Authors comments on ‘Technical note: A high-resolution inverse modelling technique for estimating surface CO2 fluxes based on the NIES-TM – FLEXPART coupled transport model and its adjoint’

Authors are grateful to reviewers for work they have done reading the paper and preparing the comments. The comments were very useful for realizing the deficiencies in the manuscript and the presented study and will serve as expert guidelines for further improvements.

We prepared revisions to the manuscript reflecting the review comments, as summarized below. Original review comments appear in black, and replies in color.

Replies to comments by Anonymous reviewer #1

RC1

The paper is a technical note, reporting on minor improvements in the setup already described in Belikov et al., 2016. The improvements seem to be exclusively technical (some improvements in the memory management, and the use of a different tool to derive the adjoint code, which would fit better in GMD than in ACP). Furthermore there is not demonstration that it achieves any better results of performance compared to that setup or to comparable inverse models. In fact, the only results presented are a series of model-data mismatches, which do not demonstrate much, beyond the fact that the model is indeed able to improve the fit to observations (the contrary would be very worrying!). Finally, I don’t think that the setup is adequate for what it aims to achieve (it makes no sense to optimize fluxes at a 0.1° resolution with covariance lengths of 500 km). For these reasons, I unfortunately, cannot recommend that paper for further publication in ACP.

1 Major comments:

• The modeling setup is an evolution of the one used in Belikov et al., 2016. An ACP technical note should summarize “new developments, significant advances, or novel aspects of experimental and theoretical methods and techniques that are relevant for scientific investigations within the scope of the journal” (https://www.atmospheric-chemistry-andphysics.net/about/manuscript_types.html). If the developments introduced represent such a significant advance, I think it should be demonstrated (at least by a comparison with the old setup). The only results that are presented are model observation mismatches (with assimilated and validation data), compared with those achieved by CarbonTracker. It’s a useful diagnostic, but definitely not a result, and not a proof that an inversion performs better than another one (overfitting the observations with unrealistic flux adjustments is possible). If the improvements are limited to performance, then not only this should be more explicitly mentioned, but also I think ACP is not the good journal for this.

Response: It is true that we present technical development, which is still not in perfect shape and needs more improvements and tuning, constructed amid limited supply of the necessary components. But, on the way we developed several techniques that can be of interest to readers. To make the study more complete we add section on optimized fluxes in the revised manuscript.

• I don't understand the possible interest of optimizing fluxes at a 0.1◦ Resolution when 500 km correlation distances are used. With these correlations the flux adjustments patterns will span several hundreds of contiguous grid cells (500 km is almost 5◦ at the equator): this is just a wast of resources, the same flux adjustments can be achieved with a lower resolution inversion and proper accounting of the model representation error (or at least that's my intuition. I would be happy to be proven wrong, but nothing in the paper does that). So the whole setup seems to just add a layer of complexity (and potentially of biases), without clear performance or accuracy advantages. It could the need for NIES of an inversion system capable of assimilating satellite data at a high resolution, but this is not what is presented here. There is no guarantee that the inversion would still be feasible with shorter correlation length and a lot more data.

Reply: It should be noted that the flux corrections are not smoothed with 500 km scale filter. The correlations are applied to a smooth field of a scaling factor which is multiplied by high-resolution flux uncertainty field to give flux corrections, which is explained in more detail in the text added to sect 4.2.

"This design of covariance operator helps preserving high resolution structure of the resultant flux corrections, given by $x = L \cdot z = u_F \cdot (L_{xy} \otimes L_t) \cdot m \cdot z$, as it can be factored into multiple of uncertainty u_F and scaling factor $S = (L_{xy} \otimes L_t) \cdot m \cdot z$ as $x = u_F \cdot S$. While the scaling factor S is smoothed with covariance length of 500 km, the original structure of spatial heterogeneity of surface flux uncertainty u_F is still present at original high resolution in the optimized flux corrections x. "
With given resolution of fluxes and transport model of 0.1 degree, inversion can be done at various settings independently from resolution of transport model, prior fluxes and flux uncertainties, such as using 22 regions globally, 1x1 degree regions or 0.1 degree grid. The reason we do inversion at the same resolution as flux dimension is technical, as it is simpler to implement.

Minor comments:
p3, l9: A disadvantage . . . , is addressed . . . aggregated flux regions: there is something missing in that sentence

Corrected, by starting new sentence.

p3, l 15: thus → this

Corrected.

p6, l26: isn't it dangerous to base the uncertainty on GPP? (what about winter time, when GPP is near zero . . . there would still be uncertainty on the respiration term . . .

Reply. That is useful notice, the respiration field would serve as better base for biospheric flux uncertainty, due to better seasonal coverage, but we did not have it at high resolution (now of course, a number of alternative datasets is made available by Jung et al. 2020, Jones et al. 2016)

p8: If x is the optimized flux (as stated in line 6), then the minimum of the second term of Equation 1 is obtained for ||x|| = 0 (the flux is minimized, not its distance to its prior). Or "x" is a flux correction, but then it is the first term of the equation that is wrong, Either way, I don't think that it's what the authors meant . . .

Reply: The paragraph is revised. The first term used in place of observations a residual misfit *r* - difference between observations and simulation with prior fluxes, as was introduced on page 7, lines 22-26

p9, l20: isn't it a problem to have all footprints stopping at 0 GMT, regardless of when they started? That means that some footprints will systematically span longer (I guess up to 4 days?) depending on their origin longitude?

Reply: Even if some footprints take more time it is accounted for when the sensitivity is recorded, those with longer time will produce more signal. To clarify, added sentence: 'The coupling time is set to be within 2 to 3 days before observation event..'

p11, l16: What is the "implicit diffusion with directional splitting"? I think it's technical enough to be worth a more detailed explanation. The rest of the paragraph is dedicated to explaining the merits of that diffusion algorithm, so I assume it's an important part of the setup, but if I wanted to reproduce it, I would have no idea how to do it (based on what's written in the paper).

Reply: We add more text as explanation for use of diffusion to approximate covariance.

p12, l19: "our report is limited to technical development" ==> I don't think that it's the aim of ACP then . . .

Reply: Historically, there have been some publications on similar line in ACP, in a format of technical note.

p12, l23: Three-month → Three-months

Corrected

p14, l23: the spatial resolution is roughly 500 km (i.e. the length of the prescribed correlations), not 0.1◦ . The system might technically be ready for inversions at 0.1◦ , but given the way it is setup, I doubt that the optimized fluxes would be any different if the optimization was done at a 1◦ or 2◦ resolution. Of course, I would be happy to be proven wrong, but the authors don't even try . .

Reply. To counter impression that the optimized flux resolution could be 500 km, explanation was added on composition of posterior flux correction, in section 4.2 (prior covariance) and a figure in section 5.2 (posterior fluxes)

RC3

I initially suggested rejecting the paper, since I thought it was just a minor evolution of a previous work from the same team (e.g. Belikov et al., 2016) and because of the insufficient quality (and quantity!) of the results presented. The authors have clarified in a reply that the difference with Belikov et al., 2016 was larger than what I had understood. I thank them for this clarification and apologize for the confusion.
Even though the sections 1 to 4 (introduction and methods) are well written, it remains difficult to give good gradings to paper (especially for the "scientific quality" and "presentation quality" criterias) as the results presented are absolutely insufficient to prove that the model is working as expected, and more generally to support some of the main claims of the conclusions (p15, l5-7). I am however willing to change my recommendation to major revisions, in the light of the clarifications provided by the authors.

> “Detailed reply to author’s comment
> The review comment states: “The paper is a technical note, reporting on minor improvements in the setup already described in Belikov et al., 2016. The improvements seem to be exclusively technical (some improvements in the memory management,and the use of a different tool to derive the adjoint code, which would fit better in GMD than in ACP). Furthermore there is not demonstration that it achieves any better results of performance compared to that setup or to comparable inverse models.”
> Author’s reply: There is some misunderstanding about developments made since the mentioned paper. It should be noted that in a paper by Belikov et al., (2016), there was no attempt to do the inversion, instead, it focused on development of forward coupled model (at lower resolution of 1 degree), its adjoint, the adjoint accuracy and performance. In this study, (1) the Lagrangian model resolution was increased to 0.1 degree, and necessary prior fluxes were developed; (2) Flux covariance operator was developed specifically to handle the challenges of operation at high spatial resolution; (3) Iterative optimization technique was implemented and multiple (time consuming) inversion trials were performed before achieving reported results.”

Ok, noted. But then I would suggest to make this much clearer in the paper. The paragraph starting at line 15 on page 3 is particularly confusing.

Reply. The paragraph starting at line 15 on page 3 was revised.

> “The review comment: “In fact, the only results presented are a series of

> model-data mismatches, which do not demonstrate much, beyond the fact that the model is indeed able to improve the fit to observations (the contrary would be very worrying!). "Author's reply: Still, do demonstrate that the technical development is valid, and the inverse model does work, showing the fit to the observations is desirable."

What is your criteria to say that "the inverse model does work"? If the criteria is just to "improve the fit to the observations", then indeed it works, and you demonstrate it. But the aim of doing an inversion is to find the optimal value for the optimized parameters, in you case the CO_2 fluxes (given the information from the prior and from the observations), which you formalize as finding the vector that minimizes J(x) in Equation 1:

> 1. How can I be sure that the posterior fluxes indeed minimize J(x)? There are certainly vectors that improve the fit to the observations but increase the value of J, which your system could find if it malfunctions). Even if it works properly, how do you know that 45 iterations is enough?

Reply: To clarify, the comment that inversion works only relates to computational side not to the physical result for fluxes, the flux biases and biases in simulated concentrations. That should require an extra work to improve. Fortunately, we never see a cost function increasing in the output of M1QN3 minimizer. We added a text: Figure A3 in Appendix presents the cost function reduction in case of optimizing fluxes for 2011 and completing 61 gradient calls. The cost function reduction declines nearly exponentially, by almost 3 times for each 10 gradient calls completed. Relative improvement between 41 and 61 gradient calls is 1.5% of the total reduction from the first to the 61 gradient calls.

> 2. And how can I be sure that the value of x that minimizes J is indeed closer to reality than the prior? (if your transport model is strongly biased, or if the uncertainties matrices are not adapted, you can totally end up with posterior fluxes that are worse than the prior).

Could you be more specific? What is the cost of a NIES-TM-FLEXPART inversion, compared to a NIES-TM inversion? And how different are the posterior fluxes (and how better are they)?

Reply: We planned to do more improvements before producing the flux estimate that can be used for scientific objectives, this paper is meant to introduce a technique. A figure comparing prior, posterior and CarbonTracker was added. Added a relative CPU time information as: "The fraction of CPU time spent on running the Eulerian component of the coupled transport model is 82%, on the Lagrangian component 1%, and on covariance 17%."

> "The covariance scale is a tunable parameter, it can be set according to information content available in the observations. Many inverse modeling studies (eg Chevallier et al, 2010) do not assume the current observing network provides enough information to constrain the land biosphere fluxes globally at a higher resolution than 500 km. It is mentioned (Chevallier et al, 2010) that with shorter covariance scales the model may take more iterations to converge. Accordingly, the transport model resolution is often higher here than the effective resolution of the inverse model. The rationale for using higher resolution inversion in comparison to lower resolution, such as using large regions, is to reduce aggregation error (Kaminsky et al., 2001)."

I am not questioning the benefit of solving the fluxes at a higher resolution, just the adequacy of that combination of covariance scales and transport/optimization resolution. If your observations limit you to an effective resolution of 500 km, then I think that you don't need such a high resolution transport model. And you had enough observations to require such a high resolution transport (if you have enough obs. to use covariance scales of e.g. 50 km), would your inversions still be feasible (it would require much more iterations). If you can demonstrate that you get much better results (fluxes, not just observation fits!) with 500km/0.1˚ than with e.g. 500km/1˚, then it could be a strong incentive from other groups to implement a similar coupling, so I think that it would be really interesting to do that comparison properly.

Reply. It is possible that coupled 1 deg model will perform similarly to 0.1 deg model for estimating the land biosphere and ocean fluxes. However, the push for higher resolution is driven by a need to estimate the anthropogenic emissions. Our study is a step in that direction (as was mentioned in introduction, page 4 lines 6-7), and the high resolution has to be used there even when the land biosphere and ocean fluxes are better estimated with 1 degree resolution model. Thus, best direction for such objective is to keep improving the 0.1 deg setup, making effort to reduce the biases in the land biosphere and ocean fluxes. To make the direction clear, we reformulate the sentence in the introduction as: "The objective of this study is optimizing the natural CO_2 fluxes in order to provide a background for estimating the fossil CO_2 emissions where the advantage of high-resolution approach is more evident."

Replies to comments by Anonymous reviewer #2

RC2

The 'Technical note' by Maksyutov and co-authors describes the technical details of a CO2 flux inversion technique based on the coupled Eulerian-Lagrangian transport model NIES-TM-FLEXPART. The coupled system operates at a high spatial resolution of CO2 fluxes of 0.1˚ x 0.1˚ globally and also attempts a flux inversion at this resolution. As such the approach is novel and promising. The paper is well structured and written. The performance of the inversion is documented by time series comparisons/evaluations of different data sources assimilated and not assimilated by the system and a comparison to another independent inversion system. However, the inverted flux fields are never shown/discussed in the manuscript, which makes it rather difficult to judge if the inversion yielded reasonable results. Even though the manuscript is a 'technical note', it would be very beneficial to overcome this shortcoming (detailed suggestion below) before publication.

Major comments

Section 5: Evaluating the performance of an inverse modelling system is not straightforward. Restricting this evaluation to a comparison of model skill for the assimilated and additional independent concentration time series is not sufficient. By over-fitting the flux fields a very good agreement of the posterior concentrations with the observations may be achieved but the flux fields may contain unrealistic detail in order to achieve this. Given the large degree of freedom in the fluxes, as indicated by the fact that grid and time resolved fluxes are inverted from a relatively limited set of observations, there seems to be a high risk for the presented inversion setup to over-fit the solution. Since no flux fields are presented, it is impossible to judge this possibility. Therefore, I would encourage the authors to extend their discussion of results in section 5 to include a brief analysis of the obtained flux fields. I can see that the authors have planed this for a later publication and, hence, I don't think this needs to be very quantitative here, but the presented flux fields should document the validity of the inversion approach. A qualitative comparison with flux fields obtained from CarbonTracker (as done for the concentration time series) would also be beneficial.

Reply: To illustrate the flux adjustments by inversion, we add the flux comparison figure for selected regions, comparing prior, posterior and CarbonTracker fluxes.

Minor comments

P3,L18: Resolution of coupled Eulerian-Lagrangian models. I find it a bit misleading to say that the transport in these models is run at a resolution of (as in the cited publication) 1 km. Yes, technically the transport is not run at any fixed resolution in the Lagrangian sense, but the driving meteorology is still determining what scales of motion can be correctly resolved by the model. The Lagrangian model may still have some skill in the sub-resolved range, but basically it degenerates to a Gaussian plume at these scales with constant wind direction, speed and dispersion characteristics. This fact is not sufficiently highlighted throughout the manuscript. Another example is P4, where the transport system for the current study is introduced. The driving meteorology is 1.25°x1.25° this is certainly not sufficient for a detailed transport description in complex, mountainous terrain, but also not for coastal areas. Since sites from both environments are contained in the list of sites used for assimilation, this limitation should be discussed in more detail. Next to spatial resolution also temporal resolution is important for a transport description in the mentioned environments. Although in the following, the use of observations was restricted to certain times of the day, these observations still carry the transport history for a longer period and if temporal variability in the transport is not sufficiently described may lead to biases in the simulated concentrations as well.

Reply. We revised the sentence at P3, L18 and added a note on resolution in model description section.

P5,L13f: If I understand this correctly, there is no diurnal cycle of the biospheric CO2 flux considered in the model setup. How much is this simplification limiting the model performance? Not all sites used for the inversion are remote coastal sites but are surrounded by dense vegetation. How much does the constant diurnal flux and the restriction to afternoon observations introduce a bias in the flux inversion? In general, I have the feeling that the low temporal resolution of fluxes does not keep up with their high spatial resolution in the current setup.

Reply. Following the comment, we estimated the impact of using daily constant flux in place of hourly varying by running forward simulation with hourly and daily mean versions of SIB model fluxes (Denning et al. 1996), as used in Transcom intercomparison (Law et al. 2008), and added a Figure showing the simulated seasonal mean biases. It appears the hourly fluxes produce 0.5 – 1 ppm lower simulated CO_2 in summer with respect to daily mean fluxes. Thus, it will be clearly helpful to apply diurnally varying fluxes in place of those with daily temporal resolution.

P6,L8: How much do the posteriori fluxes actually depend on the chosen biospheric flux climatology? Given the large year-to-year variability in biospheric fluxes, is it sufficient to operate with a climatology of prior fluxes? Was this evaluated by choosing a different averaging interval or even individual year for the prior climatology.

Reply. Use of climatology leads to degradation of prior simulation, but it allows for extending the simulations to years when the simulation is not available (we only had the data till 2010). For some regions, inverse corrections to prior are substantial, and look like exceeding the interannual variability. Added a note: “Although the use of climatology in place of original fluxes degrades the prior, the posterior fluxes show significant departures from the prior, thus reducing the impact of prior variations.”

Section 3.3: Are all biomass burning emissions considered to be released at the model surface or was any kind of vertical emission profile used? Again, this may be crucial when considering transport simulations at the mesoscale.

Reply. Added a text ‘The fluxes are input to the model at the surface, which may lead to underestimation of injection height for strong burning events and occasional overestimation of biomass burning signal simulated at surface stations.’

P7, L9 and P9, L21: According to the first text location only afternoon samples were considered for the inversion. However, according to the second location daily average footprints of the Lagrangian model were applied. If that is really the case the footprints are not representative for the used observations, largely neglecting temporal variability once again!

Reply: To avoid confusion, removed 'daily average', added sentence: 'The flux footprints are saved at daily or hourly timestep, depending on available surface fluxes.'

P9, L25ff: The description of the forward model steps. From this description it is not clear to me how the concentration increments from the different models are added to avoid double counting of the fluxes (step 2.c). Shouldn't the Eulerian model use different fluxes than the Lagrangian (cropping those that are covered by the Lagrangian model)?

Reply: Added clarification: 'For each observation event, the fluxes used in Eulerian and Lagrangian components are separated by coupling time, so that there is no double counting of fluxes for the same date and time in the coupled model simulation.'

P10, L24: Initial conditions are used from an optimised run from the previous year. But then the question remains how the previous year was initialised. Was this done with a spin-up run?

Reply: Added: 'When the optimised fields are not available, the output of multiyear spin-up simulation is used, with same adjustment to South Pole observations.'

Section 5.1: Why is RMSE used as the sole estimator of model performance? RMSE will decrease even if only the baseline fits better after optimisation. Most of the regional flux information, however, is stored in the peak concentrations, for which a more robust performance estimator could be a bias corrected RMSE or the coefficient of determination. The bias should be reported as well. Taylor skill score could be another performance parameter that would be more suited to focus more on the short term variability.

Reply: We added a site bias data to the plot.

Sites: It would be useful to see the sites and the aircraft locations on a map. Would help to judge which areas are not well covered by assimilated observations in comparison to validation data. Such a plot should also contain the information of flask vs. continuous observations.

Reply: Site map added, with separate symbols for flask, continuous and aircraft observations.

Technical comments

P3,L10: Start a new sentence after '... Kaminski et al. (2001). This is addressed by ..."

Corrected

P12,L14: ’previous’ instead of ’pervious’.

Corrected

Figure 1, caption: Label as ’Examples of ...’

Corrected

Figure 3+4: Since the x-axis is not along a continuous variable, I would suggest to not include lines in the plot or even use a barplot instead of symbols. The lines are just confusing and have no physical meaning

Reply: revised as suggested

References

Jones, L. A., Kimball, J. S., Reichle, R. H., Madani, N., Glassy, J., Ardizzone, J. V., Colliander, A., Cleverly, J., Desai, A. R., Eamus, D., Euskirchen, E. S., Hutley, L., Macfarlane, C., and Scott, R. L.: The SMAP Level 4 Carbon Product for Monitoring Ecosystem Land–Atmosphere CO_2 Exchange, IEEE Transactions on Geoscience and Remote Sensing, 55, 6517-6532, 10.1109/TGRS.2017.2729343, 2017.

Submitted in behalf of the authors by Shamil Maksyutov, Sep 15, 2020