The authors present an OSSE study of the capability of ICOS $^{14}CO_2$ 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 $^{14}CO_2$ 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 $^{14}CO_2$ 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 $\Delta^{14}C$. 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 $^{14}CO_2$ budget is beyond the scope of this study. It would be necessary to study the impact of uncertainties in other $^{14}CO_2$ fluxes on the $^{14}CO_2$ gradients, to precisely quantify the corresponding errors when converting the gradients of atmospheric $^{14}CO_2$ measurements into $FFCO_2$ 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 °×1.8 ° resolution, which is slightly higher than LMDZv4 used in our study, to simulate their continental-scale influences on $\Delta^{14}C$. Their results showed that nuclear enrichment may cause an impact of -0.9 [-0.6, -1.4] ppm in $FFCO_2$ for Orleans, France (48.8 °N, 2.5 °E) and an impact of -0.7 [-0.4, -1.3] ppm for Heidelberg if nuclear ^{14}C 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 $FFCO_2$ signals caused by ignoring other ^{14}C fluxes including cosmogenic production, ^{14}C disequilibrium between atmosphere and biosphere and between atmosphere and ocean, and ^{14}C source from nuclear power plant (the estimate of ^{14}C 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 $^{14}CO_2$ sources in translating atmospheric measurements of $^{14}CO_2$ into $FFCO_2$ is mainly from terrestrial biosphere, whereas the contributions from ocean CO_2 exchange and cosmogenic production of

^{14}C 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 ^{14}C 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 $^{14}CO_2$ other than the dilution of ^{14}C in CO_2 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 $^{14}CO_2$ fluxes may cause some impact on the interpretation of $FFCO_2$ using atmospheric $^{14}CO_2$ 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 $^{14}CO_2$ fluxes other than the dilution of ^{14}C in CO_2 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 $^{14}CO_2$ fluxes. But this study aims at understanding how the inversion system behaves when dealing with uncertainties in the $FFCO_2$ 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 $^{14}CO_2$ fluxes will be needed to refine those numbers.

We will better explain the assumptions underlying the conversion of $^{14}CO_2$ into $FFCO_2$ 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). We will add some discussions in Sect. 4.2 to recall the fact that the uncertainties associated with the signals from $^{14}CO_2$ fluxes other than the dilution of ^{14}C in CO_2 by fossil fuel emissions are expected to be relatively small compared to the uncertainties in the atmospheric $^{14}CO_2$ caused by fossil fuel emissions and by other types of observation errors. We will also add some discussions in Sect. 4.2 to emphasize that this study aims at providing some understanding of the system behavior rather than a precise quantification of UR that would apply when conducting real data applications. Lastly, we will add in the manuscript that future studies will be needed to quantify the impact from uncertainties in the nuclear plant emissions, cosmogenic production, biogenic fluxes, etc., on the inversion of $FFCO_2$ emissions.

References:

Graven, H. D. and Gruber, N.: Continental-scale enrichment of atmospheric $^{14}CO_2$ 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 $^{14}CO_2$ 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 $^{14}CO_2$, 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 CO2 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 will revise the sentence in lines 160 and 249 to explicitly explain that the "observations" $\mathbf{y}_o$ are $FFCO_2$ gradients to JFJ.

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 $\mathbf{x}^b_{i,m}$-$\mathbf{x}^t_{i,m}$ is denoted $\boldsymbol{\varepsilon}^b_{i,m}$ and the posterior misfit $x^a_{i,m}$-$x^t_{i,m}$ is denoted $\boldsymbol{\varepsilon}^a_{i,m}$. We compute a misfit reduction for each region-month emission budget as the relative difference between the prior and posterior misfits:

$MR_{i,m}=1-\boldsymbol{\varepsilon}^a_{i,m}/\boldsymbol{\varepsilon}^b_{i,m}$

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.

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

In all cases MR values could theoretically range between –infinity 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 will revise the sentences in line 209-213 and add two equations for the computation of UR and MR at monthly 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 °×2.5 °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 will clarify the rational for the NET233 network in Sect. 2.1.

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 CO_2, and the impact of transport errors.

Response:

The present protocol for almost all of the ICOS $^{14}CO_2$ sites is to fill continuously two-week samples. And some sites fill continuously daily/weekly atmospheric $^{14}CO_2$. 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 $FFCO_2$, 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 will revise the sentence in lines 233-235 to mention 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 CO_2 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 CO_2: determining fossil fuel CO_2 over Europe using Jungfraujoch 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 $\Delta^{14}CO_2$ will translate into two week average FF CO2? What the two week average $\Delta^{14}CO_2$ represents depends on the method of collection; an open tray will fix CO_2 proportional to the partial pressure of CO_2, while a bubbled trap will fix all the CO_2 in the ingested air. The former represents average FF CO_2 weighted by the total CO_2 mole fraction, while the latter represents average FF CO_2 over two weeks. Which one applies for the ICOS protocol?

<u>Response:</u>

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 m^3 of air during the two week period. CO_2 is collected by chemical absorption in CO_2-free NaOH solution in the so-called Raschig-tube samplers. In this context, the $FFCO_2$ represents the average $FFCO_2$ over two weeks (the second way that the reviewer mentioned).

We will clarify the fact that our definition of $FFCO_2$ corresponds to the average afternoon $FFCO_2$ over the course of two weeks here.

References:

Levin, I., Munnich, K. O., Weiss, W.: The effect of anthropogenic CO_2 and ^{14}C 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 $FFCO_2$ 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-CO_2 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 $\mathbf{H}_{distr}^{PKU}$. The mismatch between the temporal profiles of synthetic true emission and $\mathbf{H}_{distr}^{PKU}$ 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 $FFCO_2$ 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.

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

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 $\mathbf{H}^{prac}= \mathbf{H}_{samp}^{coloc}\ \mathbf{H}_{transp}^{LMDZ}\ \mathbf{H}_{distr}^{PKU}$ and we also wrote in line 364 "The synthetic observations are generated using $\mathbf{x}^{IER\text{-}EDG}$ and the operator $\mathbf{H}^{OSSE}=\mathbf{H}_{samp}^{coloc}\mathbf{H}_{transp}^{LMDZ}\mathbf{H}_{distr}^{IER\text{-}EDG}$, which relies on the same $\mathbf{H}_{samp}^{coloc}$ and $\mathbf{H}_{transp}$ 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 $\mathbf{H}_{distr}^{PKU}$ and $\mathbf{H}_{distr}^{IER\text{-}EDG}$, respectively. We will clarify such a use of the "practical" term to name $\mathbf{H}^{prac}$ in Sect. 2.2.2.

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 will revise the sentence in lines 295-296 to emphasize the fact that it is applied to all regions over the globe, not only to regions in Europe.

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 $FFCO_2$ 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 will add a discussion about the temporal auto-correlation of the transport error in Sect. 2.2.2, when introducing the configuration of transport error.

References:

Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T. and Ciais, P.: A European summertime CO_2 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., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguí, 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.: CO_2 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 CO_2 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 CO_2 measurement network for estimating the biogenic CO_2 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 CO_2 sources and sinks using ensemble model simulations, Biogeosciences, 6(6), 1089–1102, 2009.

Locatelli, R., Bousquet, P., Hourdin, F., Saunois, M., Cozic, A., Couvreux, 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 CO_2 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 CO_2 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 CO_2 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 CO_2 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 CO_2 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 **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 will revise the paragraphs in Sect. 2.2.1 to clarify the use of one realization of observations.

References:

Basu, S., Miller, J. B. and Lehman, S.: Separation of biospheric and fossil fuel fluxes of CO_2 by atmospheric inversion of CO2 and $^{14}CO_2$ 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 will 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.

11. Line 381: I think the reference to Figure 4(d) should actually be to 4(e). Likewise, in line 384, the refences 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 $FFCO_2$ 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 $FFCO_2$ 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 will rewrite Sect. 4.3, lines 541-555 to present the inconsistencies between UR and MR and discuss the possible causes of the negative scores of MR.

References:

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

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 $\mathbf{AH}^T\mathbf{R}^{-1}(\mathbf{y}_o - \mathbf{Hx}^b)$, so that the covariance matrix for the corrections is $\mathbf{AH}^T\mathbf{R}^{-1}(\mathbf{HBH}^T+\mathbf{R})\mathbf{R}^{-1}\mathbf{HA}$, 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 will emphasize this by updating the discussions.

15. Line 484: In real trend detection situations, the transport will vary year by year, as will the disequilibrium and nuclear fluxes of ^{14}C. 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 ^{14}C 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 **H** and the uncertainties in the transport modeling in **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 will add some discussions to admit that the changing transport may have an impact on the trend detection but quantifying this impact would require detailed studies and further investigation.

References:

Aulagnier, C., Rayner, P., Ciais, P., Vautard, R., Rivier, L. and Ramonet, M.: Is the recent build-up of atmospheric CO_2 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-Rémy, I. and Necki, J. N.: A recent build-up of atmospheric CO_2 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 will revise the sentence here to avoid the confusion.

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 $FFCO_2$ inversion in Europe is the atmospheric diffusion of the $FFCO_2$ 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 $FFCO_2$ 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 CO_2 fluxes assimilating total CO_2 concentrations which only target at natural CO_2 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 CO_2 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 $^{14}CO_2$ fluxes, the situation could be different and could pose problems for the boundary conditions of the regional scale systems. There could be large $^{14}CO_2$ fluxes from Atlantic Ocean and other adjacent continents like Asian Russia which can cause significant spatial patterns of $^{14}CO_2$ within Europe.

This will be better discussed in Sect. 4.2.

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

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

where R = 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 cov(**Y**) (in the review's equations, it's **R**) is not diagonal or if the diagonal terms (variances of the variables in ***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 cov(**Y**) is diagonal and the terms in ***y*** all have equal variance, that is cov(**Y**)=σ^2**I** where σ^2 is the variance for all variables and **I** is an identity matrix. The equivalence between our equations and those proposed by the reviewer can be proved:

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

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

In order to avoid any confusion, we will add the equations proposed by the reviewer.

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 will change Figure 2 by removing the lines country and state boundaries, as the reviewer suggested, and will also mention the fact we repeatedly use 12 colors for non-adjacent regions in the caption.