Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2019-54-RC1, 2019

Geoscientific Model Development Discussions
Open Access

Interactive comment on "A dual-pass carbon cycle data assimilation system to estimate surface CO_2 fluxes and 3D atmospheric CO_2 concentrations from spaceborne measurements of atmospheric CO_2" *by* Rui Han and Xiangjun Tian

Anonymous Referee #1

Received and published: 28 June 2019

The manuscript "A dual-pass carbon cycle data assimilation system to estimate surface CO2 fluxes and 3D atmospheric CO2 concentrations from spaceborne measurements of atmospheric CO2" by Han and Tian discusses a two step global CO2 natural flux inversion approach applied sequentially to relatively short (14-day) inversion cycles partitioning the full period of analysis (1 year). For each inversion cycle, the first step consists in optimizing the CO2 initial condition, and the second step consists in optimizing the CO2 fluxes. The system seems to strongly rely on the choice of a shorter assimilation window (at the beginning of the inversion cycle) for the first step, the initial

Printer-friendly version

Discussion paper

condition being constrained by a subset of the observations, while the fluxes are constrained using all observations of the inversion cycle in the second step (for which the assimilation window is the inversion cycle).

The authors attempt at demonstrating the advantages of such an approach using OSSEs with the assimilation of pseudo OCO-2 data, and comparing the results of Tan-Tracker(v1) (that uses such an approach and the NLS-4DVar system) to that of Tan-Tracker (v0) (which uses the POD-4DVar system and which optimizes the CO2 initial condition and fluxes simultaneously for each inversion cycle, without splitting the inversion into two steps).

I see severe issues in this study, in particular:

- regarding the theory: I can not understand how the split of the inversion cycle into 2 steps ("passes") could be an improvement. It raises theoretical issues, at least regarding the assimilation of some data twice. And it should hamper the proper distinction between the errors from the prior initial conditions and from the prior fluxes. The need to use a shorter assimilation window for the first pass is an indication of this limitation. Controlling both the CO2 initial condition and fluxes together, using the transport model and the prior uncertainties to drive the balance between corrections to the initial condition and to the surface fluxes, should lead to more robust results and is much more satisfying in terms of theory. The "dual pass" could be seen as a pragmatic way of controlling "manually" this balance (by playing with the length of the assimilation window for the first "pass"), but refining the set-up of the prior uncertainties in the initial condition and in the fluxes is a much proper way for such a control.

- in practice: my understanding is that the comparison between Tan-Tracker(v1) and Tan-Tracker(v0) in section 3 is completely biased. For the direct comparison in section 3.3.1 and 3.3.2, TTv1 uses 14-day inversion cycles while TTv0 uses 7-day inversion cycles, and, more critically, TTv1 uses 3 iterations for the minimization of the cost function, while TTv0 uses 1 iteration only (they also use different localization radius and in practice, different systems asking for different parameters). Therefore, there is no reason to think that this comparison says something about the "dual pass" approach itself. Actually, TTv0 seems to provide results that are extremely similar to that of TTv1 using 1 iteration and 14-day inversion cycles (see Figure 5b vs.Figure 9b) ! One could even assume that it provides better results than TTv1 using 1 iteration and 7-day inversion cycles (since results are better with 14-day cycles than with 7-day cycles for TTv1), not speaking about using 2000 km localization radius. My understanding is thus that the authors have misinterpreted their experiments and results.

The use of very short inversion cycles (here 14-days) exacerbates the problem of the corrections to initial conditions. For CO2 inversions, the use of short inversion cycles can hardly be seen as an advantage. Most of state of the art systems, especially global ones, use very long inversion cycles (1 year and more) to avoid breaking the link between uncertainties in the fluxes in an area and the errors at a remote location a long time later (which have to be solved for in the initial condition when using short inversion cycles while the target of inversion is a better estimate of the fluxes). I guess that the ensemble approach explains the need for short inversion cycles and maybe why results with 14-day inversion cycles are better than with 30-day inversion cycles. However, the manuscript does not attempt at explaining it.

The actual inversion system (i.e. the NLS-4DVar system, over which lies the TanTracker (v1) framework, and which is the actual code proposed in the "code and data availability" section) has already been detailed in past publications involving the second author. The section 2.2 is just the duplication of material from Tian et al. (2018), Zhang and Tian (2017), Tian and Feng (2015) and even in Tian et al. (2011). It thus cannot be a strong topic of this new manuscript, nor the overall changes from Tan-Tracker v0 to Tan-Tracker v1 i.e. from POD-4DVar with a single "pass" to NLS-4DVar with a "dual pass". Regarding the specific analysis of this paper on this NLS-4DVar system, I see in Figure 9 that the optimal number of iterations found for the minimization of the cost function, when testing 1,2 and 3 iterations is 3. But 3 is still much smaller to the typical number of iterations usually used for such minimizations, and the experiments and analysis of this paper do not show that the minimization has converged after 3 iterations (Figure 9b even imply the opposite).

My opinion is thus this manuscript should be rejected.

It is important to note that the authors forget to say that their system is a global inversion system (and even to say that it inverts natural CO2 fluxes, until the details of the equations clarifies it) which should strongly influence the way the problem of the initial condition should be tackled, the choice of the inversion cycles and of the data assimilation windows, the size of the ensembles and the number of iterations used for the inversions. . . in a more general way, the authors ignore the influence of the specific framework of their inversions -domain, resolution, data assimilated- on their results and on their choices of values for the inversion parameters.

I add, without entering into too much details about it, that the quality of the text is not sufficient for a scientific publication. The abstract already gives a good illustration of the confusing way with which this manuscript is written. The authors insert a lot of technical jargon from the data assimilation community, but they actually misuse many of the corresponding terms (few examples: "surface flux inversion measurement", "flux assimilation" to speak about flux inversion assimilating CO2 data, the alternative use of "background" or "prior" to speak about the same thing, "one-step iteration", "atmospheric chemical transmission mode", "ensemble-based hybrid assimilation algorithm". . .), leading to meaningless of confusing sentences. The abusive use of words and mathematical notations that seem more complicated than needed (or that they just forget to define, such as Py in eq 15) severely hampers the clarity of the paper. Most of the introduction sounds like a random sampling of references to past inversions, with meaningless comments (like "the surface carbon flux inversion method, obtained by combining model and atmospheric CO2 information, has made great progress in carbon cycle data assimilation", "For example, CarbonTracker is a well-designedÂăcarbon assimilation system", "Basu et al. showed that satellite data provided an effective constraint for surface carbon source-sink inversion"...). It hardly provides clues about the specific topic of this paper. The analyzes in section 3.3 lack of depth and of hindsight on the significance and scope of the OSSEs and of the conclusions.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2019-54, 2019.