Response to Anonymous Referee #1 (RC1)

March 2024

We would like to thank Referee #1 for his/her comments on our work. Referee #1 has raised some interesting points, which we would like to deal with as early as possible, especially in case her/her comments influence other reviewers. We believe this response can resolve the apparent paradoxes that the Referee raises. We reproduce quotes from Referee #1's report (in italics), and respond.

- *Without being new, the article is interesting, stimulating, educational and generally well written.*
  - Thank you. We do believe, however, that the article is new as this is the first time (to our knowledge) that a spectral scheme has been implemented in a full flux inversion system.
- *The first paradox of the presentation is the motivation: "The spectral method is very efficient. It is applicable to systems with very high resolutions, where existing methods that explicitly represent the B-matrix would not be feasible (Appendix D)" (l. 377-379 and similar sentences in the rest of the text). Actually, looking at Figure D1 about the cost of the two approaches, one can see that the authors' illustration is on the lower end (L = 32, small resolution), while current results with an explicit representation reach the high end (L seems to be about 400 in doi:10.22541/au.171052488.85903583/v1). Where is the better efficiency argued in l. 66?*
  - The tests done in the paper are indeed for low resolution $L = 32$, but these are purely to show that the method works.
  - Looking at Fig. D1 for $L \sim 400$, the spectral method is 3-4 orders of magnitude more efficient to use than the explicit approach (compare dashed lines in Fig. D1), and even more so to setup (assuming an eigenvalue decomposition is used to find the square-root of the B-matrix – compare the continuous lines in Fig. D1).
- *The authors seem to ignore that the explicit representation is simplified by numerous zero correlations in the 2D flux errors: no space-time correlations, ..., no correlations between land and ocean for surface flux errors, ...*
  - Thank you. This was an oversight, and we believe the paradox can be resolved. We have revisited Appendix D concerning the costs of the spectral method compared to explicit representations of the B-matrix. By "explicit", we have now included three possibilities of explicit representation of the B-matrix (rather than one in the manuscript). (i) An explicit representation of B in the entire state space, which has $n = n_x n_y n_z + n_x n_y (T+1)$ elements (n_x longitudes, n_y latitudes, $T+1$ months). This B-matrix has n^2 elements; this is what is meant by the explicit matrix in the current manuscript. Our oversight was to neglect two more compact alternatives, which we believe, especially the latter, will address Referee #1's point. (ii) Separate explicit representations of the parts of the B-matrix that associated with the initial concentration, c, and with the flux, ρ. These B-matrices have $(n_x n_y n_z)^2$ and $(n_x n_y (T+1))^2$ elements respectively. (iii) The same as (ii), but where the vertical/horizontal and temporal/horizontal parts of the correlations are separable (e.g. total flux correlation = temporal correlation $\times$ horizontal correlation), which leads to use of the Kronecker product mentioned by the Referee. Although (ii) and (iii) are more efficient than (i), they are still much less efficient than the spectral method. We have shown the workings in a possible replacement to Table D1 and to Fig. D1 below.
  - For example, running an inversion at the resolution of ERA-5 reanalysis (1440×720, corresponding to $L \approx 720$) let's compare the cost of using explicit form (iii) above (the most efficient of the explicit representations) to the cost of the spectral method. Let us put some numbers to our argument and consider the flux field only. The following are costs associated with setting-up the square-root matrices:
    - Explicit form (iii) for ρ: $\sim 10^{18}$ operations.
    - Spectral form for ρ: $\sim 10^{7}$ operations.

- The following are costs associated with each variational iteration:
    * Explicit form (iii) for ρ: $\sim 10^{13}$ operations.
    * Spectral form for ρ: $\sim 10^{7}$ operations.
- We hope these numbers provide a clear justification for the usefulness of the spectral method at high resolutions.

- *... (hidden cost of the remark made in l. 407-408 that suggests duplicating the control vector when using the spectral method)*
    - We think the above remark concerns how our spectral method can be adapted to decouple land and sea points, by duplicating the control vector. Although this duplicating doubles the cost of the spectral method, it is an almost negligible increase compared to any of the explicit schemes mentioned.
- *... the authors find that assigning spatial prior error correlations for the initial state degrades the inversion (see their embarrassed explanation in l. 417-420).*
    - We are actually intrigued, rather than embarrassed about this finding. This result will hopefully spark some more investigation concerning the role of biases between the observations and the forecasts, so we can make this clearer in any revision.
- *Atmospheric inversions suffer from edge effects and it is usual to cut off both ends: in contrast to NWP, obtaining the optimal initial state is not strategic and therefore the representation of its prior uncertainty can be simplified.*
    - We wanted to keep the edge effect in order to study the effect of changing the representation of the B-matrix for the initial concentration. Flux inversion is fundamentally affected by the initial conditions, so this part of the B-matrix deserves some attention. Whether the ends are cut or not in practice is a concern of any application of the method over and above the testing done in this paper.
- *The second paradox of the paper is related. The objective of the method is to facilitate the resolution increase, but the detail of the increments is blurred by the horizontal reconfiguration operator Rh.*
    - This blurring is only marginal, and anyway reduces with increased resolution (interpolation from the high-resolution grid required by the spectral transform to another high-resolution grid required by the model), so we do not regard this as a significant issue, or a paradox. In fact it would not be needed with a model that has the same grid as the spectral transform.
- There are other comments of Referee #1 that he/she describes as minor. These can be dealt with in any revision, if requested by the Editor.

	Object	How determination of related objects scale with system size	How cost of use and storage scale with system size
1	$\mathbf{B}_{\mathrm{d}}$	—	$n_x n_y \left[n_z + T + 1\right]$
2	explicit $\mathbf{B}$-matrix (i) (c, ρ coupled)	$\mathbf{B}^{1/2}$: $\{n_x n_y \left[n_z + (T+1)\right]\}^3$	$\{n_x n_y \left[n_z + (T+1)\right]\}^2$
3	explicit $\mathbf{B}$-matrix (ii) (c, ρ uncoupled)	$\{n_x n_y n_z\}^3 + \{n_x n_y (T+1)\}^3$	$\{n_x n_y n_z\}^2 + \{n_x n_y (T+1)\}^2$
4	explicit separable $\mathbf{B}$-matrix (iii) (c, ρ uncoupled)	$(T+1)^3 + (n_x n_y)^3 + n_z^3 + (n_x n_y)^3$	$n_x n_y (T+1) +$ $(n_x n_y)^2 (T+1) + (T+1)^2 n_x n_y +$ $n_x n_y n_z + (n_x n_y)^2 n_z + n_z^2 n_x n_y$
5	$\mathbf{\Sigma}_c^{\mathrm{b}}$	—	$n_x n_y n_z$
6	$\mathbf{\Sigma}_{\rho^0}^{\mathrm{b}}, \ldots, \mathbf{\Sigma}_{\rho^T}^{\mathrm{b}}$	—	$n_x n_y (T+1)$
7	$\mathbf{\Lambda}_{\mathrm{h}c}^{1/2}$	$L^2 n_z$	$L n_z$
8	$\mathbf{\Lambda}_{\mathrm{h}\rho}^{1/2}$	$L^2 (T+1)$	$L(T+1)$
9	$\mathbf{S}_{\mathrm{h}}$	—	$(T+1+n_z) \times$ $\left[(L+1)L + (2L+1)\log_2(2L)\right]$
10	$\mathbf{R}_{\mathrm{h}}$	—	negligible
11	$\mathbf{F}_{\mathrm{v}c}$	$\mathbf{F}_{\mathrm{v}c}$, $\mathbf{\Lambda}_{\mathrm{v}c}^{1/2}$: $n_y n_z^3$	$2 n_x n_y n_z^2$
12	$\mathbf{\Lambda}_{\mathrm{v}c}^{1/2}$		$n_x n_y n_z$
13	$\mathbf{\Xi}^{-1}$	—	negligible
14	$\mathbf{F}_{\mathrm{t}\rho}$	$(T+1)^3$	$2 n_x n_y (T+1)^2$
15	$\mathbf{\Lambda}_{\mathrm{t}\rho}^{1/2}$		$n_x n_y (T+1)$

Tab. 1: Adapted from Table D1 in the paper. How the cost of various **B**-matrix model components scale with L, n_z, and T. The 'explicit' schemes are given in terms of the number of longitudes and latitudes, which are related to L via $n_x = 2L+1$, $n_y = L+1$. The third column reflects the cost of computing the component, where that requires the eigenvalues and eigenvectors of a matrix of size n (the computation is assumed to scale as $\mathcal{O}(n^3)$). The fourth column reflects the cost of storage and use of the component in each variational iteration. In the case of $\mathbf{S}_{\mathrm{h}}$, this has a part which is a fast Fourier transform, which scales as $(2L+1)\log_2(2L+1)$.

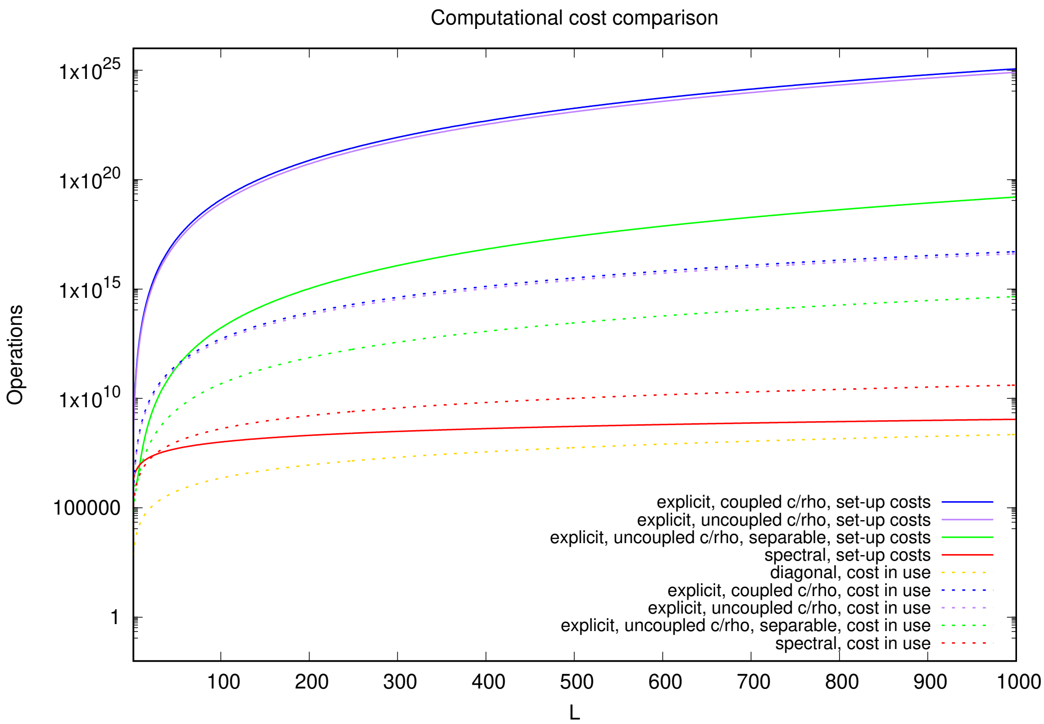

Fig. 1: (new Fig. D1) Plot of how the cost of various **B**-matrix model components scale with the total wavenumber, L. The continuous lines show how the cost of computing the components of **B** scale with L (i.e. setup costs, third column of Table 1), and the dashed lines show how the cost of using the components scale in each variational iteration (fourth column). The blue, purple, and green lines are respectively for the explicit **B**-matrix representations (i), (ii), and (iii) (rows 2, 3, and 4 in the Table), the red lines are for the spectral **B**-matrix representation (sum of rows 5 to 15), and the gold line is for the diagonal **B**-matrix (row 1 in the Table). All curves assume $n_z = 100$ vertical levels and $T = 12$ months.