Thanks for the careful reading and insightful comments. We have carefully revised the manuscript accordingly.

The authors present an improved EnKF approach to consistently estimate atmospheric CO2 concentrations and surface fluxes from satellite and surface GHG observations. This approach is computationally very efficient and has been shown to be able to well reproduce the 'true flux' in OSSE experiments. Overall, this manuscript is clearly written, and the results are meaningful. But I think revisions are needed to address some concerns.

Major comments:

This approach estimates atmospheric CO2 concentrations and surface fluxes simultaneously. But I don't see much assessment of the quality of the resulting CO2 concentrations by comparing with the 'true' (model) atmosphere etc. I can see some benefits from additional constraint on global atmospheric CO2 mass on the a posteriori flux estimate. It is interesting to know how the imposed mass constraint will affect the horizontal and vertical CO2 distributions in a long (such as 1 or 2 years) run. Inconsistency is a potential concern, when adjustments from global atmospheric CO2 mass conservation (instead of the more solid mass balance between total surface emission/sink and atmospheric CO2 mass) are applied only on CO2 distributions but does not the flux distributions accordingly. Regions with poor constraints (such as the boreal Eurasia) can be used to 'dump' the mass imbalance of other better constrained regions, leading to degraded agreements with the 'true' fluxes (see for example the boreal Eurasia in Figure 9 & 10).
Response: Thanks for the critical and insightful comment. First, the main purpose of applying CEnKF is to improve the SCF estimation without influencing the CO_2 forecast. We show that the imposed mass constraint has a small effect on the CO2 distributions for each season (Fig. A1). As the reviewer points out that there are significant biases over the Eurasia boreal region. However, we can see that both EXP-L and EXP-LC show a relatively large CO_2 bias (Fig. A1), which indirectly shows that the CEnKF has less impact on the CO2 concentration over the Eurasia boreal region. And it is the LETKF that causes the bias. Although, the regional budget estimation of EXP-L is better over Eurasia boreal region (Fig. 11). This is mainly due to the large dipole deviation of EXP-L that reduce the regional budget error. And the estimated SCF of EXP-LC is better than EXP-L over the north-east of Eurasia.

In my opinion, deviations between a posteriori and the 'true' flux (see for example Figures 9 and 10) are still significant over many regions, in particular over, northern high latitudes. Our understanding of the global carbon cycle has been hindered by unquantified discrepancies in the posterior fluxes inferred by different top-down flux inversion models. I think, robustness is now more important than the computational speed.
Response: Thanks for the comments. I agree that there are significant deviations over some regions at the grid scale. Even though there are many inversion systems, including COLA, that estimate the flux at grid scale, but most of the systems have enough faith only on the continental scale like TRANSCOM/OCO2MIP regions. This is mainly because of the sparse distribution of the observation network. However, this is not saying that the grid point estimation is meanless. It reduces the aggregation error and gives much more insights. And most of the deviations are in a

dipole pattern. When aggregating the flux to the OCO2MIP regions, the precision is significantly improved (Fig. 10). Moreover, computationally fast means we can run at higher spatial resolution, which could reduce the transport error and improve the estimation.

It would be interesting to know whether the agreement with the 'true' can be further improved, for example, by using a longer window or using a larger ensemble etc. If possible, it is also interesting to know how the traditional top-down inversion will perform in those OSSEs. It become increasingly important to understand the discrepancies between different approaches.
Response: Thanks for the comment. We are also interested in improving the results and comparing COLA with other traditional inversion methods. During the last three months, we have discussed with the inversion community and got a lot of feedback. However, it is not practical for us to conduct traditional inversion, and it is also not the focus of this paper. To address the concern and to know the performance of COLA, we have conducted read-data assimilation experiments and submitted the results to the OCO2MIP. Some preliminary results of COLA are posted to the OCO2MIP official websites (https://gml.noaa.gov/ccgg/OCO2_v10mip). The preliminary comparison shows that COLA is well consistent with the MIP ensembles (CT, CAMS, CMS-Flux, etc.), which indirectly validate our method and robustness of COLA system. We believe the further analysis of COLA among the MIP could help us improve the results and give more insights on how the COLA method compared with traditional methods.

Minor comments:

As pointed out by other reviewers, some sentences are ambiguous or poorly structured. There are also a lot of typos for example: Line 304, Page 14: 'the accululation of the annual global imbalances …' I think the manuscript will benefit from a careful revision.
Response: Thanks for the suggestions. We have carefully revised the manuscript. And it has been polished by a native speaker from AJE (America Journal Expert).

No prior or posterior uncertainties presented in most of Tables and Figures such as Figures 5 and 6, and 9. Uncertainty is an important part of data assimilation product, which are particularly useful for us to assess whether improvements are substantial.
Response: Thanks for the suggestions. The error bars are added to Figure 9 and 10.

Figure 10. I'd like to see the difference shown as percentage of the true fluxes.
Response: Thanks for the suggestion. Because there are many grids with a very small value, even the difference is also small, the percentage of the true fluxes could be very large, such as in Northern Africa (Fig. R1). We prefer showing the difference in the main text.

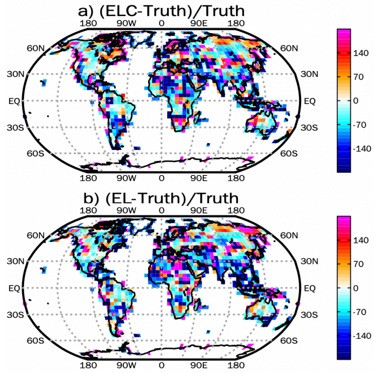

Figure R1: The percentage of the difference of EXP-LC and EXP-L compared with the truth.

Figure 11. Will the increments improve or degrade the agreement with 'true' model concentrations? Response: Thanks for the comment. The CEnKF increment has a small impact on the model concentration at the seasonal scale, that it is hard to say whether the CEnKF will improve or degrade the agreement (Fig. A1). And we do not expect an improved model concentration since the purpose of CEnKF is to improve the SCF.