Atmospheric Chemistry and Physics Discussions

Atmos. Chem. Phys. Discuss., referee comment RC1
https://doi.org/10.5194/acp-2022-183-RC1, 2022

Comment on acp-2022-183

Anonymous Referee #1

Referee comment on "Global and regional carbon budget for 2015–2020 inferred from OCO-2 based on an ensemble Kalman filter coupled with GEOS-Chem" by Yawen Kong et al., Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2022-183-RC1, 2022

The authors present a new framework for performing atmospheric flux invesions. This framework is based on a four-dimensional local ensemble transform Kalman filter (4D-LETKF), with GEOS-Chem used as the transport model. The method is applied to version 10r column-averaged CO_2 retrievals from the OCO-2 satellite over the five year period from 2015 to 2020. The estimated fluxes were found to be broadly consistent with those from other flux inversion systems. This represents the first application of a 4D-LETKF algorithm to perform an atmospheric inversion using OCO-2 data.

I find the paper to be well written and clear in general. The method is novel because it combines a high-dimensional grid-based parameterisation with an ensemble Kalman filtering approach. I appreciate that the authors performed sensitivity experiments to better understand what is driving the results. My main critique is that I find the mathematical description of the method lacking. My secondary critique relates to the lack of a discussion of the advantages or disadvantages of the method in comparison to other inversion systems.

Major comments

- I found it difficult to understand the method based on Section 2.1. Here are my specific questions:
  - How are the ensemble members initialised?
  - The matrix B appears in equation (1) but not in equations (2)-(5). Is B the error structure for x^b that is mentioned in line 115? How does B affect the posterior state if it does not appear in the calculations?
  - What is contained within the vector x^b (and x^a, and so on)? Is it the control variables (scaling factors) for the whole assimilation period (7 days), or is it just for the first day? If it's just the first day, what are the implied values for the next 6 days, which will affect the modelled concentrations y^{b(i)}? Are these assumed to be equal to the prior mean? I looked at Figure~1 but I still could not understand

what was happening.
    - Related to the last point, the calculation of \bar{x}^b is described as "the average optimized result from the two previous time steps and a fixed value of one". Does this calculation apply to the new day entering the assimilation period, or to all the days?
    - The modelled concentrations y^{b(i)} must also depend on state values from before the assimilation period. Are these set to the posterior mean, or are they different from each ensemble member? What is assumed exactly?
    - I find the notation regarding \bar{x} a little confusing. Is this the unweighted average of the ensemble members? I ask because \bar{x}^a and \bar{x}^b are not unweighted averages, so the notation is a little bit inconsistent.
    - How does the localisation length work? It is stated that "y^o contains the assimilated OCO-2 XCO_2 within the assimilation window and localization length". Since the state vector contains every grid cell for a day, how can any observations be excluded by the localization length?
- I think it would help for the authors to discuss how their method compares to other methods. For example, a convention al 4D-Var system has a similar state space and a similar cost function. What, in the authors view, are the advantages of their method? I think just a short discussion of the most common methods and how they compare qualitatively to the authors method would be enough.

Minor comments

Line 151, what does the word ``integrated'' mean here? Does it mean that the flux field was shifted to have annual mean zero? How was this done?