Lund, March 10th, 2024

Carlos Gómez-Ortiz
Department of Physical Geography and Ecosystem Science
Lund University
Sweden

Editor and Referees assigned to Research article EGUSPHERE-2023-2215.
Atmospheric Chemistry and Physics (ACP)

Dear Editor and Referees,

We thank both referees for their insightful comments and constructive suggestions. They have contributed to significantly improving the manuscript and its discussion. In the following, we address their comments point-by-point. We use text in italics to repeat the referees' comments, normal text for our response, and the marked-up text from the manuscript showing the changes applied.

RC1

RC1: The manuscript delves into the pivotal task of independently estimating and verifying regional and national fossil CO2 emissions, employing the Lund University Modular Inversion Algorithm (LUMIA) for assimilating in situ observations over Europe. The study's foundation lies in the assimilation of data from the Integrated Carbon Observation System (ICOS) network, a crucial aspect that warrants attention. However, the paper falls short in clearly articulating the novel contributions it brings to the existing body of inverse modeling studies. The reliance on the ICOS network is evident, but the connection to prior studies and the specific advancements provided by this investigation are not well-defined.

A critical aspect to address is the lack of clarity on why earlier studies might have failed or why they were conceptually, or due to data limitations, unable to address the question posed in the title. The reviewer suggests considering aspects such as bio and oceanic recycling of dispersed 14C, which could have been potential challenges or gaps in previous research. Providing insights into these aspects would enhance the reader's understanding of the study's significance in addressing potential limitations or gaps in existing literature.

We have extended the Introduction (L75-109 and 136-139 of the revised manuscript) and put the objective of our study in place with the existing body of inverse modeling studies focusing on the estimation of both fossil CO_2 emissions and terrestrial CO_2 fluxes using a multi-tracer (CO_2 and $\Delta^{14}CO_2$) approach. Clearly, the existing body of literature is very limited regarding such inversions on a continental scale, which is the specific objective of our study here. We do not claim that earlier studies have failed or could not address the problem of constraining fossil CO_2 emissions over Europe using $\Delta^{14}CO_2$ observations. In fact, to our knowledge, there is only one other existing study (Wang et al., 2018a) that addressed this specific problem. Our approach here is based on a very different modeling set-up than that of Wang et al. (2018) (e.g. transport model, resolution, $\Delta^{14}CO_2$ modeling approach), and hence contributes to the estimation of model uncertainty.

Besides, this is the first time LUMIA is used in a multi-tracer approach and this manuscript serves as a model description reference paper for future studies on some of the important questions raised by the referee such as the impact of the terrestrial disequilibrium on the inferred fluxes.

~~. Nevertheless.~~ For instance, Levin and Karstens (2007) present an observational approach to estimate hourly regional fossil fuel CO_2 offsets at a continental site (Heidelberg, Germany), using weekly mean $^{14}CO_2$-based fossil fuel CO_2 mixing ratios and CO observations. On a larger scale, Levin et al. (2008) examine monthly mean $^{14}CO_2$ observations from two German stations (Schauinsland and Heidelberg), compared against background measurements from Jungfraujoch, to assess the regional fossil fuel CO_2 surplus and emphasize the importance of high-precision radiocarbon measurements for quantifying fossil fuel CO_2 contributions at a regional scale in Europe. The study by Miller et al. (2012) explores the relationship between fossil fuel CO_2 emissions and enhancements in atmospheric concentrations of $^{14}CO_2$ and other anthropogenic trace gases. Utilizing a six-year dataset from vertical profiles in the northeast U.S., they separate the fossil and natural components of atmospheric CO_2 using apparent emission ratios of various gases to fossil fuel CO_2, offering observationally-based estimates of national emissions and comparing these with inventory-based estimates. Turnbull et al. (2015) use measurements of CO_2, $^{14}CO_2$, and CO from multiple sampling towers around Indianapolis, U.S., to differentiate fossil fuel CO_2 from background levels in an urban environment and evaluate the consistency of a bottom-up emission product. More recently, by using radiocarbon observations in CH_4 ($\Delta^{14}CH_4$) and CO_2 ($\Delta^{14}CO_2$) over London, Zazzeri et al. (2023) reveal that fossil fractions of CH_4 and atmospheric concentrations of fossil CO_2 are consistently higher than those predicted by simulations using emission products such as EDGAR. This discrepancy highlights the potential of $^{14}CO_2$ measurements to refine our understanding of fossil and biospheric CO_2 and CH_4 partitioning in urban settings, especially when the influence of nuclear power plants is minimal.

Nevertheless, large-scale four-dimensional inversion systems have only recently ~~included~~ begun to include $\Delta^{14}CO_2$ as an additional tracer to constrain fossil CO_2 emissions~~(Basu et al., 2016, 2020; Wang et al., 2018). Results from Observing System Simulation Experiments (OSSE)based on synthetic observations , assuming the current as well as an anticipated future network of~~ . Basu et al. (2016) introduced a novel dual-tracer atmospheric inversion technique that differentiates between biospheric and fossil fuel CO_2 fluxes using atmospheric CO_2 and $\Delta^{14}CO_2$ ~~measurement stations, have shown the high potential of constraining fossil~~ measurements over the U.S. This method not only allows for the estimation of monthly regional fossil fuel CO_2 fluxes but also addresses biases in biospheric flux estimates that occur when using traditional CO_2-only inversion methods with fixed fossil fuel flux assumptions. Their approach represents a significant advancement in quantifying regional and national fossil fuel emissions from atmospheric observations. Building upon this study, Basu et al. (2020) presented a more focused analysis in providing national and sub-national-scale estimates of fossil fuel CO_2 emissions, using an extensive observation database of both CO_2 and $\Delta^{14}CO_2$. Graven et al. (2018) conducted an in-depth analysis of fossil fuel CO_2 emissions in California, utilizing atmospheric observations from nine sites and employing the Weather Research and Forecasting model along with the Stochastic Time-Inverted Lagrangian Transport model (WRF-STILT). The research integrates measurements of CO_2 concentration and $\Delta^{14}CO_2$, uniquely combining these observations with high-resolution emission data from Vulcan v2.2 and EDGARv4.2, aiming to refine estimates of regional fossil fuel CO_2 emissions and explore the impact of various factors such as nuclear industry emissions and air-sea exchanges on atmospheric CO_2 levels. In Europe, Wang et al. (2018) evaluated the potential of a $\Delta^{14}CO_2$ observation network for estimating regional fossil fuel CO_2 emissions ~~over North America (Basu et al., 2016, 2020) and Europe (Wang et al., 2018). Having~~ through atmospheric inversions. They examined the effectiveness of different network configurations, from minimal to very dense setups, in reducing uncertainties in fossil CO_2 emissions estimation. The study used synthetic observations and the LMDZv4 global transport model, paying special attention to representation and aggregation errors. Establishing a network of both CO_2 and $\Delta^{14}CO_2$ measurement stations

Europe. For this purpose, we expanded the ~~LUMIA~~ Lund University Modular Inversion Algorithm (LUMIA) system (Monteil and Scholze, 2021) to perform simultaneous inversions of atmospheric CO_2 and $\Delta^{14}CO_2$, thus optimizing ~~the~~ fossil emissions, natural fluxes, and ~~the~~ isotopic disequilibrium. We perform ~~observing system simulation experiments~~ Observing System Simulation Experiments (OSSEs), recreating the current state of the ICOS network and its sampling strategy, and using different flux products (as priors and true values) to demonstrate the performance of the inversion scheme and show its capabilities. We begin by assessing the impact of oceanic fluxes on the total CO_2 and $\Delta^{14}CO_2$ concentrations. Then, we evaluate the impact of adding $\Delta^{14}CO_2$ observations on the estimation of fossil CO_2 emissions by comparing the model's ability to recover true fluxes starting from a prior flux set to zero. Finally, with a more realistic setup, i.e., prior, we evaluate the impact of the prescribed fossil CO_2 flux uncertainty and the impact of the terrestrial isotopic disequilibrium product.

Despite these concerns, the paper effectively demonstrates LUMIA's capabilities in well-sampled regions, showcasing its potential for accurate estimation of fossil CO2 budgets. The challenges faced in regions with low sampling coverage are acknowledged, shedding light on the limitations of the applied methodology in certain contexts.

Furthermore, the study underscores the importance of a reliable prior estimate of terrestrial isotopic disequilibrium, emphasizing the need to minimize uncertainties for robust posterior fossil CO2 flux estimates. This aspect adds valuable insights to the methodology used in estimating and verifying fossil CO2 emissions.

In summary, while the study contributes valuable information regarding fossil CO2 emissions, addressing the critique by explicitly stating the novel results in relation to prior studies, highlighting potential limitations, and discussing alternative explanations, particularly related to the bio and oceanic recycling of dispersed 14C, would significantly strengthen the paper.

As mentioned above we have extended the Introduction to put our study in context to previous studies and explained the novel aspect (LUMIA as a multi-tracer inversion system) of the manuscript. We also highlight the potential limitations and open questions of employing our system for estimating fossil CO_2 emissions. We believe this paper should serve as a model description reference and will address the open questions such as $\Delta^{14}CO_2$ sampling strategies, terrestrial disequilibrium, and $\Delta^{14}CO_2$ emissions from nuclear power plants in detail in a follow-up study. In addition to the adjustments made to the introduction in this regard, we updated the discussion to further comment on this regard.

Specific points:

Prior to publication, a number of points need to be addressed, both in terms of content and methodological description. In detail:

Help the reader to make the paper more self-consistent. References to previous LUMIA papers and others (e.g. Chatterjee and Michalak, 2013; Rayner et al., 2019; Scholze et al., 2017) should not be overused as a substitute for a more comprehensive description of the inversion model. See below for specific points.

RC1: L120: What is a shifted delta? This should be defined in a mathematically sound way.

We are not certain what exactly the referee means by 'shifted delta' because we do not use this terminology in the manuscript. We explained the meaning of capital delta (Δ) in lines 110 to 113 of the original manuscript. Nevertheless, we have rephrased the explanation of $\Delta^{14}CO_2$ and clarified the delta notation in L152-153 of the revised manuscript.

RC1: There is further confusion in the transition from the rather traditional formulation of the cost function (6) (line 172) to the use of a space-time covariance matrix B (see also sloppy use of simple or bolded notation of B). Some specific details:

We updated Equation 6 and the subsequent use of the matrix terms B, H, and R to straight bolded notation, consistent with other studies. The use of a space-time covariance matrix is completely standard in the research field, and our paper does not depart from the norm in that respect (see e.g. Broquet et al., (2011), Monteil & Scholze, (2021), Munassar et al., (2023))

$$J(x) = \frac{1}{2}(x - x_b)^T \mathbf{B}^{-1}(x - x_b) + \frac{1}{2}\left(\mathbf{H}x - y\right)^T \mathbf{R}^{-1}\left(\mathbf{H}x - y\right) \quad (6)$$

where ~~B~~ $\mathbf{B}$ is the prior uncertainty covariance matrix, and ~~R~~ $\mathbf{R}$ is the observational uncertainty covariance matrix, controlling the weight of each observation and target variable in the optimization. The iterative procedure searches for the value of x that ~~minimizes~~ minimize $J(x)$, i.e. the value of x for which the gradient ($\nabla_x J$) is equal to zero. The observation operator $H(x)$ can be expressed as the Jacobian matrix ~~Hx~~ $\mathbf{H}x$ that stores the sensitivity of each observation to each control vector element (Monteil and Scholze, 2021).

RC1: L192-195: Confusing description of eqn (7): Adding a matrix TTXTH to a vector Fc0? Please rephrase and give more explanatory details. As it stands, this is of little use. Define T and X more precisely!

There is indeed a small error: $\mathbf{F}_c$ with shape (n^t_{mod}, n^h_{mod}) is a matrix and not a vector. We corrected this in L236 of the revised manuscript. The remainder of the text is correct.

$$F\mathbf{F}_c = F\mathbf{F}^0_c + \mathbf{T}_T \mathbf{X}^c_x \mathbf{T}_H \quad (7)$$

where ~~F_c is the vector~~ $\mathbf{F}_c$ is the matrix containing gridded emissions for the category c, with prior value ~~F^0_c~~ $\mathbf{F}^0_c$. The matrix

RC1: L202: Why spatio-temporal for the diagonal matrix elements? This is just autocorrelation. Spatial correlations between two different locations are off-diagonal elements. Is this a block diagonal construction?

The sentence refers to the spatiotemporal pattern of the uncertainties (i.e. the spatiotemporal distribution of the variances), not of their correlations. The text seems clear enough to us (and to the second referee), so we chose not to modify it.

RC1: L203: Explain the temporal formulation together with the related segmentation of the vectors x / F (see also eq 6) for the cause of confusion, where x is used in a traditional space phase vector.

We believe that what the referee asks is exactly what is provided in L211-216 of the original manuscript. Lines 202-210 describe in a generic way the process used to achieve a target's overall uncertainty, independently of what they are based upon.

RC1: L207: Is the scaling controlled by some a posteriori technique? For systematic approaches see Talagrand, O. 1997 Assimilation of observations; an introduction. J. Meteorol. Soc. Japan, 75; Desroziers et al, 2005, QJRMS, Diagnosis of observation, background and analysis-error statistics in observation space.

No, the scaling is set (prescribed) by us. The target uncertainties are given in Table 2 (as indicated in line 223 of the original manuscript).

RC1: L214: This optimisation time step of one week should be explained in more detail above, along with the optimisation concept. Again, the implications for the construction of the state vector x are not sufficiently addressed to be understood.

The optimization time step (now replaced in the text by 'optimization interval' to not be confused with the model time step) refers to the temporal resolution of the flux adjustments: the state vector contains weekly offsets to the prior emissions, whereas the prior emissions are hourly: the emissions within an optimization interval will be adjusted by the same value. This is already described in Section 2.3.1 (L185-191 of the original manuscript). There is no implication on the construction of the state vector (rather, the opposite: the way the state vector is constructed determines how the **B** matrix should be constructed).

We would like to point out that the overall approach is very standard in the field, similar to what has been used in other studies, both with the same inversion system (e.g. Monteil et al. (2020), Monteil & Scholze (2021), Munassar et al. (2023)) and with other inversion systems (e.g. (Basu et al., 2016) (TM5), Broquet et al. (2011, 2013) (PYVAR-CHIMERE)), therefore, it does not seem appropriate to re-describe this in details. We describe what makes the specificity of our implementation of that approach, but we think that we do not need to explain the fundamentals of inversions here and refer to other studies for a more fundamental understanding.

RC1: Figure 3: Caption and subscript of the left panel show 1-hour integration backward in time. Is there evidence of hypersonic winds?

The caption indicating a 1-hour sampling integration for CO_2 does not imply the presence of hypersonic winds. Instead, the 1-hour integration time refers to the period over which atmospheric data are collected and synthesized to represent the CO_2 sample at a specific moment. The plume or back trajectory displayed in the maps by the black to orange colors represents the sensitivity of the sampled atmospheric CO_2 to the surface fluxes over the 14 days before the sampling time (or the starting sampling time in the case of $\Delta^{14}CO_2$). This methodology is standard in atmospheric inversions for capturing the influence of regional fluxes on a sampling site and does not suggest unusual wind speeds.

We updated Figure 3 and removed the text "Integration: 1h" from the left panel to not generate further confusion. We added the text "Continuous samples" to the left panel and "Integrated samples" to the right panel, and updated the caption.

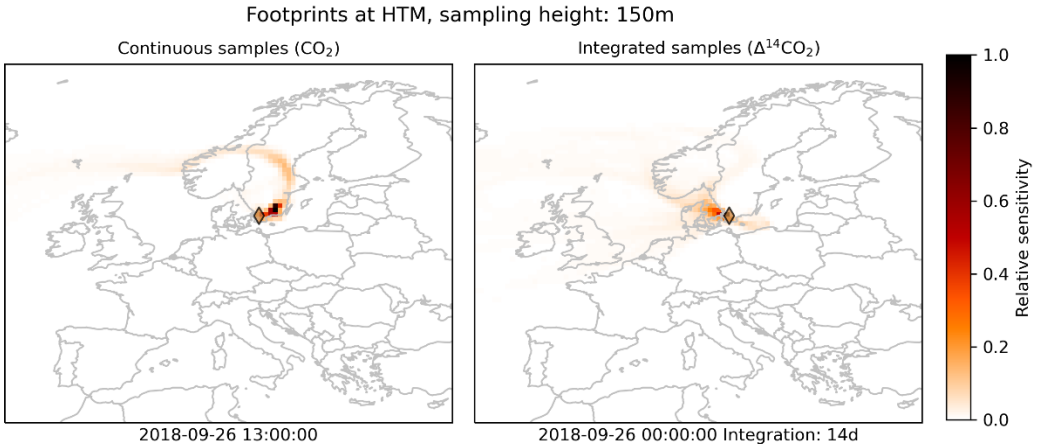

Figure 3. ~~Example~~ Examples of so-called (pre-calculated~~observations~~) footprints for CO_2 (left) and $\Delta^{14}CO_2$ (right) at the Hyltemossa ICOS station. ~~$\Delta^{14}CO_2$ (~~The maps show the sensitivity of the respective atmospheric tracer at the sampling site to the surface fluxes over the regional domain up to two weeks before the observation. The left panel displays the sensitivity of CO_2 at the indicated sampling time and shows influences by surface fluxes from the North Atlantic through Scandinavia, while the right ~~) has an~~ panel demonstrates the dispersed sensitivity of a 14-day integrated $\Delta^{14}CO_2$ sample across Northwestern Europe and the Baltic region. The two maps illustrate the distinct spatial integration ~~time~~ of ~~14 days~~the two tracers over time.

RC1: Lines 266-269: The statement is difficult to reconcile with the displayed extension in the graphics. Please clarify the meaning of 'm(?) 3'.

As explained before, the purpose of the statement and the figure is to show and describe the simulated sensitivity of the two types of samples used in the study for a specific sampling time: continuous for CO_2 and integrated for $\Delta^{14}CO_2$. 'm(?) 3' is a typo and was removed from the text (see L313 of the revised manuscript).

RC1: Lines 282-283: Do you conduct experiments with identical twins instead of OSSEs?

In our study, we performed OSSEs. To clarify this, we modified the text in L330-337 of the revised manuscript.

355 is sampling the free troposphere. ~~This data selection is not strictly necessary for this study since we assume a perfect~~ For our OSSEs, we use the same transport model (~~the same model is used~~ i.e. the pre-computed observation footprints from FLEXPART) to generate the synthetic observations and perform the inversions~~). However,~~ . Therefore, this data selection is not strictly necessary for this study, but we want to replicate the conditions of a real inversion. ~~We~~ Since we are using the same background concentration for the synthetic observations and the simulated prior and posterior observations (i.e. we are
360 assuming a perfect boundary condition), we simplify the calculation of it by computing a smoothed and detrended weekly (for CO_2) and monthly (for $\Delta^{14}CO_2$) average of the real observations (ICOS et al., 2023) for each sampling site. For sampling sites, for which there are for some reason no real observations for the year 2018 in the ICOS database (e.g. $\Delta^{14}CO_2$ measurements were not yet implemented or were not yet part of ICOS), we took the observations from the nearest year available to calculate the background.

OSSEs are the reference term in the field of study for this kind of sensitivity as shown in other studies such as Basu et al. (2016, 2020), Philip et al. (2019), Wang et al. (2018).

RC1: Please declare sources for the values estimated along external information in L300-301.

The text was modified in L357-365 of the revised manuscript to give more clarity.

385 flux uncertainties (See Section 2.3.1). Since our main purpose in this study is to demonstrate that our multi-tracer inversion system is capable of estimating both the fossil CO_2 emissions and natural CO_2 fluxes, we choose prior uncertainty values that are reasonable and consistent with other studies. The prior uncertainties are assigned as follows: ~~50% to 150% (0.1~~ for F_{ff}, we use the difference between the annual budgets for the whole study domain from ODIAC (1.26 PgC yr^{-1}~~to~~) (Oda and Maksyutov, 2020) and from an emissions product based on EDGARv4.3 (1.47 PgC yr^{-1}) (Gerbig and Koch, 2021b)
390 as a reference to define its uncertainty (Basu et al., 2016). We use 150% (0.3 PgC yr^{-1}) of the difference ~~in the annual budget (0.21~~ as the base uncertainty for all the experiments, and we select two extra values to evaluate the impact of the prescribed uncertainty on the inversion: 50% of the difference (0.1 PgC yr^{-1}~~) of EDGAR and ODIAC for F_{ff} to evaluate its impact in the inversion, 10% (0.37 PgC yr^{-1}) of the annual negative values for F_{bio}~~ and the exact difference of 0.21 PgC yr^{-1} ~~(where production is higher than respiration) 100%~~). For F_{bio} we choose the 25% (0.37 PgC yr^{-1}) of the monthly prior
395 (Monteil and Scholze, 2021), and 30% (0.22 PgC yr^{-1}) of the annual budget for F_{biodis} (Basu et al., 2020). We optimize all the

The same methodology was implemented by Basu et al. (2016), in which they defined the uncertainty for the fossil fuel fluxes as the spread among a series of emission inventories and products in the U.S. (CarbonTracker, VULCAN, and ODIAC). However, we need to point out that, while this is a reasonable value, we intend to demonstrate that our system works and can perform multi-tracer inversions using observations of CO_2 and $\Delta^{14}CO_2$.

RC1: In L305-309, are there errors of representativity?

In our inversion system, the observation error represents both the instrumental error and the model representation error. In the case of the CO_2 observations, the observation error is mainly composed of the error of representativity, since the instrumental error is very small (in the order of 0.1 ppm), and by calculating the moving standard deviation in a weekly window we calculate an error proportional to how rapidly CO_2 varies (for background sites it will be small while for polluted sites it will be larger). For $\Delta^{14}CO_2$, the opposite is true. The instrumental error is larger than the error of representativity, therefore, we pick a value of 0.8 ppm in $C\Delta^{14}C$ units (1.91±0.05‰ $\Delta^{14}CO_2$).

Note: There was confusion generated by the value used as the observation error of $\Delta^{14}CO_2$ and the unit conversion from ppm to ‰ that was pointed out by RC2 and was included in the manuscript.

We modify the text in L367-375 of the revised manuscript to clarify this.

(500 points). To set up the observation ~~uncertainty,~~ error, which includes the instrumental and the representativity errors, we use different methods for the CO_2 and the $\Delta^{14}CO_2$. For CO_2, where the error of representativity is usually larger than the instrumental error, we apply a weekly moving standard deviation to each observation i.e. the prior ~~uncertainty~~ error of each observation is equal to the standard deviation of the observations in a time window of ± 3.5 days around that observation. ~~The prior uncertainty for the~~ In this way, we account for the changes in the CO_2 ~~observations ranges from 0.91 to 215.5 ppm~~concentrations according to the local site conditions. For instance, at a background station such as Jungfraujoch (JFJ) on the top of the Swiss Alpes, the observation error ranges from 0.9 to 29.2ppm (mean value of 9.3 ± 4.0ppm), while at polluted sites such as Saclay (SAC) just outside Paris the CO_2 concentrations change rapidly and the error ranges from 5.9 to 215.5ppm (mean value of 55.8 ± 40.7ppm). For $\Delta^{14}CO_2$ ~~,~~on the other hand, the instrumental error is larger than the representativity error, we use a constant value of ~~1.5‰ for~~ 0.8ppm in $C\Delta^{14}C$ units or 1.91 ± 0.05‰ in $\Delta^{14}CO_2$ units, calculated using Equation 8.

RC1: The caption in Figure 4 is of little value, please add a significantly extended description to the 8 panels.

We extended the description and updated the figure to enhance its value.

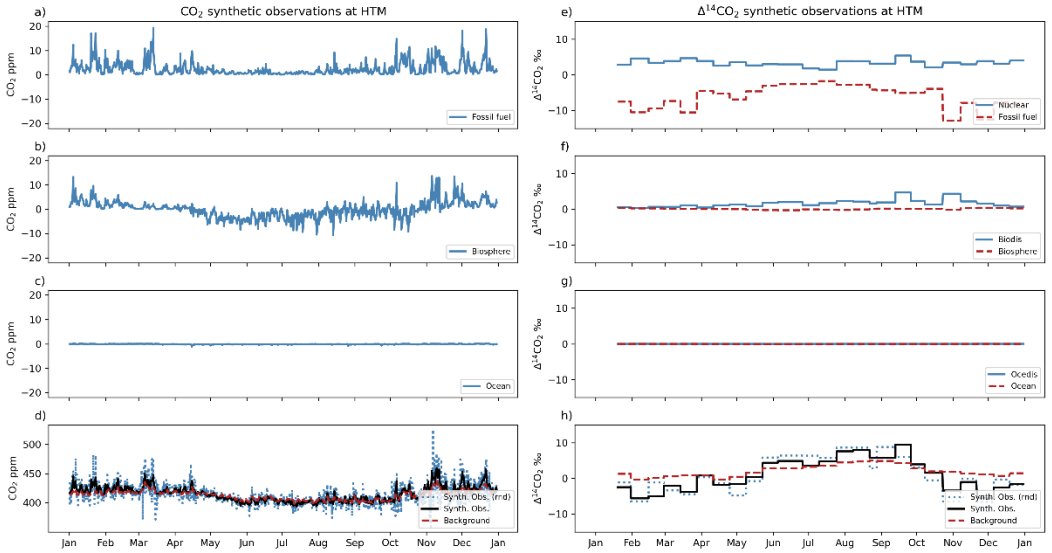

Figure 4. Synthetic ~~observation time-series for Hyltemossa (~~observations of CO_2 and $\Delta^{14}CO_2$ at the HTM ~~)~~station over a one-year period. Panels a) to d) display CO_2 concentration variations due to different sources: a) fossil fuel, b) biosphere, c) ocean, and d) combined synthetic observations with random perturbation (blue dotted line) against the background concentrations (red dashed line). Panels e) to h) illustrate $\Delta^{14}CO_2$ variations: e) nuclear and fossil fuel, f) biospheric disequilibrium and biosphere, g) ocean disequilibrium and ocean, and h) total synthetic observations with random perturbation (blue dotted line) compared to the background (red dashed line). The blue solid and dashed lines represent the synthetic observations without and with random noise added, respectively.

RC1: In Section 4.2.1: The subsection could benefit from additional analyses on the air mass pathways for other stations, following the scheme provided in Fig. 3 for a single station. The following footprints aim to explain the reasons for the varying performances across the individual regional domains that were analyzed.

Figure 3 is a "snapshot" of the sensitivity for one specific hourly CO_2 and integrated $\Delta^{14}CO_2$ sample at one sampling station. Figure 2 is a more statistical representation of the overall sensitivity in the study. In regions where there is a high sensitivity (as shown in Figure 2 for Western/Central Europe where most yellow clusters are located), there is a better constraint of

the emissions as shown in Figure 6. An expanded analysis, as suggested, would require evaluating individual timesteps for each station to construct a detailed understanding of the air mass pathways. This would involve a substantial increase in computational resources and time, as each station's footprints need to be analyzed for each timestep to ascertain the variances in air mass influences. Moreover, the spatial and temporal resolution of the data, along with the model's inherent limitations in resolving complex atmospheric processes, may introduce additional uncertainties. It is not possible, and to our understanding also not needed, to provide the footprint for every observation and every site.

RC1: In line 382, please define the term 'pixel-level'.

We replaced the term “pixel-level” with “grid cell level” (L441) which is the correct term.

RC1: Additionally, in line 386, please specify which samples were used to obtain the statistics. It would be clearer to state 'average values of...' and 'standard deviation' with respect to the underlying samples.

We updated Equation 9 to make it consistent with the metrics used in the previous subsection (4.2.1 Retrieval of the monthly and regional time series).

We additionally updated Figure 7 to show the individual maps of the posterior RMSE for both experiments and both categories and updated the text to include the changes in the equation and the figure.

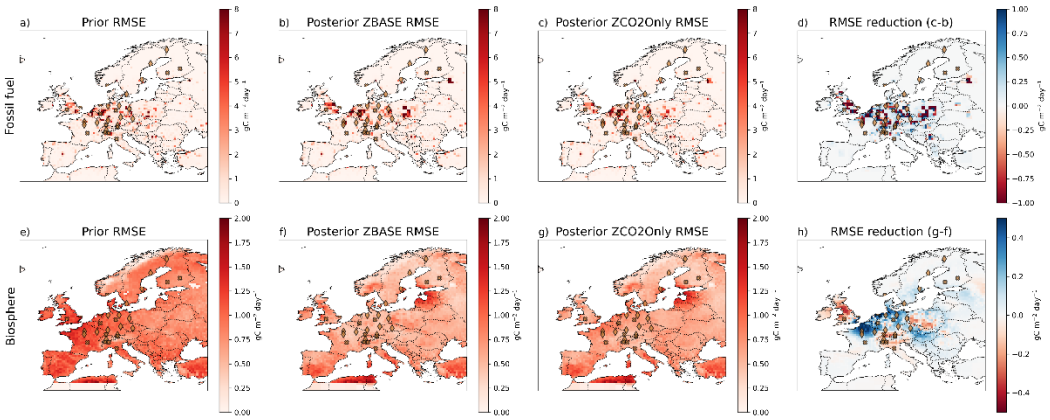

Figure 7. Spatial error of fossil CO_2 (a to d) and biosphere (e to h) for the ZBASE and ZCO2Only experiments. a) and ~~e~~c) show the prior RMSE for F_{ff} and F_{bio}, respectively, ~~and~~ b) and f) show the posterior RMSE for ZBASE, c) and g) show the posterior RMSE for ZCO2Only, and d) and h) show the ~~relative~~ RMSE reduction (see Equation 9) for fossil and biosphere. In Figures ~~b~~d) and ~~d~~h), positive values (in blue) show the pixels where ZBASE performs better than ZCO2Only (i.e. adding ~~$\Delta^{14}C$~~ $\Delta^{14}CO_2$ observations improves the posterior estimates), and negative values (in red) where ZCO2Only performs better than ZBASE. Crosses and diamonds represent stations that only measure CO_2 and those that additionally measure ~~$\Delta^{14}C$~~$\Delta^{14}CO_2$, respectively.

RC1: Lastly, please revise line 400. In the northern part of

The sentence was changed to “[…] the northern part of Western/Central Europe, Denmark, and southern Sweden, as well as some areas in Eastern Europe.” (see L456 of the revised manuscript).

RC1: L475: the conclusion drawn that the inversion process has effectively adjusted the model outputs, bringing them closer to the true observations, is scientifically poor. It is recommended to provide more quantitative evaluation to support this claim.

The whole section was completely modified following this comment and comments from RC2. We included the analysis of a polluted station (Saclay), along with JFJ (background), and added Figures 13 and 14, and Table 5 showing the performance metrics.

4.3.3 The observational space

Finally, we analyze the model's performance in the observational space ~~, which is crucial for evaluating its effectiveness. Figure 12compares the prior and posterior concentrations from the BASE0.1 experimentwith the corresponding synthetic observations for all sampling stations and , as a representative example, for the Jungfraujoch (JFJ)station. Examining the correlation coefficients, we find that the prior concentrations already correlate significantly with the synthetic observations . The correlation coefficients for the prior estimates are 0.61 for JFJ~~ at all sampling stations aggregated together, one polluted station (Saclay, SAC) (Figure 12) and one background station (Jungfraujoch, JFJ) (Figures 12 and 13) for the BASE experiment. We calculate two performance metrics: the correlation coefficient between the synthetic observations and the prior and posterior simulated concentrations for all the sites and individually for the two sites selected, and the χ^2 for the overall simulation (Table 5). The histograms in Figure 12 show the mismatches between the synthetic observations and the prior and posterior concentrations. For the CO_2 concentrations at all sites (Figure 12~~b)and 0.92 (Figure 12~~a)~~for all stations, indicating a reasonable correlation with the synthetic observations~~, the histogram shows a distribution centered around zero for both prior and posterior mismatches with a standard deviation of 14.2 and 13.4, respectively (see Table 5), indicating systematic deviations from the

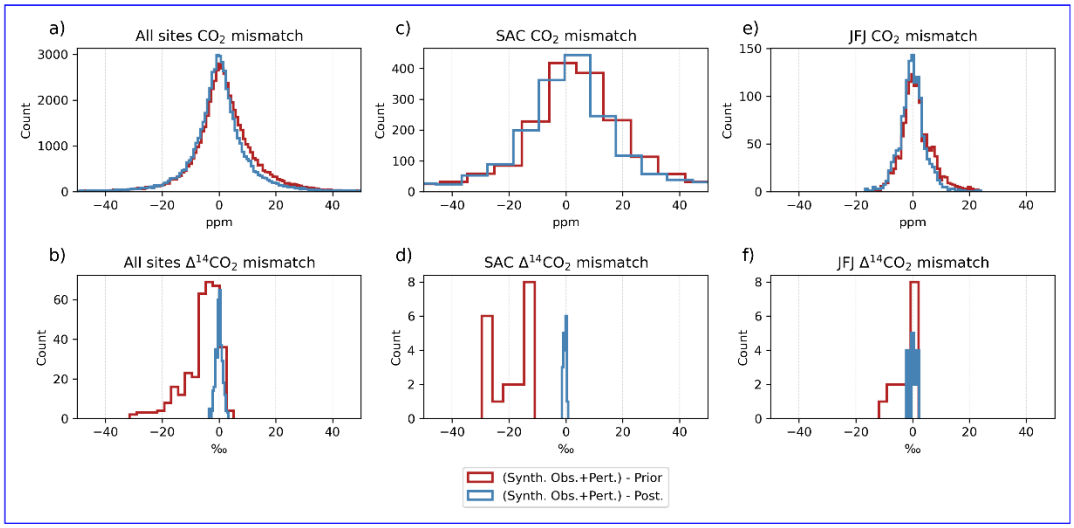

Figure 12. Mismatches between the synthetic observations and the prior (red) and posterior (blue) concentrations for all the sampling stations, Saclay (SAC) and Jungfraujoch (JFJ) for CO_2 (a, c and e) and for $\Delta^{14}CO_2$ (b, d and f). All prior and posterior concentrations correspond to the BASE experiment.

observed values. The posterior mismatch has a slightly tighter distribution, suggesting a small improvement in the model after adjustments as reflected in the correlation coefficient (Table 5). At Saclay (Figure 12c), the mismatch distribution is wider than the aggregate of all sites, which could suggest greater variability or larger errors at this particular site. The posterior adjustment has not significantly tightened the distribution, indicating that the model adjustments did not perform as well at this site as they did on average across all sites. On the other hand, the ~~posterior concentrations exhibit a slight enhancement in the correlation coefficients compared to the prior concentrations: 0.71 for JFJ (Figure 12b) and 0.98 (Figure 12e)for all stations, indicating a refinement in the model's ability to reproduce the synthetic observations accurately. This improvement in correlation coefficients is also reflected in the mismatch plots. For example, the mismatch between the posterior concentrations and the synthetic observations of $\Delta^{14}C$ at all stations~~ distribution in Jungfraujoch (Figure 12e) is much tighter than in all sites and SAC, with the posterior mismatch displaying a slight improvement in precision as evidenced by the narrower spread. However, when comparing the posterior time series with the synthetic observations before adding the random perturbation (Figures 13a and 13c), there is a better agreement between them than with the prior values, especially during periods of higher variability (April to July at SAC, and April to September at JFJ).

The $\Delta^{14}CO_2$ synthetic observations are in general better fitted by the posterior than CO_2 at all sites, SAC and JFJ (Table 5). In all cases, the prior distribution is displaced to the negative values, indicating that the prior simulated values are in general

higher than the synthetic observations as shown for the whole period at SAC (Figure 12c) shows a narrower distribution around zero compared to the previous mismatch. This suggests that the inversion process has effectively adjusted the model outputs, bringing them closer to the true observations. Analyzing the concentration time series at the JFJ station d) and from July to November at JFJ (Figure 12f). These larger prior concentrations are mainly caused by the prior terrestrial disequilibrium flux from July to November, and by the nuclear production fluxes throughout the year, which is significantly larger at Saclay (Figure 14). However, the posterior mismatches showed a much narrower spread around zero at all sites (Figure 12e and b), Saclay (Figure 12d), and Jungfraujoch (Figure 12f) , we observe that that is evident in the time series at both sites where the posterior closely follows the posterior concentrations agree better with the synthetic observationsthan the prior concentrations. This improvement is particularly notable for $\Delta^{14}C$, indicating that the inversion has successfully captured the dynamics of this tracer. Lastly, we consider the , and supported by the correlation coefficients (Table 5).

The reported χ^2 values of 1.77 for the prior and posterior concentrations across all sampling stations and observations. The prior 1.06 for the posterior across all sites and samples suggest a substantial improvement in the model's performance in adjusting the prior concentrations to the synthetic observations. A χ^2 value is 1.52, indicating some discrepancy of 1.77 for the prior indicates that there were significant discrepancies between the prior concentrations and the synthetic observations. However, the posterior This is consistent with the broader spread of mismatches in the histograms for both SAC and JFJ sites, as well as the apparent overestimation of $\Delta^{14}CO_2$ content in the time series. The improvement to a χ^2 value improves to 1.00, indicating a closer match between the posterior concentrations and the observed data. These results confirm that the inversion process has effectively improved the model's performance in the observational space. of 1.06 for the posterior indicates a better fit to the synthetic observations that are likely to be reflective of the underlying data patterns while still maintaining some degree of generalizability without overfitting the data.

Table 5. Performance metrics (correlation coefficient R, standard deviation and χ^2) for all sites, Saclay (SAC), and Jungfraujoch (JFJ).

		Prior		Posterior	
		R	σ	R	σ
All sites	CO_2	0.64	14.2	0.68	13.4
	$\Delta^{14}CO_2$	0.72	6.4	0.99	1.2
SAC	CO_2	0.56	31.9	0.59	31.1
	$\Delta^{14}CO_2$	0.63	6.8	0.99	0.5
JFJ	CO_2	0.65	5.5	0.74	4.5
	$\Delta^{14}CO_2$	0.75	4.2	0.84	1.5
χ^2		1.77		1.06	

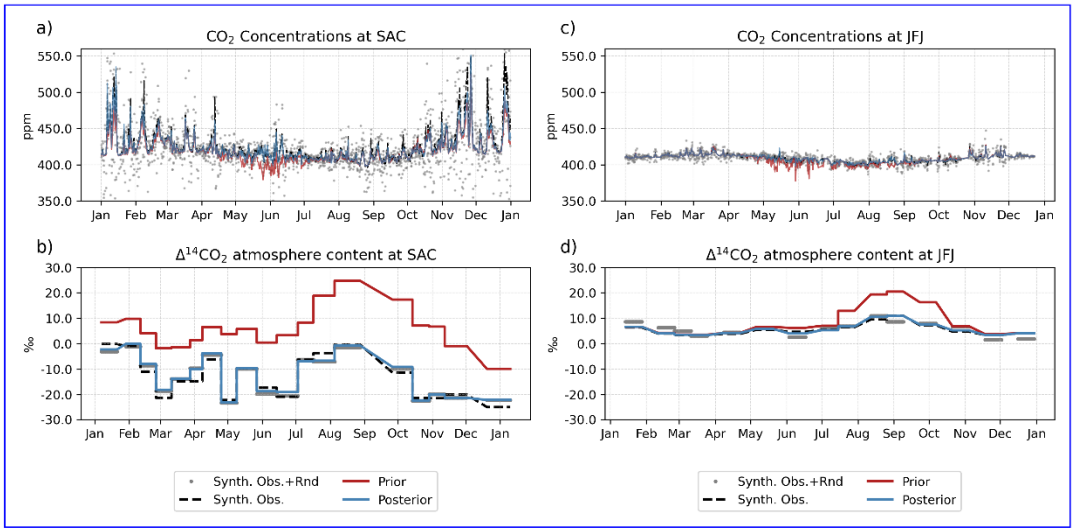

Figure 13. ~~Mismatches between the synthetic observations and the prior (red) and posterior (blue) concentrations for all the sampling stations and at Jungfraujoch ICOS station for~~ Concentration time series of CO_2 (a and ~~b~~c) and ~~for $\Delta^{14}C$~~ $\Delta^{14}CO_2$ (~~c~~b and d) ~~. The right panel shows the time series of CO_2~~ at Saclay (~~e~~SAC) and ~~$\Delta^{14}C$~~ Jungfraujoch (~~f~~JFJ)~~at Jungfraujoch~~, respectively. All prior and posterior concentrations correspond to the ~~BASE0.1~~ BASE experiment.

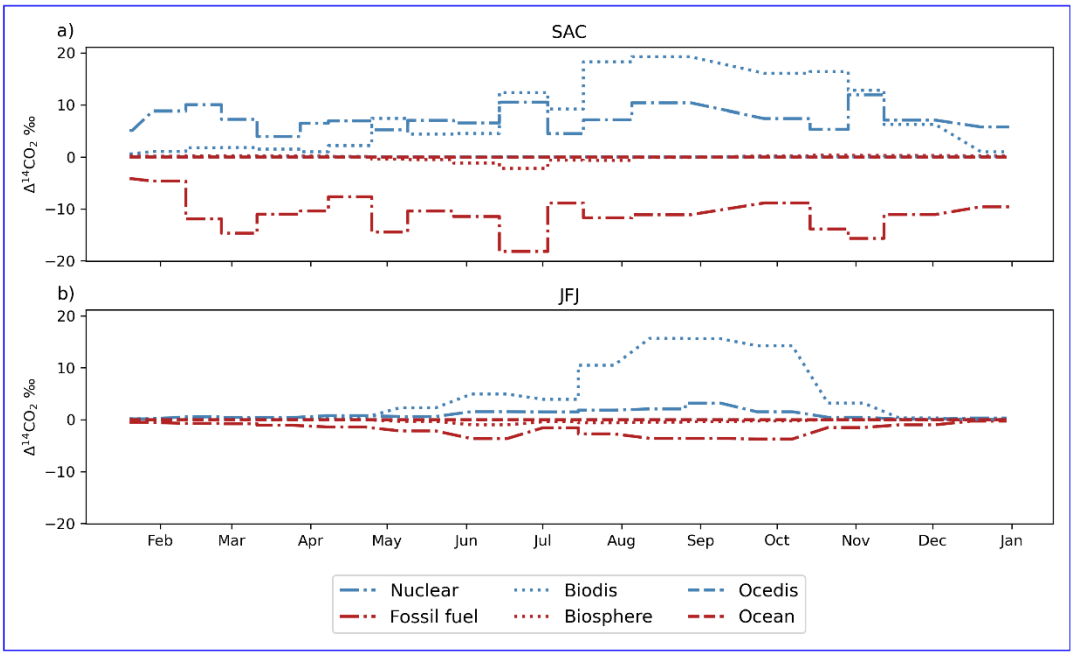

Figure 14. Contribution of each category to the prior $\Delta^{14}CO_2$ simulated concentrations at Saclay (a) and Jungfraujoch (b).

RC1: In line 480, the text asks how the Chi2 validation is performed by designing B. It is unclear what is meant by 'designing B'. The text could be improved by providing more context and explanation for the question being asked.

We do not understand the referee's comment. The term 'designing B' does not appear in this section, nor the whole document.

RC1: The options for the validation method are Talagrand, O. 1997 Assimilation of observations and Desroziers et al, 2005, QJRMS, Diagnosis of observation, background and analysis-error statistics in observation space.

We find his suggestions out of the scope of this study.

RC1: "To improve numerical accuracy, it may be necessary to normalize the fluxes or implement regional scaling factors." A sound approach involves preconditioning techniques from minimization methods. Please provide your comments on this suggestion.

The preconditioning technique was implemented since the initial development of LUMIA (see Monteil & Scholze (2021)). We realized that this was not mentioned anywhere in the text and we included it in section "2.3.1 Construction of the control vector (*x*)" (see L240-242 of the revised manuscript).

> 0.5°) to each optimized time-step t_{opt} and cluster p_{mod}. To reduce the number of iterations and large matrix multiplications, the optimization is performed on a preconditioned control vector $\omega = \mathbf{B}^{1/2}(x - x_b)$. More information about the preconditioning can be found in Monteil and Scholze (2021).

This suggestion was completely removed from the discussion by the recommendation of the second referee.

RC2

The main question of the paper by Gomez-Ortiz et al. is already in the title: "Can ∆14CO2 observations assist atmospheric inversions in constraining Europe's fossil CO2 emission budget?" The authors employ the Lund University Modular Inversion Algorithm (LUMIA) to estimate fossil CO2 emissions and natural fluxes by simultaneously inverting in-situ observations of CO2 and ∆14CO2 across Europe. They evaluate the system's performance through a series of Observing System Simulation Experiments (OSSEs). Their main result is that in regions with dense observation networks, like Western/Central Europe, LUMIA, with the inclusion of ∆14CO2 observations, can reconstruct the emissions' time series and the fossil and biogenic CO2 fluxes. However, in regions with lower observation coverage, such as Southern Europe and the British Isles, estimating fossil CO2 emissions is less successful.

The paper discusses a current and significant topic, showcasing the LUMIA system's capability to utilize both CO2 and 14CO2 data simultaneously. However, the manuscript's discussion of the results is often too descriptive and fails to address the underlying processes. As a result, the study requires major revisions before it can be published.

Overarching comments:

There are certain parts in the manuscript where the language needs to be more accurate and specific. For instance, the national emission data that is reported to the UNFCCC cannot be compared with the emission inventories that are distributed spatially and temporally. However, both are referred to as "inventories" in the text. For experienced readers, the context might be

clear, but for new readers, the language needs to be more precise and differentiated in many places.

We updated the manuscript to use the correct language.

The manuscript contains many estimates of posterior CO2 emissions at regional and national levels, but it fails to mention the uncertainties associated with these estimates. To address this issue, an ensemble approach can be used that takes into account different realizations of synthetic observations and various prior uncertainties.

We performed a Monte Carlo ensemble of 100 members to calculate the posterior uncertainties and updated Figures 8 and 10 and the corresponding results and discussion.

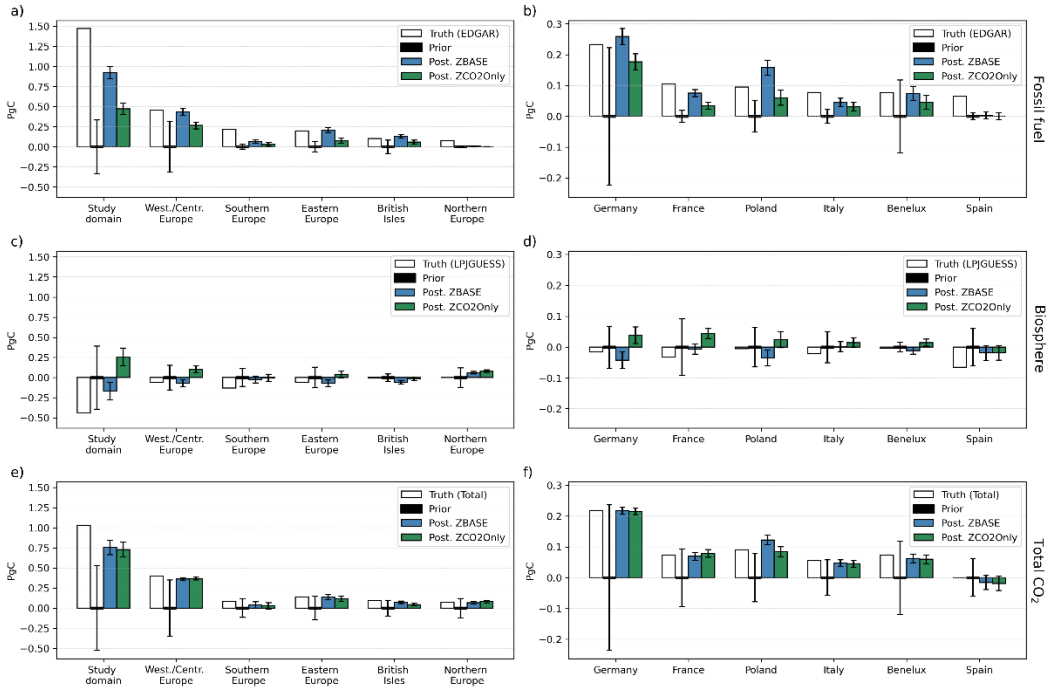

Figure 8. ~~Annual~~ True, prior and posterior annual budgets of fossil (a-b), biosphere (c-d) and total CO_2 (e-f) for the study domain, the sub-regions (right), and some of the largest European countries by area (left). The white bars show the true ~~emissions~~ annual budgets based on ~~the~~ EDGAR ~~emission inventory~~and LPJ-GUESS flux products. The ~~red~~ black bars ~~(horizontal hatching) are for reference and~~ represent ~~fluxes according to~~ the ~~ODIAC for fossil CO_2 (a and b)~~prior value, ~~VPRM for biosphere (c and d), and the sum of the two for total CO_2 (e and f)~~0 PgC. The blue ~~, green,~~ and ~~gray~~ green bars show the posterior ~~fossil fuel, biosphere, and total CO_2 fluxes for the~~ budgets of ZBASE ~~(grid hatching)~~ and ZCO2Only~~(diagonal hatching) experiments~~, respectively. The ~~red line represents~~ error bars represent the prior ~~value, 0 PgC~~and posterior uncertainty calculated with a Monte Carlo ensemble of 100 members.

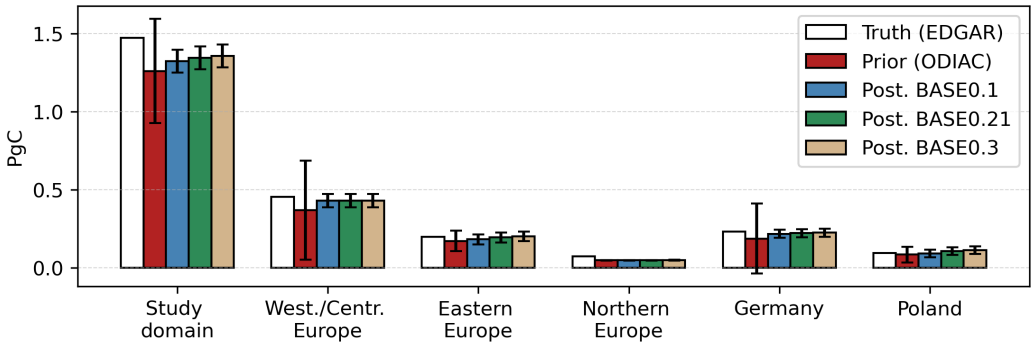

Figure 10. Total annual fossil CO_2 emissions for the study domain, Western/Central Europe, Eastern Europe, Northern Europe, Germany, and Poland. The white bars show the true emissions based on the EDGAR emission ~~inventory~~database. The red bars show the prior fluxes based on the ODIAC ~~emissions inventory~~emission data product. The blue, green and tan bars show the posterior fossil CO_2 emissions for the BASE0.1~~(grid hatching)~~, BASE0.21~~(diagonal hatching)~~, and BASE0.3 ~~(cross hatching)~~ experiments, respectively. The error bars represent the prior and posterior uncertainty calculated with a Monte Carlo ensemble of 100 members.

The OSSE assumes a constant nuclear 14C contamination. However, as the authors write in their summary, this does not reflect reality. An OSSE study on 14CO2 in Europe is predestined to analyse the influences of variable nuclear contamination. The authors should calculate an additional scenario for this purpose.

We mentioned this in the 'Discussion' and 'Conclusions and future perspectives' sections in the original manuscript. We feel that adding a scenario for this purpose is beyond the scope of this manuscript (which serves as a model description reference paper). A more detailed analysis of the influence of a range of influencing variables (including nuclear contamination) will be investigated in a follow-up study.

Below are general remarks on specific sections. Note that these comments do not cover all minor issues such as grammatical errors, missing words, or imprecise wording. After addressing the general remarks, a second review should take care of these smaller issues.

Section 1:

RC2: The literature cited lacks clarity and omits fundamental publications, particularly regarding the basic 14C cycle, or is citing them only indirectly.

The Introduction section was thoroughly updated to satisfy the concerns of both referees regarding existing literature in the field of using radiocarbon observations for estimating fossil CO_2 emissions (L75-109 and 136-139 of the revised manuscript).

. Nevertheless. For instance, Levin and Karstens (2007) present an observational approach to estimate hourly regional fossil fuel CO_2 offsets at a continental site (Heidelberg, Germany), using weekly mean $^{14}CO_2$-based fossil fuel CO_2 mixing ratios and CO observations. On a larger scale, Levin et al. (2008) examine monthly mean $^{14}CO_2$ observations from two German stations (Schauinsland and Heidelberg), compared against background measurements from Jungfraujoch, to assess the regional fossil fuel CO_2 surplus and emphasize the importance of high-precision radiocarbon measurements for quantifying fossil fuel CO_2 contributions at a regional scale in Europe. The study by Miller et al. (2012) explores the relationship between fossil fuel CO_2 emissions and enhancements in atmospheric concentrations of $^{14}CO_2$ and other anthropogenic trace gases. Utilizing a six-year dataset from vertical profiles in the northeast U.S., they separate the fossil and natural components of atmospheric CO_2 using apparent emission ratios of various gases to fossil fuel CO_2, offering observationally-based estimates of national emissions and comparing these with inventory-based estimates. Turnbull et al. (2015) use measurements of CO_2, $^{14}CO_2$, and CO from multiple sampling towers around Indianapolis, U.S., to differentiate fossil fuel CO_2 from background levels in an urban environment and evaluate the consistency of a bottom-up emission product. More recently, by using radiocarbon observations in CH_4 ($\Delta^{14}CH_4$) and CO_2 ($\Delta^{14}CO_2$) over London, Zazzeri et al. (2023) reveal that fossil fractions of CH_4 and atmospheric concentrations of fossil CO_2 are consistently higher than those predicted by simulations using emission products such as EDGAR. This discrepancy highlights the potential of $^{14}CO_2$ measurements to refine our understanding of fossil and biospheric CO_2 and CH_4 partitioning in urban settings, especially when the influence of nuclear power plants is minimal.

Nevertheless, large-scale four-dimensional inversion systems have only recently included begun to include $\Delta^{14}CO_2$ as an additional tracer to constrain fossil CO_2 emissions(Basu et al., 2016, 2020; Wang et al., 2018). Results from Observing System Simulation Experiments (OSSE)based on synthetic observations , assuming the current as well as an anticipated future network of . Basu et al. (2016) introduced a novel dual-tracer atmospheric inversion technique that differentiates between biospheric and fossil fuel CO_2 fluxes using atmospheric CO_2 and $\Delta^{14}CO_2$ measurement stations, have shown the high potential of constraining fossil measurements over the U.S. This method not only allows for the estimation of monthly regional fossil fuel CO_2 fluxes but also addresses biases in biospheric flux estimates that occur when using traditional CO_2-only inversion methods with fixed fossil fuel flux assumptions. Their approach represents a significant advancement in quantifying regional and national fossil fuel emissions from atmospheric observations. Building upon this study, Basu et al. (2020) presented a more focused analysis in providing national and sub-national-scale estimates of fossil fuel CO_2 emissions, using an extensive observation database of both CO_2 and $\Delta^{14}CO_2$. Graven et al. (2018) conducted an in-depth analysis of fossil fuel CO_2 emissions in California, utilizing atmospheric observations from nine sites and employing the Weather Research and Forecasting model along with the Stochastic Time-Inverted Lagrangian Transport model (WRF-STILT). The research integrates measurements of CO_2 concentration and $\Delta^{14}CO_2$, uniquely combining these observations with high-resolution emission data from Vulcan v2.2 and EDGARv4.2, aiming to refine estimates of regional fossil fuel CO_2 emissions and explore the impact of various factors such as nuclear industry emissions and air-sea exchanges on atmospheric CO_2 levels. In Europe, Wang et al. (2018) evaluated the potential of a $\Delta^{14}CO_2$ observation network for estimating regional fossil fuel CO_2 emissions over North America (Basu et al., 2016, 2020) and Europe (Wang et al., 2018). Having through atmospheric inversions. They examined the effectiveness of different network configurations, from minimal to very dense setups, in reducing uncertainties in fossil CO_2 emissions estimation. The study used synthetic observations and the LMDZv4 global transport model, paying special attention to representation and aggregation errors. Establishing a network of both CO_2 and $\Delta^{14}CO_2$ measurement stations Europe. For this purpose, we expanded the LUMIA Lund University Modular Inversion Algorithm (LUMIA) system (Monteil and Scholze, 2021) to perform simultaneous inversions of atmospheric CO_2 and $\Delta^{14}CO_2$, thus optimizing the fossil emissions, natural fluxes, and the isotopic disequilibrium. We perform observing system simulation experiments Observing System Simulation Experiments (OSSEs), recreating the current state of the ICOS network and its sampling strategy, and using different flux products (as priors and true values) to demonstrate the performance of the inversion scheme and show its capabilities. We begin by assessing the impact of oceanic fluxes on the total CO_2 and $\Delta^{14}CO_2$ concentrations. Then, we evaluate the impact of adding $\Delta^{14}CO_2$ observations on the estimation of fossil CO_2 emissions by comparing the model's ability to recover true fluxes starting from a prior flux set to zero. Finally, with a more realistic setup, i.e., prior, we evaluate the impact of the prescribed fossil CO_2 flux uncertainty and the impact of the terrestrial isotopic disequilibrium product.

Section 2.1:

RC2: The regional model is presented, but not a word is said about the boundary conditions and how these were realised.

We added an explanation of the boundary condition calculation to Section 3.3 (L331-337 of the revised manuscript) and commented on the perfect transport and perfect boundary conditions in the discussion (L630-652 of the revised manuscript).

not strictly necessary for this study, but we want to replicate the conditions of a real inversion. ~~We~~ Since we are using the same background concentration for the synthetic observations and the simulated prior and posterior observations (i.e. we are assuming a perfect boundary condition), we simplify the calculation of it by computing a smoothed and detrended weekly (for CO_2) and monthly (for $\Delta^{14}CO_2$) average of the real observations (ICOS et al., 2023) for each sampling site. For sampling sites, for which there are for some reason no real observations for the year 2018 in the ICOS database (e.g. $\Delta^{14}CO_2$ measurements were not yet implemented or were not yet part of ICOS), we took the observations from the nearest year available to calculate the background.

Discussion

The Observing System Simulation Experiment (OSSE) framework used in this study assumes a perfect realization of atmospheric transport and mixing processes by employing the same transport model across the simulations. This assumption simplifies the complex nature of atmospheric dynamics and is a common approach to limit the scope of variability in such studies. However, it is crucial to acknowledge that this simplification overlooks one of the largest sources of uncertainty in atmospheric inverse modeling: the accurate representation of atmospheric transport and mixing processes. The variability and uncertainty in atmospheric transport can significantly impact the estimation of greenhouse gas sources and sinks. As demonstrated by Schuh et al. (2019), inconsistencies in transport simulations can introduce systematic biases in surface flux estimations, which can be as substantial as 1.7 PgC $year^{-1}$ for large zonal bands. In a study by Munassar et al. (2023), in which multiple combinations of global and regional models were tested using two different inversion frameworks (LUMIA and CarboScope-Regional (CSR)), they found that using a different regional transport (FLEXPART and STILT (Stochastic Time-Inverted Lagrangian Transport)) model can cause differences in the posterior NEE annual budget of 0.51 PgC $year^{-1}$. This highlights the sensitivity of inversion-derived emission estimates to the accuracy of the transport model used and emphasizes the critical role that transport uncertainty plays across global flux inversion systems.

Furthermore, the assumption of perfect boundary conditions in the model presents another significant simplification. Boundary conditions in atmospheric modeling can greatly influence the concentration gradients and flux estimates, and their mischaracterization can propagate errors throughout the model domain. Coming back to the study by Munassar et al. (2023), the use of a different global transport model (TM3 and TM5) for the estimation of the boundary condition can cause discrepancies in the posterior annual budget as large as 0.23 PgC $year^{-1}$. Errors in these aspects of the transport model could lead to skewed emission estimates. Given these considerations, the presented results should be interpreted with caution, understanding that the true uncertainty in atmospheric inverse modeling is likely understated in these OSSEs. It underscores the need for more comprehensive approaches that account for transport model uncertainties, such as employing ensemble modeling techniques that incorporate multiple transport models and boundary conditions to better capture the inherent uncertainties in atmospheric dynamics (Locatelli et al., 2015; Aleksankina et al., 2018).

RC2: Eq. 3: What assumptions are made here concerning the d13C?

Since the nuclear emissions have a resolution of 1 year, we assumed $\delta^{13}C$ as the global atmospheric average value reported by NOAA, as used in studies such as Basu et al. (2016). We included this explanation in the revised manuscript in L176-177.

Section 2.3.2:

RC2: Please begin this section by pointing out that this construction of the B matrix is about determining their spatiotemporal structure and that the absolute magnitude of the uncertainty is afterwards scaled with the reported uncertainties.

The text was modified as suggested in L244-245 of the revised manuscript.

> **2.3.2 Construction of the prior error covariance matrix (B)**
>
> Our matrix B is constructed such that we first determine the spatio-temporal structure of the uncertainties, which is then scaled to match the reported uncertainties. Since we are optimizing for offsets, the prior control vector x_b contains only zeros (so

Section 3:

RC2: In the introduction of the OSSE, you should point out that this OSSE uses the same transport model, and thus, a perfect realisation of the atmospheric transport and mixing processes is assumed. This ignores one of the largest sources of uncertainty existing for inversions.

We included this in the introduction of Section 3 as suggested by the referee (L331-337 of the revised manuscript), and commented on this in the discussion (L630-652 of the revised manuscript).

Section 3.3.

RC2: Where do the synthetic background concentrations for CO2 and Δ14CO2 come from? Was TM5 used for the background? I can't find anything about this in the manuscript.
As mentioned in a previous answer to the referee, we added the explanation on the background concentration calculation in Section 3.3.

L274: There are no gaps during sampling that can be attributed to the calibration.

We agree with the referee and removed this from the text (see L321-322 of the revised manuscript).

> In this way, we account for the sampling gaps and the differences in integration times commonly produced due to ~~calibrations,~~ maintenance, and general operational eventualities. For stations with the number of observations, N_{obs}, equal to zero in Table

L285: CΔ14C -> Δ14C as you also show the 14C in Fig 4 and not the C14C. How large was the random perturbation which was added to the data?

We updated $C\Delta^{14}C$ with $\Delta^{14}CO_2$ and added a sentence explaining the calculation of the random perturbation added to the synthetic observations (see L341-343 of the revised manuscript).

site, observation time and tracer. To weaken the assumption of a perfect transport and boundary condition, we add a random perturbation to the synthetic observations~~to weaken the assumption of a perfect transport.~~. This random perturbation is equal to $y^* = y + \varepsilon \times \xi$, where y is the synthetic observation, ε is the observation error (both the instrumental and representativity errors, see Section 3.3.1 below), and ξ is a standard normal random vector. In this way, the added perturbation is based on the observation error. Figure 4 shows the synthetic CO_2 ~~concentration~~ and $\Delta^{14}CO_2$ observation time-series and the components

Section 3.3.1:

What is the motivation behind the definition of the prior uncertainties?

For Fbio and Fbiodis, this is not motivated in detail. What is the rationale for setting the bio error as 10% of negative Fbio flows? Likewise, for the 30% for the Fbiodis?

We updated the text in L357-365 of the revised manuscript to better motivate the choice of our prior uncertainties. We also modified the explanation of the F_{bio} uncertainty to be more accurate. As described in the manuscript, our aim with this study is to demonstrate the capabilities of the multi-tracer LUMIA system. Therefore, we focused on selecting uncertainty values that are reasonable and consistent with other studies.

flux uncertainties (See Section 2.3.1). Since our main purpose in this study is to demonstrate that our multi-tracer inversion system is capable of estimating both the fossil CO_2 emissions and natural CO_2 fluxes, we choose prior uncertainty values that are reasonable and consistent with other studies. The prior uncertainties are assigned as follows: ~~50% to 150% (0.1~~ for F_{ff}, we use the difference between the annual budgets for the whole study domain from ODIAC (1.26 PgC yr^{-1}~~to~~) (Oda and Maksyutov, 2020) and from an emissions product based on EDGARv4.3 (1.47 PgC yr^{-1}) (Gerbig and Koch, 2021b) as a reference to define its uncertainty (Basu et al., 2016). We use 150% (0.3 PgC yr^{-1}) of the difference ~~in the annual budget (0.21~~ as the base uncertainty for all the experiments, and we select two extra values to evaluate the impact of the prescribed uncertainty on the inversion: 50% of the difference (0.1 PgC yr^{-1}~~) of EDGAR and ODIAC for F_{ff} to evaluate its impact in the inversion, 10% (0.37 PgC yr^{-1}) of the annual negative values for F_{bio}~~ and the exact difference of 0.21 PgC yr^{-1} (~~where production is higher than respiration)~~ 100%). For F_{bio} we choose the 25% (0.37 PgC yr^{-1}) of the monthly prior (Monteil and Scholze, 2021), and 30% (0.22 PgC yr^{-1}) of the annual budget for F_{biodis} (Basu et al., 2020). We optimize all the

The definition of the prior Fff uncertainty as the difference between EDGAR and ODIAC, where EDGAR is used as truth and ODIAC as prior, is clearer. However, it FORCES the inversion in the BASE scenario to fully utilise the prior uncertainty budget to arrive at the "truth". In the 01Base sensitivity run, it is then even more "expensive" for the inversion algorithm to return to the truth. Some discussion on this would be welcome.

Following the referee's suggestion, we commented on this in the discussion in L539-600 of the revised manuscript.

The ~~realistic approach shows us~~ BASE experiments, in which we use realistic prior fluxes, show that the posterior fossil CO_2 emissions are not very sensitive to the prescribed prior uncertainty in regions with a dense sampling network~~. This is a positive result since the prior uncertainty is difficult to define both in magnitude and in spatial and temporal structure. Basu et al. (2016), for example, defined the prior uncertainty as the inter-prior spread (i.e. the difference among multiple prior~~ , even when using a low prior F_{ff} uncertainty in which case it is more difficult for the inversion algorithm to recover the true fluxes. As we have observed in previous studies using LUMIA (Monteil et al., 2020; Monteil and Scholze, 2021), the cost of fitting the observations dominates the total cost function value. In this sense, the relative value of the prior uncertainty of F_{bio} against F_{ff} ~~products), and as they pointed out in the study, this inter-prior spread is comparable in magnitude with the annual average NEE estimated by CarbonTracker~~is going to significantly impact the spatio-temporal distribution of flux adjustments, but the total uncertainty of the fluxes is of lesser importance since the model has enough freedom to adjust the data. In other words, the error structure and how is it set up for the different flux categories, is going to have more impact than the total prior uncertainty. Both Basu et al. (2016) and Wang et al. (2018) highlight the importance of a regional horizontal correlation and

Uncertainties of observations:

Defining the uncertainty of the observations as the standard deviation of the observations over a 7-day window is incorrect. Over such a long period, the standard deviation of the concentration is dominated by the variable transport and mixing processes in the atmosphere. This has nothing to do with the uncertainty of the measurements. In the manuscript (p.14 l.307), this leads to uncertainties of the CO2 measurements varying from 1 to 215ppm! (See also Table A1.)

The selected constant measurement uncertainty of 1.5‰, however, is too optimistic and should be replaced by 2‰. How does the (0.8 ppm ‰) error in Table 2 result?

With regard to the 14C measurement error, it would also be extremely interesting for an OSSE to illuminate the difference between 1.5 and 2 ‰ measurement accuracy.

Indeed, our aim with this definition of the CO_2 observation error is to represent the variability of the transport and mixing processes in the atmosphere, which can be larger in polluted sites in contrast with background sites. In our inversion system, the observation error represents both the instrumental error and the model representation error. In the case of the CO_2 observations, the observation error is mainly composed of the error of representativity, since the instrumental error is very small (in the order of 0.1 ppm), and by calculating the moving standard deviation in a weekly window we calculate an error proportional to how rapidly CO_2 varies.

For $\Delta^{14}CO_2$, in which the instrumental error is larger than the error of representativity, we selected a fixed value of 0.8 ppm $C\Delta^{14}C$ which translates to a mean value of 1.91±0.05‰ $\Delta^{14}CO_2$ using Equation 8 (closer to the 2‰ error suggested by the referee), and not 1.5‰ as previously stated in the original manuscript. Nevertheless, we run new inversions using an observation error of 0.9 ppm $C\Delta^{14}C$ (2.15±0.05‰ $\Delta^{14}CO_2$) and 1.0 ppm $C\Delta^{14}C$ (2.38±0.06‰ $\Delta^{14}CO_2$), the last one to show the impact of an approximately 0.5‰ in the observation error, as suggested by the referee. There are small differences in the results, but we consider them not to be significant enough and do not change the discussion and conclusions of the manuscript. Replicating Figures 5, 6, 9, and 11 of the manuscript:

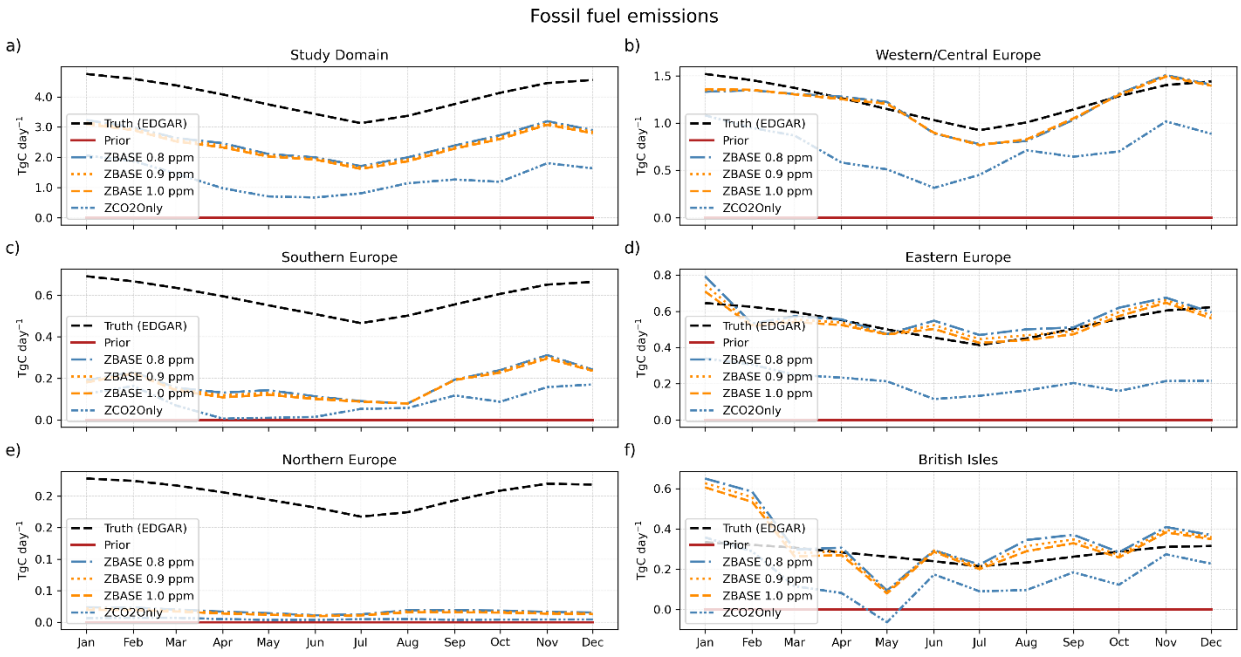

Figure 5.

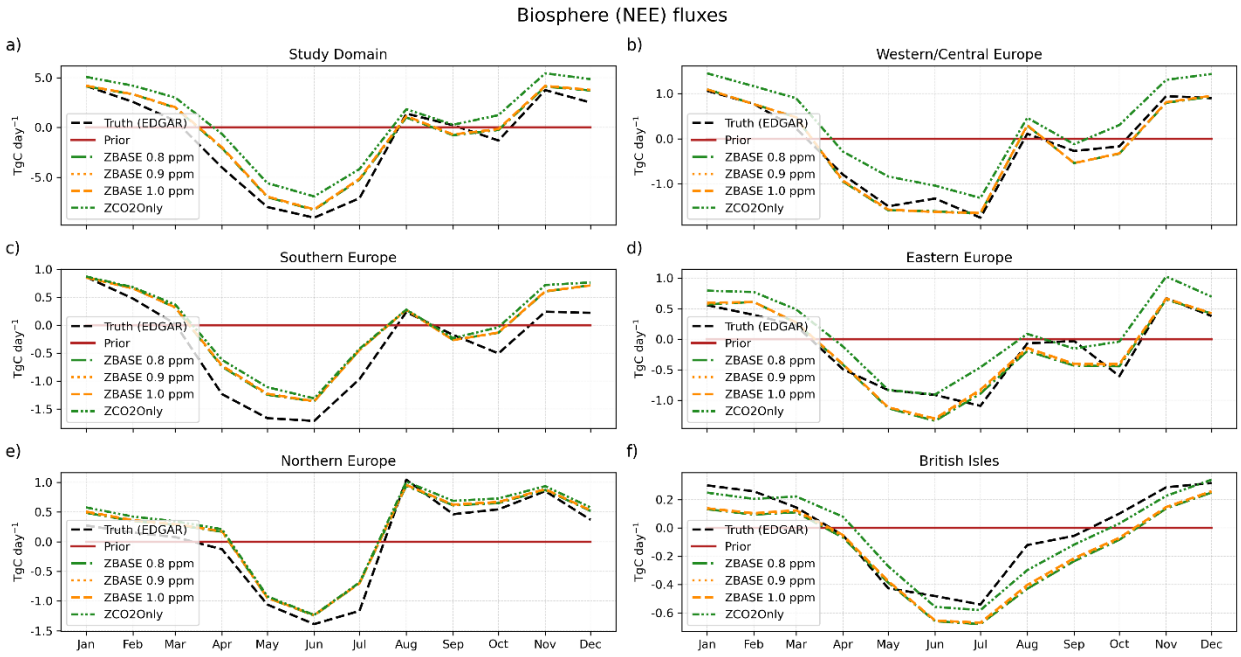

Figure 6.

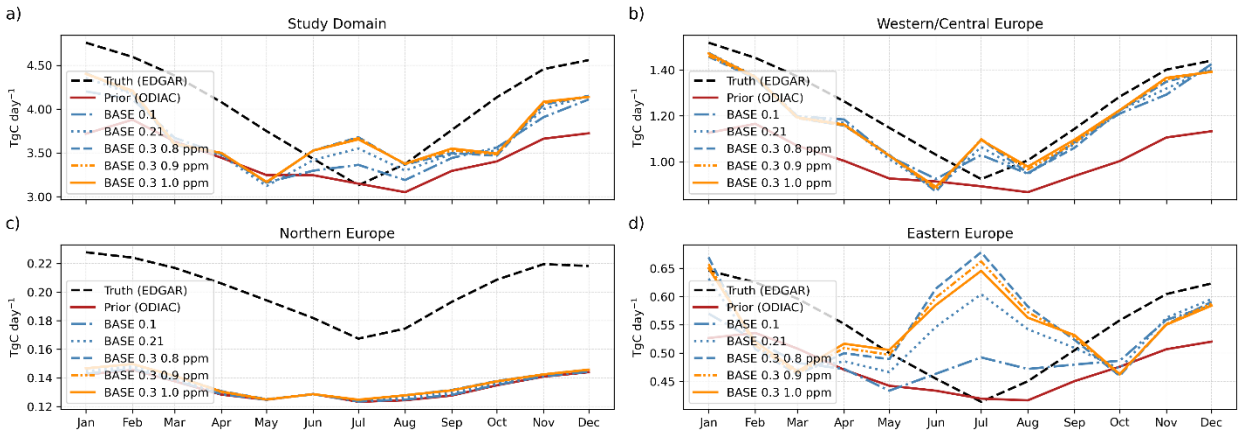

Figure 9.

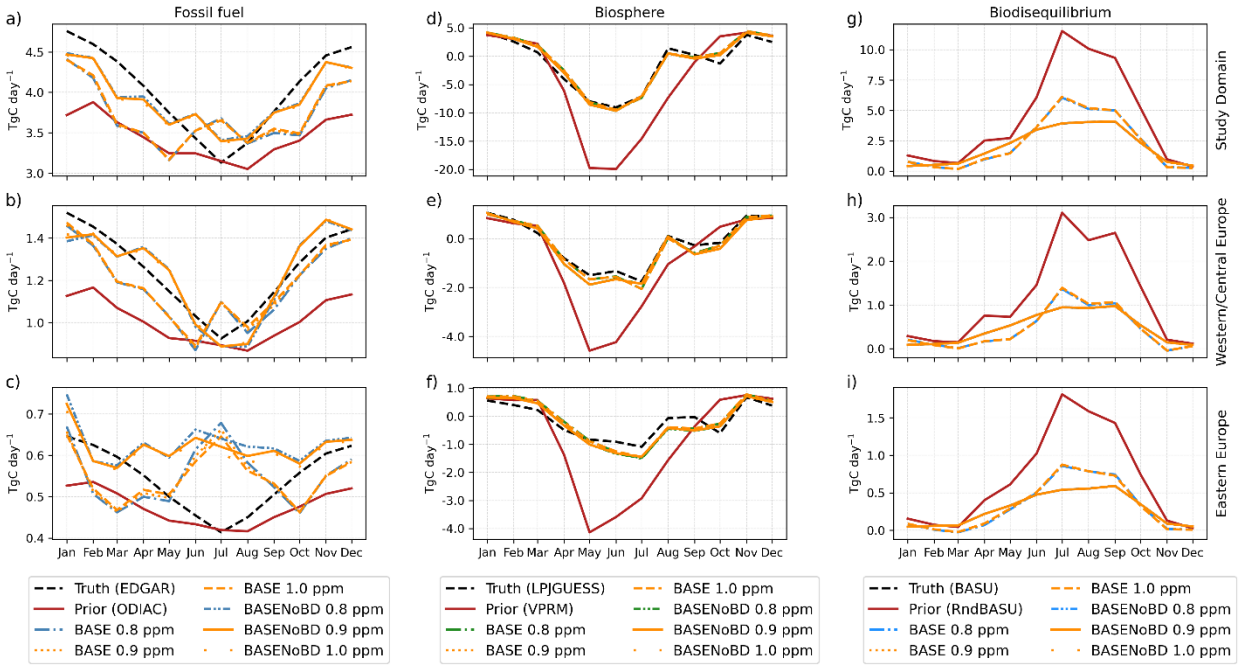

Figure 11.

We modify the text in L367-375 of the revised manuscript to clarify this.

(500 points). To set up the observation ~~uncertainty,~~ error, which includes the instrumental and the representativity errors, we use different methods for the CO_2 and the $\Delta^{14}CO_2$. For CO_2, where the error of representativity is usually larger than the instrumental error, we apply a weekly moving standard deviation to each observation i.e. the prior ~~uncertainty~~ error of each observation is equal to the standard deviation of the observations in a time window of ± 3.5 days around that observation. ~~The prior uncertainty for the~~ In this way, we account for the changes in the CO_2 ~~observations ranges from 0.91 to 215.5 ppm~~concentrations according to the local site conditions. For instance, at a background station such as Jungfraujoch (JFJ) on the top of the Swiss Alpes, the observation error ranges from 0.9 to 29.2ppm (mean value of 9.3 ± 4.0ppm), while at polluted sites such as Saclay (SAC) just outside Paris the CO_2 concentrations change rapidly and the error ranges from 5.9 to 215.5ppm (mean value of 55.8 ± 40.7ppm). For $\Delta^{14}CO_2$ ~~,~~ on the other hand, the instrumental error is larger than the representativity error, we use a constant value of ~~1.5‰ for~~ 0.8ppm in $C\Delta^{14}C$ units or 1.91 ± 0.05‰ in $\Delta^{14}CO_2$ units, calculated using Equation 8.

Section 4.2.1

This section is long and difficult to read... much is obvious and should be tried to be presented in a shorter and more concise way. For example, the numbers are all given in Table 4 and, therefore, do not need to be included in the text.

We updated the whole subsection following the referee's recommendation.

4.2.1 Retrieval of the monthly and regional time series

~~The posterior fossil CO_2 (F_{ff}) time series show heterogeneous results across the regions and experiments (Figure 5) in contrast~~
445 ~~with the biosphere fluxes (F_{bio}), where there is , in general, a good~~ In general, there is a closer agreement between the truth and the posterior time series for the ~~two experiments in all regions~~ (ZBASE and ZCO2Only experiments across all regions for the biosphere fluxes (F_{bio}) (Figure 6) ~~. Starting with~~ in contrast to the fossil CO_2 emissions (F_{ff}) (Figure 5). In the study domain, the ~~posterior ZBASE (adding $\Delta^{14}C$ observations) performs better than the~~ inclusion of $\Delta^{14}CO_2$ observations in the ZBASE experiment yields better performance than ZCO2Only for both flux categories. ~~For F_{ff} (Figure 5a), both experiments~~
450 ~~show a negative bias and follow the seasonality, albeit ZBASE is closer~~ Specifically, ZBASE exhibits closer alignment to the posterior ~~than ZCO2Only and therefore has a lower root mean square error (RMSE): 1.51 TgC day^{-1} versus 2.75 TgC day^{-1}, respectively~~ with a lower RMSE (see Table 4)~~. Posterior biosphere fluxes, on the other hand,~~ , indicating a better fit of the seasonality for F_{ff}. Similarly, the posterior biosphere fluxes more closely follow the true time series ~~closer than~~ than the fossil CO_2 emissions in both experiments, with ~~a positive bias (Figure 6a). Once again, ZBASE performs better than~~ ZBASE
455 outperforming ZCO2Only ~~most of the year and presents smaller RMSE (ZBASE = 1.12 TgC day^{-1}, ZCO2Only = 2.12 TgC day^{-1})~~ in terms of RMSE and BIAS values~~(ZBASE = 0.74, ZCO2Only = 1.90) (see Table 4). Before continuing with the regional results, it is important to mention the characteristics of the regions regarding the coverage of the sampling stations . In total, we consider 33 stations, all of them measuring CO_2 and 15 measuring additionally $\Delta^{14}C$. Most of the stations are located in .~~

The regional analysis reflects the influence of the coverage by sampling stations on the inversion outcomes. Western/Central Europe, benefiting from the highest number of stations (18 ~~stations~~out of 33 stations considered in this study, 10 of them measuring both tracers), ~~followed by Northern Europe with eight stations, four measuring both tracers, Southern Europe with four stations measuring CO_2 only, the British Isles with two (CO_2) and Eastern Europe with one (CO_2 and $\Delta^{14}C$) (see Figure 1). We find the best posterior time series in Western/Central Europe, which, as we already mentioned, also has the largest number of stations. The posterior fossil CO_2~~ shows the best alignment between the posterior and true time series for F_{ff}, especially in the ZBASE experiment ~~fit closely with the true time series, while ZCO2Only shows a pronounced bias~~ (Figure 5b)~~as in the case of the study domain for both experiments. Likewise, the posterior biosphere shows better results when adding $\Delta^{14}C$ observations (ZBASE)than without them (ZCO2Only), in which the latter has a positive bias most of the year (Figure 6b). Eastern Europe~~ , while ZCO2Only shows pronounced RMSE and BIAS values (Table 4). Conversely, regions like Eastern Europe (one station measuring both tracers) and the British Isles ~~show a posterior ZBASE fossil~~ (two stations measuring only CO_2~~close to the truth despite their low coverageof sampling stations. Eastern Europe shows the best results during the year for ZBASE, and ZCO2Only follows the tendency from the last regions (Figure 5d). The British Isles show a posterior ZBASE fossil CO_2 with more differences from~~), despite their lower station coverage, exhibit posterior ZBASE F_{ff} time series that closely approximate the truth, ~~particularly at the beginning of the year , where the posterior surpasses the truth in almost 100% of its value(Figure 5f), resulting in a similar RMSE but a lower BIAS than ZCO2Only (Table 4)~~with Eastern Europe showing consistent performance throughout the year (panels d and f in Figure 5). However, the posterior ZBASE biosphere fluxes in these ~~two~~ regions do not ~~show a good fit to the truth as~~ align as closely with the true values as observed in e.g. Western/Central Europe (panels d and f in Figure 6). In Eastern Europe, the posterior ZBASE shows big differences with the truth during May, June (maximum difference of 0.42 TgC day^{-1}), and later in September, while ZCO2Only shows a better fit during these months but a positive bias the rest of the year (Figure 6d). In contrast, the posterior biosphere flux from the ZCO2Only experiment shows a better fit to the truth than the ZBASE one in the British Isles (Table 4). ~~The ZBASE experiment shows a negative bias most of the year, except from March to May (Figure 6f).~~

Lastly, Southern and Northern Europe show similar results despite their differences: Northern Europe has better coverage of sampling stations, and its annual truth fossil CO_2 emissions are lower (an average of 0.20 TgC day^{-1} against 0.59 TgC day^{-1}). In both regions, the posterior ~~fossil CO_2~~ F_{ff} of the two experiments is far from the truth (Figures 5c and 5e)~~. The posterior biosphere~~ , while the posterior F_{bio} of both regions and experiments is close to each other, with Northern Europe showing a better fit to the truth than Southern Europe, in which the posterior shows a more pronounced positive bias along the year (Figures 6c and 6e).

Section 4.2.2:

L380: I thought the errors in the Z simulations were larger than indicated in Table 2?

As we mentioned in the caption of Table 2 and at the beginning of Section 3.3.1, the same uncertainty and error correlation values were used across all the inversions (experiments) in the study.

To Fig 7:

RC2: From Eq. 9 I understand that it is a relative RMSE reduction. If this is true, then why are the units in Fig 7.d g/(m2day)?

We updated Equation 9 to make it consistent with the metrics used in the previous subsection (4.2.1 Retrieval of the monthly and regional time series) and the units shown in Figure 7.

$$RMSE_{\text{reduction}} = ((RMSE^{apos}_{\text{ZCO2Only}} - RMSE^{apos}_{\text{ZBASE}}) - \mu)/\sigma \quad (9)$$

~~where μ is the average value of the difference between the two RMSEs and σ its standard deviation.~~ Here, positive values

RC2: The multi-pole structure in the biospheric RMSE reduction in Fig. 7d is striking. However, this is not discussed in the text. In Fig 7c one can see that the prior is spatially relatively smooth. However, since the results of ZBASE and ZCO2 are only mixed in the RMSE_reduction (7d), it is not possible to recognise whether this multipole structure arises from an inversion or from the interaction of the two. In any case, the multipole structure should be discussed more, also regarding a possible overfitting. The formation of the multipole structure cannot be due to the station distribution alone, as there are no stations in South-East Europe, and similar RSME_reductions are achieved in Central-West Europe.

We updated Figure 7 to show the individual maps of the posterior RMSE for both experiments and both categories and updated the text to comment on the multipole structure. In general terms, the regions where the biosphere fluxes are poorly constrained show higher RMSE values conforming dipoles in the map. We include a comment on this in Section 4.2.2 (see L453-454 of the revised manuscript) and the Discussion (L607-609).

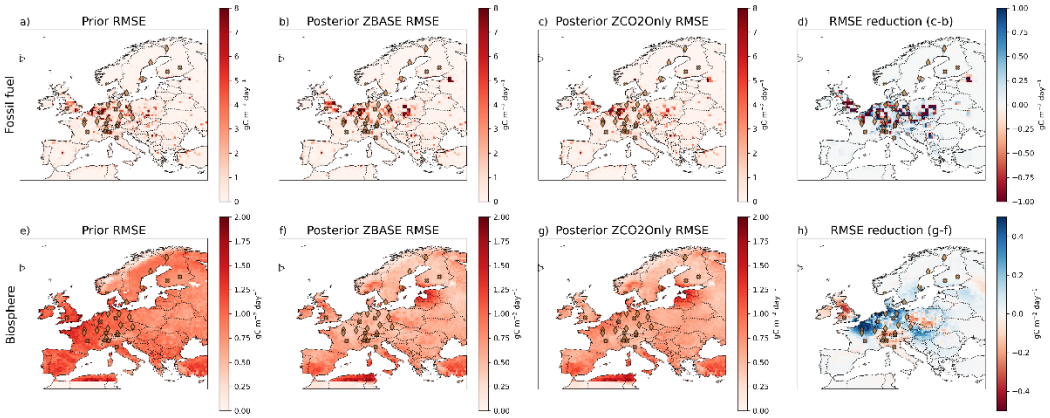

Figure 7. Spatial error of fossil CO_2 (a to d) and biosphere (e to h) for the ZBASE and ZCO2Only experiments. a) and ~~e~~c) show the prior RMSE for F_{ff} and F_{bio}, respectively, ~~and~~ b) and f) show the posterior RMSE for ZBASE, c) and g) show the posterior RMSE for ZCO2Only, and d) and h) show the ~~relative~~ RMSE reduction (see Equation 9) for fossil and biosphere. In Figures ~~b~~d) and ~~d~~h), positive values (in blue) show the pixels where ZBASE performs better than ZCO2Only (i.e. adding ~~$\Delta^{14}C$~~ $\Delta^{14}CO_2$ observations improves the posterior estimates), and negative values (in red) where ZCO2Only performs better than ZBASE. Crosses and diamonds represent stations that only measure CO_2 and those that additionally measure ~~$\Delta^{14}C$~~$\Delta^{14}CO_2$, respectively.

Section 4.2.2

510 observations (Figure 7~~b~~d). For the biosphere fluxes, ~~despite~~ the posterior RMSE maps (Figures 7f and 7g) show the regions that are poorly constrained due to the absence of observations such as the southern part of the domain and the Baltic States.

Discussion

the scale that is relevant given the model setup, as we demonstrate through the time series and annual budget results. Such

700 spatial misattribution is illustrated in the spatial RMSE reduction results for the biosphere fluxes. We can clearly identify the formation of dipoles (clusters of larger RMSE values) in regions with no observations such as the southern part of the study domain and the Baltic States indicating that these areas are underconstrained.

RC2: The first sentence of the caption of Fig7D refers to 8 images... but only 4 are shown.

We updated the figure and the corresponding caption in the revised manuscript as shown in the previous answer.

RC2: L389: …show for each location…?

We removed the whole sentence since we found it rather confusing.

CO_2 observations (ZCO2Only). ~~Since the prior here is zero, the prior RMSE maps (Figures 7a and 7c) show the locations where fluxes have their larger values (either negative or positive for biosphere) during the year.~~ For fossil fuel, we find ~~higher~~ larger

Section: 4.2.3

RC2: All data in Figure 8 are presented without any uncertainty information. A suitable measure for determining uncertainty should be considered and added here.

We performed a Monte Carlo simulation ensemble of 100 members to calculate the posterior uncertainty. We added this uncertainty to Figures 8 and 10. We also removed the bar "Ref. (ODIAC)" from Figure 8 to not cause any confusion, and updated the caption accordingly.

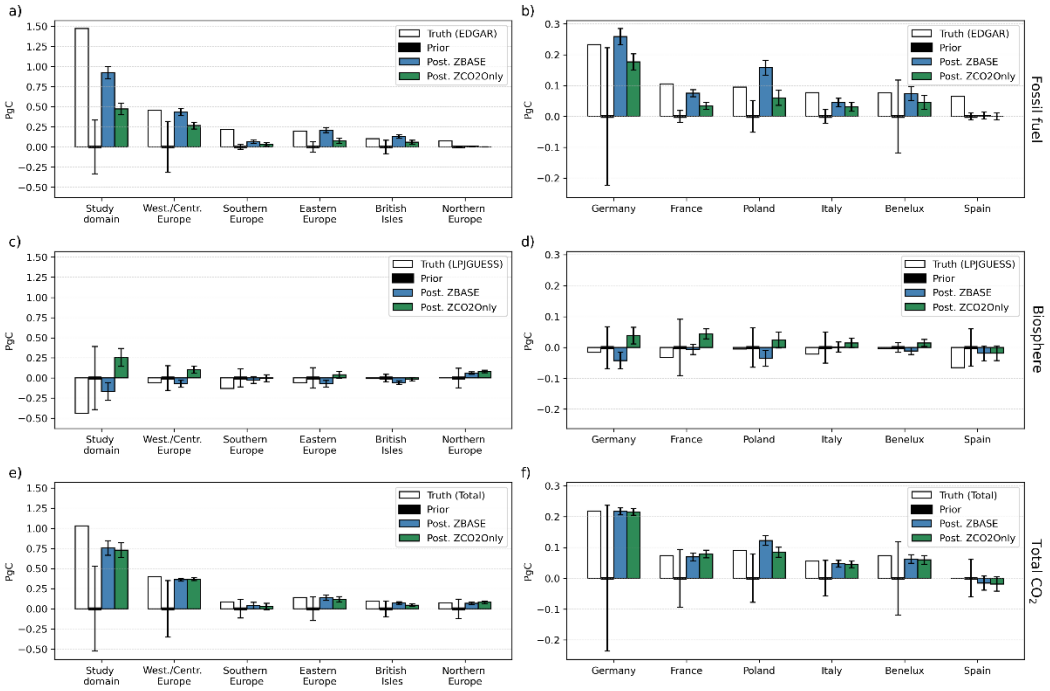

Figure 8. ~~Annual~~ True, prior and posterior annual budgets of fossil (a-b), biosphere (c-d) and total CO_2 (e-f) for the study domain, the sub-regions (right), and some of the largest European countries by area (left). The white bars show the true ~~emissions~~ annual budgets based on ~~the~~ EDGAR ~~emission inventory~~and LPJ-GUESS flux products. The ~~red~~ black bars ~~(horizontal hatching) are for reference and~~ represent ~~fluxes according to~~ the ~~ODIAC for fossil CO_2 (a and b)~~prior value. ~~VPRM for biosphere (c and d), and the sum of the two for total CO_2 (e and f)~~0 PgC. The blue ~~, green,~~ and ~~gray~~ green bars show the posterior ~~fossil fuel, biosphere, and total CO_2 fluxes for the~~ budgets of ZBASE ~~(grid hatching)~~ and ZCO2Only~~(diagonal hatching) experiments~~, respectively. The ~~red line represents~~ error bars represent the prior ~~value, 0 PgC~~and posterior uncertainty calculated with a Monte Carlo ensemble of 100 members.

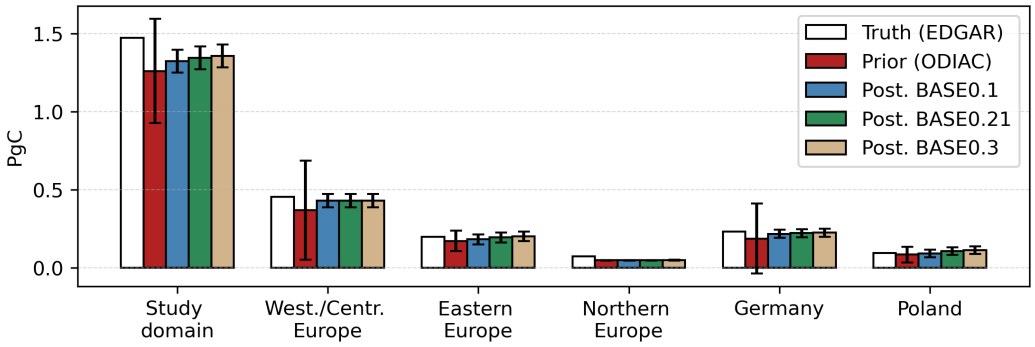

Figure 10. Total annual fossil CO_2 emissions for the study domain, Western/Central Europe, Eastern Europe, Northern Europe, Germany, and Poland. The white bars show the true emissions based on the EDGAR emission ~~inventory~~database. The red bars show the prior fluxes based on the ODIAC ~~emissions inventory~~emission data product. The blue, green and tan bars show the posterior fossil CO_2 emissions for the BASE0.1~~(grid hatching)~~, BASE0.21~~(diagonal hatching)~~, and BASE0.3 ~~(cross hatching)~~ experiments, respectively. The error bars represent the prior and posterior uncertainty calculated with a Monte Carlo ensemble of 100 members.

RC2: What is the reason for the significant underestimation of the total CO2 fluxes in the study domain in both the ZBASE and the ZCO2only simulation (Fig 8e)? Is this an indication that the uncertainties in the prior fluxes are too small?

The main reason is under-sampling in the whole study domain. As mentioned in the manuscript (and as can be seen in Figure 8) for Western/Central Europe and its countries (Germany, France, and Benelux), where there is a better network coverage (and therefore more samples) the inversion is able to estimate posterior values close to the true values, while in Southern Europe (and Spain) with a sparse sampling network this is not the case.

Section 4.3.1

RC2: L437: The summer deviation in Western/Central and Eastern Europe are nearly of the same magnitude. The analysis of the annual cycles should go into a little more depth. Is it due to an incorrect BIO prior? Or is an incorrect assumption on the 14C signature of the heterotrophic respiration? Is it due to the stronger dilution of the fossil emissions and, therefore, the smaller signal-to-noise ratio of the 14C measurements? The analysis needs to go into more depth here.

The prior terrestrial isotopic disequilibrium flux is on purpose incorrect with the aim of showing the impact that it can have on the estimation of fossil CO_2 emissions. We commented on this in the discussion in L611-618 of the revised manuscript.

(Figure 11). The prior terrestrial isotopic disequilibrium flux in our experiments is on purpose incorrect with the aim of showing the impact that it can have in the estimation of fossil CO_2 emissions. As shown in Figure 11, the maximum difference between the prior and the true F_{biodis} is of the same order of magnitude for Western/Central Europe (2.1 TgC day^{-1}) and Eastern Europe (1.3 TgC day^{-1}) in July. For F_{ff}, however, the difference between the prior and truth is about one order of magnitude larger for Western/Central Europe compared to Eastern Europe (0.03 vs 0.005 TgC day^{-1}). This larger difference causes a stronger dilution of the fossil emissions in Eastern Europe, and therefore essentially lowers the signal-to-noise ratio of the $\Delta^{14}CO_2$ measurements, and added to the lower network coverage compared to Western/Central Europe, a poorer constrain of the fossil CO_2 emissions. According to Turnbull et al. (2009), one of the main contributors to atmospheric $\Delta^{14}CO_2$ is heterotrophic

RC2: L444: Make a reference to Fig. 10.

The reference was included in the text (see L504 of the revised manuscript).

RC2: L447: "… with BASE0.3 having the highest recovery of 92%.". This sentence is slightly misleading. Even in this scenario, the improvement in the DIFFERENCE of the Prior is only around 50%. The figure of 92% probably refers to the total emission, of which approx. 80% will certainly already be "recovered" by the prior.

We agree with the referee. In the text, we are referring to the recovery of the difference between prior and truth, while the percentage corresponds to the recovery of the total budget. We correct this in L505-507 of the revised manuscript.

by all three experiments, with a recovery ranging from 30% for BASE0.1 to 45% for BASE0.3~~having the highest recovery, 92%~~. In Western/Central Europe, the three experiments recover 96% of the truth (around 71% of the difference between true and prior), similar to Germany, where the recovery ranges from 94% for BASE0.1 to 97% for BASE0.3 (68% to 82% of the difference). As we find in the time series (Figure 9d), the prior annual budget is very close to the truth both in Eastern

Section 4.3.2

This section lacks the desired discussion of the effects. The summer overestimation of the Fbiodis flux is probably reflected in the summer maximum of the fossil flux. However, this is not discussed. This is also clearly shown in the BASEnoBD control experiment. The fact that Eastern Europe does not improve in this control experiment should not come as a surprise given the assumed station distribution.
At this point, the fundamental question arises as to how the third unknown, Fbiodis (in addition to FffCO2 and FbioCO2), can be robustly derived from the two observed variables CO2 and 14C?

These suggestions, together with the ones for section 4.3., were added to the discussion in L611-618 of the revised manuscript

Section 4.3.3

Why was JFJ selected as a representative station? JFJ certainly has by far the lowest ffCO2 contribution in Central/Western Europe. Also, the fact that the prior correlation for JFJ is only 0.61, whereas it is 0.92 for all stations, shows that JFJ is not a representative station.

We kept JFJ and added Saclay (SAC) to the analysis to show the parallel between a background and a polluted station. This section was rewritten completely and we added Figures 13 and 14 and Table 5 to complement the results analysis.

4.3.3 The observational space

Finally, we analyze the model's performance in the observational space ~~, which is crucial for evaluating its effectiveness. Figure 12compares the prior and posterior concentrations from the BASE0.1 experimentwith the corresponding synthetic observations for all sampling stations and , as a representative example, for the Jungfraujoch (JFJ)station. Examining the correlation coefficients, we find that the prior concentrations already correlate significantly with the synthetic observations . The correlation coefficients for the prior estimates are 0.61 for JFJ~~ at all sampling stations aggregated together, one polluted station (Saclay, SAC) (Figure 12) and one background station (Jungfraujoch, JFJ) (Figures 12 and 13) for the BASE experiment. We calculate two performance metrics: the correlation coefficient between the synthetic observations and the prior and posterior simulated concentrations for all the sites and individually for the two sites selected, and the χ^2 for the overall simulation (Table 5). The histograms in Figure 12 show the mismatches between the synthetic observations and the prior and posterior concentrations. For the CO_2 concentrations at all sites (Figure 12~~b)and 0.92 (Figure 12~~a)~~for all stations, indicating a reasonable correlation with the synthetic observations~~, the histogram shows a distribution centered around zero for both prior and posterior mismatches with a standard deviation of 14.2 and 13.4, respectively (see Table 5), indicating systematic deviations from the

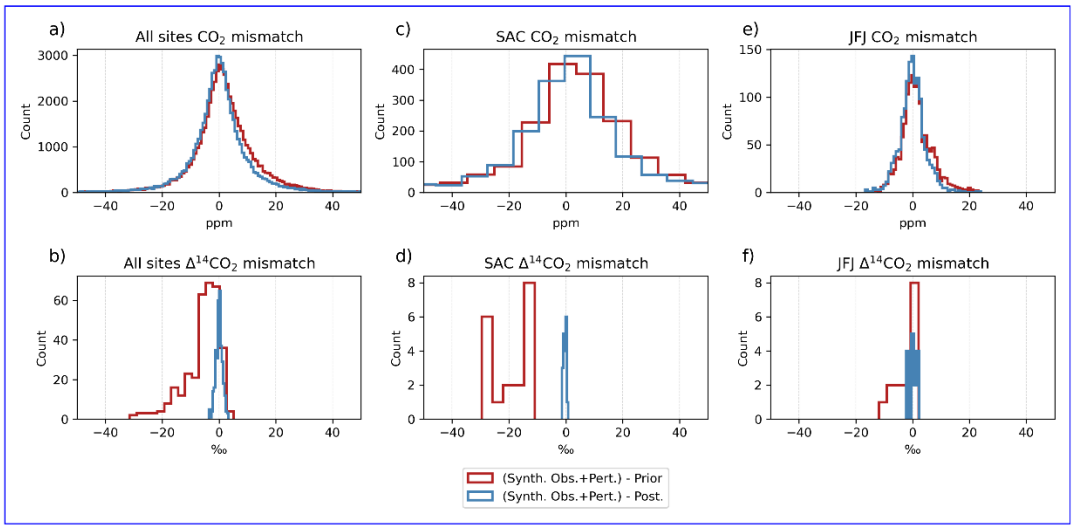

Figure 12. Mismatches between the synthetic observations and the prior (red) and posterior (blue) concentrations for all the sampling stations, Saclay (SAC) and Jungfraujoch (JFJ) for CO_2 (a, c and e) and for $\Delta^{14}CO_2$ (b, d and f). All prior and posterior concentrations correspond to the BASE experiment.

observed values. The posterior mismatch has a slightly tighter distribution, suggesting a small improvement in the model after adjustments as reflected in the correlation coefficient (Table 5). At Saclay (Figure 12c), the mismatch distribution is wider than the aggregate of all sites, which could suggest greater variability or larger errors at this particular site. The posterior adjustment has not significantly tightened the distribution, indicating that the model adjustments did not perform as well at this site as they did on average across all sites. On the other hand, the ~~posterior concentrations exhibit a slight enhancement in the correlation coefficients compared to the prior concentrations: 0.71 for JFJ (Figure 12b) and 0.98 (Figure 12e)for all stations, indicating a refinement in the model's ability to reproduce the synthetic observations accurately. This improvement in correlation coefficients is also reflected in the mismatch plots. For example, the mismatch between the posterior concentrations and the synthetic observations of $\Delta^{14}C$ at all stations~~ distribution in Jungfraujoch (Figure 12e) is much tighter than in all sites and SAC, with the posterior mismatch displaying a slight improvement in precision as evidenced by the narrower spread. However, when comparing the posterior time series with the synthetic observations before adding the random perturbation (Figures 13a and 13c), there is a better agreement between them than with the prior values, especially during periods of higher variability (April to July at SAC, and April to September at JFJ).

The $\Delta^{14}CO_2$ synthetic observations are in general better fitted by the posterior than CO_2 at all sites, SAC and JFJ (Table 5). In all cases, the prior distribution is displaced to the negative values, indicating that the prior simulated values are in general

higher than the synthetic observations as shown for the whole period at SAC (Figure 12c) shows a narrower distribution around zero compared to the previous mismatch. This suggests that the inversion process has effectively adjusted the model outputs, bringing them closer to the true observations. Analyzing the concentration time series at the JFJ station d) and from July to November at JFJ (Figure 12f). These larger prior concentrations are mainly caused by the prior terrestrial disequilibrium flux from July to November, and by the nuclear production fluxes throughout the year, which is significantly larger at Saclay (Figure 14). However, the posterior mismatches showed a much narrower spread around zero at all sites (Figure 12e and b), Saclay (Figure 12d), and Jungfraujoch (Figure 12f) , we observe that that is evident in the time series at both sites where the posterior closely follows the posterior concentrations agree better with the synthetic observationsthan the prior concentrations. This improvement is particularly notable for $\Delta^{14}C$, indicating that the inversion has successfully captured the dynamics of this tracer. Lastly, we consider the , and supported by the correlation coefficients (Table 5).

The reported χ^2 values of 1.77 for the prior and posterior concentrations across all sampling stations and observations. The prior 1.06 for the posterior across all sites and samples suggest a substantial improvement in the model's performance in adjusting the prior concentrations to the synthetic observations. A χ^2 value is 1.52, indicating some discrepancy of 1.77 for the prior indicates that there were significant discrepancies between the prior concentrations and the synthetic observations. However, the posterior This is consistent with the broader spread of mismatches in the histograms for both SAC and JFJ sites, as well as the apparent overestimation of $\Delta^{14}CO_2$ content in the time series. The improvement to a χ^2 value improves to 1.00, indicating a closer match between the posterior concentrations and the observed data. These results confirm that the inversion process has effectively improved the model's performance in the observational space. of 1.06 for the posterior indicates a better fit to the synthetic observations that are likely to be reflective of the underlying data patterns while still maintaining some degree of generalizability without overfitting the data.

Table 5. Performance metrics (correlation coefficient R, standard deviation and χ^2) for all sites, Saclay (SAC), and Jungfraujoch (JFJ).

		Prior		Posterior	
		R	σ	R	σ
All sites	CO_2	0.64	14.2	0.68	13.4
	$\Delta^{14}CO_2$	0.72	6.4	0.99	1.2
SAC	CO_2	0.56	31.9	0.59	31.1
	$\Delta^{14}CO_2$	0.63	6.8	0.99	0.5
JFJ	CO_2	0.65	5.5	0.74	4.5
	$\Delta^{14}CO_2$	0.75	4.2	0.84	1.5
χ^2		1.77		1.06	

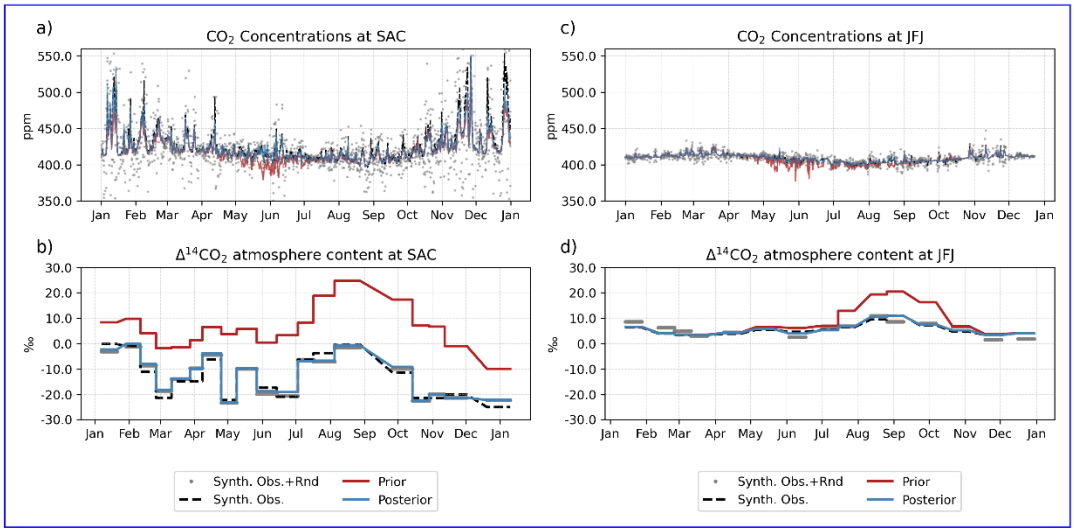

Figure 13. ~~Mismatches between the synthetic observations and the prior (red) and posterior (blue) concentrations for all the sampling stations and at Jungfraujoch ICOS station for~~ Concentration time series of CO_2 (a and ~~b~~c) and ~~for $\Delta^{14}C$~~ $\Delta^{14}CO_2$ (~~c~~b and d) ~~. The right panel shows the time series of CO_2~~ at Saclay (~~e~~SAC) and ~~$\Delta^{14}C$~~ Jungfraujoch (~~f~~JFJ)~~at Jungfraujoch~~, respectively. All prior and posterior concentrations correspond to the ~~BASE0.1~~ BASE experiment.

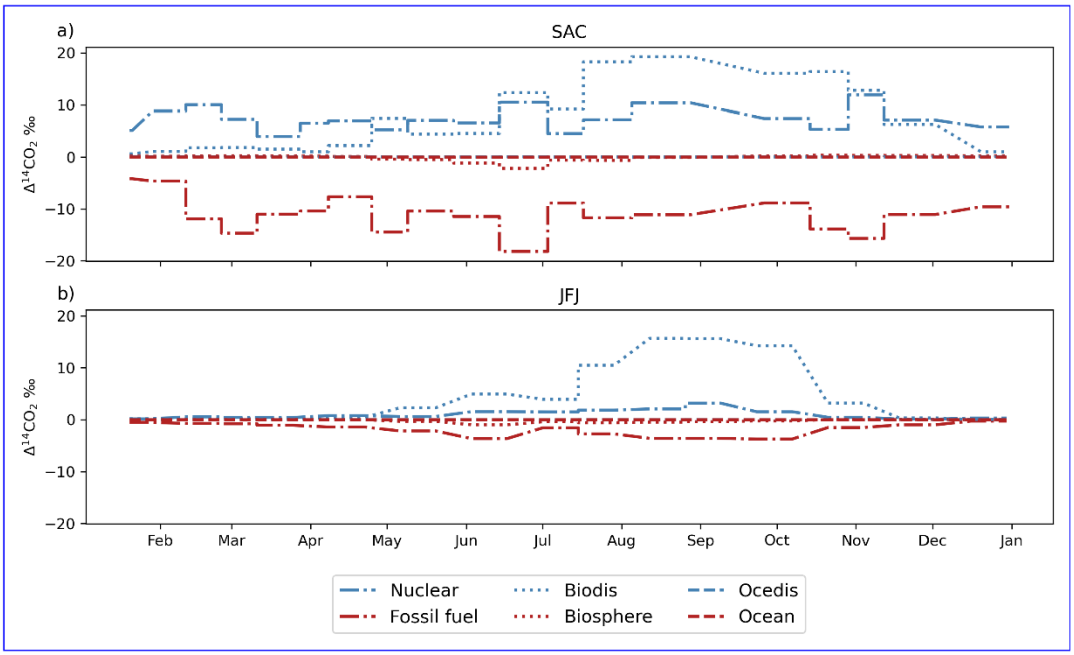

Figure 14. Contribution of each category to the prior $\Delta^{14}CO_2$ simulated concentrations at Saclay (a) and Jungfraujoch (b).

Fig 10e: The Synth Obs cannot be seen in the way they are plotted. Please change the line style to allow every piece of information to be seen.

We updated the figure to improve the visualization as shown in the answer above.

The random perturbation of the synth CO2 also appears to be too large in this plot, at least for measurement errors (see comments above). If the uncertainties of a real transport model error

should also be represented by the large uncertainties, then a fundamental revision of section 3.3 on error determination is required.

As mentioned in a comment before, the observation error is the aggregate of the measurement or instrumental error and the error of representativity.

The conclusions drawn at the end of section 4.3.3 are fairly trivial. They merely show that the inversion optimisation algorithm works as it should, but this does not imply that the results are correct.

We acknowledged this comment when rewriting this section.

Unfortunately, the authors do not address the interesting mismatch of the prior 14C observations. Fig 12c shows a pronounced overestimation in 14C. The JFJ plot suggests that this overestimation occurs in summer/autumn and, therefore, likely originates from Fbiodis. All of this reflects to the previous results, but the discussion was not very in-depth. Fig. 12f, suggests that the Fbiodis fluxes are strongly overestimated. Here it would be interesting to know whether this is due to the randomly altered 14C signature of heterotrophic respiration or to the very different fluxes of VRPM and ORCHIDEE.

We added Figure 14 to show the influence of each category on the total prior $\Delta^{14}CO_2$ content. The main reason for this overestimation indeed originated from the prior F_{biodis}, followed by the prior F_{nuc}, which we found to have a large impact on stations surrounded by nuclear facilities such as SAC.

Section 5: Discussion

The discussion section is more like a summary and a comparison with previous literature. Unfortunately, there is no real in-depth discussion of the results.

We updated the Discussion section including the referee's concern about the assumption of perfect transport, perfect boundary condition, prior and posterior uncertainties, spatial distribution, and the impact of the prior terrestrial isotopic disequilibrium product.

L508: Fig10 -> Fig9, Fig9 -> Fig 10 , Fig 11-> ?

We updated the references to the figures.

L520: How will such a scaling workaround solve the problem of small signals? This approach does not change the measurement uncertainty or the observational signal-to-noise ratio. The authors mention that this approach might be problematic due to noise. Thus, I recommend not making this suggestion at all.

We removed this suggestion from the manuscript and updated the text in L586-592 of the revised manuscript.

sions in an OSSE setup given the current CO_2 and $\Delta^{14}CO_2$ observation networks. ~~A~~ We think that a more likely explanation is the difference in the magnitude of the fossil CO_2 emissions in ~~the~~ this region against the natural fluxes~~and other regions. A workaround can be the normalization~~. This can be seen by the differences in the seasonal amplitude of the fluxes~~or the implementation of regional scaling factors that allow having similar magnitudes across the study domain but still, the impact from other sources of noise, such as the background concentration (boundary condition) , which includes the fluxes transported from other regions within the domain, can make it difficult to solve~~. In Western/Central Europe F_{bio} and F_{ff} are of a similar order of magnitude (2.81 TgC day^{-1} for F_{bio} and 0.6 TgC day^{-1} for F_{ff}) (see Figures 5 and 6). In contrast, in Northern Europe, there is a tenfold difference in the seasonal amplitude of the two fluxes: 2.44 TgC day^{-1} for F_{bio} and 0.06 TgC day^{-1} for F_{ff}. In addition, the prior uncertainty for F_{ff} (0.002 PgC year^{-1}) in this region is much lower compared to F_{bio} (0.12 PgC year^{-1}) in Northern Europe.

L524: The realistic… -> The more realistic…

We modified this in the text.

Appendix A:

Table A1: The numbers for the CO2 Obs Error are unrealistic (see above).

We commented on this in the answers above.

References

Basu, S., Lehman, S. J., Miller, J. B., Andrews, A. E., Sweeney, C., Gurney, K. R., Xu, X., Southon, J., & Tans, P. P. (2020). Estimating US fossil fuel CO2 emissions from measurements of 14C in atmospheric CO2. *Proceedings of the National Academy of Sciences*, *117*(24), 13300–13307. https://doi.org/10.1073/pnas.1919032117

Basu, S., Miller, J. B., & Lehman, S. (2016). 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. https://doi.org/10.5194/acp-16-5665-2016

Broquet, G., Chevallier, F., BrÉon, F. M., Kadygrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morguí, J. A., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R. L., Vermeulen, A. T., Yver, C., & Ciais, P. (2013). Regional inversion of CO2 ecosystem fluxes from atmospheric measurements: Reliability of the uncertainty estimates. *Atmospheric Chemistry and Physics*, *13*(17), 9039–9056. https://doi.org/10.5194/ACP-13-9039-2013

Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T., & Ciais, P. (2011). A European summertime CO2 biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements. *Journal of Geophysical Research: Atmospheres*, *116*(D23), 23303. https://doi.org/10.1029/2011JD016202

Monteil, G., Broquet, G., Scholze, M., Lang, M., Karstens, U., Gerbig, C., Koch, F.-T., Smith, N. E., Thompson, R. L., Luijkx, I. T., White, E., Meesters, A., Ciais, P., Ganesan, A. L., Manning, A., Mischurow, M., Peters, W., Peylin, P., Tarniewicz, J., … Walton, E. M. (2020). The regional European atmospheric transport inversion comparison, EUROCOM: first

results on European-wide terrestrial carbon fluxes for the period 2006–2015. *Atmos. Chem. Phys.*, *20*(20), 12063–12091. https://doi.org/10.5194/acp-20-12063-2020

Monteil, G., & Scholze, M. (2021). Regional CO2 inversions with LUMIA, the Lund University Modular Inversion Algorithm, v1.0. *Geosci. Model Dev.*, *14*(6), 3383–3406. https://doi.org/10.5194/gmd-14-3383-2021

Munassar, S., Monteil, G., Scholze, M., Karstens, U., Rödenbeck, C., Koch, F. T., Totsche, K. U., & Gerbig, C. (2023). Why do inverse models disagree? A case study with two European CO2 inversions. *Atmospheric Chemistry and Physics*, *23*(4), 2813–2828. https://doi.org/10.5194/ACP-23-2813-2023

Philip, S., Johnson, M. S., Potter, C., Genovesse, V., Baker, D. F., Haynes, K. D., Henze, D. K., Liu, J., & Poulter, B. (2019). Prior biosphere model impact on global terrestrial CO2 fluxes estimated from OCO-2 retrievals. *Atmospheric Chemistry and Physics*, *19*(20), 13267–13287. https://doi.org/10.5194/ACP-19-13267-2019

Wang, Y., Broquet, G., Ciais, P., Chevallier, F., Vogel, F., Wu, L., Yin, Y., Wang, R., & Tao, S. (2018a). Potential of European 14CO2 observation network to estimate the fossil fuel CO2 emissions via atmospheric inversions. *Atmos. Chem. Phys.*, *18*(6), 4229–4250. https://doi.org/10.5194/acp-18-4229-2018