Response to Referee #1

1 General Comments

The authors provide a novel re-imaging of the canonical RIP technique in the LETKF through the no-cost smoothing approach in order to provide a better spin-up of the ensemble accuracy. The method is used to estimate surface carbon fluxes through a sophisticated model using real-world observations. Quantifying Earth's is an essential part of our modern understanding of the Earth's climate.

The approach is certainly novel and has the potential to be applicable to a wide range of geophysical—and large-scale in general—data assimilation problems.

Thank you very much for your constructive and insightful comments.

2 Specific Comments

Section 3: It is not immediately clear what error is being measured, and why it is not a time-averaged RMSE.

Thank you very much for the comment. We did use the time-average RMSE. The RMSE value in Table1 and the sub-titles of Figures 3 and 5 are the time-averaged global mean RMSEs. We clarify this in the revised manuscript.

Section 2.2 talks about throwing away 'low-quality' observations. As the authors no doubt know, the Burns Effect places a significant burden on making sure that no unique correct observations are discarded. The reviewer would like to see more justification for this sort of heuristic.

Thank you very much for the comment. The approach to aggregate OCO-2 has been described by Basu et al 2018. We have included a brief description and cited this reference in the revised manuscript.

The observation network from Basu et al 2018 had used to produce pseudo-observations that are somehow statistically representative of real OCO2 data. Since we conducted our study under the OSSE framework, these observations are idealized. The side effects in the real OCO-2 dataset generated by aggregating observation or throwing away 'low-quality' observations will not be carried over to our pseudo-observations.

In Section 2.2 again, the reviewer would like to see more justification of this type of pseudo-observation.

See response to previous comment.

In section 2.5: As far as the reviewer can tell the paper never explicitly mentions the size of the component space (nor frankly any roughly estimated dynamical properties of either the system or the model) of the model, thus it is impossible to gauge the sufficiency of the ensemble size.

Thank you this comment. We have clarified in the revised manuscript that the experiment (system) include one state variable (co2) and one parameter (surface carbon flux) at each grid point. When we talk about ensemble size, one must consider both sets of variables. Given that the choice of ensemble size is determined heuristically, with due consideration to model-assimilation system complexity and required computational time. We did evaluate the ensemble size through a series of

simulations and determined that a sample size of 20 is reasonable for our assimilation experiments because the similar experiment but with 80-member ensemble size showed only slight improvement of assimilation quality (figure not shown) but dramatically increased the computational cost.We have clarified this point in the revised manuscript.

In Section 2.6 Additional clarification about additive inflation being randomly selected from the nature run would be appreciated.
Thank you for this comment. We have added more details on the additive inflation method in the revised manuscript.

In the conclusion the authors talk about advantages over that of 4D-LETKF, but omit to mention vanilla 4D techniques, which are still state-of-the-art, and against which such computationally intensive smoothing would have to compete.
Thank you very much for the useful comment. We do agree that 4D-LETKF is a state-of-the-art technique, though it is not our focus in this paper. We have added some discussion related to 4D technique, as per your suggestion, to the revised manuscript.

3 Technical Corrections
The overall document could use some basic proofreading to address fundamental grammatical and lexical issues. A non exhaustive list:

- p11 l1: LETKF
- p11 l11: LETKF
- p12 l5: the first two months
- Honestly all of section 2.6

Thanks for comments. We have made these changes in the revised manuscript.

Response to referee 2

It was really fun to read this informative manuscript which well describes its goal and methodologies. Authors introduce interesting methodology to use different length of observation window (OW) from that of assimilation window (AW) for estimating surface carbon fluxes (SCF) which does not have enough observations to be well constrained.

Thank you very much for your constructive and insightful comments.

However, it would be great to improve the manuscript responding to the following points. 1) This study does not assimilate other available observation datasets of atmospheric CO_2 such as GV+, GOSAT, etc. Authors need to explore a possible sensitivity of AW/OW lengths to the observation density. Since the current experiments includes column mixed OCO data only, you may need much longer OW. If you include more observations like GV+ (direct information, not like column-mixed information) and GOSAT, it may results in quite different RMSEs from AW/OW length experiments (Table 1). One can guess that you may need much shorter length of OW in the case with more observations including direct in-situ CO_2 concentration data. Also, this study incorporates very low resolution of the numerical model. Increasing model resolution increases the number of unknowns while you can use much dense remote sensing data (with proper thinning/superobing). In that case, the ratio of GV+ data contents to column mixed remote sensing data contents would drop, and then there would be another possible sensitivity of AW/OW lengths.

Thank you very much for your thoughtful suggestion. We did test the proposed assimilation system with data from GV+ network, and found consistent results. The optimal AW/OW length is still 1 day/7 days, and very stable. We have included the results in the revised manuscript.
We speculate the optimal length of OW is mainly decided based on the time scale of model response to the SCF signal.

2) In addition, the horizontal localization scale sets too small (150km) although the horizontal resolution of the model is very coarse. If it is not just typo, the exceptional setting of horizontal localization scale will cause high frequency errors of SCF estimates with 6-hr AW. Therefore, authors should check whether the conclusion is still valid with reasonable setting of the horizontal localization scale (_1000-2000km). This reviewer doubts that greater horizontal localization scale may give good enough SCF results even with 6-h AW.

Thank you very much for pointing out the typos which are fixed. The horizonal localization scale is 15000km. We tested the sensitivity of the assimilation system to the horizontal localization scale and found the extremely large horizonal localization perform best for the coarse observation coverage situation like OCO2 GV+.

3) Experimental setting includes slowly varying parameters, SCF that have only seasonal variation without diurnal cycle. Authors need to explain whether this long OW will be good for estimating SCF that fluctuates from day to night every day.

Thank you very much for this constructive comments. Our approach works best for the slowly varying parameters. It is not optimum for estimating SCF variation for sub-daily to daily time scales because it smooths out those variations due to long OW. In other words, the OW should be shorter than the signal time scale. We have added some text to clarify this point in the revised manuscript.

[Specific comments]
1. p.7, line 5: “at every land grid point” means authors only correct SFC over the land? not ocean? Please clarify it.

Thank you very much for pointing out this typo. We have corrected this and it should be for every grid point.

2. p. 7: Since this study set the horizontal resolution of the model very low, the observation data of OCO-2 were aggregated. Please give more detailed explanation about how to aggregate the observations.

The approach to aggregate OCO-2 has been described by Basu et al 2018. We have included a brief description and cited this reference in the revised manuscript.

3. p. 9, lines 19-20: A regular 4D-LETKF has 1.5 times longer forecast than the assimilation time window. e.g. if you have 6-h cycles of 4D-LETKF, you need 9-h forecast. Please correct this sentence.

Thank you very much for the comments. We have remove the confusing sentence from the manuscript.

4. p.10, lines 15-16: This reviewer cannot fully agree with the statement about the sensitivity of enKF DA to the ensemble size. It would be great to give any reference to support this statement, or to modify it carefully.

We agree, and edited the text based on our results. We did do the same test with 80-member, and the estimation results showed only slight improvement. We have stated this in the revised manuscript.

5. p.10, line 20: When authors use more than 400km horizontal resolution, a horizontal localization radius should be about 1500 km as a standard deviation of Gaussian localization function. This reviewer hopes that 150km is just typo. Otherwise, please seriously answer the major comment 2 above.

Please see the answer to (3)

6. p.12, line 20: Does the experiments include a diurnal cycle? If not, please correct the sentence from "mainly on" to "only on", or appropriately.
The assimilation experiments did include diurnal cycle for meteorological drivers as well as the fossil emission flux component of "truth" flux datasets. We think that it is too ambitious to claim that the diurnal cycle can be resolved . We have revised the manuscript accordingly

7. p.13, line 5: "deviations of estimates from the "truth" incases" cannot be clearly found from the figure.

Thank you very much. We have edited the text to remove the statement

8. Figure 6: Please give more detailed information that you show as a result. How did you define summer and winter (which months are they)? In addition, agreement of your estimates with true state looks amazing. But, it would be great if you additionally show how far your prior states of SCF were at the very initial time.
The summer and winter is JJA and DJF. It has listed on the figure subtitles. We will also add corresponding descripting in the revised manuscript.
The initial ensemble is random picked from model nature run. SO SCF prior (ensemble mean) at initial time is around the annual mean "truth" SCF. The ensemble spread is around the seasonal cycle of SCF .

9. p.18, lines 25-27: This statement needs to be modified carefully. The new assimilation method can be useful for the parameter estimation with EnKF when the observations are too limited to constrain the parameters well and the parameters have slow and smooth variation in time and space, respectively. For example, if your parameters have very rapid temporal variation, long OW may not work well as the SCF case in this manuscript. In that sense, the statement should be revised.

Thank you very much for this constructive comment. We have edited the text accordingly as "It is worth noting that our approach works best for estimating parameters that vary slowly over moderate time scales. It is not optimum for estimating SCF variation for short time-scales such as sub-daily to daily because the variations shorter than OWs are filtered out. "

[Technical corrections]
1. Figures does not have subtitle of (a), (b), etc, although authors explain the subfigures in that way. It would be good to explicitly mark them.

Thank you very much. We have added the subtitle for the figures accordingly.