

Interactive comment on “A regional carbon flux data assimilation system and its preliminary evaluation in East Asia” by Z. Peng et al.

Z. Peng et al.

pengzhen@nju.edu.cn

Response to Reviewer #2's comments:

We are very grateful to the reviewer for his helpful and careful reviews. The thoughtful comments (in bold style below) have helped us improve the manuscript greatly. In particular, following the reviewer's suggestion, we have changed many expressions what are improper, supplemented a paragraph to address the initialization of the OSSEs, and rewritten most of the summary to emphasize the newness in our work.

1. Paper Title: you are not only doing CO₂ flux, but also doing CO₂ initial condition. Should the title reflect this?

Yes, CFI-CMAQ can optimize both the surface CO₂ fluxes and CO₂ initial conditions by assimilating observations. So we changed the title of this paper as: A Regional Carbon Data Assimilation System and its Preliminary Evaluation in East Asia.

2. P20346, Line 2: what do you mean by “finer scales”? Your OSSE is done with 64km x 64 km resolution, hard to justify “finer scales”. Maybe just say “grid scales”.

We want to develop a regional surface CO₂ flux inversion system to optimize the surface CO₂ fluxes. Compared to the surface CO₂ flux inversion system based on global model, the regional surface CO₂ flux inversion system may has finer scales. But in this manuscript, our main task is to optimize the surface CO₂ fluxes at grid scales. So, “grid scales” is better than “finer scales” and we changed this expression in

Line 1, Page 2.

3. Line 4: “simultaneously assimilating”. Should use “analyzing”. The word “assimilation” should be applied to observations, not quantities to be analyzed.

4. Line 4: “simultaneously”. This is not really true because flux and concentration are analyzed sequentially. It is more appropriate to directly say here that EnKF for analyzing CO₂ initial conditions, EnKS for analyzing CO₂ flux.

We changed the expression as: a regional surface CO₂ flux inversion system (Carbon Flux Inversion system and Community Multi-scale Air Quality, CFI-CMAQ) has been developed by applying the ensemble Kalman filter (EnKF) to constrain the CO₂ concentrations and applying the ensemble Kalman smoother (EnKS) to optimize the surface CO₂ flux. See details in Line 1 to 6, Page 2.

5. P20350, Line 23: GOSAT XCO should read “XCO₂”?

Yes. We have corrected this mistake in the revised manuscript in Line 6, Page 7.

6. P20351, Eq. (1): The statement here is confusing with the OSSE part, where F₀ is referred to as the truth. Overall, superscripts/notations used in the paper are confusing, not all consistent. “prior”, “background”, and “forecast” typically represent the same thing in the data assimilation framework. Also, does surface flux have vertical variation? If not, it should just be the function of (x,y,t), not (x,y,z,t). If so, it needs to be stated clearly.

The superscripts/notations used in the first draft of our manuscript were not all consistent really. In the revised version, they are standard. $F^*(x, y, z, t)$ (refer to as F_t^*) was served as the prescribed net CO₂ surface flux in formula (1) in Page 7 and the corresponding symbol has been changed. In this study, it was generated by formula (25) (Page 17). In addition, the superscript p, f, and a are standard.

Among them, the superscript p refers to the prior. It was used in the following variables:

① $F^p(x, y, z, t)$ (refer to as F_t^p): the prior surface CO₂ flux. It was generated by Eq. (24) (Page 16) in this study. In all the OSSEs in this study, F_t^p was assumed as the true surface CO₂ flux.

② $\lambda_{i,t/t-1}^p$: the prior values of the linear scaling factors. We have addressed the way to generate $\lambda_{i,t/t-1}^p$ in Line 16, Page 9 to Line 3, Page 10.

③ $C^p(x, y, z, t)$ (refer to as C_t^p): the artificial true CO₂ concentration fields. Forced by F_t^p , the RAMS-CMAQ model was run to produce the artificial true CO₂ concentration fields C_t^p from 1 January 2010 to 30 March 2010. It was addressed in Line 18 to 20, Page 16.

④ $X_{CO_2}^p$ or y_t^{obs} : the artificial GOSAT observations, which were generated by substituting C_t^p into Eq. (19). It was addressed in Line 20 to 21, Page 16.

The superscript f refers to the forecast or the background. It was used in the following variables:

② $\hat{C}_i^f(x, y, z, t)$ (referred to as $\hat{C}_{i,t}^f$): which was generated by applying CMAQ to integrate from time $t-1$ to t forced by F_t^* with $C_i^a(x, y, z, t-1)$ as initial conditions. It was used to generate $\lambda_{i,t}^p$. It was addressed in Line 17 to 20, Page 9.

$$\textcircled{2} \quad \overline{\hat{C}_{i,t}^f} : \overline{\hat{C}_{i,t}^f} = \frac{1}{N} \sum_{i=1}^N \hat{C}_{i,t}^f$$

③ $C_i^f(x, y, z, t)$ (refer to as $C_{i,t}^f$): the i th ensemble member of the background concentration fields. CMAQ integrates from time $t-1$ to t forced by $F_{i,t/t-1}^a$ with $C_i^a(x, y, z, t-1)$ as initial conditions. It was addressed in Line 21 to 22, Page 10.

$$\textcircled{4} \quad \overline{C_t^f} : \text{the ensemble mean of } C_{i,t}^f. \quad \overline{C_t^f} = \frac{1}{N} \sum_{i=1}^N C_{i,t}^f.$$

⑤ $C^f(x, y, z, t)$ (refer to as C_t^f): the background (wrong) CO₂ concentration

fields. Forced by F_t^* , the RAMS-CMAQ model was run to produce these CO₂ concentration fields from 1 January 2010 to 30 March 2010. That was addressed in Line 17 to 19, Page 17.

⑥ $X_{CO_2}^f$: the column-averaged concentrations of C_t^f at the GOSAT X_{CO_2} locations, which were generated by substituting C_t^f into Eq. (18). It was addressed in Line 19 to 20, Page 17.

The superscript *a* refers to the analysis. It was used in the following variables:

① $\lambda_{i,j|t-1}^a$: analyzed quantities from the previous assimilation cycle at time j , $|t-1$ means that these factors have been optimized by using observations at time $t-1$.

② $F_{i,j|t-1}^a$: analyzed fluxes from the previous assimilation cycle at time j .

③ \overline{F}_t^a : the ensemble mean values of the assimilated fluxes, which are before the next smoother window and will not be updated by the succeeding observations. We regarded them as the final optimized fluxes. It was addressed in Line 11 to 13, Page 12.

④ $C_{i,t}^a$: the i th member of the assimilated CO₂ concentrations fields.

⑤ \overline{C}_t^a : the ensemble mean values of the assimilated CO₂ concentrations fields, which is regarded as the final analyzing concentration field.

⑥ $X_{CO_2}^a$: the column-averaged concentrations of \overline{C}_t^a at the GOSAT X_{CO_2} locations, which were generated by substituting \overline{C}_t^a into Eq. (18).

Besides, the surface fluxes have vertical variation. We explained it in Line 12 to 16, Page 16.

7. Line 12: “exchanges”, why not use the word “fluxes”. It sounds like you are talking about a different quantity.

We have changed this expression in the manuscript.

8. P20352, Eq. (2) uses “M”, but figure 2 uses “M+1”.

We have corrected this equation.

9. P20353, Line 4: not very clear what is the “signal-to-noise” problem, and how it is resolved in this study.

The difference between our dynamical model and the one used in CarbonTracker (Peters et al, 2007) is in the way to set values for $\lambda_{i,t|t-1}^p$. In CarbonTracker, all $\lambda_{i,t}^p$ are set to 1. So the persistence dynamical model is only the smoothing operator. In our study, the CO₂ ensemble forecasts of the atmospheric transport model are used to calculate the values for $\lambda_{i,t|t-1}^p$. So the persistence dynamical model in our study is associated the smoothing operator with the atmospheric transport model. We have discussed this difference briefly in Line 4 to Line 14, Page 10. Besides, we designed another OSSE to illustrate the limitation by only using the smoothing operator as the persistence dynamical model to generate all future scaling factors in Line 9, Page 19 to Line 6, Page 20. Then we discussed the assimilated results in Line 22, Page 26 to Line 3, Page 28.

We have addressed how the “signal-to-noise” problem arises of the reference OSSE in Line 4, Page 27 to Line 16, Page 27. And then explained how it is resolved by describing the way $\lambda_{i,t|t-1}^p$ are updated by associating with the atmospheric transport model in CFI-CMAQ in Line 21, Page 27 to Line 3, Page 28. Please see details in the revised manuscript.

10. P 20358, Line 20: does simulated observations consider observation error?

The artificial observations $X_{CO_2}^p$ used in this study did not have observation errors though the measurement errors are set to 1.5 ppmv in the EnKS and EnKF updating equations. We have considered observation error when generated artificial observations before. But when we assimilated these artificial observations with observation errors, we cannot get effective assimilation results. That is to say, the impacts of assimilating artificial observations with observation errors on CO₂

simulations and surface CO₂ fluxes are negligible. When we compared the values of $X_{CO_2}^p$, which have no observation errors, with $X_{CO_2}^f$, it showed that the maximum values of $X_{CO_2}^f - X_{CO_2}^p$ can only reach 2 ppmv in the east and south of China (see Fig. 3e) though F_t^* is about 1.8 times as F_t^p and the magnitude of the difference between C_t^p and C_t^f was at least 6 ppmv at model level 1 in the east and south of China (see Fig. 4d). While in most model domain, the magnitudes of the difference between $X_{CO_2}^p$ and $X_{CO_2}^f$ are less than 0.5 ppmv (see Fig. 3e). So if we add errors (1.5 ppmv) to $X_{CO_2}^p$ to generate the artificial observations, the errors are too strong to extract the effective signal. However, at this stage, the uncertainties of the ACOS GOSAT X_{CO2} retrievals range from 0.7 to 1.5 ppmv (Osterman et al., 2011). So further works are needed to assimilate satellite retrievals with so large errors. But in this study, we had to neglect the observation errors when generate artificial observations.

11. Eq. (22): do you need ensemble of Fb, like in Eq. (1)? Not clear how initial ensemble was created for EnKF/EnKS.

Yes. We have changed the symbols in the revised manuscript. The ensemble of the fluxes are calculated in Eq. (1).

In CFI-CMAQ, only the ensemble of background concentration fields $C_i^f(x, y, z, 0)$ need to be initialized at $t = 0$. We supplemented a paragraph to address the initialization of the OSSEs in Line 14, Page 18 to Line 8, Page 19.

12. Line 25: “random number” needs to be more specific, distribution, mean, variance etc.?

δ was a standard normal distribution time series at each grid in the integration period of our numerical experiment. We addressed this in Line 16 to 17, Page 17.

13. P20359, Line 5: why so big number “70” for Beta inflation factor? Any explanation?

The values of the ensemble spread of $\lambda_{i,t|t-1}^p$ before inflating are very small. (ranging from 0 to 0.08 in most area at model-level 1, see Fig. 11b). We addressed

that in detail in Line 20, Page 25, to Line 7, Page 26 and added Fig. 11 to illustrate.

14. “Lag-window”, is that same as “smoother window”?

Yes. But we have changed this expression in Line 8 to 9, Page 18.

15. Line 9: needs to specify the year of OSSE.

All the numerical experiments started on 1 January 2010 and ended on 30 March 2010. We have specified the year in the manuscript (see in Line 12 to 13, Page 18).

16. Better to state that the goal of OSSE is to retrieve the true flux F_0 from given true observations and “wrong” flux F_b .

We added this statement in Line 22, Page 17, to Line 1, Page 18.

17. What is the frequency for EnKF cycling? How frequent GOSAT data are available?

If there are some observations, CMAQ stop integrating, and the assimilation part start to assimilate the observations. In all the OSSEs, we assimilated artificial observations $X_{CO_2}^p$ about three times a day since GOSAT has about three orbits in the study model domain. We added this statement in Line 1 to 3, Page 17. Besides, we added some description of the ACOS GOSAT X_{CO_2} retrievals in Line 1 to 13, Page 17.

18. P20360, Line 18: “near” should read “close”; “trues” should be “true”.

We have changed the expression in Line 16, Page 23.

19. Line 24,25: no experiment was performed without EnKF step. It is not clear how you can separate impact of EnKF step (concentration analysis) and EnKS step (flux analysis).

The performance of the EnKF subsection will be greatly influenced by the validation of the EnKS subsection, or vice versa. We have addressed this statement in Line 3 to 6, Page 20. And in Line 20, Page 21, we corrected the expression: All the results illustrated that CFI-CMAQ can provide a convincing CO_2 initial analysis fields for CO_2 flux inversion.

20. Line 28: “prior true : : :”, strange wording.

We have corrected the expression in Line 2, Page 22.

21. P20361, Line 6: when you say Fa and Fb (and Ca and Cb), do you refer to the ensemble mean values? Need to be clearly stated.

Yes. In the revised manuscript, the symbols are standard. F_t^* is the first-guess net CO₂ surface flux. $\overline{F_t^a}$ is the ensemble mean values of the assimilated fluxes, which is regarded as the final optimized flux. See the description of other symbols in Question 6.

22. Line 20,21: from 0.5 to 0.65, here should point out that these values are consistent with F0/Fb=1.8+delta.

We address these statement in Line 1, Page 23.

23. Line 22, 23: ratios should be strictly related to 1.8+delta, why related to strong diurnal variation?

For a certain time, $F_t^p / F_t^* = 1/1.8 + \delta$. But for the ratios of the monthly mean F_t^p to the monthly mean F_t^* , we calculate like this: $\text{Ratios} = \frac{1}{n} \sum_{\text{Feb}} F_t^p / \frac{1}{n} \sum_{\text{Feb}} F_t^*$ (n is the number of the flux), which are equal to $\frac{\sum_{\text{Feb}} F_t^p}{\sum_{\text{Feb}} F_t^*} = 1 / (1.8 + \frac{\sum_{\text{Feb}} \delta F_t^p}{\sum_{\text{Feb}} F_t^p})$. So the values of the ratios are related to the ratios of $\sum_{\text{Feb}} \delta F_t^p$ to $\sum_{\text{Feb}} F_t^p$. We explained why the ratios are related to strong diurnal variation in Indo-China Peninsula in Line 18, Page 23 to 7, Page 24. Please see details.

24. P20362, Line 12-13: “: : excessive impact of assimilation”, this sentence is not clear.

We used a wrong expression. The assimilated time series were much smaller than the true time series in Beijing. In another words, CFI-CMAQ failed to show improvements at Beijing. One of the possible reasons is that the impact of advection transport of CO₂ is ignored during the procedure of CO₂ flux inversion. We addressed this in Line 1 to 13, Page 25.

25. P20363, Line 8: you state “: : similar to Kang et al. (2011, 2012) and

Tian et al. (2013)”, but not very clearly describe what is really new in your work.

We rewrote this paragraph. Please see the detail in Line 10 to 18, Page 28.

26. Fig 3: there are two (d).

We have corrected this mistake in Fig.3.

27. Fig 4, Fig 5, Fig 7, Fig 8, Fig 9: need to use better font for color bar to display.

We have edited all this figures. Please see details in the revised manuscript.