

Responses and revisions to reviewers after revision 1 for: 2025-6352, EnviFlux (v1.0): a simplified surface flux inversion tool based on four-dimensional variational data assimilation (4D-Var)

Ross Bannister, NCEO and Department of Meteorology, University of Reading, UK

July 9, 2026

This document is a summary of the Reviewers' comments, my previous replies, and descriptions of my revisions.

Responses and revisions for Reviewer 1

The following is a copy of my previous replies to Reviewer 1. His/her comments are in italic, my original responses are in Roman and added action points to produce revision 1 of the manuscript are in red.

The author presents a useful “toy model” data assimilation environment for flux estimation, complementing a long history of such toy models in the dynamical literature. The model description is appropriate and the presented case studies are interesting, and I recommend publication after the following comments are addressed. My comments are from my perspective as a practitioner of emissions flux inversions (i.e., an enthusiastic potential user of EnviFlux), and are primarily about the extent to which the full range of problems arising in my work could be investigated with this tool.

Many thanks to the reviewer for their time and effort in providing valuable feedback. Their insights are very interesting and gratefully received.

What is the computational performance of EnviFlux, in terms that are meaningful for a typical practitioner? For example, what would it take to infer daily fluxes at roughly 2x2.5 degree resolution for a year? How many cores? How much wall time, memory, etc?

The inversions in the paper are for a global grid of longitude $\times$ latitude $\times$ levels = $65 \times 33 \times 56$, which is about $5.5^\circ \times 5.5^\circ$ in the horizontal. The runs are performed for 100 days of monthly fluxes (although memory is reserved for 16 months). These take around six hours wall time on a single core and use about 64 MB of RAM. Looking at $2.0^\circ \times 2.5^\circ$, this is 180×72 horizontal grid points. Since EnviFlux uses a global spectral representation of the background error covariance matrix, the grid is designed to be consistent with this. This uses a “maximum degree” wavenumber, L , which has an equivalent grid of size $(2L + 1) \times (L + 1)$. A value $L = 80$ gives roughly the same number of horizontal grid points as 180×72 , but the configuration consistent with this is 161×81 or about $2.25^\circ \times 2.25^\circ$. Assuming the same number of vertical levels (56), this grid has about $161 \times 81 / (65 \times 33) \sim 6$ times more grid points than the configuration used in the paper. For daily fluxes, this requires $365/16 \sim 23$ more times in the state vector.

- The RAM required is estimated to be $6 \times 23 \times 64 \text{ MB} \sim 9 \text{ GB}$.
- The wall time is estimated to be $6 \times 365/100 \sim 22$ hours.

The wall time estimate assumes the same number of variational iterations (20). These are quite high running costs, but the main purpose of EnviFlux is to research new ideas and methods at lower-resolution than the ‘real-world’ counterparts.

There is no parallelisation implemented in EnviFlux v1.0, but implementing this in the most computationally intensive parts of the system (the forward and adjoint operators) would obviously reduce the wall time. I would implement this with a shared memory parallelisation such as OpenMP.

Actions: Added appendix giving a revised quantification of the resources required to run EnviFlux (referred to near starts of Sects. 4 and 5). Done for the model resolution applied in the paper, and for the higher resolution equivalent to that mentioned by the reviewer.

Clarify the spatial domain(s) supported by EnviFlux (e.g., does it have to be global). Regional simulations with boundary conditions imposed by a global flux inversion are a very important case in, for example, methane, and introduce huge questions in error quantification (e.g., to what extent do we alias boundary condition error onto in-domain emissions estimates); it would be enormously useful to investigate this with a low-cost tool like EnviFlux.

The spatial domains supported by EnviFlux Vn 1.0 are global with number of longitudes $\times$ latitudes $= (2L + 1) \times (L + 1)$, where L is the spectral maximum degree wavenumber. EnviFlux could be adapted for a local domain by (i) altering the grid data structure, the input/output and inner-product routines; (ii) calling a limited area model (and adjoint) which is driven by boundary conditions; (iii) altering the background error covariance model; and (iv) exploiting the existing descent algorithm to do the minimisation. This is a moderate piece of work, but would be possible.

Actions: Word, “globally” added to the abstract, “global” to the introduction. Statement of the possibility of modifying the software for local domains added to the summary, Sect. 6.

What are the spatial and temporal resolutions of the case studies presented, and what are the range of possible resolutions?

The case studies presented in the paper use $L = 32$, which is a global grid of size $(2L + 1) \times (L + 1) = \text{longitude} \times \text{latitude} = 65 \times 33$. There are 56 levels. It is possible to set L to any value that the memory and netCDF libraries permit.

Action: Information made more prominent (near start of Sect. 4), and error at top of Table 1 corrected.

How are wind fields specified? Could users prescribe assimilated wind fields?

The driving winds are stored as a set of netCDF files – one set of u, v, w 3D fields per major time step. These are interpolated from ERA-5 reanalyses. In principle, these could be derived from any reasonable source.

Action: Word, “prescribed” added after Eq. (1), connection between advection and wind fields made in bullet point 1 of Sect. 2.2, mention of ERA-5 added to Table 1 (including reference).

The semi-Lagrangian approach makes simulating chemistry tricky, making this tool (at least at global resolutions) most useful for long-lived species. However, for problems like methane and N_2O , it is necessary to account for 3D atmospheric sink terms due to tropospheric OH and stratospheric loss processes respectively. I would be very interested in exploring the sensitivity of inversion approaches to characterizing this sort of loss field. Could EnviFlux be adapted to such problems?

Such sinks could be introduced in EnviFlux via prescribed loss rates. It is an interesting question how a semi-Lagrangian method would work with processes like the OH sink. The semi-Lagrangian method is used in EnviFlux for its stability, but at the cost of lack of strict conservation. One could approximate the methane loss terms in a semi-Lagrangian scheme by treating the advection completely separately from the chemistry (so that the chemical sink of methane happens at the air parcels’ destination points and neglecting contributions over the back-trajectories). It would also be possible to add atmospheric loss rates to the state vector so that the sinks (due to OH and methane oxidation, e.g.) could form parts of the state vector, in addition to the surface fluxes. Toy systems like EnviFlux could be useful to do exploratory work in this area, especially to help understand how the observations divide information between the surface fluxes and the atmospheric sinks.

Action: Brief discussion added to the summary section.

Section 3.2: observational error correlation has proven to be a big problem in my own satellite-based data assimilation. Shared surface characteristics (e.g. mountains) make for systematic errors that are irreducible due to averaging. How hard would it be to specify more complex observational error distributions with off-diagonal terms?

The principle of dealing with correlated $\mathbf{R}$ is relatively straightforward, but the practice is complicated due to determining which observation errors are correlated and by how much, and cost implications for large numbers of observations. If correlated observations can be batched, this would increase the numerical efficiency of accounting for them.

I would adopt the following strategy to calculate the gradient of the cost function. Assuming that observation errors (even at different times) are correlated, the gradient calculation would no longer have the form of Eq. (6) of the manuscript, which assumes uncorrelated observation error across different times:

$$\nabla_{\delta\mathbf{v}}J(\delta\mathbf{v}) = \delta\mathbf{v} - \mathbf{B}^{\tau/2} \sum_{k=0}^{N_T} \left(\mathbf{M}_{0 \rightarrow k}^{\{\mathbf{x}^b + \mathbf{B}^{1/2}\delta\mathbf{v}\}_\rho} \right)^\top \mathbf{H}_k^\top \mathbf{R}_k^{-1} \left(\mathbf{y}_k - \mathbf{H}_k \mathbf{M}_{0 \rightarrow k}^{\{\mathbf{x}^b + \mathbf{B}^{1/2}\delta\mathbf{v}\}_\rho} \left[\chi^b(0) + \{\mathbf{B}^{1/2}\delta\mathbf{v}\}_\chi \right] \right).$$

Instead, the gradient would have the form:

$$\nabla_{\delta\mathbf{v}}J(\delta\mathbf{v}) = \delta\mathbf{v} - \mathbf{B}^{\tau/2} \times \underbrace{\left(\mathbf{M}_0^\top \mathbf{H}_0^\top \quad \cdots \quad \mathbf{M}_{N_T}^\top \mathbf{H}_{N_T}^\top \right) \left(\mathbf{R} \right)^{-1} \begin{pmatrix} \mathbf{y}_0 - \mathbf{H}_0 \mathbf{M}_0 \left[\chi^b(0) + \{\mathbf{B}^{1/2}\delta\mathbf{v}\}_\chi \right] \\ \vdots \\ \mathbf{y}_{N_T} - \mathbf{H}_{N_T} \mathbf{M}_{N_T} \left[\chi^b(0) + \{\mathbf{B}^{1/2}\delta\mathbf{v}\}_\chi \right] \end{pmatrix}}_{\mathbf{q}}},$$

where $\mathbf{M}_k$ is shorthand here for $\mathbf{M}_{0 \rightarrow k}^{\{\mathbf{x}^b + \mathbf{B}^{1/2}\delta\mathbf{v}\}_\rho}$ in the first equation, $\mathbf{R}$ is the observation error covariance for all observations, and the other notation is explained in the paper. The key aspect of the last equation is that all observations are represented together, allowing observation errors to couple different elements in observation space. Either $\mathbf{R}$ can be written in terms of its eigenvalue decomposition, so $\mathbf{R}^{-1} = \mathbf{F}\mathbf{\Lambda}^{-1}\mathbf{F}^\top$ to evaluate the above, or the equation $\mathbf{p} = \mathbf{R}\mathbf{q}$ can be solved for $\mathbf{q}$, where $\mathbf{p}$ and $\mathbf{q}$ are the observation space vector as indicated above. The eigenvalue decomposition contains information on the condition number of $\mathbf{R}$, which may need to be monitored.

Action: Brief discussion added to the summary section.

Section 3.3: How easy would it be to trade out this background matrix for other approaches that might better approximate the varied operational systems used by potential EnviFlux users?

All what the EnviFlux algorithm needs is a square-root representation of the background error covariance matrix, which could be something very different to that implemented. Examples are where the correlation lengthscales depend on land or sea, or where the land and sea points are uncorrelated. If the resolution is low, then a horizontal correlation matrix could be explicitly constructed and then ‘square-rooted’ with an eigenvalue or Cholesky decomposition. If one moves away from the spectral scheme that is implemented in v1.0, then there would no longer be any need to have a grid with $(2L+1) \times (L+1)$ dimensions (see above).

Action: Brief discussion added to the summary section.

Lines 411-414: This is a very interesting discrepancy. In my very rough experience, satellite data is more “reliable” because of the representation error of surface sites, and because satellite data is generally less sensitive dilution in a mischaracterized PBL (the convective transport explanation of Basu et al. is also plausible). Or at least, this is the way I have thought about it without rigorously investigating the issue. To what extent could these sorts of errors in convection/PBLH be studied with EnviFlux?

In the work shown in the manuscript there is no unaccounted transport (in experiments when the wind factor $\eta = 1$) and there is no representativity error. This is likely to be the reason for the discrepancy with Basu et al. (2018). These issues could be studied in EnviFlux by running a much higher resolution model to make the synthetic observations (the nature run), but maintain the lower resolution model in the assimilation. The nature run could also include the effects of sub-grid transport.

Actions: Added brief discussion of representativeness error and how this could be studied in EnviFlux (near end of Sect. 5.3).

Responses and revisions for Reviewer 2

The following is a copy of my previous replies to Reviewer 2. His/her comments are in *italic*, my original responses are in Roman and added action points to produce revision 1 of the manuscript are in blue

In this study, the author has developed a simple flux inversion analysis tool, named as EnviFlux, for estimating surface fluxes of atmospheric constituents. It is no doubt that this kind of tool is useful, because a flux inversion tool that is used practically for inverse analysis is computationally expensive. It is sometimes difficult to use a practically used system for several sensitivity studies, though those sensitivity experiments are needed to evaluate reliability of inversion results. I have found this simple “toy model” scientifically sound, however I cannot recommend this manuscript for publication because of two major reasons described below.

Many thanks to the reviewer for their time and effort in providing valuable feedback. I am pleased that the reviewer has described this model as “scientifically sound”. I am very surprised in their belief that the paper should not be published, and on their assignment of some lower scores, and their belief that the work is not novel. I think these are harsh opinions which I disagree with, and I hope The Editor will read my detailed replies and decide to proceed with paper revisions. The reviewer has, however, raised some interesting issues and I can defend the work against their comments with some modifications to the manuscript. I will be as concise as possible. [Please see below for specific actions.](#)

1. No information of computational efficiency or useful facility as a tool

Given the information described in this manuscript, this model seems not to have any advantage other than simplicity compared to other existing models (e.g., lacking turbulent and cloud convection processes, not assuring tracer masses). However, the most important things, how simple and easy to use as a tool this model is, are not well described. How fast this model can run should be explicitly discussed compared with existing models. I do not think just using the semi-Lagrangian scheme cannot make the model faster than others as this technique is often used in existing models. In addition, not implementing turbulent or cumulus convection scheme may make the model fast, but it does not provide novelty of the study, because other models can do the same thing by just turning off those processes.

The paper concentrates on the scientific set-up of the model and inverse method, with some demonstrations, I can provide information on run time, memory, etc (see e.g. my response to Reviewer 1). I am not aware of any other such inverse model that can be run on a low-spec computer, such as a basic laptop, as other such systems rely on high-performance computing infrastructure. My implementation of the semi-Lagrangian method is especially efficient as after it has run once, it stores the trajectories’ departure points, making it extremely fast thereafter (“boosting”). This is especially advantageous as the model and adjoint are run many times. I can document this (and the resources needed to run the system) in the revision. Turbulence and convection parametrisations can be added in a future development, but the emphasis here is on studying the limitations of inverse modelling per se, which can be done most simply without these processes at this stage.

Actions: Added mention of “boosting” to list item 1 in Sect. 2.2, and added appendix giving a quantification of the resources required to run EnviFlux (referred to near starts of Sects. 4 and 5).

2. Too simple experimental design

Usually, during development of this kind of inversion analysis model, one should perform ideal experiments to see whether the model appropriately works or not. In that sense, the experiments the author showed in this manuscript are not novel.

I do not agree that experiments are not novel. I believe, in particular, that the experiments demonstrating the ability of the 3D inverse model to identify the positions and strengths of the a-priori unknown source and sink, are novel. In addition, the work shows the consequences of having observation biases and model error. I am not aware such experiments have been done before (the reviewer has not provided any citations to support their claim that the work is not novel).

Action: Added “novel” to the start of Sect. 4 to remind reader that the work is new.

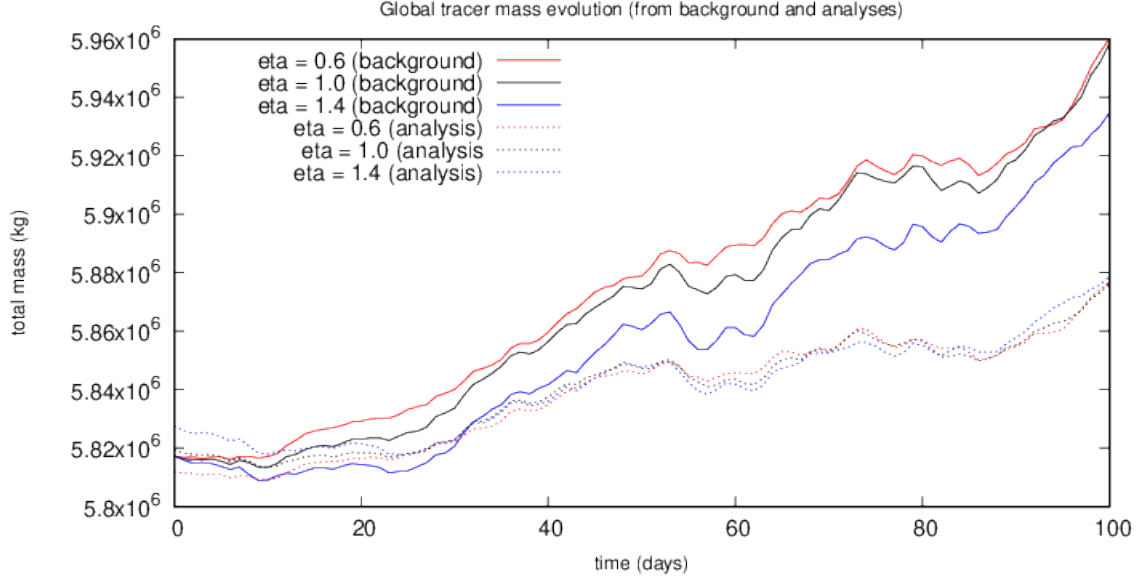


Figure 1: Total tracer amounts for backgrounds (solid) and analyses (dotted) with three different η values: 0.6, 1.0, and 1.4. The factor η multiplies the vertical winds in each respective assimilation/forecast experiment.

Furthermore, scaling vertical wind speeds, the author applied to see the effects of model errors, is too simplistic and the results may not provide scientifically useful information. If the atmosphere is considered as hydrostatic (this assumption is usually valid), horizontal wind velocities should accordingly be corrected when vertical velocity is changed. Just changing vertical velocities would deteriorate the consistency with the continuity equation and induce considerable errors in mass conservation. It means that unexpected sources or sinks would occur in the atmosphere, which is not suitable for simulations of long-lived species such as CH_4 .

The vertical wind scaling $w \rightarrow \eta w$, $\eta \sim \mathcal{O}(1)$ is meant to be a simple way of adding model error. It is true that scaling only the vertical wind will lead to extra sources and sinks appearing, which would have the form

$$\rho w(1 - \eta) \frac{\partial \chi}{\partial z},$$

where ρ is the air density (not affected by η being different from unity, see below), χ is the tracer mixing ratio, and z vertical distance. Indeed, the whole point of these experiments is to ask how the system compensates for such model error. (Anomalous sources and sinks can result from erroneous use of convective parametrisation in systems where that is applied, so this set-up is relevant.)

One can see the compensating effect in Fig. 1 here. The solid lines are the evolutions of global mass of tracer with three different values of η (see the caption), each based on the same background (a-priori) state (made up of initial conditions and surface fluxes). The values are quite different between the runs, showing the effect of the erroneous sources. The dotted lines are based on the respective analyses, each found by assimilating the same total column observations. Each analysis integration converges in time to similar total tracer masses even though each has a different realisation of model error. The source and sink distributions are different in each integration to compensate for the model error. This is a fundamental issue in inverse modelling, which EnviFlux can be used to investigate further (ideas are given in Sect. 6 of the manuscript).

In terms of mass conservation and hydrostatic balance, the consequences of $\eta \neq 1$ are not as profound as the reviewer states as the vertical wind scaling is applied only to the tracer transport and not to the air mass constituting ρ . The analogy in NWP for instance would be to solve the main prognostic equations (including for the air density) using w , but advecting the tracer with ηw . The anomalous sources and sinks of the tracer would have negligible

effect on hydrostatic balance since the tracer has negligible mass contribution. Figure 1 here is not added to the paper.

Actions: Erroneous source effect mentioned in Sect. 4.2 and added new Appendix F. (Fig. here is only for reply to Reviewer and not in the revision.)

In addition, the author did not consider atmospheric sink of CH₄ due to oxidation with OH, which is one of the most important features of atmospheric CH₄.

The paper is not meant to be an analysis of methane, but instead a methane-like tracer (having similar patterns of sources and sinks in the latter part of the paper, but no methane oxidation with hydroxyl radical) so that the inverse problem can be studied in its simplest form.

Action: Have emphasised this in Sect. 2.1.

Minor comments:

1. L25-26: CO₂, CH₄, N₂O; because they appear first, they should be fully spelled-out. **Done.**
2. L34: That is true also for 4D-Var here. **Done.**
3. L42: That is true also for CO here. **Done.**
4. L76: Please specify what χ is here (e.g., mass concentration, mole fraction, mixing ratio). **Done.**
5. L77: Please elaborate how to derive wind velocities and diffusion coefficients. Are they taken from reanalysis data? **Mentioned in list item 1 in Sect. 2.2 (winds); diffusion coefficients are zero, mentioned in list item 2.**
6. L78: “ $u = (u \ v \ w)$ ” should be a transpose of a vector, i.e. “ $u = (u \ v \ w)^T$ ” **Done.**
7. L82: “ppb” should be explained as “parts per billion”. In addition, this is not usually used as the unit of mixing ratio, but as that of dry air mole fraction. **No longer referring to ppb, but to μgkg^{-1} consistent with mixing ratio.**
8. L86: What does “ τ_ρ ” mean? **Added.**
9. L110: “For the experiments shown in this paper, no explicit diffusion is applied” This is a critical flaw of the experiments, which makes the results unsound. **Comments added to list item 2 in Sect. 2.2 (the semi-Lagrangian scheme has numerical diffusion, I believe results are sound).**
10. L117: “EnviFlux does not include (sub-grid) turbulent or convective transport process” I think that if a model does not include these processes, it cannot appropriately simulate any atmospheric constituent transport. This flaw critically affects its inversion results. **The system is not meant to simulate the real-world, but is a simplified, but useful, tool to study such inverse problems. Have emphasised this in Sect. 2.1.**
11. L165: Does “total column amount (TCA)” mean vertically integrated amount? Usually, satellite products provide column averaged amount. **Confirmed, added “integrated” to Sect. 3.2.**
12. L169: “The TCA observations are found from a discretization of ...” The author should note that satellite data are made with averaging kernels, which should also be incorporated in the model. **Text added to Sect. 3.2.**
13. Table 1: Please elaborate the experimental settings of Table 1 in detail in the main text. Please explain the validity of experimental length of 100 days and source time-step of 30 days (while temporal correlation scale is 3 months, I’m wondering if these settings are reasonable). “ $n_x = 33$, $n_y = 65$ ” looks strange to me, because the number of longitudes is usually larger than that of latitude. Furthermore, how long the assimilation window was set should be clarified. Please explain the validity of the number of 4D-Var iterations (=25). Is it enough to converge the parameters? **Numbers of points in lon/lat directions corrected near top of Table 1 (note latexdiff has not highlighted this change). 100 days, 30 days, 3 month timescales, and 20 4D-Var iterations (wrongly given as 25 in the first version) are justified near the top of Sect. 4.**

14. L256-257: *“Here, each is taken as a point location in the horizontal. . .” Is the satellite orbit assumed here derived from GOSAT? Are data gaps due to clouds (typically existing in the tropics) considered?* [Brief comment added just before Sect. 4.1.](#)
15. L443: *“a simplified tool” From my point of view, the model looks unmatured rather than simple. If the author claims this simplicity is advantageous, that should be elaborated.* [Hope this is answered by my replies to previous comments.](#)
16. L446: *“cheaply without the need for high performance computing” To claim this, the author should describe what kind of machine was used and specify wall-clock time for each experiment. Furthermore, the author should also demonstrate those computational utilities by comparing with other existing models.* [As mentioned, appendix F is added \(see above\) and computer spec. given in Table E1 caption. I do not think it is reasonable to require me to add computational resources needed by other inverse modelling systems. I have some experience with another system, and I know it was not designed to be run on anything but a high performance computing platform \(even for relatively low-res\), making it complex and resource hungry.](#)

I believe I have addressed all the Reviewer’s criticisms and I have argued that the system is a useful tool and that the paper is a valid scientific study, and with the revisions, will make a interesting contribution to GMD.

Other changes made

1. Included $L = 32$ to Table 1.
2. Removed Observation error variance specifications heading from Table 1 as it is unnecessary.
3. Added missing $\mathbf{H}_k^\top$ in Eq. (6).
4. Fig. 6 caption: removed incorrect units (note latexdiff has not highlighted this change).
5. Fig. 8 caption: corrected units.
6. Fig. 9 caption: corrected units.
7. Fig. 12 caption: corrected units.
8. Corrected “panel (b)” to “panel (a)” in Sect. 5.3.
9. Split conclusions section into two parts (summary/conclusions and future work).
10. Add “used to model background error covariances” to Appendix A title.
11. Added items to acknowledgements.
12. Corrected mistake: total column observations have units μgm^{-2} .