Referee #1

We would like to thank the anonymous referee for his/her comprehensive review and valuable suggestions. These suggestions help us to present our results more clearly. In response, we have made changes according to the referee's suggestions and replied to all comments point by point. All the page and line number for corrections are referred to the revised manuscript, while the page and line number from original reviews are kept intact.

General comments.

Authors present estimates of regional carbon dioxide flux variability based on assimilating GOSAT satellite observations of CO_2 with ensemble-based data assimilation system. The estimated CO_2 fluxes where evaluated by comparison to indexes of climate variability, and published top-down and bottom-up estimates. The analysis of the carbon cycle variability and comparison with data on climate variability makes a strong point of the study. On the other hand, the description of the ensemble-based data assimilation system can be improved. The paper is well written and can be accepted after minor revisions addressing the review suggestions.

Detailed comments.

Lines 130-139 Suggest clarifying, what becomes a state vector to optimize, currently it is implicit. Some details emerge much later on Lines 358-366, when uncertainties are discussed.

Response: Thank you for this suggestion. In this study, the terrestrial ecosystem (BIO) and ocean (OCN) carbon fluxes are treated as state vector and optimized. Indeed, as you said, the state variables had been mentioned in two places in the article. The first place is in the section of system description, and the second is in the section of "Experimental Design". In the first place, we are introducing the current system (GCASv2) that we have improved, we set 4 state vector schemes in this system for different applications: 1) only the BIO flux is state vector; 2) both BIO and OCN fluxes are treated as state vectors; 3) the BIO, OCN and FOSSIL fluxes are optimized at the same time; and 4) only net flux is optimized. In this study, we chose to optimize both BIO and OCN, which were introduced in the section of "Experimental Design". To further clarify the state vector of this study, we added a sentence of "*In this study, the second scheme was selected.*" at the end of the 2nd paragraph in section 2.1 (see Line 178, Page 7).

Lines 232-236 The logic behind selecting 1-week data assimilation window doesn't look solid, as the other ensemble-based assimilation systems use longer window in order of 12 weeks, (Peters et al. 2005, Feng et al. 2009, Jacobson et al. 2020. The notice that there was a problem reproducing CO_2 growth rate with a longer window in Zhang et al (2015) doesn't look like a strong argument, if considered in comparison with other studies.

Response: Many thanks for this suggestion. We have added more discussions about the assimilation window, and shown the mean observation (only GOSAT XCO_2) number (Figure S2) during the study period that each grid could have within the 1 week assimilation window and the 3000 km localization scale. We also conduct a test in the year of 2010 for different DA windows (1, 2 and 4 weeks) and evaluate the posterior results using surface observations (see Table 1). We have revised that paragraph (see Lines 303-307, Lines 309-340, Pages 11-12) as follows:

"The DA window is set to one week in GCASv2, which is the same as before. Theoretically, a longer DA window is better, because CO_2 is a stable species. The longer window, the farther CO_2 will be transported. In this way, more observation stations will sense the flux change of one area, and thus more observations can be used to optimize the flux of that place. Therefore, many previous ensemble-based assimilation systems used a longer DA window (e.g., Peters et al. 2005, Feng et al. 2009, Jacobson et al. 2020). However, the farther away, the weaker signal the stations can sense. Bruhwiler et al. (2005) clearly shown that a pulse traveling from a faraway place would contribute relatively little signal compared to recent pulses from nearby source regions. In addition, Limited by the method of EnKF, this weak signal will be masked by the method's own unphysical signal (spurious correlation), and in order to reduce this influence, we must increase the ensembles, thereby greatly increasing the computational cost. Miyazaki et al. (2011) tested the differences of 3 days and 7 days DA windows, and pointed that with a longer DA window, more observation data will be available to constrain the surface flux, but a longer window can make the effect of model error more obvious. Thus, the assimilation result can be improved as long as the observations with spurious correlations can be neglected. However, spurious correlations can be more serious with increases in the DA window, because of a limited number of ensembles. As a result, a longer window is not necessarily better than a shorter window system. To avoid the influence of spurious signals, Kang et al. (2012) used a very short DA window (6 hours) in their assimilation system (LETKF_C) and pointed out that the flux inversion with a long window (3 weeks) is not as accurate as the one obtained with a 6 h DA window, particularly in smaller-scale structures. During the development of GCASv1, Zhang et al. (2015) tested different DA windows and found that the longer the window, the larger optimized terrestrial carbon sink will be, resulting in a smaller optimized annual atmospheric CO_2 growth rate as compared to the observed rate. Considering the fact that at present, due to the release of satellite XCO_2 retrievals like GOSAT and OCO-2, the atmospheric CO_2 observations and coverages have increased significantly compared to before, which means that we do not need to extend the DA window to include more observation data now. Figure S2 shows the mean super observation (see section 2.1.1, only GOSAT XCO_2) numbers during the study period that each grid could have within the 1-week DA window and a localization scale (3000 km, see the next paragraph). In most land areas and pan-tropical waters, each grid can already have more than 3 super observations. On average, each grid over the land could has 4 super observations. Two sensitivity tests in 2010 were conducted using 2- and 4- weeks DA windows but the same localization scale, the results are shown in Table S3. When the length of DA window increases from 1 week to 4 weeks, the mean super observation number increases from 4 to 9, accordingly, the inverted global BIO flux increased from -4.16 PgC yr^{-1} to -4.49 PgC yr^{-1}, resulting in a larger deviation of the simulated and observed atmospheric CO_2 growth rate (AGR) and larger simulation error against the surface observations. Therefore, we still use the 1-week DA window in GCASv2."

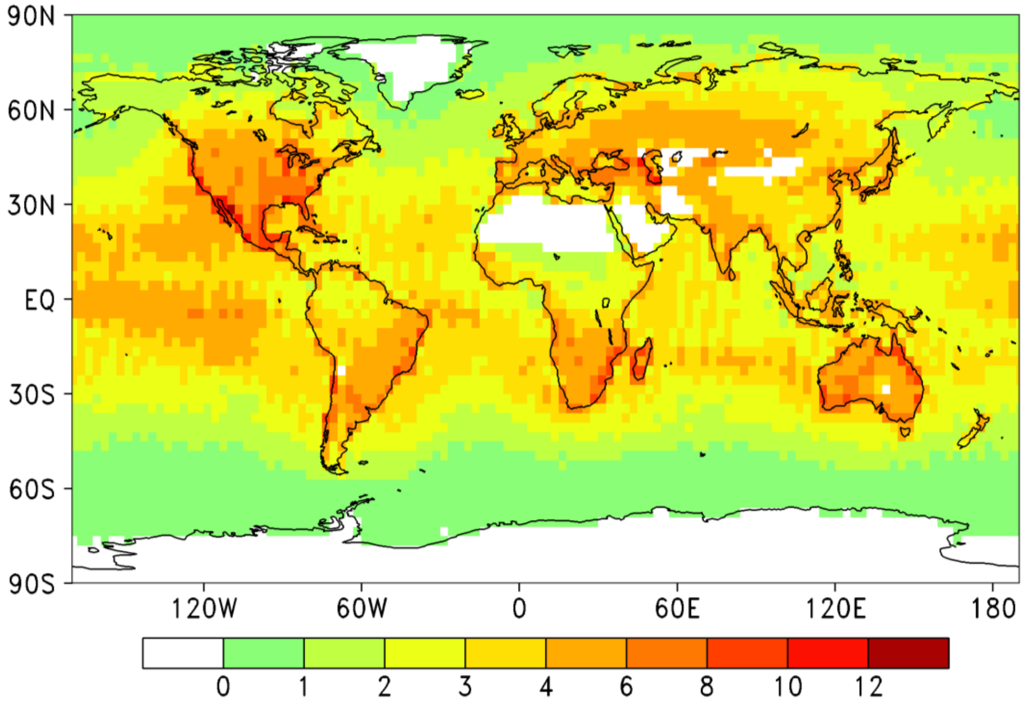

Figure 1. Mean observation numbers within a DA window (1 week) during May 2009 ~ Dec 2015 (This figure has been added in the revised Supporting Information, and named as Figure S2)

Table 1. Results of sensitivity tests in the year of 2010 (1week, 2weeks and 4weeks are three additional experiments using 1 week, 2 weeks, and 4 weeks assimilation windows, respectively) (This Table has been added in the revised Supporting Information, and named as Table S4)

		Prior	1 week	2 weeks	4 weeks
Super Obs. Num. per window	Total	-	730	1039	1360
	Each grid over land	-	4	6	9
Global Flux (PgC/yr)	BIO	-2.07	-4.16	-4.46	-4.49
	OCN	-2.08	-2.33	-2.32	-2.35
	FOSSIL	9.07	9.07	9.07	9.07
	Net	7.25	4.91	4.62	4.55
Regional Flux (PgC/yr)	North America Boreal	-0.29	-0.43	-0.41	-0.35
	North America Temperate	-0.42	-1.25	-1.75	-2.41
	Tropical South America	-0.17	-0.26	-0.32	-0.27
	Temperate South America	-0.24	-0.4	-0.36	-0.19
	Northern Afirca	0.21	0.32	0.36	0.62
	Southern Africa	0.22	-0.3	-0.59	-1.04
	Boreal Asia	-0.4	-0.46	-0.3	0.11
	Temperate Asia	-0.3	-0.29	-0.15	-0.06

	Southeast Asia	-0.29	-0.23	-0.21	-0.2
	Australia	-0.17	-0.4	-0.48	-0.53
	Europe	-0.19	-0.41	-0.21	-0.12
independent evaluation	BIAS	1.43	-0.44	-0.4	-0.38
	MAE	1.92	1.37	1.39	1.51
	RMSE	2.36	2.11	2.18	2.39
Deviation from the observed AGR ($PgC\ yr^{-1}$)		2.08	-0.26	-0.55	-0.62

Technical corrections

Lines 119-120 Need to clarify, written that fluxes "are perturbed with a Gaussian random distribution" – better add more detail on whether perturbation is applied independently to each grid or over regions.

Response: Thank you! We have rewritten that sentence (see Lines 123-124, Page 5), as follows:

"the prior fluxes of X^b in each grid are independently perturbed with a Gaussian random distribution"

Line 216 As resolutions of the transport model and fluxes are apparently different, suggest writing which of them are referred as 'model grids'.

Response: Thanks for this suggestion. We have changed 'model grids' as 'transport model grids' (see Lines 268 and 269, Page 10).

Line 584 Revise 'a very stronger carbon sink' as 'a stronger carbon sink' or 'a very strong carbon sink'

Response: Thanks! We have changed 'a very stronger carbon sink' as 'a very strong carbon sink' (see Line 857, Page 29).

Line 594 Suggest revising 'weak' to 'weaker'

Response: Thanks! We have changed 'weak' to 'weaker' (see Line 870, Page 30).