Reply to Review #2

I thank the authors for their work addressing my comments. I believe that the text is clearer now and that the reader will have a better opportunity to understand the work done here and its significance. While the authors have generally done a good job addressing my questions, I still have some worries and questions about the work that I would like addressed before giving the thumbs up for publication.

Thank you for the helping us to improve the manuscript.

In my original comments (at line 595) I questioned the small uncertainty reductions (under 10%) seen in Figure 3 for many regions, especially those in the tropics, and speculated that the uncertainties for the grouped regions might not be being calculated correctly.

In Figure 3, we are showing uncertainties for individual 84 regions (used in the inversion), not for the grouped regions. We plotted the monthly flux uncertainties, averaged over the analysis period 2001-2020, for both a priori (as input to the inversion model) and a posteriori (o/p of the inversion) fluxes. It is not uncommon to find low uncertainty reduction by inversion based on the Bayesian a posteriori flux uncertainty for each inversion regions (e.g., Gurney et al., 2002), and more importantly there is "difficulty" to calculate Bayesian posterior uncertainty for most inversions accurately/appropriately (we have faced this since working on Thompson et al., Nature Comm, 2016, and also in GCP Global carbon budget analysis, series of ESSD papers). Due to such difficulties, we have developed the ensemble inversion approach by accounting uncertainties arising from prior fluxes (CASA vs VISIT land, Takahashi vs JMA ocean) and input parameters of the inverse model (PFU, MDU).

In light of your comments, we have added a few lines of clarifications in the main manuscript:

Section 2.5: Performance of inversion using a posteriori uncertainty
"The inverse model output monthly means flux corrections and a posteriori flux uncertainties for each of the 84 regions, and the full error covariance matrix of dimension 24192 × 24192 (=84 regions × 12 month × number of year). The monthly time and spatial covariances are accounted for flux uncertainty calculation when annual mean values are calculated for aggregated regions or global budgets. In the aggregation scheme, the larger regions have to follow the boundaries of 84 regions, contrary to the method proposed in section 2.6 by using ensemble inversions where ensemble spreads can be calculated for any region of interest.

We use flux uncertainty reduction (FUR, in %), based on the mean values without time aggregation …"

Section 3.2: in the final sentence
"The ensemble spread is much lower (Table 3; MIROC4-ACTM columns) compared to the inversion predicted flux uncertainties, which are in the range of 1.4 and 0.7 PgC yr^{-1} for the global land and ocean, respectively, even after accounting for the monthly time and spatial covariances (vary from low values of 0.8 and…"

Similarly, in my comment on lines 349-350, I suggested that the 1.4 PgC/yr uncertainty on the global land flux was a few times higher than it should be. The authors, in their response, gave a list of a posteriori uncertainties at the scale of Transcom regions (lines 113-142 of the response) that, rather than squelching my worries have instead exacerbated them. Along with the uncertainties for the individual Transcom regions, they give uncertainties for the global land and global ocean total, along with the global land+ocean total: those for land and the land+ocean total are up around 8 PgC/year -- those values seem to be in conflict with the 1.4 PgC/year value given in the paper, so which is correct? If these uncertainties are indeed up around 8 PgC/year, that would suggest that the correlations in the a posteriori covariance matrix are not being considered (since taking the sum of the squares of the uncertainties for the individual Transcom regions given in lines 113-142 of the response gives about (8 PgC/year)^2).

Here we calculated the aggregated fluxes and flux uncertainties for annual mean values (the covariances accounted for) before taking the long-term means which are given in the text. Thus the 1.4 PgC/yr uncertainty is correct for the annual mean fluxes (temporal covariances accounted).

We have rechecked the calculation of aggregated region flux uncertainties. In the reply, what we produced was for monthly flux values – only spatial covariances accounted for but not the temporal covariance. We believed those a posteriori values, without accounting for temporal covariances, are more relevant to compare with the a priori flux uncertainties (as is done for FUR calculation).

The 1.4 PgC yr^{-1} uncertainty on the global land flux is several times higher because we use much larger prior flux uncertainty compared to inversion like TransCom. This was already discussed in the manuscript as follows (line#406-409)
"The ensemble spread is much lower (Table 3; MIROC4-ACTM columns) compared to the inversion predicted flux uncertainties, which are in the range of 1.4 and 0.7 PgC yr^{-1} for global land and ocean, respectively (vary from low values of 0.8 and 0.5 PgC yr^{-1} for gpp_v2 cases to 1.6 and 0.9 PgC yr^{-1} for the gpp_v4 inversions)."

For "gpp_v2", the uncertainty is as low as 0.8, close to the commonly reported values. Note that in gpp_v2 inversion PFUs for some regions are up to 2 PgC/yr, which is still higher than the TrransCom inversions.

We hope this clarifies your doubt. Sorry for the confusion.

Another possibility is that there is a problem with the posterior covariance matrix that they are using for the calculation: for example, if the correlations given by the off-diagonal elements were computed incorrectly. In my original comments (lines 183-187), I asked about one point that might lead to that covariance matrix to be calculated incorrectly: if the Green's function relating fluxes at a given time to mixing ratios at later times were to be truncated too soon. The authors responded that they run the Greens' functions for each flux pulse out for four years. This is long enough to capture the spread of the input pulse to the point where it have negligible latitudinal gradients and is typical what has been done in previous inversions. However, what the authors do not address (and what could cause problems that might result in an incorrect covariance matrix) is what is assumed for the influence of those fluxes at times after those four years: is zero influence put in the matrix (bad), is the fully spread-out value of about 0.4 ppm / (PgC/year) put in (better), or some exponential decay to the spread out value (even better)? The authors point to the original Rayner et al code that they have modified for use here, but don't explicitly address this issue. If they put zeroes in matrix J for all years after Year 4 instead of a better spread-out value, I could see how the correlations in the posterior covariance could be too low and the a posteriori uncertainties would be wrong. The original Rayner et al code may have been used for only a short span (four years) such that this point would not have been an issue. As things stand at the moment, I do not have confidence that the posterior covariance, upon which so many of the uncertainties discussed here rest, is being calculated correctly.

Greens functions decay nicely to the common value after 47 months in MIROC4-ACTM forward simulation of the monthly pulse functions, and we have kept it constant afterwards in the inversion code. We have not applied "exponential decay to the spread-out value", but that will preassembly of minor importance for CO_2 as it has no chemical loss present in the atmosphere. We only anticipate minor decay after four years due to the mixing through the whole atmosphere by slow transport from troposphere to stratosphere.

As stated earlier, we think the flux uncertainties are calculated correctly but the presentation varies based on whether or not temporal covariances are included in the annual/long-term mean flux uncertainties.

In Section 2.4, we add following line for clarification (line #211).
The elements of **J** for later months are kept constant at the value of 48th month.

Another issue about which I asked for clarification in my original comments (lines 70-81 and 694-967) was why an additional forward run with the optimized fluxes was needed (i.e., why the impact of the fluxes on concentrations through the J matrix did not fully represent the effect). The authors give some vague generalities but do not explicitly say what the answer is. They also delete some text. However, it is possible that the answer (not given) is that there is no influence whatsoever after four years from a given set of fluxes (i.e. zeroes in the J matrix after Year 4). As mentioned above, this would have an impact on the a posteriori covariances calculated.

We are not using all the measurements in the inversion system (only 50 sites are used), as measurement data gaps produce artifacts in the flux interannual variability. The forward simulation of the fluxes is needed for validation of the a posteriori fluxes, in particular using the large numbers of aircraft (or satellite) measurements. (It is only in the data assimilation system that we can get full 4-D concentration field, but not in the case of batch inversion in our case).

We calculate and store the J-matrix only for selected fixed sites, which could be used in inversion. Thus when the a posteriori fluxes are to evaluated with independent measurements (those not used in inversion), we need to simulate the CO_2 concentrations using the inversion corrected fluxes. This is done for GCP-CO_2 submissions of MIROC4-ACTM (since 2018), or other multi-model assessments (Gaubert et al., BG, 2019; Long et al., Science, 2021).

Some of the texts were deleted for cleaning up the discussion in the first submission relating to further development of inversion evaluation metric, which was found to be confusing to both the reviewers. Hope that we are not missing any significant text here.

I would need to have answers to these questions before I will feel confident in the results presented here. Errors in the J matrix would affect not only the uncertainties calculated in this manuscript, but also the flux estimates themselves, and could materially change some conclusions presented.

As we have shown in several figures and tables that the inversions are able to come to common solutions of a posteriori fluxes from extreme a priori fluxes, we have a strong feeling regarding the validity of the fluxes. Direct comparison of the ensemble means fluxes with IPCC-AR6 (Table 3) and RECCAP-1 fluxes (Table 4) also raise the confidence.