very dense network of 233 sites (NET233), in which two sites are placed in each European land pixel of the LMDZv4 transport model (Fig. 1c). The NET233 network is denser than NET17 and NET43 in the high emitting regions, e.g. Germany, and also covers the region that is not well sampled by NET17 and NET43. However, the location of its 233 sites is not intended to be optimal since the emissions have a very heterogeneous spatial distribution. Their homogeneous spreads allow us to reduce the impact of representation and aggregation errors (Trappert and Snieder, 1996; Kaminski et al., 2001) and to assess the impact of having a dense network for all control regions.

165 The high-altitude station Jungfraujoch (JFJ) at 3450 meter above sea level (masl) in Switzerland samples free tropospheric
air over Europe, assumed to be representative of the “background” concentration. In all the three configurations of the
observation network, JFJ is chosen as the reference station. In this study, we assimilate gradients of FFCO₂ between other sites
and JFJ in the inversion. Measurements at other sites than JFJ are all assumed to be made at 100 meter above the ground level
(magl), the typical height of ICOS tall towers (Kadygrov et al., 2015; Marquis and Tans, 2008).

170 Wang et al. (2017) have already made a detailed characterization of the distributions of representation errors at the sites
considered here and characterized two types of stations based on the population density of the grid cells within which a station
is located and on the locations of large point sources (e.g. large power plants). All the sites in different networks are thus
categorized as “urban” or “rural” sites according to their results. In the NET233 network, the two sites in each land pixel of
the transport model are assumed to be one urban and one rural distant by more than 200 km in order to combine data for the
175 structures of representation errors are different (i.e. which have a different view in terms of the scale of FFCO₂ emissions).
Any of the transport model pixels provides such locations since having areas of nearly 10⁵ km² (Wang et al. 2017).

2.2 Configuration of the inversion system

The assessment of the potential of different networks to constrain fossil fuel emissions is based on the inversion
framework presented by Wang et al. (2017). In this section we summarize the main elements of this framework for which the
180 details can be found in Wang et al. (2017).

2.2.1 Theoretical framework of the Bayesian inversion and diagnostics of the inversion performance in OSSEs

The inversion relies on a Bayesian statistical framework. The estimate of the fossil fuel emission budgets at monthly and
regional scales over one year, called hereafter the control variables \mathbf{x} , is corrected from a prior knowledge of these variables
 \mathbf{x}^b (that from a gridded inventory covering the globe). This correction is based on (i) a set of gradients of FFCO₂ mixing ratios
185 between the different measurement sites and JFJ sampled during the afternoon (see Sect. 2.2.2) across Europe, called hereafter
the “observations” \mathbf{y}_o , (ii) the observation operator \mathbf{H} linking \mathbf{y} with \mathbf{x} based on the spatial and temporal distribution of the
emissions within a control region and within a month, on a linear CO₂ atmospheric transport model, and on the sampling of
the gradients between the corresponding sites and (iii and iv) a modeling of the covariances \mathbf{B} and \mathbf{R} of the distributions of the
uncertainties in the prior estimate and of the observation errors. The observation error is a combination of the measurement
190 error, the errors from the model transport, representation and aggregation errors. In this study, we ignore the impact on the
FFCO₂ gradients from the transport model initial conditions that are not controlled by the inversion since it is assumed to be
negligible (Wang et al., 2017). Assuming that the prior uncertainties and observation errors are uncorrelated with each other
and have unbiased and Gaussian statistical distributions, the statistical distribution of the estimate of \mathbf{x} , given \mathbf{x}^b and \mathbf{y}_o , is also
unbiased and Gaussian, and its corresponding mean \mathbf{x}^a and covariance matrix \mathbf{A} are given by:

195 $\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1}$ (1)

$\mathbf{x}^a = \mathbf{x}^b + \mathbf{A} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y}_o - \mathbf{H} \mathbf{x}^b)$ (2)

where T and $^{-1}$ denote the transpose and inverse of a matrix respectively.

Equation (1) shows that \mathbf{A} depend on neither the value of the observations \mathbf{y}_o nor the prior emission budgets \mathbf{x}^b themselves, but rather on the prior and observation error covariance matrices, on the observation times and locations (through the definition of \mathbf{H} corresponding to the \mathbf{y} -space) and on the observation operator. Equation (2) shows that the actual value of \mathbf{x}^a also depends on the observations \mathbf{y}_o and on the prior emission budgets \mathbf{x}^b .

A common performance indicator is the theoretical uncertainty reduction (UR) for specific budgets of the fossil fuel emissions (at control or larger space and time scales), defined by:

$$\text{UR} = 1 - \frac{\sigma^a}{\sigma^b}$$
 (3)

205 where σ^a and σ^b are the standard deviations of the posterior and prior uncertainties in the corresponding budget of emissions. Such an indicator can directly be derived from the modeling of \mathbf{B} and from the theoretical computation of \mathbf{A} by Eq. (1). Of note is that the scores of uncertainty and of UR given in this study will refer to the standard deviation of the theoretical uncertainty in a specific emission budget.

However, if the modeling of \mathbf{B} and \mathbf{R} does not match the actual statistics of the prior and observation uncertainties, or if the theoretical framework of the inversion (assuming that all sources of uncertainty have unbiased and Gaussian distributions, that prior and observation errors are uncorrelated and that the observation operator is linear) is not well satisfied, such a theoretical computation of UR may not reflect the actual performance of the inversion. Wang et al. (2017) derived the statistics of the different components of the observation errors for the same inversion framework as used here. Their statistics of the representation and aggregation errors were based on the comparison of transport model simulations made at high and low spatial resolutions. They highlighted the fact that the distribution of these errors depart from purely Gaussian distributions, and that their covariances can hardly be characterized by the relatively simple models traditionally used in atmospheric inversion systems. In this study, we thus test the inversion system with OSSEs using synthetic truth and errors to build \mathbf{x}^b and \mathbf{y}_o that better reflect the type of observation errors found by Wang et al. (2017). We use Eq. (2) to derive the estimates of \mathbf{x}^a and we analyze the misfits between \mathbf{x}^b and \mathbf{x}^a against the synthetic true emission budgets \mathbf{x}^t . This leads us to define an alternative indicator of the inversion performance, called misfit reduction (MR) hereafter. While this indicator does not provide an exhaustive statistical view of the uncertainty in the inverted emissions, it is used to evaluate the confidence in the more complete (with a full covariance estimate rather than just a realization of the distribution) but more theoretical computation of the posterior uncertainties and of the UR based on Eq. (1). We write the MR for specific budgets of the fossil fuel emissions (at control or larger space and time scales) as follows:

225
$$\text{MR} = 1 - \frac{\varepsilon^a}{\varepsilon^b}$$
 (4)

where ε^a and ε^b are the posterior and prior misfits between the inverted and prior emission budgets against true values for the

corresponding emission budgets. MR range from negative values (when the inversion deteriorates the precision of the estimation) to 1 (or “100%”, when the inversion provides a perfect estimate of the emissions).

We focus on uncertainties and misfits at both monthly and annual scales. However, we ~~can only~~ have only one practical realization for \mathbf{x}^b , \mathbf{y}_o and \mathbf{x}^a following the protocol of that is presented in (see Sect. 2.3). Therefore, the assessment of the performance of the inversion for a given region-month using the corresponding score of MR may be over- or under-estimated due to the lack of sampling of the prior and observation errors. Consequently, at monthly scale, in order to strengthen the evaluation of the theoretical uncertainties based on these single realizations of the prior and posterior misfits, we compare, for a given region, the quadratic mean of the twelve monthly misfits (called “monthly misfits” without mention of a specific month in Sect. 3) to the quadratic mean of the standard deviations of the twelve monthly uncertainties (called “monthly uncertainties” without mention of a specific month in Sect. 3), which characterizes the average monthly uncertainties over the year. This computation implicitly assumes that the twelve monthly misfits through a year follow the same statistical distribution, and represent twelve independent realizations of this distribution. In such a situation, the comparison between the averages of the prior and posterior monthly misfits give a good indications of the error reduction that should not be highly skewed by sampling errors. In the result section, for a given region i , UR and MR scores derived at the “monthly” scale without mention to a specific month will correspond to the relative difference between the prior and posterior values of these average monthly uncertainties and misfits from a whole year of inversion:

$$UR_i = 1 - \frac{\sqrt{\frac{1}{12} \sum_{m=1}^{12} (\sigma_{i,m}^a)^2}}{\sqrt{\frac{1}{12} \sum_{m=1}^{12} (\sigma_{i,m}^b)^2}} \quad (5)$$

$$MR_i = 1 - \frac{\sqrt{\frac{1}{12} \sum_{m=1}^{12} (\epsilon_{i,m}^a)^2}}{\sqrt{\frac{1}{12} \sum_{m=1}^{12} (\epsilon_{i,m}^b)^2}} \quad (6)$$

~~(following equations similar to Eq. (3) and (4)).~~ At the annual scale, the diagnostics of UR will have to be compared to MR values for single realizations of the annual misfits. In addition, we discuss the scores of the relative uncertainty and misfit, defined as the ratios of the absolute uncertainties and misfits to the absolute prior emission budgets.

2.2.2 Practical setup

Control vector

The inversion system has a global coverage and controls monthly budgets of FFCO₂ emissions for a set of regions during the year 2007. The map of these regions is given in Fig. 2a. The space discretization of regions is higher where emissions are the largest in Europe (area of interest, Fig. 2b) and also in the US and China. In other areas with lower emissions or where observational data to further constrain the prior emissions are lacking (Fig. 2a and Table S1), the size of the control regions is

255 much larger, and can reach that of a continent. The spatial resolution of the control vector (a region) in Central and Eastern Europe corresponds to the typical size of a medium-sized European country, but in western Europe apart from Spain, Portugal and Ireland, where emissions are the highest, the control variables correspond to sub-national regions (e.g. southern and northern UK, southern and northern Italy, western and eastern Germany, western and eastern France in Fig. 2b). Monthly emissions over the ocean are included in the control vector, but the ocean is considered as one large region. In total, the world is divided into 54 land regions and 1 ocean region (Table S1). The inversion solves for the 12 monthly budgets of emissions for these regions, but not for the spatio-temporal distributions within each region and month. In our framework, choosing year 2007 for the inversion only impacts the meteorological conditions and thus the atmospheric transport conditions. We assume that the atmospheric transport conditions in 2007 are representative of average conditions. We also ignore the impact of inter-annual variations of FFCO₂ emissions, which is usually less than 4% (Levin and Rödenbeck, 2008), and of their prior uncertainty (see below the configuration of the prior uncertainty matrix, which is a function of the emissions).

Time selection of data to be assimilated

270 Current atmospheric ¹⁴CO₂ samples in Europe are usually filled continuously over the course of two weeks (Vogel et al., 2013; Levin et al., 2013). However, state-of-art inversion systems generally make use of data during the afternoon only, due to limitations of transport models in simulating night-time mixing ratios near the ground. Given the ability to have an intermittent filling of air samples for ¹⁴C analysis (Turnbull et al., 2016; Levin et al., 2008), we thus define the observations to be selectively sampled only during the afternoon (12:00-18:00 local time). Since the cost of the ¹⁴CO₂ analysis of one sample is presently high, monitoring of ¹⁴CO₂ (and thus FFCO₂) during a whole year favors the choice of integrated samples at the weekly to 2-week scale (Levin et al., 1980; Turnbull et al., 2009; Vogel et al., 2013). In this study, we first consider 2-week integrated afternoon data. More precisely, we first consider 2-week averages of afternoon FFCO₂ gradients with respect to JFJ.

275 In addition, we present tests with daily afternoon ~~data~~gradients, for which the correspondingsuch a sampling scheme would be more costly. Sampling FFCO₂ observations at high temporal resolution should decrease the weight of the random errors on longer time scales, which should improve the potential of the inversions of monthly to annual emission budgets. While inversions are conducted with 2-week samplings for the three networks, daily sampling is tested for NET43 only, which is sufficient to evaluate the usefulness of high frequency sampling.

280 Observation operator

The atmospheric FFCO₂ mixing ratios are influenced by the 3-D initial FFCO₂ distribution, and by surface emissions during the year. In this study, the inversion rescales all emissions during one year (here 2007) and we ignored initial conditions on January 1st which are rapidly transported out of Europe and do not cause subsequent FFCO₂ gradients between European sites (Wang et al., 2017). The observation operator is restricted to a matrix \mathbf{H} which consists of a chain of three sub-operators, $\mathbf{H}=\mathbf{H}_{\text{samp}}\mathbf{H}_{\text{transp}}\mathbf{H}_{\text{distr}}$, where $\mathbf{H}_{\text{distr}}$ distributes regional monthly emission budgets into a gridded emission map at the resolution of the transport model, $\mathbf{H}_{\text{transp}}$ is the atmospheric transport model, and \mathbf{H}_{samp} samples the FFCO₂ gradients with respect to JFJ corresponding to the observation vector from the transport model outputs (Wang et al. 2017).

We use the high-resolution (0.1 °) annual FFCO₂ emission map from the PKU-CO₂ inventory in the year 2007 (Wang et al., 2013) to distribute the emissions in space within each region. PKU-CO₂ is an annual emission map with no temporal profile, so that the modelled temporal distribution in $\mathbf{H}_{\text{distr}}$ is flat between months. This implementation of $\mathbf{H}_{\text{distr}}$ is denoted $\mathbf{H}_{\text{distr}}^{\text{PKU}}$.

The off-line version of the general circulation model of ~~Laboratoire de Météorologie Dynamique-LMDZv4 (Hourdin et al., 2006)~~ forms $\mathbf{H}_{\text{transp}}$. Atmospheric transport simulations was nudged to analyzed wind fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim, Dee et al., 2011) for the year 2007. We denote this implementation of $\mathbf{H}_{\text{transp}}$ by $\mathbf{H}_{\text{transp}}^{\text{LMDZ}}$.

The sampling of FFCO₂ gradients relies on the extraction of individual simulated mixing ratio data at the measurement locations and chosen temporal sampling frequency, followed by the computation of differences (gradients) between time series of FFCO₂ mixing ratios at each site and that at the JFJ reference site. The mixing ratio data for a given site is sampled at the chosen sampling height in the transport model grid cell containing this site. We recall that the sampling height is 100 m above ground level (magl), the 1st level of LMDZv4, except for JFJ being at 3450 m above sea level (masl), the 6th level. The resulting implementation of \mathbf{H}_{samp} is denoted $\mathbf{H}_{\text{samp}}^{\text{coloc}}$.

In sum, the ~~practical~~-observation operator used ~~for in the practical configuration of the~~ inversions system is defined by $\mathbf{H}^{\text{prac}} = \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{PKU}}$.

Prior error covariance matrix

Emission estimates from inventories are limited to annual and national scales and rarely provide systematic assessments of uncertainties. There are a limited number of datasets providing emission maps at higher spatial/temporal resolutions. Although there have been some efforts to compare such FFCO₂ emission maps (Macknick et al., 2009; Ciais et al., 2010; Andres et al., 2012; Andres et al., 2016), the ability to characterize the uncertainties of an emission inventory is limited, especially for sub-national and sub-annual scales. In this study, we use different streams of information to model the prior emission uncertainty covariance matrix \mathbf{B} and we use two different configurations of this matrix in the inversions.

The first configuration of the \mathbf{B} matrix, called here notional or $\mathbf{B}^{\text{notion}}$, is related to the notional estimates of (1-sigma) uncertainties for national emissions claimed by inventory compilers to range from 1-2.5% for the USA (US EPA, 2015), 2%-7% for European countries (Andres et al., 2014; Ballantyne et al., 2015) to 7.5-10% for China (Gregg et al., 2008; Liu et al., 2015). However, Ciais et al. (2010) found that the ratios between geographically distributed emission maps, even after correction for inconsistencies and aggregated at national scale, ranged from 0.86 to 1.5, which is larger than the uncertainties claimed by inventory compilers. In this study, the prior uncertainty covariance $\mathbf{B}^{\text{notion}}$ of monthly emissions is set up based on three constraints: 1) the relative uncertainty in annual emission equals 10% for US and European national budgets, 15% for China, and 10% for individual control regions outside US, Europe and China; 2) uncertainties in monthly emissions have a 2-month exponentially decaying temporal auto-correlation, and 3) spatial correlations between uncertainties in monthly emissions across adjacent regions within the same country are fixed to -0.2, a negative value to account for the fact that sub-national emissions are usually disaggregated from national inventories, so that a positive bias in part of a country must be

compensated by a negative one in another. All other spatial correlations in $\mathbf{B}^{\text{notion}}$ are assumed to be null, and the overall correlation matrix in $\mathbf{B}^{\text{notion}}$ is derived from the Kronecker product of temporal and spatial correlation matrices (assuming that the correlation between two control variables are given by the product of the spatial and temporal correlations between the two corresponding control regions and the two corresponding time window respectively). The full computation of $\mathbf{B}^{\text{notion}}$ is detailed in Appendix A. With this setting, prior uncertainties in monthly emissions can exceed 10% and be as large as 30% for some sub-national control regions.

The second configuration of the \mathbf{B} matrix, known as empirical or $\mathbf{B}^{\text{empiric}}$, is based on the empirical derivation of the statistics of the differences between two spatially gridded emission maps (which will be used to define the prior and true estimate of emissions in the OSSEs, see Sect. 2.3). The two maps are PKU-CO₂ (Wang et al., 2013, <http://inventory.pku.edu.cn/>) and IER-EDG (available at <http://carbones.ier.uni-stuttgart.de/wms/index.html>), both corresponding to the year 2007. The IER-EDG map combined EDGAR annual map with country specific temporal profiles (monthly, daily and hourly) from IER. In general, the differences in annual emissions from the control regions in Europe between these two emission maps range from 3% to 20%, except for the Balkans where they reach up to 44%. We assume that there is no spatial correlation of the prior uncertainty between different control regions. For each control region of the globe, the statistics of the difference between the monthly emission budgets from the two maps aggregated at the control region scale are fitted by a covariance model that combines four different covariance matrices, with exponentially decaying temporal correlations at time scales of 1 month, 3 months, and 6 months for the first three ones respectively, and a full temporal correlation over the year for the fourth one (representing the annual bias on the prior emissions). The mathematical formulation for this computation and the full derivation of $\mathbf{B}^{\text{empiric}}$ is detailed in the Appendix B. ~~In addition, we assume that there is no spatial correlation of the prior uncertainty between different control regions. Again, the overall correlation matrix in $\mathbf{B}^{\text{empiric}}$ is derived from the Kronecker product between the temporal and spatial correlation matrices.~~

$\mathbf{B}^{\text{empiric}}$ is built using an error covariance model which cannot perfectly characterize the structure of the differences between the PKU-CO₂ and IER-EDG budgets at the control resolution, which will be used to derive realistic \mathbf{x}^b and \mathbf{x}^t respectively and thus the “actual prior errors“ in the OSSEs with synthetic data (see Sect. 2.3). However, by construction, $\mathbf{B}^{\text{empiric}}$ better fits these errors in our OSSEs than the $\mathbf{B}^{\text{notion}}$ matrix in terms of both the standard deviation of the uncertainty at the 1 month / regional scale and the temporal correlations. The differences between the results of the inversions using either $\mathbf{B}^{\text{empiric}}$ or $\mathbf{B}^{\text{notion}}$ will be used to give an estimate of the range of the inversion skills as a function of different assumptions regarding the prior uncertainty in emission budgets.

Observation error covariance matrix

Wang et al. (2017) derived estimates of the observation errors in FFCO₂ gradients across Europe when using the same inverse modeling framework as in this study. They analyzed four sources of observation errors (i.e. sources of misfits when comparing the modeled to the measured FFCO₂ gradients other than the uncertainties in the estimates of the emission budgets at the 1-month and regional scale), one related to the FFCO₂ data, three to the observation operator:

1) The measurement error ϵ_i on FFCO₂ gradients is simply assumed to be 1 ppm with no temporal and spatial correlations, which corresponds to the typical precision of the analysis of air samples by accelerator mass spectrometry (AMS) for ¹⁴CO₂ (2‰-3‰) (Hammer et al., 2016; Turnbull et al., 2014).

2) The representation error ϵ_r arises from the mismatch between the coarse resolution of modelled emissions and concentrations in the observation operator (here the transport model) and the spatial variability of the actual emissions and concentrations.

3) The transport errors ϵ_t is due to discretized and simplified equations for modeling transport, using a given meteorological forcing in practice.

4) The aggregation error ϵ_a arises from the mismatch between the control resolution (budgets of regions in each month) and the resolution of the emission modeled in the observation operator (here the transport model). It reflects uncertainties in $\mathbf{H}_{\text{distr}}$.

In this study, we use the estimates of the standard deviations and of the correlation functions for these different types of observation errors from Wang et al. (2017) to set up the \mathbf{R} matrix. Wang et al. (2017) sampled representation and aggregation errors by using simulations with a mesoscale (with higher resolution than LMDZv4) regional transport model and by degrading the spatial and temporal resolution of the emission maps in the input of this model and in the output FFCO₂. Based on these samples, the standard deviation of ϵ_r was characterized by a function of season and on whether a station is “urban” or “rural” (see Sect. 2.1). For ϵ_a , the standard deviation for spring/summer and autumn/winter were derived. The standard deviation of the transport error at a given site is set-up proportional to the temporal standard deviation of the 1-year long time-series of the high-frequency variability of the detrended and deseasonalized simulated daily mean afternoon mixing ratios in the grid cell of the transport model, at which the sites are located. Such an estimation of transport error which relies on some results from Peylin et al. (2011) aims at representing the typical value for global transport models, not that of the specific transport model used in this study. The temporal ~~and spatial~~ auto-correlations in the representation and aggregation errors were characterized by Wang et al. (2017) using the sum of a long-term component and a short-term component: $r(\Delta t) = a \times e^{-\Delta t/b} + (1-a) \times e^{-\Delta t/c}$ where Δt is the timelag (in days) and a, b, c are parameters optimized by regressions against the samples of the errors. Furthermore, Following Wang et al. (2017), we do not include temporal auto-correlations in the transport error for simulated daily to 2-week mean afternoon FFCO₂ gradients, since previous studies of the auto-correlations of the transport errors have not evidenced that they should be significant at daily scale (Lin and Gerbig, 2005; Lauvaux, 2009; Broquet et al., 2011). This choice follows the corresponding discussion by Wang et al. (2017) and implicitly ignores that transport model errors likely bear long-term components (often referred to as “biases”, Miller et al., 2015) even when being dominated by components on short timescales. The corresponding values of the standard deviation and the modelling of temporal autocorrelation of the observation errors for 2-week/daily mean afternoon FFCO₂ gradients are listed in Table S3 and Table S4.

A simpler account of the spatial correlations in the observation errors is derived from the diagnostics of Wang et al. (2017). We do not account for the spatial correlation in the representation error, as the scale of the spatial correlation according to Wang et al. (2017), i.e. 55-89 km, is much smaller than the size of the grid cells of the global transport model $\mathbf{H}_{\text{transp}}^{\text{LMDZ}}$ used for the inversion. When there are more than two sites located in the same grid cell of the transport model, we consider that the

aggregation errors and the transport errors are fully correlated between these sites, according to the definition by Wang et al. (2017). We do not account for spatial correlations between aggregation errors for measurements made at sites in different grid cells, because the scale of the spatial correlation is 171 km and is smaller than the size of the grid cell, according to Wang et al. (2017). Finally, we do not account for spatial correlations between transport errors or measurements made at sites in different grid cells.

account for the spatial correlation in the \mathbf{R} matrix, as the scale of the spatial correlation is smaller than the size of the grid cells of the global transport model $\mathbf{H}_{\text{transp}}^{\text{LMDZ}}$ used for the inversion. The standard deviation of the transport error at a given site is assumed proportional to the temporal standard deviation of the 1-year long time series of the high-frequency variability of the detrended and deseasonalized simulated daily mean afternoon mixing ratios in the grid cell of the transport model, at which the sites are located. The corresponding values of the standard deviation and the modelling of temporal autocorrelation of the observation errors for 2-week/daily mean afternoon FFCO₂ gradients are listed in Table S3 and Table S4.

Assuming that all these sources of errors are independent from each other and have Gaussian and unbiased distributions: i.e. $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_i)$, $\boldsymbol{\varepsilon}_r \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_r)$, $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$, $\boldsymbol{\varepsilon}_a \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\text{adistr}})$, \mathbf{R} is given by the sum of the covariance matrices corresponding to each of them: $\mathbf{R} = \mathbf{R}_i + \mathbf{R}_r + \mathbf{R}_t + \mathbf{R}_a$.

2.3 Configurations of the OSSEs

In this study, we consider two types of OSSEs corresponding to the two configurations of prior error covariance matrix $\mathbf{B}^{\text{notion}}$ and $\mathbf{B}^{\text{empiric}}$. The first OSSEs use $\mathbf{B}^{\text{notion}}$ (called here INV-N), while the second type of OSSEs use $\mathbf{B}^{\text{empiric}}$ (called here INV-E). As discussed in Sect. 2.2.1, in both types of OSSEs, the theoretical computation of the posterior uncertainty and UR is based on Eq. (1). These diagnostics would perfectly characterize the performance of the system if the prior uncertainty and the observation errors have Gaussian and unbiased distributions that are perfectly characterized by the set-up of the prior uncertainty covariance matrix \mathbf{B} and observation error \mathbf{R} in the inversion system. In both types of OSSEs, these diagnostics are evaluated based on a practical application of Eq. (2), and on the analysis of posterior misfits and MR, with a synthetic truth (true emissions and true observation operator) and observations that are generated in a similar way as in Wang et al. (2017). Here, the “actual” prior and observation errors have a complex origin and structure which is not perfectly adapted to the unbiased and Gaussian assumptions and are not perfectly reflected by the set-up of the prior uncertainty covariance matrix \mathbf{B} and observation error covariance matrix \mathbf{R} in the inversion system, even in INV-E where $\mathbf{B} = \mathbf{B}^{\text{empiric}}$ and \mathbf{R} are fitted to the “actual” prior and observation errors. Of note is that in INV-N, $\mathbf{B}^{\text{notion}}$ has significant inconsistencies with the actual differences between \mathbf{x}^b and \mathbf{x}^t , so that, in this experiment, the analysis of the posterior misfits and MR will be used to evaluate the performance of the inversion when using a poor configuration of the prior uncertainty covariance matrix in the inversion system in addition to accounting for errors which hardly fit with the assumption that their distribution is Gaussian and unbiased. This corresponds to situations for which there is little knowledge about the uncertainties in the inventories used for inversions with real data. The analysis of misfits and MR in INV-N is thus more pessimistic than that in INV-E.

In the OSSEs, the synthetic prior estimate of the regional/monthly emissions \mathbf{x}^b is built based on the emissions from PKU-
 420 CO₂ (\mathbf{x}^{PKU} hereafter). The synthetic true emission budgets and synthetic observations are modelled using a realistic
 representation the “actual” emission budgets \mathbf{x}^l and of the “actual” $\mathbf{H}_{\text{distr}}$ operator based on the relatively independent IER-
 EDG inventory. The synthetic true regional/monthly emissions and the synthetic true $\mathbf{H}_{\text{distr}}$ operator are thus referred to as $\mathbf{x}^{\text{IER-EDG}}$
 and $\mathbf{H}_{\text{distr}}^{\text{IER-EDG}}$ hereafter. The synthetic observations are generated using $\mathbf{x}^{\text{IER-EDG}}$ and the operator
 $\mathbf{H}^{\text{OSSE}} = \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{IER-EDG}}$, which relies on the same $\mathbf{H}_{\text{samp}}^{\text{coloc}}$ and $\mathbf{H}_{\text{transp}}$ operators as the \mathbf{H}^{prac} observation operator
 425 used in the inversion system. Consequently, the difference between \mathbf{H}^{OSSE} and \mathbf{H}^{prac} underlies aggregation errors only. Therefore,
 in order to account for the transport, representation and measurement errors, the data $\mathbf{H}^{\text{OSSE}} \mathbf{x}^{\text{IER-EDG}}$ are perturbed following
 the statistics of the corresponding errors as detailed in Sect. 2.2.2.

The parameters of the two inversion configurations are summarized in Table 1 and Fig. 3. All the combinations of
 networks and data temporal sampling described in Sect. 2.1 and 2.2.2 are tested with the two configurations of OSSEs. The
 430 resulting eight OSSEs are listed in Table 2.

3. Results

3.1 Assessment of the performance of inversions when using the NET17/NET43 and 2-week integrated sampling

3.1.1 Analysis of the results at the regional and monthly scale

Figure 4 shows the URs of monthly emissions using the NET17 and NET43 networks and 2-week sampling (N-17W, E-
 435 17W, N-43W and E-43W in Table 2). With NET17, INV-N and INV-E inversions show similar spatial patterns of UR scores.
 The largest UR occurs in region western Germany, being 34% for inversion N-17W and 38% for E-17W. The URs are also
 significant in eastern Germany for both inversions. This stems from the fact that several stations are located around and within
 these regions and that the emission in these regions are higher than those in other regions. Moderate UR values are found for
 Benelux (12%) and Eastern France (15%) in inversion E-17W and the UR values elsewhere are marginal. Going from NET17
 440 to NET43 adds a significant increase (improvement) of the UR for southern UK (from 3% to 23%), northern Italy (from 3%
 to 18%) and Eastern Europe (from 2% to 15%) in INV-N (Fig. 4d4e). The increase of UR in E-43W, compared with the UR in
 E-17W, mainly occurs in eastern France (from 16% to 33%) and the Balkans (from 3% to 13%). Because the added stations
 in NET43, compared to NET17, are mostly located outside Germany, the URs over western and eastern Germany are not
 significantly improved (Fig. 4d4e and 4e4g). Despite their different URs for specific regions, both types of inversions highlight
 445 the overall increase in the UR for western European regions by increasing the number of sites from NET17 to NET43.

The differences in the spatial patterns of UR between INV-N and INV-E inversions shown in Fig. 4 reveal the high
 sensitivity of UR to the configuration of the prior uncertainties. Figures 5a and 5b show the prior uncertainties associated with
 the two configurations of $\mathbf{B}^{\text{notion}}$ and $\mathbf{B}^{\text{empiric}}$. The regions where these uncertainties and thus the potential for reducing these

uncertainties from the inversion are the highest are very different between $\mathbf{B}^{\text{notion}}$ and $\mathbf{B}^{\text{empiric}}$. For example, $\mathbf{B}^{\text{empiric}}$ defines a
450 much larger uncertainty than $\mathbf{B}^{\text{notion}}$ over eastern France (43% vs 16%) while the opposite is true for southern UK (4% vs 14%).
As a result, the UR of eastern France is 33% in E-43W and 8% in N-43W, and the UR of southern UK is 2% in E-43W and
23% in N-43W.

Complementing the uncertainty reduction, Fig. 5 shows the prior and posterior uncertainties and provides insight into the
precision of the estimates of monthly FFCO₂ emissions after inversion with NET17 and NET43 and 2-week sampling. For
455 example, using NET17, uncertainties in monthly FFCO₂ emissions are reduced from 29% (or 17%) in the prior estimates to
17% (or 9%) in the posterior estimates for western Germany in INV-N (or INV-E). Using additional sites in NET43 reduces
the uncertainties in monthly FFCO₂ emissions in southern UK from 25% in the prior estimates to 19% in the posterior estimates
in INV-N, and reduces the uncertainties in monthly FFCO₂ emissions in eastern France from 44% in the prior estimates to 29%
in the posterior estimates in INV-E. Like the UR, posterior uncertainties and their spatial variations are different between INV-
460 N and INV-E inversions, and demonstrate a strong dependence on the choice of $\mathbf{B}=\mathbf{B}^{\text{notion}}$ or $\mathbf{B}=\mathbf{B}^{\text{empiric}}$.

The scores of the MR and misfits of monthly emissions in both inversions using NET17 and NET43 and 2-week sampling
are shown in Fig. 4 (b, d, f, h) and Fig. 5 (b, d, f, h, j, l). In INV-E, there are slight differences between posterior misfits and
uncertainties, and between MR and UR. For example, for E-43W, the MR (21%) for Iberian Peninsula is larger than the UR
(5%), while the MR (40%) for western Germany is slightly smaller than the UR (47%). Despite such differences, the spatial
465 patterns of the MRs in Fig. 4 and posterior misfits in Fig. 5 are close to those of the URs and posterior uncertainties. On the
contrary, there are large differences between the statistics of posterior misfits and posterior uncertainties, and between MRs
and URs in INV-N. In some regions, such as southern UK (MR= -0.9 in N-17W and MR= -1.4 in N-43W) and northern Italy
(MR= -0.4 in N-17W and MR= -1.5 in N-43W), the MRs are negative and far below zero. This means that the posterior misfits
are even larger than the prior misfits (comparing Fig. 5f and 5j with Fig. 5b), and thus a degradation of the emission estimates
470 from the inversion is seen in these regions when assimilating FFCO₂ data. This suggests that the theoretical computation of
posterior uncertainty poorly characterizes the actual performance of the inversion in practice when the configuration of the
prior uncertainty covariance matrix and the actual prior errors are not consistent.

Figure 6 shows the correlations in the prior and posterior uncertainties in monthly emissions from different regions, and
their differences in inversions N-43W and E-43W. After assimilating the observations, the change of correlations mainly occurs
475 among regions that have large URs. In both inversions, there are negative correlations between the posterior uncertainties in
monthly emissions from some neighboring regions, in particular between western Germany and eastern Germany (from -0.27
to -0.18 depending on the months). The negative correlations between the posterior uncertainties in monthly emissions of
different regions indicate that NET43 brings a strong constraint on the budgets over a large area but does not separate individual
regions so well. At the same time, the temporal correlations in the posterior uncertainties between different months for a given
480 region also change after the inversion. For example, in INV-N, temporal correlations between posterior uncertainties in
monthly emissions for a specific region are smaller than those between prior uncertainties for that region when the time lag is

smaller than 3 months, while they are larger than the ones in prior uncertainties when the time lag exceeds 3 months (Fig. 6e). Because our setup of $\mathbf{B}^{\text{notion}}$ only considers an exponentially decaying temporal correlation with a correlation length of 2 months (Sect. 2.2.2), these longer term correlations in monthly posterior uncertainties must hence be driven by the temporal correlations in observation error, which contains a long-term component (see Sect. 2.2.2). On the contrary, in INV-E where $\mathbf{B}^{\text{empiric}}$ includes a component with annual-scale temporal correlations, the temporal correlations between posterior uncertainties in the monthly emissions are smaller than those between prior uncertainties. The analysis of the correlations in the prior or posterior uncertainties from N-17W and E-17W leads to very similar conclusions, but is not shown here.

3.1.2 Analysis for annual emissions

We compare the performance of different inversions to constrain annual mean FFCO₂ emissions. Corresponding UR and MR values are shown in Fig. 7. The patterns and values of UR for annual emissions are very similar to those at monthly scale (Fig. 4). High URs and MRs occur mostly in regions where the observation networks are dense and the emissions are high. For example, up to 47% UR is achieved for annual emissions in western Germany when using network NET43 and 2-week sampling. As a result, the posterior uncertainties of annual fossil fuel emissions, when using NET43 with 2-week sampling, are 10% (or 4%) for southern UK, 8% (or 8%) for western Germany and 15% (or 28%) for eastern France in INV-N (or INV-E).

Both the spatial spread and the magnitude of the MR of annual emissions in INV-E (Fig. 7d and 7h) are larger than those of the UR. The differences between MR and UR are much larger at annual than at monthly scale (when comparing Fig. 4 and 7). The cause of the discrepancy between UR and MR was presented in Sect. 2.2.1, and it may have a larger impact at the annual scale than at the monthly scale due to the evaluation of annual UR scores to annual MR values corresponding to single realizations of the misfits. In INV-N, the spatial spread and the magnitude of the MR are still significantly different from those of the UR and the MRs for some regions are still negative and far below zero.

3.2 Impact of using daily measurements and using a dense observation network

Figure 8 shows the URs and MRs of monthly emissions from inversions using NET43 and daily sampling, and from inversions using NET233 network and 2-week sampling (N-43D, E-43D, N-233W and E-233W in Table 2). When using NET43 and daily sampling, the URs of monthly emissions are generally larger (improved) than when using 2-week sampling for all regions. The differences between the UR values of monthly emissions with daily and with 2-week sampling are larger (meaning more improvement with daily sampling) over the regions where the network is dense and the emissions are high. For instance, the UR of monthly emissions for western Germany are as high as 62% (or 67%) in INV-N (or INV-E). When using the much denser NET233 network yet with a lower 2-week sampling (Fig. 8 d-f), we found that UR of monthly emissions in some regions that were poorly sampled by networks NET17 and NET43 are largely improved. For instance, UR value in Eastern Europe is 36% in N-233W (compared with 15% in N-43W) and is 73% in the Balkans in E-233W (compared with 13%

in E-43W). In principle, large regions tend to encompass more sites and to be surrounded by more sites than small regions, and thus may have more observations to improve their estimates of emissions. However, in both N-233W and E-233W, the URs for regions with a large area like northern Europe are still limited to below 5%. Large URs are identified over the regions whose absolute uncertainties are high, revealing the important roles of the absolute prior uncertainties when using the coarse-resolution transport model in the inversion of FFCO₂ emissions over Europe. The scores of MR match relatively well those of UR only in E-43D and E-233W (INV-E inversions) but not in N-43D and N-233W (INV-N inversions) (comparing Fig. 8d versus Fig. 8c, and Fig. 8h versus Fig. 8g). Even though the temporal frequency or spatial coverage of the sampling of the FFCO₂ mixing ratios are largely improved using NET43 and daily sampling, or NET233 and 2 week sampling, the MRs are still negative and below zero for a large number of regions in Europe.

4. Discussion

4.1 Implication for long-term trend detection of fossil fuel emissions

In the Copenhagen conference of parties, the European Union (EU) set up the goal to decrease its emissions (in CO₂ equivalents) by 80%–95% below 1990 by 2050 (European commission, 2010). In 2015, the EU Intended Nationally Determined Contribution (INDC) submitted to the UNFCCC set a target of 40% domestic greenhouse gas emissions reduction below 1990 levels by 2030. These targets translate into annual reductions compared to 1990 of roughly 1% per year in the 2020s, 1.5% in the decade from 2020 until 2030, and 2% in the two decades until 2050 (European commission, 2010). Levin and Rödenbeck (2008) showed that, taking into account the inter-annual variations of the atmospheric transport, changes of 7–26% between two consecutive 5-year averages of FFCO₂ emissions in south-western Germany could be detected at the 95% confidence level with monthly mean gradients of ¹⁴CO₂ observations between two stations (Schauinsland and Heidelberg) and the reference site JFJ. Such a detectability skill is clearly insufficient to support the “verification” of 1–2% annual change of emissions per year (meaning 5–10% changes between two consecutive 5-year averages) corresponding to the EU targets. Here, we evaluate the skill to detect trends when using the much larger ¹⁴CO₂ networks and the atmospheric inversion framework detailed in this study.

The uncertainty in the trend of FFCO₂ emissions calculated from the linear regression of a series of annual estimates, is independent of this trend itself (see Appendix C). This allows us to extrapolate posterior uncertainties in annual emissions from this study to investigate the detectability of emissions trends. Assuming that the absolute values of the standard deviations of the uncertainties in annual emissions of different years (in Tg/year) are identical and that these uncertainties are fully independent and ignoring the changes in transport on decadal scales (Aulagnier et al. 2009; Ramonet et al., 2010), we calculate the uncertainty in relative trends for different time lengths as a function of the posterior uncertainty in annual emissions (Table 3). Here, the relative trend is defined as the ratio of the linear regression slope of emissions to the emission in the base year. Using NET17 or NET43 and 2-week sampling, the posterior uncertainty in annual emissions of some well-sampled regions,

e.g. Germany, is largely below 10% (Sect. 3.1.2). In this case, given Table 3, the uncertainty in the relative trends over 20 years is in the range of 0.27% yr⁻¹ to 0.43% yr⁻¹. However, the uncertainty in trend estimation over 10 years would be 1% yr⁻¹. The EU target of 1-2% annual reduction, could thus be verified using NET17 or NET43 in these well-sampled regions over a period of 20 years but not over a period of 10 years. For other regions with sparser coverage of stations, either the posterior uncertainty in annual emissions are much larger than 10% (e.g. in Ireland and Balkans in INV-E) or the URs (or MRs) of annual emissions are marginal (meaning no improvement in the estimate of annual emissions from the inversion), so that the verification of the trend in these regions based on the inversion framework of our study is thus challenging.

Our assumption that the posterior uncertainties in annual emissions have the same amplitude from year to year should not strongly drive the results, so the results here give a good indication of the level of uncertainty in the trend detection for a typical level of uncertainty at the annual scale. However, changes of the transport from year to year or on decadal scales (Aulagnier et al. 2009; Ramonet et al., 2010) may change the level of the sensitivity of the observations to the emissions, i.e., the level of the atmospheric constraint of the inversions which leads to uncertainty reduction, and thus the level of posterior uncertainties on the same timescales. A more complex model accounting for varying levels of annual posterior uncertainties may thus be useful to refine the quantification of the uncertainty in the trends. Of note is that the level of uncertainties in the trends could be increased if the modeling framework accounts for the trends in the transport or in the sources of ¹⁴CO₂ other than the fossil fuel emissions. Such trends in the modeling errors may have to be considered for applications with real data.

4.2 Adequacy of large-scale atmospheric inversion for the monitoring of fossil fuel emissions and potential improvements of the inversion skills

In this study, we showed that given the NET17 ¹⁴CO₂ measurement station network, the potential of our atmospheric inversion of fossil fuel emissions at large scale using a coarse-resolution model is limited (Fig. 4 and Fig. 5). When using the denser NET43 network and 2-week sampling and assimilating ~1000 measurements per year, the potential of the inversion system is improved, yet mainly over high emitting regions. In particular, Sect. 3 indicates that the inversion can significantly reduce the uncertainties and misfits in the estimate of monthly emission budgets for large or high emitting regions, even though the observation operator used by the inversion assumes flat temporal profiles for the emissions while the true emissions have diurnal, weekly and seasonal temporal profiles. This confirms that the two-week mean afternoon ¹⁴CO₂ samplings integrate the atmospheric signal transported from both daytime and nighttime emissions across Europe which can be filtered from the signal from local emissions to provide large-scale information on the emissions.

We paid attention (as compared to previous OSSEs published for the USA) to account for aggregation and representation errors, which is the reason why our inversions do not provide as impressive error reductions (uncertainty and misfit) as ~~those~~ the misfit reduction of Ray et al. (2014) and Basu et al. (2016). However, we still did not account for all sources of uncertainty. Indeed, we assumed that atmospheric FFCO₂ gradients can be derived from the ¹⁴CO₂ measurements with a precision of 1 ppm. This 1-ppm standard deviation approximately corresponds to the errors in the atmospheric measurements and ignores

uncertainties in the conversion of $^{14}\text{CO}_2$ and CO_2 measurements into FFCO₂. Uncertainties in various fluxes that influence atmospheric $^{14}\text{CO}_2$, such as those from cosmogenic production, ocean, biosphere and nuclear facilities, bring systematic errors to the conversion of ^{14}C measurements into FFCO₂ (Lehman et al., 2013; Vogel et al., 2013). For example, over land regions, heterotrophic respiration is expected to be one of the main contributors to the large-scale signals of atmospheric $^{14}\text{CO}_2$ (Turnbull et al., 2009). Over regions like some areas of Europe, ^{14}C emissions from nuclear facilities may have even larger influences than plant and heterotrophic respiration in some areas (Graven and Gruber, 2011). The level of uncertainties in these fluxes and how much their influences from these fluxes will introduce additional errors in the FFCO₂ gradients will introduce additional errors remains to be quantified. According to the simulations by Graven and Gruber (2011), Turnbull et al. (2009) and Miller et al. (2012), one can expect that the impact of signals from the uncertainties associated in the estimate of these fluxes, on the conversion of atmospheric $^{14}\text{CO}_2$ measurements to FFCO₂, are typically below 1 ppm, i.e. much smaller than the observation errors that have been accounted for in this study, justifying that we have ignored these fluxes. However these signals may have complex spatial and temporal patterns leading to significant impact on the quantification of the inversion performances. Uncertainties in the trends of these fluxes could also impact that in the fossil fuel trend detection. Therefore, in future studies, especially if working with real data, the impacts from uncertainties in the $^{14}\text{CO}_2$ fluxes other than the anthropogenic fossil fuel emissions need to be investigated and accounted for by modelling all these $^{14}\text{CO}_2$ fluxes, their atmospheric $^{14}\text{CO}_2$ signals and associated uncertainties.

In Sect. 3.3, we explored the concept of having more observations assimilated in the inversion system by increasing the sampling frequency and expanding the observational network. Wang et al. (2017) showed that because the representation error, aggregation error and the prior FFCO₂ errors have very similar error structures in time, it is difficult to use daily sampling to filter uncertainties in the prior estimate of the emissions. However, we showed that when using NET43 and daily sampling, the UR of monthly emissions is still much larger than using 2-week sampling. This stems from the fact that having daily sampling decreases the weight of the measurement errors at the 2-week to annual scales, which are assumed not to have temporal autocorrelations. We also tested the concept of extending the observation network to a very dense configuration, NET233, with a wide coverage across Europe. It exhibits a significant increase in the UR of monthly emissions across Europe, especially over Eastern Europe. Emissions in Northern Europe, however, remains poorly constrained. This illustrates the limitation of using a coarse resolution transport model to quantify fossil fuel emissions. Such a limitation is attributed to the following facts: 1) the observation error in the inversions are larger than the prior FFCO₂ error (typically 0.21 ppm for 2-week mean afternoon FFCO₂ gradients and 0.49 ppm for daily mean afternoon FFCO₂ gradients, Wang et al., 2017); and 2) the observation errors bear complex temporal and spatial correlations which are close to the prior FFCO₂ errors (Wang et al., 2017). Such a result illustrates the need for using a suitable observation error characterization (here based on the results from Wang et al., 2017) to prevent the stations having a full coverage of information on the emissions in the model framework shown here even when the observation network is as dense as NET233. A proper account for the observation errors and their temporal and spatial correlations avoid overestimating the potential of the atmospheric inversion in OSSES when using a coarse resolution

transport model.

610 This study provides understanding of the inversion behavior and sensitivity to network density, but the precise
quantification of the performance of the inversion is largely dependent on the spatial resolution of the transport model. Wang
et al. (2017) showed that the representation error contributes the most to the observation errors, followed by the transport and
measurement errors. Following the definition of the observation errors in Wang et al. (2017) and in this study, ~~the~~
representation and the transport error are highly dependent on the transport model resolution. Increasing the transport model
615 resolution will reduce the representation errors and (potentially) reduce the transport error, if topography effects and synoptic
variations are better simulated by finer-resolution models. We thus assume that using a regional mesoscale transport model
with higher resolution than LMDZv4 (like for the regional scale natural flux inversions in Kadygrov et al., 2015; Broquet et
al., 2013; Gourdji et al., 2012; Lauvaux et al., 2008) should be the most efficient way to improve the results from atmospheric
inversion of FFCO₂ emissions at regional scale. A proper quantification of the change of representation and transport error as
620 a function of spatial resolution, and of the impact of this change on the performance of the inversion system would require a
series of transport models and inversions at varying spatial resolution which are out of the scope of this study but which would
be worth being investigated in the future.

However, unlike such regional transport models, a global transport model can propagate uncertainties in emissions in
other continents to Europe and thus allow to account for them when estimating the European emissions. To quantify the impact
625 of the uncertainties in emissions from other continents, we conducted additional inversions that only solve for emissions in
European regions ignoring those of other continents. The results show that fossil fuel emissions from other continents have
negligible impacts on UR, MR and posterior emission budgets of European regions (the relative differences between these
estimates being smaller than 1%; not shown). This indicates that the inversion system mainly exploits the signals of the
gradients between the European sites to constrain the European emissions, and the incoming FFCO₂ over the European air-
630 shed from emissions outside the European continent, results in very small FFCO₂ gradients between JFJ and other stations in
Europe. As a result, it highlights the possibility of using a mesoscale regional transport model and a regional inversion
framework to derive monthly and national scale emission budgets from ¹⁴CO₂ networks in Europe. In such a framework, the
uncertainties in the signals of fossil fuel emissions from remote emissions outside Europe could be neglected or coarsely
accounted for by controlling the regional transport model boundary conditions. However, such a conclusion may need to be
635 re-evaluated when processing real data and accounting for uncertainties in other types of ¹⁴CO₂ fluxes, since, e.g., parts of the
Atlantic ocean fluxes may have a significant signature on the European ¹⁴CO₂ gradients.

4.3 The need for good estimates of the uncertainties in the prior estimate of the emissions from inventories

The inconsistencies between the posterior misfits and the theoretical computation of posterior uncertainties, and between
the scores of MR and UR in INV-N inversions indicate that the theoretical computation of posterior uncertainty is not sufficient
640 to characterize the actual performance of the inversion, especially when the prior uncertainty covariance matrix does not

capture the actual error statistics of the prior estimate of the emissions. Moreover, in INV-N, there is a degradation of the emission estimates for many regions, characterized by negative and far-below-zero MRs in Sect. 3. This degradation occurs even when using daily measurements or the network NET233. A first explanation is that the signature of the errors in the prior emission estimates in the FFCO₂ fields has a smaller amplitude than the observation errors and thus the ability to filter this information for a proper correction of the emissions strongly relies on the knowledge of the prior uncertainty covariance. If **B** misses the amplitude and the temporal and spatial correlations of the actual errors, the system can translate observation errors into corrections to the emissions. Furthermore, some of the region-months are poorly constrained by the observations (due to the meteorological conditions and/or to the observation network spatial distribution), and the corrections to such region months is imposed by the extrapolation of the corrections to other region-months following the uncertainty structures characterized by **B**. If those structures do not represent the actual errors correctly, the system could apply corrections with a wrong sign or amplitude to the poorly observed region-months. A similar problem occurs when the network can constrain the sum of the budgets for several region-months but not the individual budgets of these region months (due to being too coarse). If the structure of **B** is wrong, the repartition of the constraint from the observations between these different region-months can be erroneous. All these analysis reveal the difficulty to capture the signatures of uncertainties in the prior emission estimate from the assimilated prior-model data misfits in our specific inverse modeling problem and thus to derive good corrections when the prior uncertainty covariance matrix is not configured properly.

In such a situation, only a precise configuration of the prior uncertainty covariance matrix can support the filtering of the prior errors. Consequently, even though both $\mathbf{B}^{\text{empiric}}$ and $\mathbf{B}^{\text{notion}}$ are derived from realistic assumptions on the uncertainties in the inventories, and to some different extent, from the analysis of inventory maps, the inconsistencies between these two matrices lead, in general, to positive MR when using the former and negative ones when using the latter.

In real applications, having such a good fit between the configuration of the prior uncertainty covariance matrix in the inversion system as between $\mathbf{B}^{\text{empiric}}$ and the synthetic prior errors in our OSSEs could appear to be unlikely, especially since the difference between $\mathbf{B}^{\text{empiric}}$ and $\mathbf{B}^{\text{notion}}$ illustrates the range of assumption we could have on the uncertainties in the existing inventories. Consequently, in order to improve the estimate of FFCO₂ emissions, on the one hand, more detailed and systematic evaluations of the uncertainty in the FFCO₂ emission inventories and of their potential temporal/spatial correlations (Andres et al., 2014; 2016b) would be required. On the other hand, as mentioned in Sect. 4.2, using a regional mesoscale transport model with higher resolution would reduce the representation error and (potentially) the transport error, and thus the observation error. Such a model would be needed to decrease the ratio of the observation error to the prior FFCO₂ error and thus increase the ability to filter the prior errors from the prior-model data misfits.

5. Conclusion

In this study, we present the application of a global atmospheric inversion method to quantify FFCO₂ emissions over

Europe at regional scale using three continental networks of $^{14}\text{CO}_2$ measurement sites. Its framework has been introduced by Wang et al. (2017). This method combines a prior emission estimate from an inventory with the information from atmospheric observations of FFCO₂ gradients to provide improved emission estimates with reduced uncertainties. A set of inversions are performed to test the potential of such a global atmospheric inversion system and the relevance of the large-scale inverse modeling (using coarse resolution transport model and controlling the emissions at regional scale) to monitor FFCO₂ emissions. The results show that given the 17 $^{14}\text{CO}_2$ measurement stations that are available in 2016 and the typical 2-week sampling frequency, the inversion reduces the uncertainties in monthly emission estimates for western Germany by 34% to 38%, depending on the setup of the prior uncertainty. By using a plausible network containing 43 measurement stations which is planned for the future and using 2-week sampling, one could expect higher URs of the emissions over the high emitters in Europe, e.g. eastern France (16% to 33%), southern UK (3% to 23%). In addition, given the posterior uncertainty in the emissions that could be achieved in such an inversion system, the uncertainties in the regressed trends can be significantly reduced below 1% yr⁻¹ by monitoring the FFCO₂ emissions for more than 20 years.

Increasing the number of observations assimilated in the inversion system by using daily sampling or a very dense observational network could potentially increase the UR over European regions. However, even though the inverse modeling framework used here can be assumed to be optimistic, e.g. regarding the assumption of the FFCO₂ data precision (see Sect. 2.2.2), its potential to improve the estimate of FFCO₂ emissions is often limited. The concept of using a coarse-resolution transport model in a global inversion system to solve for fossil fuel emissions of the regions whose emissions are not as high as those of Germany/France is challenged by the fact that coarse-resolution transport model can hardly filter the signature of the uncertainties in the emission budget from other signals and sources of errors within their coarse grid cells. Thus, regional high-resolution transport models could thus be required for the monitoring of FFCO₂. At the same time, the posterior estimate of the emissions are much degraded when the configuration of prior uncertainty in the inversion system is improper, implying that systematic evaluations of the uncertainties and temporal and spatial correlations in FFCO₂ emission inventories are also needed to improve the estimate of FFCO₂ emissions when applying such an inversion system to actual data.

Appendix A. Setup of $\mathbf{B}^{\text{notion}}$

The $\mathbf{B}^{\text{notion}}$ is a block diagonal matrix. The i th main diagonal block \mathbf{B}_i represents the prior uncertainty covariance of the emissions for 12 months for a given region i . Assuming the relative error δ_i for \mathbf{x}^b_i are the same for 12 months and $\mathbf{x}^b_{i,m}$ is the emission for region i and month m ($m=1$ means January, $m=12$ means December), so that the diagonal entries of the $\mathbf{B}^{\text{notion}}$ are:

$$\mathbf{B}_{i,(m,m)} = (\delta_i \mathbf{x}^b_{i,m})^2 \quad (\text{A-1})$$

The assumed 2-month temporal autocorrelations (Sect. 2.2.2), expressed by an exponential decaying function, leads to the non-diagonal entries in \mathbf{B}_i . Accordingly, the covariance between the uncertainties in the emissions of 2 months (month m and n , for instance) to be:

$$\mathbf{B}_{i,(m,n)} = e^{-\frac{|n-m|}{2}} \times (\delta_i \mathbf{x}_{i,m}^b) \times (\delta_i \mathbf{x}_{i,n}^b) \quad (\text{A-2})$$

705 If region i and region j are within the same country, the off-diagonal block $\mathbf{B}_{i,j}$ is built to account for the spatial correlation between these two regions. We assume that $\delta_i = \delta_j = \delta_{ij}$ and the spatial correlation between these two regions for a given month m is -0.2 to account for the fact that present emission estimates at such scales are generally disaggregated from national inventories, that is:

$$\mathbf{B}_{i,j,(m,m)} = -0.2 \times (\delta_{ij} \mathbf{x}_{i,m}^b) \times (\delta_{ij} \mathbf{x}_{j,m}^b) \quad (\text{A-3})$$

710 We assume that the correlation between two control variables are given by the product of the spatial and temporal correlations between the two corresponding control regions and the two months respectively. At last, the δ for each region are determined so that the prior annual emission uncertainty is satisfied, i.e. 10% for US, 10% for European countries, 15% for China and 10% for other large regions.

Appendix B. Setup of $\mathbf{B}^{\text{empiric}}$

715 The $\mathbf{B}^{\text{empiric}}$ is also a block diagonal matrix. For a given region i and a specific month m , assuming the prior control parameter corresponding to PKU-CO₂ emission is $\mathbf{x}_{i,m}^b$, the “true” value of \mathbf{x} , corresponding to IER-EDG writes $\mathbf{x}_{i,m}^t$, so that the errors of the prior monthly emissions are:

$$\Delta \mathbf{x}_{i,m} = \mathbf{x}_{i,m}^t - \mathbf{x}_{i,m}^b \quad (\text{B-1})$$

The long-term error component at annual scale ε_{ann} equals:

$$\varepsilon_{\text{ann}} = \frac{1}{12} \sum_{m=1}^{12} \Delta \mathbf{x}_{i,m} \quad (\text{B-2})$$

720 The residues are:

$$\mathbf{r}_{i,m}^{\text{ann}} = \Delta \mathbf{x}_{i,m} - \varepsilon_{\text{ann}} \quad (\text{B-3})$$

Then the 6-month variation ε_{6m} equals the standard deviation of the 6-month mean residues:

$$\varepsilon_{6m} = \text{SD of } \left(\frac{1}{6} \sum_{m=1}^6 \mathbf{r}_{i,m}^{\text{ann}}, \frac{1}{6} \sum_{m=7}^{12} \mathbf{r}_{i,m}^{\text{ann}} \right) \quad (\text{B-4})$$

Again, the residues become:

$$\begin{aligned} \mathbf{r}_{i,m}^{6m} &= \Delta \mathbf{x}_{i,m} - \varepsilon_{\text{ann}} - \frac{1}{6} \sum_{m=1}^6 \mathbf{r}_{i,m}^{\text{ann}} & (\text{if } m \leq 6) \\ \mathbf{r}_{i,m}^{6m} &= \Delta \mathbf{x}_{i,m} - \varepsilon_{\text{ann}} - \frac{1}{6} \sum_{m=7}^{12} \mathbf{r}_{i,m}^{\text{ann}} & (\text{if } m \geq 7) \end{aligned} \quad (\text{B-5})$$

725

In the same way, the 3-month variation ε_{3m} equals the standard deviation of the 3-month mean residues:

$$\varepsilon_{3m} = \text{SD of } \left(\frac{1}{3} \sum_{m=1}^3 \mathbf{r}_{i,m}^{6m}, \frac{1}{3} \sum_{m=4}^6 \mathbf{r}_{i,m}^{6m}, \frac{1}{3} \sum_{m=7}^9 \mathbf{r}_{i,m}^{6m}, \frac{1}{3} \sum_{m=10}^{12} \mathbf{r}_{i,m}^{6m} \right) \quad (\text{B-6})$$

And the corresponding residues:

$$\begin{aligned} \mathbf{r}_{i,m}^{3m} &= \Delta \mathbf{x}_{i,m} - \varepsilon_{\text{ann}} - \frac{1}{6} \sum_{m=1}^6 \mathbf{r}_{i,m}^{\text{ann}} - \frac{1}{3} \sum_{m=1}^3 \mathbf{r}_{i,m}^{6m} & (\text{if } m \leq 3) \\ \mathbf{r}_{i,m}^{3m} &= \Delta \mathbf{x}_{i,m} - \varepsilon_{\text{ann}} - \frac{1}{6} \sum_{m=1}^6 \mathbf{r}_{i,m}^{\text{ann}} - \frac{1}{3} \sum_{m=4}^6 \mathbf{r}_{i,m}^{6m} & (\text{if } 4 \leq m \leq 6) \\ \mathbf{r}_{i,m}^{3m} &= \Delta \mathbf{x}_{i,m} - \varepsilon_{\text{ann}} - \frac{1}{6} \sum_{m=7}^{12} \mathbf{r}_{i,m}^{\text{ann}} - \frac{1}{3} \sum_{m=7}^9 \mathbf{r}_{i,m}^{6m} & (\text{if } 7 \leq m \leq 9) \\ \mathbf{r}_{i,m}^{3m} &= \Delta \mathbf{x}_{i,m} - \varepsilon_{\text{ann}} - \frac{1}{6} \sum_{m=7}^{12} \mathbf{r}_{i,m}^{\text{ann}} - \frac{1}{3} \sum_{m=10}^{12} \mathbf{r}_{i,m}^{6m} & (\text{if } 10 \leq m \leq 12) \end{aligned} \quad (\text{B-7})$$

730 The 1-month variation ε_{1m} equals the standard deviation of these residues:

$$\varepsilon_{1m} = \text{SD}(\mathbf{r}_{i,m}^{3m}) \quad (\text{B-8})$$

Using such a decomposition, the root mean square of the errors (RMSE) between the prior and the “true” values $\Delta \mathbf{x}_{i,j}$ satisfy the following equation:

$$\text{RMSE}_i = \frac{1}{12} \sum_{m=1}^{12} \Delta \mathbf{x}_{i,m}^2 = \varepsilon_{\text{ann}}^2 + \varepsilon_{6m}^2 + \varepsilon_{3m}^2 + \varepsilon_{1m}^2 \quad (\text{B-9})$$

735 Finally, for the diagonal entries of the \mathbf{B} matrix corresponding to the monthly emissions of region i , they are equal to the RMSE_i , for the non-diagonal entries, the covariance between month j and month k for a given region is expressed as the sum of the products of the different variations multiplied by corresponding correlations (expressed by exponential decay functions) at different time scales:

$$\begin{aligned} \mathbf{B}_{i,(m,n)} &= \varepsilon_{\text{ann}}^2 + \varepsilon_{6m}^2 + \varepsilon_{3m}^2 + \varepsilon_{1m}^2 & (\text{if } m = n) \\ \mathbf{B}_{i,(m,n)} &= \varepsilon_{\text{ann}}^2 + e^{-\frac{|n-m|}{6}} \varepsilon_{6m}^2 + e^{-\frac{|n-m|}{3}} \varepsilon_{3m}^2 + e^{-\frac{|n-m|}{1}} \varepsilon_{1m}^2 & (\text{if } m \neq n) \end{aligned} \quad (\text{B-10})$$

740 Appendix C. Calculation of trends and corresponding uncertainties

Assuming the linear trend of the FFCO₂ emissions in an n -year period is to be calculated, which satisfies the function:

$$\mathbf{y} \approx \hat{\mathbf{y}} = \mathbf{a}\mathbf{x} + b \quad (\text{C-1})$$

where \mathbf{y} is the vector of annual emissions for the n years, $\hat{\mathbf{y}}$ is the predicted value by the regression and \mathbf{x} is the vector of corresponding years, the slope a is the linear trend we are going to calculate by linear regression. We rewrite Eq. (C-1) as follows:

745

$$\begin{bmatrix} y_1 \\ \vdots \\ y_{10} \end{bmatrix} \approx \begin{bmatrix} y_1 \\ \vdots \\ y_{10} \end{bmatrix} = \underbrace{\begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_{10} & 1 \end{bmatrix}}_{\mathbf{X}} \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{(C-2)}$$

Thus the linear trend a and the interception b can be solved using linear algebra. With the notations used in Eq. (C-2), the result of the linear regression is:

$$\mathbf{p} = (\mathbf{X}^T \text{cov}^{-1}(\mathbf{Y}) \mathbf{X})^{-1} \mathbf{X}^T \text{cov}^{-1}(\mathbf{Y}) \mathbf{y} \quad \text{(C-3)}$$

the associated uncertainties in the regression parameters in vector \mathbf{p} is thus given by the following covariance matrix:

$$\text{cov}(\mathbf{p}) = (\mathbf{X}^T \text{cov}^{-1}(\mathbf{Y}) \mathbf{X})^{-1} \mathbf{X}^T \text{cov}(\mathbf{Y}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \quad \text{(C-4)}$$

where $\text{cov}(\cdot)$ is the covariance matrix for a set of variables.

Since \mathbf{X} is a fixed matrix filled by the numbers of years and 1's, the uncertainties in the linear trend (first item in main diagonal of $\text{cov}(\mathbf{p})$), is independent of the annual emissions themselves but only depends on the uncertainties and associated correlations of annual emissions. As sketched in Fig. C1, this error covariance of \mathbf{y} should include two independent parts: 1) the uncertainties associated with the estimation of the emissions for each year in \mathbf{y} and 2) the inter-annual variability (IAV) in the detrended \mathbf{y} .

In this study, based on the time series of national annual emissions from IER-EDG, we assume a 5% IAV in the annual fossil fuel emissions for European countries. In general, this 5% IAV is the upper limit of the typical values for European countries (Levin and Rödenbeck, 2007). Ballantyne et al. (2015) assumed that in the self-reported fossil fuel emission inventories, the emission error in one year could be highly correlated with the error from the previous year by an autoregressive coefficient of 0.95, due to potential errors that are not corrected retroactively after about 20 years. However, we do not conduct a multi-year inversion to get a typical estimate of the correlations in the posterior uncertainties in annual emissions, and assume that there is no correlations between the posterior uncertainties in annual emissions. This assumption is fairly conservative, since Eq. (C-4) implies that the larger (either positive or negative) the correlations between the estimation of fossil fuel emissions from different years, the smaller the uncertainties in the regressed trends.

Acknowledgement. The authors acknowledge the support of the French Commissariat à l'énergie atomique et aux énergies alternatives (CEA). This study is co-funded by the European Commission under the EU Seventh Research Framework Programme (grant agreement no. 283080, geocarbon project). G. Broquet and F. Vogel acknowledge funding from the industrial chair BridGES (supported by the Université de Versailles Saint-Quentin-en-Yvelines, the Commissariat à l'Energie Atomique et aux Energies Renouvelables, the Centre National de la Recherche Scientifique, Thales Alenia Space and Veolia). We are also grateful to Ingeborg Levin for the useful discussions on this topic. We also would like to thank the partners of the

References

- Andres, R. J., Boden, T. A. and Higdon, D. M.: Gridded uncertainty in fossil fuel carbon dioxide emission maps, a CDIAC example, *Atmos. Chem. Phys.*, 16(23), 14979–14995, doi:10.5194/acp-16-14979-2016, 2016b.
- 780 Andres, R. J., Boden, T. A. and Marland, G.: Annual Fossil-Fuel CO₂ Emissions: Mass of Emissions Gridded by One Degree Latitude by One Degree Longitude, [online] Available from: <http://dx.doi.org/10.3334/CDIAC/ffe.ndp058.2016> (Accessed 9 August 2016), 2016a.
- Andres, R. J., Boden, T. A., and Higdon, D.: A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission, *Tellus B*, 66, 2014.
- 785 Andres, R. J., Boden, T. A., Br éon, F. M., Ciais, P., Davis, S., Erickson, D., Gregg, J. S., Jacobson, A., Marland, G., Miller, J., Oda, T., Olivier, J. G. J., Raupach, M. R., Rayner, P. and Treanton, K.: A synthesis of carbon dioxide emissions from fossil-fuel combustion, *Biogeosciences*, 9(5), 1845–1871, doi:10.5194/bg-9-1845-2012, 2012.
- Aulagnier, C., Rayner, P., Ciais, P., Vautard, R., Rivier, L. and Ramonet, M.: Is the recent build-up of atmospheric CO₂ over Europe reproduced by models. Part 2: an overview with the atmospheric mesoscale transport model CHIMERE, *Tellus B*, 62, 1, 2009.
- 790 Ballantyne, A. P., Andres, R., Houghton, R., Stocker, B. D., Wanninkhof, R., Anderegg, W., Cooper, L. A., DeGrandpre, M., Tans, P. P., Miller, J. B., Alden, C. and White, J. W. C.: Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty, *Biogeosciences*, 12(8), 2565–2584, doi:10.5194/bg-12-2565-2015, 2015.
- Basu, S., Miller, J. B. and Lehman, S.: Separation of biospheric and fossil fuel fluxes of CO₂ by atmospheric inversion of CO₂ and ¹⁴CO₂ measurements: Observation System Simulations, *Atmos. Chem. Phys.*, 16(9), 5665–5683, doi:10.5194/acp-16-5665-2016, 2016.
- Broquet, G., Chevallier, F., Br éon, F. M., Kadyrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morgu í J. A., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R. L., Vermeulen, A. T., Yver, C. and Ciais, P.: Regional inversion of CO₂ ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates, *Atmos. Chem. Phys.*, 13(17), 9039–9056, doi:10.5194/acp-13-9039-2013, 2013.
- 800 Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T. and Ciais, P.: A European summertime CO₂ biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, *J. Geophys. Res.*, 116(D23), doi:10.1029/2011jd016202, 2011.
- Cambaliza, M. O. L., Shepson, P. B., Caulton, D. R., Stirm, B., Samarov, D., Gurney, K. R., Turnbull, J., Davis, K. J., Possolo, A., Karion, A., Sweeney, C., Moser, B., Hendricks, A., Lauvaux, T., Mays, K., Whetstone, J., Huang, J., Razlivanov, I., Miles, N. L. and Richardson, S. J.: Assessment of uncertainties of an aircraft-based mass balance approach for quantifying urban greenhouse gas emissions, *Atmos. Chem. Phys.*, 14(17), 9029–9050, doi:10.5194/acp-14-9029-2014, 2014.
- 805 Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fr öhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morgu í J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S. and Worthy, D.: CO₂ surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, *J. Geophys. Res.*, 115(D21), doi:10.1029/2010jd013887, 2010.
- 810 Ciais, P., Paris, J. D., Marland, G., Peylin, P., Piao, S. L., Levin, I., Pregger, T., Scholz, Y., Friedrich, R., Rivier, L., Houwelling, S., and Schulze, E. D.: The European carbon balance. Part 1: fossil fuel emissions, *Global Change Biol*, 16, 1395-1408, 10.1111/j.1365-2486.2009.02098.x, 2010.
- 815 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., H ídm, E. V., Isaksen, I., K állberg, P., K öhler, M., Matricardi,
- 820

M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q.J.R. Meteorol. Soc.*, 137(656), 553–597, doi:10.1002/qj.828, 2011.

825 European Commission: Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A Roadmap for moving to a competitive low carbon economy in 2050, 2010.

Gourdji, S. M., Mueller, K. L., Yadav, V., Huntzinger, D. N., Andrews, A. E., Trudeau, M., Petron, G., Nehrkorn, T., Eluskiewicz, J., Henderson, J., Wen, D., Lin, J., Fischer, M., Sweeney, C. and Michalak, A. M.: North American CO₂ exchange: inter-comparison of modeled estimates with results from a fine-scale atmospheric inversion, *Biogeosciences*, 9(1), 457–475, doi:10.5194/bg-9-457-2012, 2012.

830 Graven, H. D. and Gruber, N.: Continental-scale enrichment of atmospheric ¹⁴CO₂ from the nuclear power industry: potential impact on the estimation of fossil fuel-derived CO₂, *Atmos. Chem. Phys.*, 11(23), 12339–12349, doi:10.5194/acp-11-12339-2011, 2011.

835 Gregg, J. S., Andres, R. J. and Marland, G.: China: Emissions pattern of the world leader in CO₂ emissions from fossil fuel consumption and cement production, *Geophys. Res. Lett.*, 35(8), 2008.

Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C., Geethakumar, S. and Can, S. de la R. du: High Resolution Fossil Fuel Combustion CO₂ Emission Fluxes for the United States, *Environ. Sci. Technol.*, 43(14), 5535–5541, 2009.

840 Hammer, S., Friedrich, R., Kromer, B., Cherkinsky, A., Lehman, S. J., Meijer, H. A. J., Nakamura, T., Palonen, V., Reimer, R. W., Smith, A. M., Southon, J. R., Szidat, S., Turnbull, J. and Uchida, M.: Compatibility of Atmospheric ¹⁴CO₂ Measurements: Comparing the Heidelberg Low-Level Counting Facility to International Accelerator Mass Spectrometry (AMS) Laboratories, *Radiocarbon*, 1–9, doi:10.1017/RDC.2016.62, 2016.

845 Hourdin, F., Musat, I., Bony, S., Braconnot, P., Codron, F., Dufresne, J.-L., Fairhead, L., Filiberti, M.-A., Friedlingstein, P., Grandpeix, J.-Y., Krinner, G., LeVan, P., Li, Z.-X. and Lott, F.: The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection, *Climate Dynamics*, 27(7–8), 787–813, doi:10.1007/s00382-006-0158-0, 2006.

Kadyrov, N., Broquet, G., Chevallier, F., Rivier, L., Gerbig, C. and Ciais, P.: On the potential of the ICOS atmospheric CO₂ measurement network for estimating the biogenic CO₂ budget of Europe, *Atmos. Chem. Phys.*, 15(22), 12765–12787, doi:10.5194/acp-15-12765-2015, 2015.

850 [Kaminski, T., Rayner, P. J., Heimann, M. and Enting, I. G.: On aggregation errors in atmospheric transport inversion, *J. Geophys. Res.*, 106, 4703–4715, 2001.](#)

Lauvaux, T., Uliasz, M., Sarrat, C., Chevallier, F., Bousquet, P., Lac, C., Davis, K. J., Ciais, P., Denning, A. S. and Rayner, P. J.: Mesoscale inversion: first results from the CERES campaign with synthetic data, *Atmos. Chem. Phys.*, 8, 3459–3471, 2008.

855 Lehman, S. J., Miller, J. B., Wolak, C., Southon, J., Tans, P. P., Montzka, S. A., Sweeney, C., Andrews, A., LaFranchi, B., Guilderson, T. P. and Turnbull, J. C.: Allocation of Terrestrial Carbon Sources Using ¹⁴CO₂: Methods, Measurement, and Modeling, *Radiocarbon*, 55(2–3), 1484–1495, doi:10.2458/azu_js_rc.55.16392, 2013.

Levin, I. and Rödenbeck, C.: Can the envisaged reductions of fossil fuel CO₂ emissions be detected by atmospheric observations?, *Naturwissenschaften*, 95(3), 203–208, doi:10.1007/s00114-007-0313-4, 2008.

860 Levin, I., Hammer, S., Kromer, B. and Meinhardt, F.: Radiocarbon observations in atmospheric CO₂: determining fossil fuel CO₂ over Europe using Jungfraujoch observations as background, *Sci. Total Environ.*, 391(2–3), 211–6, doi:10.1016/j.scitotenv.2007.10.019, 2008.

Levin, I., Kromer, B., Schmidt, M. and Sartorius, H.: A novel approach for independent budgeting of fossil fuel CO₂ over Europe by ¹⁴CO₂ observations, *Geophys. Res. Lett.*, 30(23), 2003.

865 Levin, I., Munnich, K. O., Weiss, W.: The effect of anthropogenic CO₂ and ¹⁴C sources on the distribution of ¹⁴C in the atmosphere, *Radiocarbon*, 22, 379–391, 1980.

[Lin, J. C. and Gerbig, C.: Accounting for the effect of transport errors on tracer inversions, *Geophys. Res. Lett.*, 32\(1\), L01802, doi:10.1029/2004GL021127, 2005.](#)

Lindenmaier, R., Dubey, M. K., Henderson, B. G., Butterfield, Z. T., Herman, J. R., Rahn, T. and Lee, S.-H.: Multiscale

- 870 observations of CO₂, ¹³CO₂, and pollutants at Four Corners for emission verification and attribution, Proc. Natl. Acad. Sci. U. S. A., 111(23), 8386–8391, doi:10.1073/pnas.1321883111, 2014.
- Liu, Z., Guan, D., Wei, W., Davis, S. J., Ciais, P., Bai, J., Peng, S., Zhang, Q., Hubacek, K., Marland, G., Andres, R. J., Crawford-Brown, D., Lin, J., Zhao, H., Hong, C., Boden, T. A., Feng, K., Peters, G. P., Xi, F., Liu, J., Li, Y., Zhao, Y., Zeng, N. and He, K.: Reduced carbon emission estimates from fossil fuel combustion and cement production in China, Nature, 524(7565), 335–338, doi:10.1038/nature14677, 2015.
- 875 Macknick, J.: Energy and carbon dioxide emission data uncertainties, International Institute for Applied Systems Analysis, Laxenburg, Austria, 2009.
- Marquis, M. and Tans, P.: Carbon Crucible, Science, 320(5875), 460–461, 2008.
- 880 [Miller, S. M., Hayek, M. N., Andrews, A. E., Fung, I. and Liu, J.: Biases in atmospheric CO₂ estimates from correlated meteorology modeling errors, Atmospheric Chemistry and Physics, 15\(5\), 2903–2914, 2015.](#)
- Miller, S. M. and Michalak, A. M.: Constraining sector-specific CO₂ and CH₄ emissions in the US, Atmos. Chem. Phys., 17(6), 3963–3985, doi:10.5194/acp-17-3963-2017, 2017.
- Oda, T. and Maksyutov, S.: A very high-resolution (1 km×1 km) global fossil fuel CO₂ emission inventory derived using a point source database and satellite observations of nighttime lights, Atmos. Chem. Phys., 11(2), 543–556, doi:10.5194/acp-11-543-2011, 2011.
- 885 Pacala, S. W., Breidenich, C., Brewer, P. G., Fung, I. Y., Gunson, M. R., Heddle, G., Law, B. E., Marland, G., Paustian, K., Prather, M., Randerson, J. T., Tans, P., Wofsy, S. C., Linn, A. M. and Sturdivant, J.: Verifying greenhouse gas emissions: methods to support international climate agreements, The National Academies Press, Washington, DC, 2010.
- Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., van der Laan-Luijkx, I. T. and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO₂ inversions, Biogeosciences, 10, 6699–6720, doi:10.5194/bg-10-6699-2013, 2013.
- 890 Pregger, T., Scholz, Y. and Friedrich, R.: Documentation of the anthropogenic GHG emission data for Europe provided in the Frame of CarboEurope GHG and CarboEurope IP, Institut für Energiewirtschaft und Rationelle Energieanwendung, Universität Stuttgart, Stuttgart, Germany, 2007.
- 895 Ramonet, M., Ciais, P., Aalto, T., Aulagnier, C., Chevallier, F., Cipriano, D., Conway, T. J., Haszpra, L., Kazan, V., Meinhardt, F., Paris, J.-D., Schmidt, M., Simmonds, P., Xueref-Rény, I. and Necki, J. N.: A recent build-up of atmospheric CO₂ over Europe. Part 1: observed signals and possible explanations, Tellus B, 62(1), 1–13, doi:10.1111/j.1600-0889.2009.00442.x, 2010.
- Ray, J., Yadav, V., Michalak, A., van Bloemen Waanders, B. and Mckenna, S. A.: A multiresolution spatial parameterization for the estimation of fossil-fuel carbon dioxide emissions via atmospheric inversions, Geosci. Model Dev., 7(5), 1901–1918, 2014.
- 900 Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L. and Yadav, V.: Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO₂ measurements, Geophys. Res. Lett., 41(12), 4381–4388, doi:10.1002/2014gl059684, 2014.
- 905 Staufer, J., Broquet, G., Br éon, F.-M., Puygrenier, V., Chevallier, F., Xueref-R ény, I., Dieudonn é E., Lopez, M., Schmidt, M., Ramonet, M., Perrussel, O., Lac, C., Wu, L. and Ciais, P.: The first 1-year-long estimate of the Paris region fossil fuel CO₂ emissions based on atmospheric inversion, Atmos. Chem. Phys., 16(22), 14703–14726, doi:10.5194/acp-16-14703-2016, 2016.
- [Snieder, R. and Trampert, J.: Inverse problems in geophysics. in Wavefield inversion. pp. 119–190, Springer., 1999.](#)
- 910 Turnbull, J. C., Keller, E. D., Baisden, T., Brailsford, G., Bromley, T., Norris, M. and Zondervan, A.: Atmospheric measurement of point source fossil CO₂ emissions, Atmos. Chem. Phys., 14(10), 5001–5014, doi:10.5194/acp-14-5001-2014, 2014.
- Turnbull, J. C., Keller, E. D., Norris, M. W. and Wiltshire, R. M.: Independent evaluation of point source fossil fuel CO₂ emissions to better than 10%, Proc. Natl. Acad. Sci. U. S. A., 113(37), 10287–10291, doi:10.1073/pnas.1602824113, 2016.
- 915 Turnbull, J. C., Miller, J. B., Lehman, S. J., Tans, P. P., Sparks, R. J. and Southon, J.: Comparison of ¹⁴CO₂, CO, and SF₆ as tracers for recently added fossil fuel CO₂ in the atmosphere and implications for biological CO₂ exchange, Geophys. Res. Lett., 33(1), doi:10.1029/2005gl024213, 2006.
- Turnbull, J., Rayner, P., Miller, J., Naegler, T., Ciais, P. and Cozic, A.: On the use of ¹⁴CO₂ as a tracer for fossil fuel CO₂:

Quantifying uncertainties using an atmospheric transport model, *J. Geophys. Res.*, 114(D22), doi:10.1029/2009jd012308, 2009.

920

US Environmental Protection Agency: U.S. Greenhouse Gas Inventory Report, [online] Available from: <http://www3.epa.gov/climatechange/ghgemissions/usinventoryreport.html> (Accessed 15 January 2016), 2015.

Vogel, F. R., Levin, I. and Worthy, D.: Implications for deriving regional fossil fuel CO₂ estimates from atmospheric observations in a hot spot of nuclear power plant ¹⁴CO₂ emissions, *Radiocarbon*, 55(2), 1556–1572, 2013.

925

Wang, R., Tao, S., Ciais, P., Shen, H. Z., Huang, Y., Chen, H., Shen, G. F., Wang, B., Li, W., Zhang, Y. Y., Lu, Y., Zhu, D., Chen, Y. C., Liu, X. P., Wang, W. T., Wang, X. L., Liu, W. X., Li, B. G. and Piao, S. L.: High-resolution mapping of combustion processes and implications for CO₂ emissions, *Atmos. Chem. Phys.*, 13(10), 5189–5203, doi:10.5194/acp-13-5189-2013, 2013.

930

Wang, Y., Broquet, G., Ciais, P., Chevallier, F., Vogel, F., Kadyrov, N., Wu, L., Yin, Y., Wang, R. and Tao, S.: Estimation of observation errors for large-scale atmospheric inversion of CO₂ emissions from fossil fuel combustion, *Tellus B*, 69(1), 1325723, doi:10.1080/16000889.2017.1325723, 2017.

Table 1 Setup and performance indicators of the two types of inversions

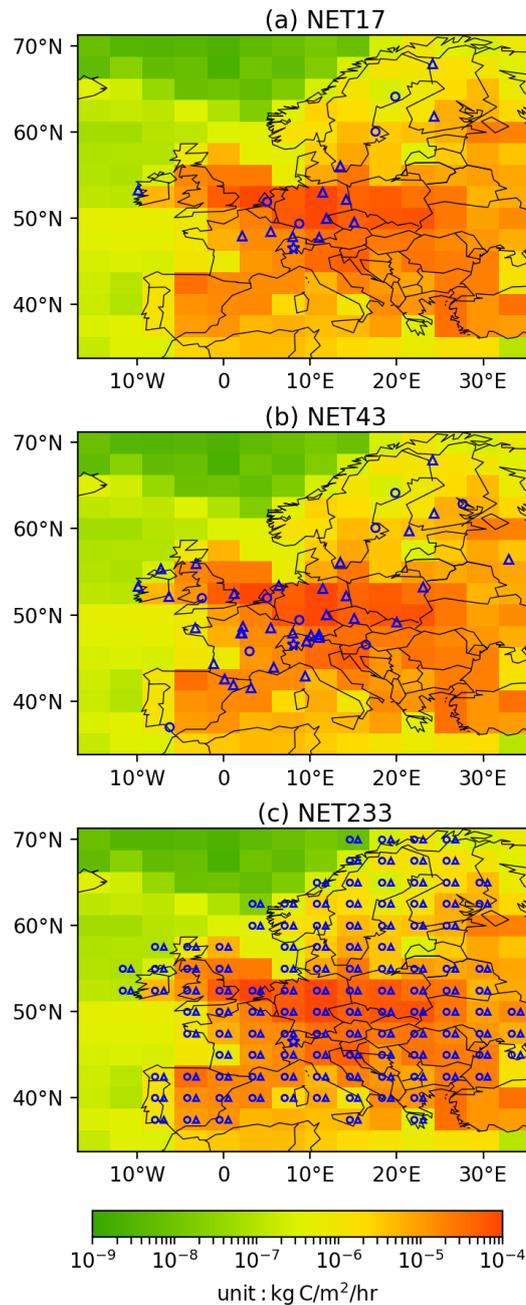
Input of inversions	INV-N	INV-E	Performance indicators
B	B ^{notion}	B ^{empiric}	A (Eq. (1))
R	R _i + R _r + R _t + R _a	R _i + R _r + R _t + R _a	UR (Eq. (3))
H	H ^{prac}	H ^{prac}	
x ^t	x ^{IER-EDG}	x ^{IER-EDG}	x ^a - x ^t (Eq. (2))
x ^b	x ^{PKU}	x ^{PKU}	MR (Eq. (4))
y ₀	H ^{OSSE} x ^t + ε _i + ε _r + ε _t	H ^{OSSE} x ^t + ε _i + ε _r + ε _t	

Table 2 Notations for the eight OSSEs.

	Number of synthetic data	INV-N	INV-E
NET17, 2-week sampling	416	N-17W	E-17W
NET43, 2-week sampling	1092	N-43W	E-43W
NET233, 2-week sampling	6032	N-43D	E-43D
NET43, 1-day sampling	15288	N-233W	E-233W

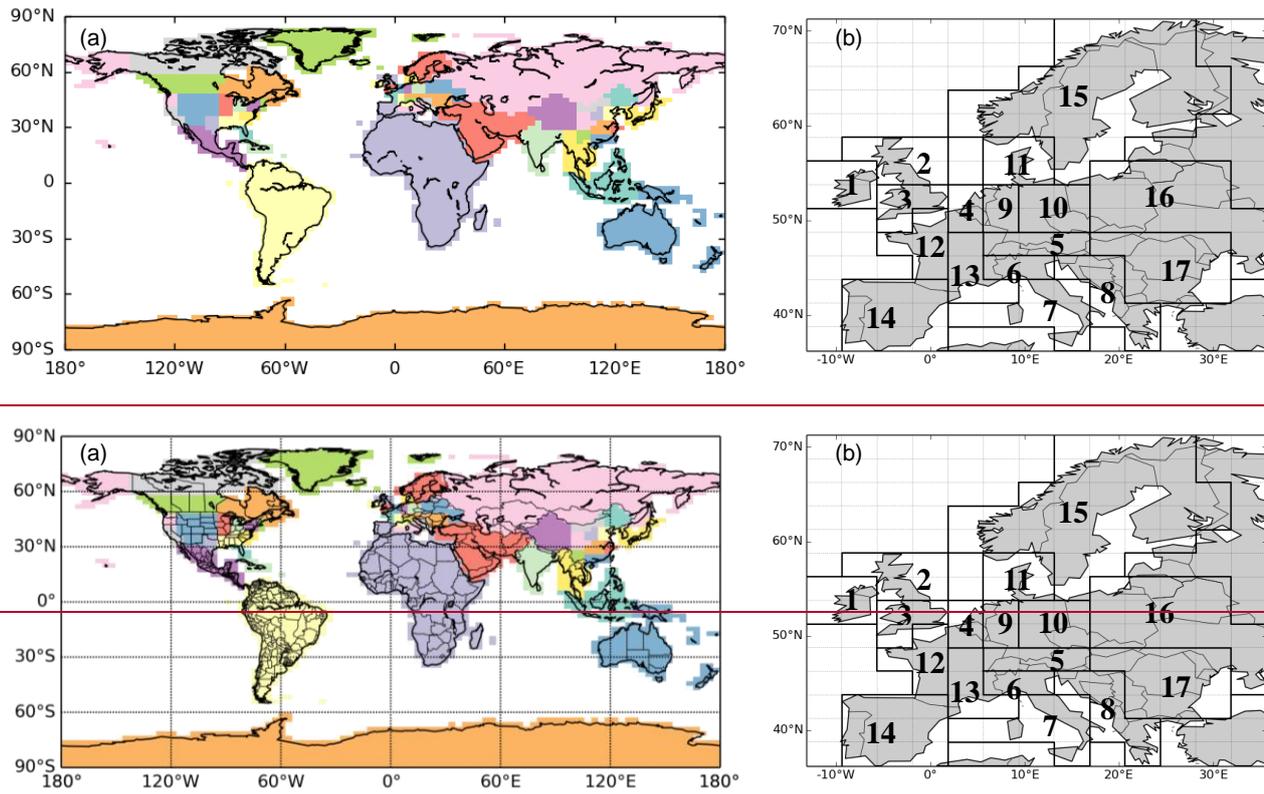
Table 3 Uncertainties in the regressed linear trends as a function of the posterior uncertainty in annual emissions. The uncertainties in the trends are defined as the ratio between the uncertainties in the linear regression slope of absolute annual emissions and the annual emission budget in the base year.

Relative posterior uncertainty in annual emissions	10-year trend	20-year trend
10%	1.2% yr ⁻¹	0.43% yr ⁻¹
5%	0.78% yr ⁻¹	0.27% yr ⁻¹
1%	0.56% yr ⁻¹	0.20% yr ⁻¹



945 **Figure 1:** Site locations for the three continental network configurations used in this study: (a) NET17, (b) NET38, and (c) NET232. Circles correspond to “urban” sites and upper triangles are “rural” sites. Urban and rural sites are categorized according to the population density of the grid cells within which the stations are located and according to the locations of large point sources. The background color map is the annual FFCO₂ emissions in 2007 at the resolution of LMDZv4 from the PKU-CO₂ inventory (Wang et al., 2013).

950



955

Figure 2: (a) Map of the 56 regions whose monthly emission budgets are controlled by the inversion; (b) zoom over the 17 control regions in Europe. In (a), we repeatedly use twelve colors for non-adjacent regions. For example, the Northern Europe, Middle East, one region in the USA and one region in China are all red. But since they are in different continents, they represent four different regions.

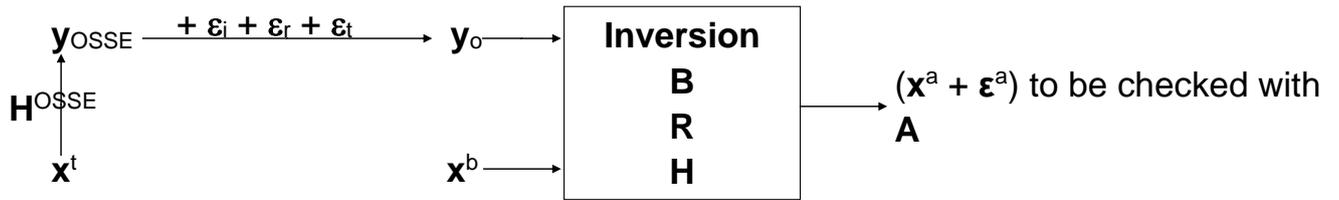


Figure 3: Schematic of the OSSEs

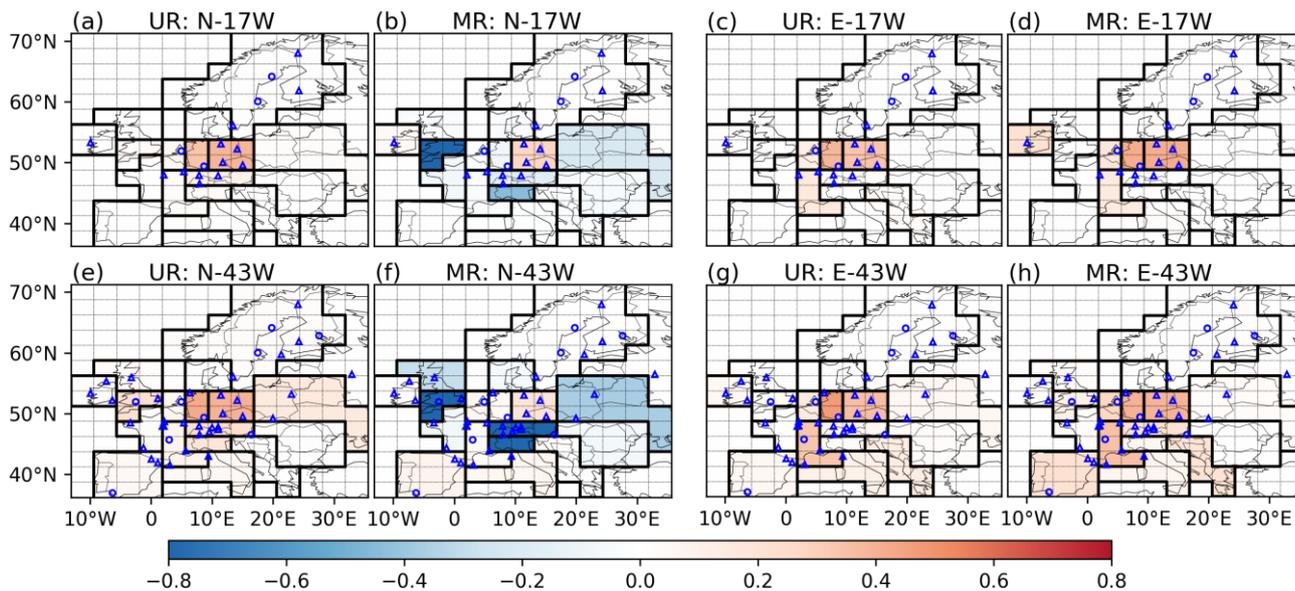
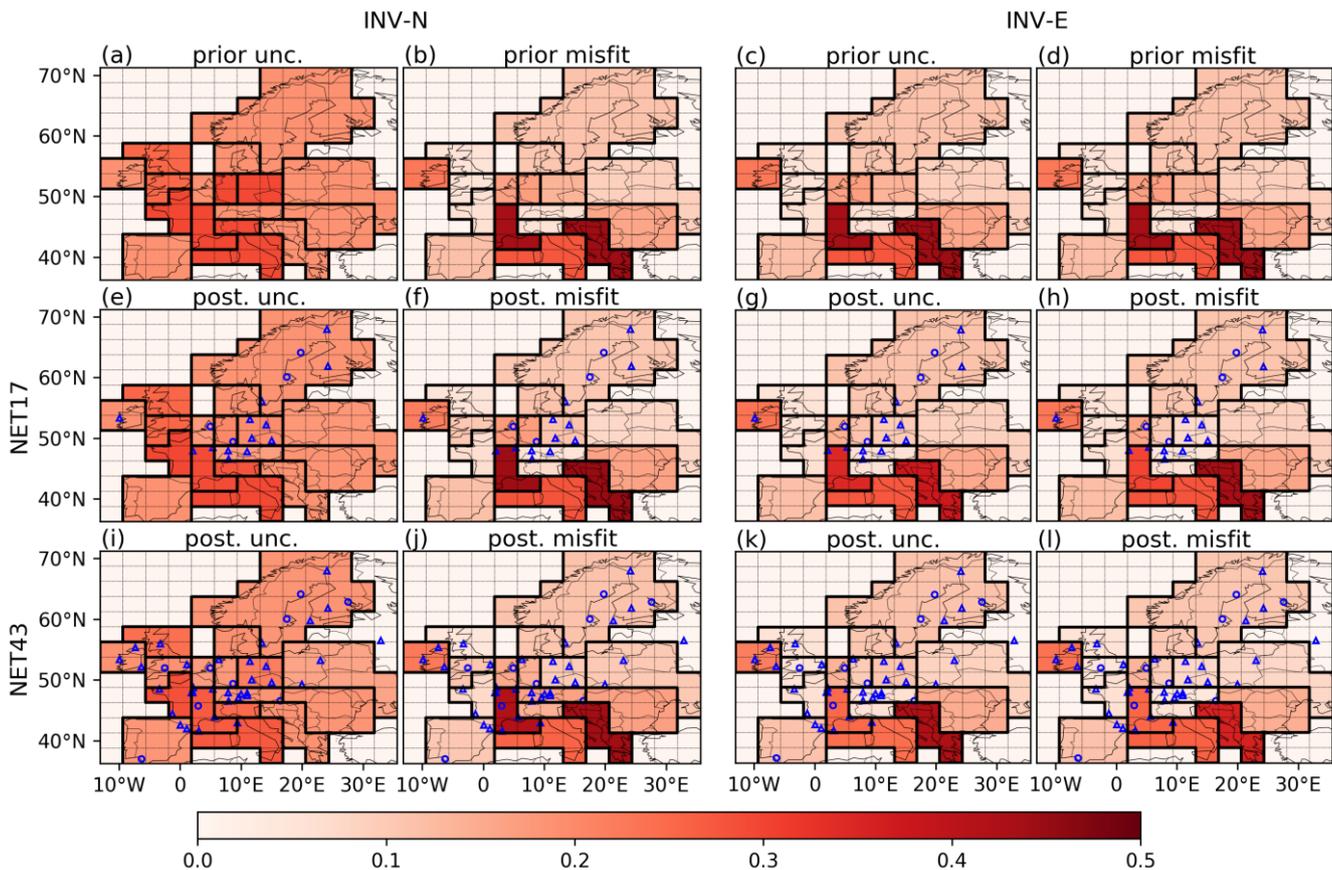


Figure 4: Average monthly uncertainty reductions and misfit reductions in FFCO₂ emissions over regions delineated by solid black lines, using the NET17 and NET43 networks and 2-week sampling for the inversions. The first and second columns are the results of INV-N inversions. The third and fourth columns are the results of INV-E inversions. The dashed lines show the grid cells of the transport model LMDZv4. The dots and triangles are 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. A value of UR and MR closer to unity means a better performance of an inversion to constrain FFCO₂ emissions in a region.

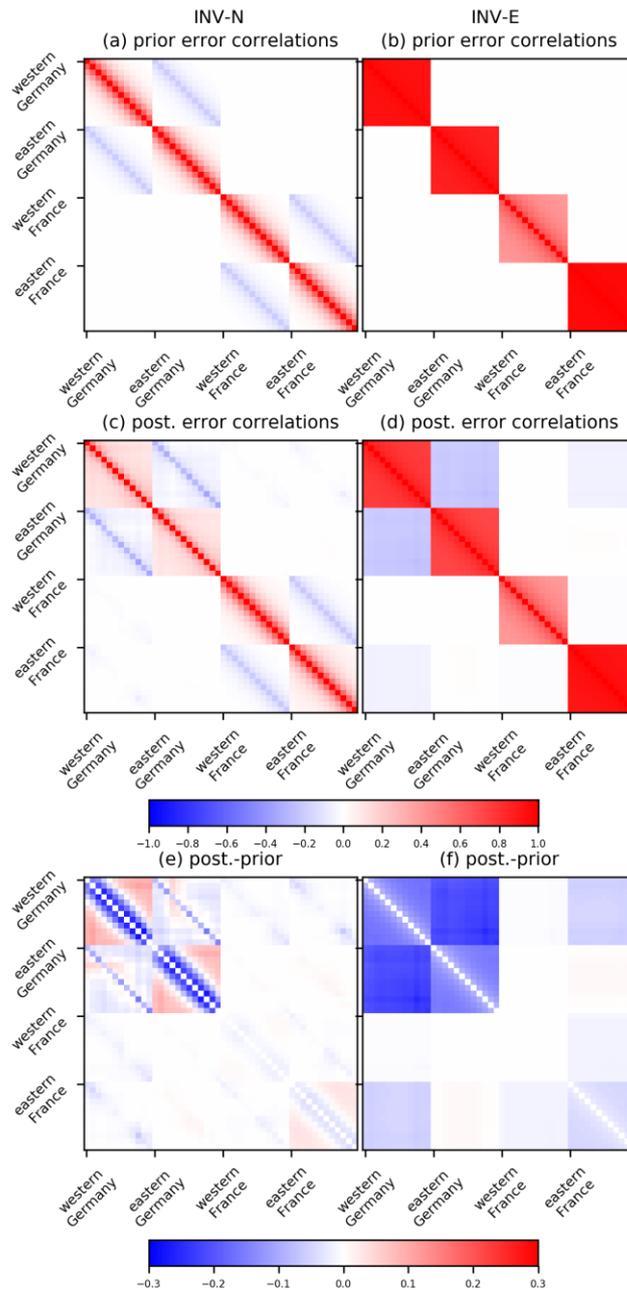
965



970

Figure 5: Average monthly relative prior and posterior uncertainties and misfits of FFCO₂ emissions over regions delineated by black lines, using the NET17 and NET43 networks and 2-week sampling for INV-N (first and second columns) and INV-E (third and fourth columns) inversions. First row shows the relative prior uncertainties and misfits. The second row shows the posterior uncertainties and misfits after assimilating 2-week mean afternoon observations from network NET17. The third row shows the posterior uncertainties and misfits after assimilating 2-week mean afternoon observations from network NET43. The dashed lines show the grid cells of the transport model LMDZv4. The dots and triangles are 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.

975



980

Figure 6: The correlation structure in the prior (first row) and posterior (second row) uncertainties in monthly regional FFCO₂ emissions for the four Germany and France regions using the NET43 network and 2-week sampling for INV-N (first column) and INV-E (second column) inversions, as well as their differences (third row). The x and y axes cover all the control region-months iterating through region first and months second (the blocks of pixels in each matrix). For clarity, we group these correlations into four regions and organize them for each region according to month indices.

985

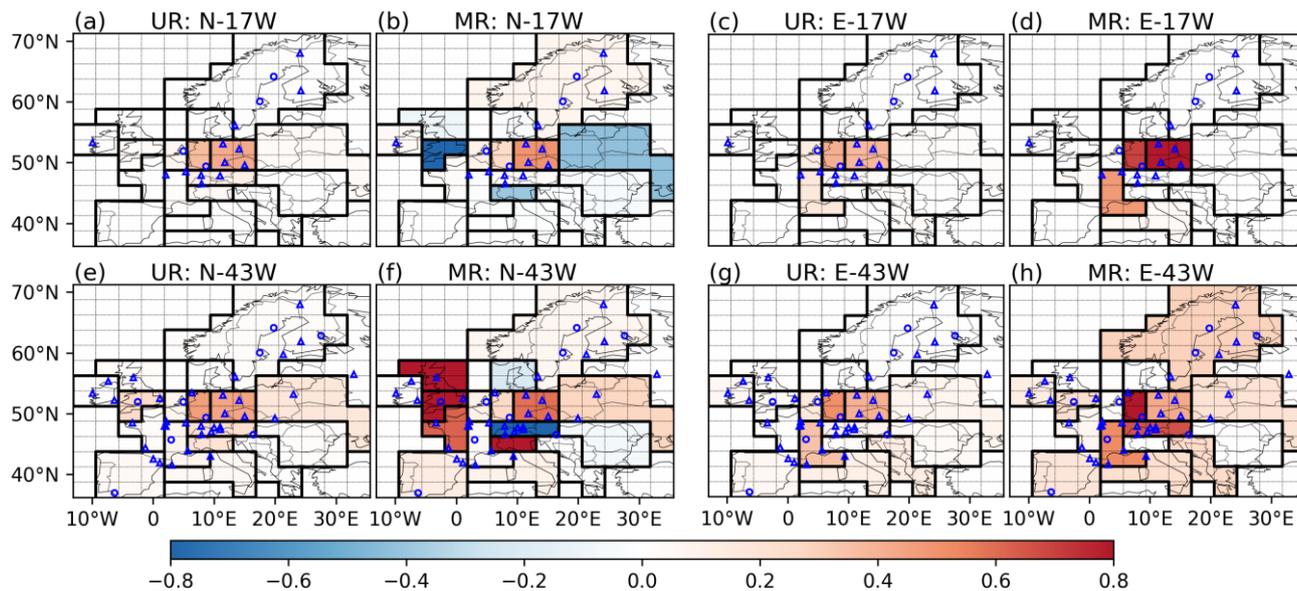
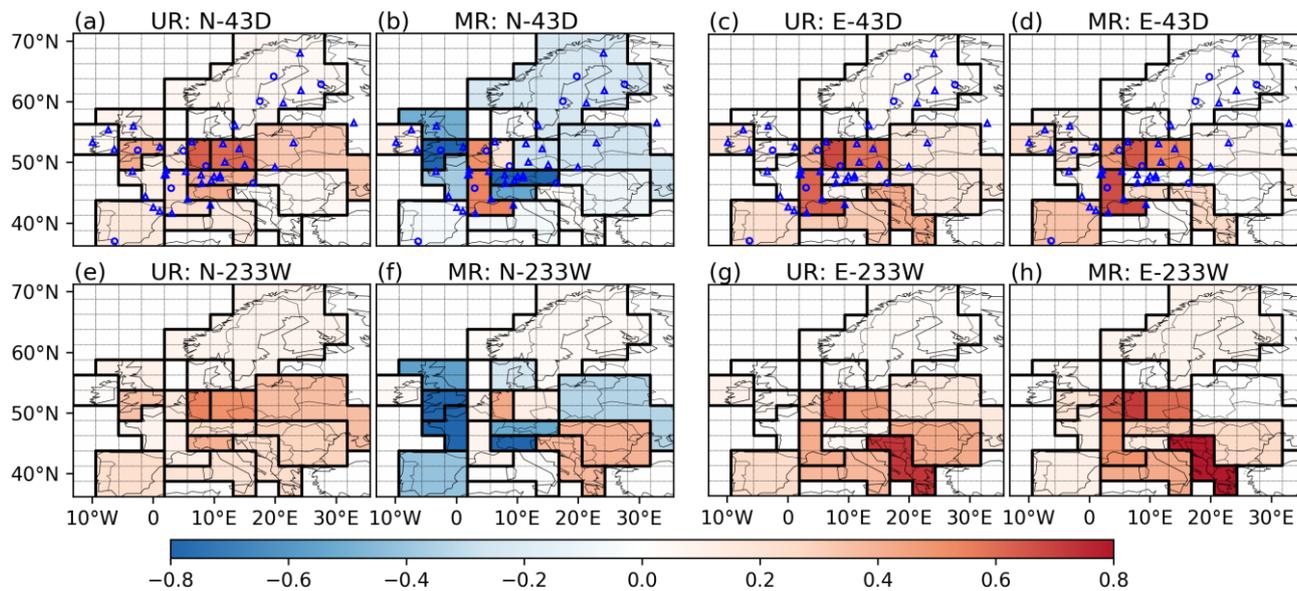


Figure 7: Uncertainty reduction (UR) and misfit reduction (MR) of annual FFCO₂ emissions over regions delineated by black lines using the NET17 and NET43 networks and 2-week sampling. The first and second columns show the results of INV-N inversions. The third and fourth columns show the results of INV-E inversions. The dashed lines show the grid cells of the transport model LMDZv4. 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. A value of UR and MR closer to unity means a better performance of an inversion to constrain FFCO₂ emissions in a region.

990



995

Figure 8: Average uncertainty and misfit reductions in the monthly FFCO₂ emissions over regions delineated by black lines using the NET43 network with daily sampling and NET233 network with 2-week sampling. The first and second columns are the results of INV-N inversions. The third and fourth columns are the results of INV-E inversions. The dashed lines show the grid cells of the transport model LMDZv4. The dots and triangles are 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. The locations of the sites in the OSSEs N-233W and E-233W are not plotted to avoid blurring the maps.

1000

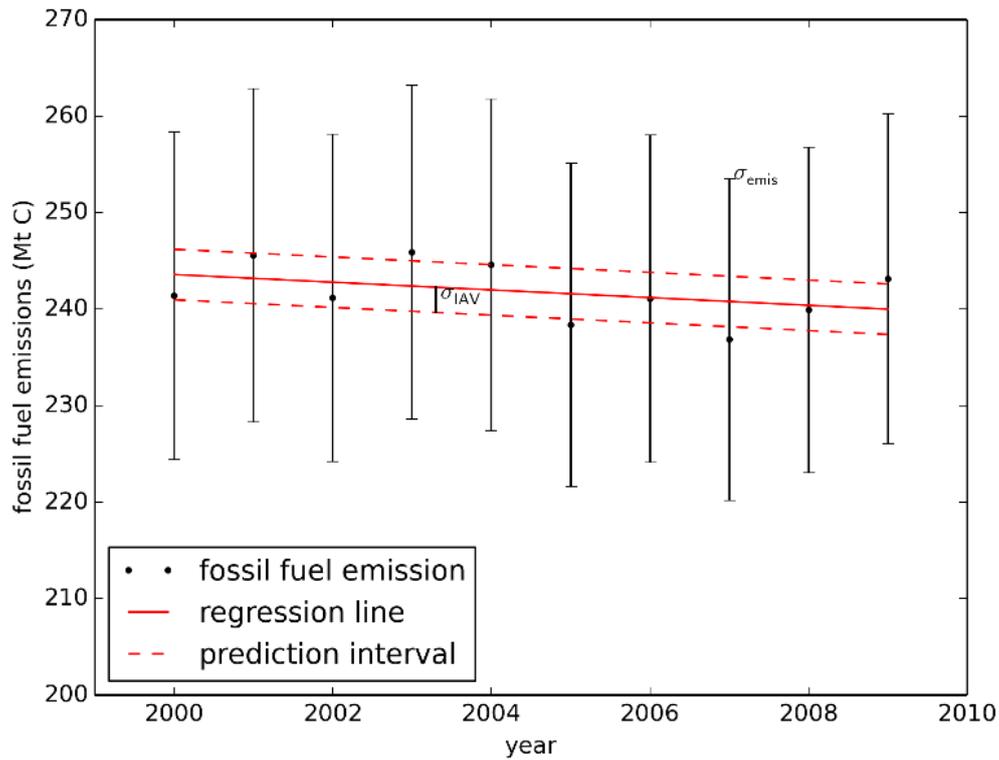


Figure C1: Annual FFCO₂ emissions from Germany in the period 2000-2009 calculated from IER-EDG.