

Manuscript number: EGUSPHERE-2025-44

Title: Improving the computation efficiency of a source-oriented chemical mechanism for the simultaneous source apportionment of ozone and secondary particulate pollutants

Reviewer comment #1:

Xu et al. develop a source-oriented chemical mechanism that with tags that is computationally efficient. A novel and useful contribution is the ability to retain the source tags through secondary chemistry with complexity that scales linearly with number of sources, rather than quadratically.

An algorithm that reduces the computational complexity from quadratic to linear with number is in itself a useful contribution (analogous complexity-linearized algorithm have been very useful in other subfields of atmospheric chemistry and physics, e.g. the superdroplet method). This approach is also well explained in section 2.1. However, the authors dedicate much of the paper to the Euler Backward Iterative (EBI) solver, where the motivation, methodology, and novelty is still unclear after several readthroughs. In a revision, the authors should make clear why EBI can't be applied to the fully expanded mechanism in V1, and make clearer what its contribution is and how the EBI relates to the linear reduction method.

I believe this paper could be a good fit for ACP (or perhaps the related journal GMD), but should address a few major issues that can be resolved with clarifying a few sections.

Below are specific questions and suggestions, divided into major and minor comments.

Response: We would like to express our sincere gratitude for the detailed and constructive feedback on our manuscript from the reviewer's positive comments and valuable suggestions. We confirm that all your concerns and suggestions have been fully addressed in the revised manuscript. This includes modifications to the main

technical issues, enhancements to the figures and tables, as well as substantial improvements to the clarity and precision of the language throughout the text. Specifically, revisions include clarifications regarding the main purpose of this study, EBI scheme's application and limitations, enhanced explanations of the proposed methods (dual-tagged reaction reduction and successive under-relaxation), corrections to equations and terminology, more detailed analysis of the dynamic under-relaxation factor, and additional discussion of error propagation over time. Below are the point-to-point responses to all of the comments (The comments are marked in black color and the responses are marked in dark blue color). Note that the following line numbers are shown in the corrected version.

Major Comments:

Comment 1: Line 22: There are a few sentences here that are copied and pasted verbatim into the conclusions. While these are the key points, I would discourage exact repeating of text in different parts of the manuscript.

Response: Thanks for the comment. The duplicated sentence has been modified to: “The new approach is based on tracking the total concentration of the source-tagged species and reduce the n^2 number of second-order reactions into $2n$ pseudo first-order reactions for chemical system with n sources, this method preserves individual species' production and loss rates, thus leading to improved computational efficiency because the total number of reactions increases linearly with the source number.” (Line 508–512) in the conclusion.

Comment 2: Lines 130-137: In this paragraph, it is unclear on how the EBI solver is incompatible with source-oriented mechanisms (without the linear reduction approach introduced in this paper). This is important motivation for the rest of the paper, but is not clear. What is a “non-typed regular mechanism”? What does “applicability” mean, specifically? By “explicit solutions”, do the authors mean closed-form solutions, e.g. algebraic solutions derived from pseudo steady-state assumptions, or explicit Euler solvers, e.g. forward Euler solvers with many small explicit timesteps? Refining the

language in a revision may help with clarity, and clearly explaining the limitations of applying EBI to source-tagged mechanisms is important for motivating this study.

Response: Thanks for the comment and sorry for the misunderstanding. The original version of the EBI scheme was designed to calculate the total concentrations of species, with reactive nitrate species strongly involve in the photochemical reactions such as NO, NO₂, O₃, HONO, their total concentration calculated as analytical solutions, by several steps including a set of complicated procedures for the non-linear algebraic equations (Hertel et al., 1993), rather than the generalized Eulerian backward scheme (equation (2)). In comparison to SMVGEAR solver, the virtue of EBI solver is speeding, as the computational cost is way lower than the former. However, the original EBI scheme could not be directly used for source-tagged mechanisms, because the analytical solution procedures for photochemical species in the EBI scheme are sophisticated, hard-coded and for total concentrations of species only, which inherently excludes source-tagged species. The words “non-typed regular mechanism” means the original mechanism, in which all species are referring to the total species regardless the source tagged, for instance the specie with name NO₂ indicates all NO₂ in the original mechanism. Contrarily, in the source-oriented mechanism, the NO₂ refers to only the NO₂ with null source tag, while the NO₂_Xi indicates the NO₂ relates to the ith source tag. Therefore, the original EBI scheme will be incompatible to the source-oriented mechanism, as the species other than null source tagged species will be excluded. Direct application of the original EBI scheme to the source-tagged mechanism results in incorrect solutions for null source tag species, even though they are included, because the underestimation of reaction rates neglects the contribution of source-tagged reactions. The “applicability” means with the capability of handling the source-oriented mechanism, the EBI solver could be used for tasks which requires timing efficiency such as air quality forecast scenario with source-oriented mechanism, if the concentrations of species related with certain emission sector are interested. Juxtaposing the words 'efficiency' and 'applicability' is not quite appropriate, as the latter is often a consequence of the former. The term “explicit solutions” refers to the

analytical solutions of total concentration of reactive nitrate species within the four special groups. These solutions are calculated directly using four sets of non-linear algebraic equations (e.g. Eq. 4a-4d in manuscript, also mentioned in Line 216). These analytical solutions, unlike implicit or iterative methods, employs a sophisticated solution procedure involving bunch of mathematical manipulation, as demonstrated in the work of Hertel et al. Revision has been made upon the mentioned points in the second last paragraph of section 1.2. (Line 140–152)

Comment 3: Equations 3a and 3b: This approach only works for bimolecular reaction rates, is that correct? It seems like the separation by source in this way is only compatible with the mass action law because of the bilinearity of $k[A_{tot}][B_{tot}]$. This would be good to acknowledge, or if compatible with other rate laws or reaction orders, perhaps include a note of how it is compatible.

Response: Thanks for the comment and sorry for the misunderstanding. This approach targets bimolecular reaction rates involving double source tagged reactants, which constitute a minority of the reactions within this source-tagged mechanism. Single tagged reactions, on the other hand, are the majority of source-tagged reactions. The reactants separation method remains valid as long as the reaction rate is proportional to the concentration or a power of the concentration of individual species (A) or a combination of species (A and B). Given the extremely low concentrations of atmospheric species, apparent mechanisms for atmospheric chemistry are typically linear or bi-linear, ensuring the compatibility of this method in current and future mechanisms. Consequently, the separation method is applicable because terms other than [A] can be lumped into the effective reaction rate constant, a common practice in atmospheric chemistry studies. Extra description added “This method is valid for reactions with rate laws of the form $rate=k[A]^m[B]^n\dots$, where the rate is proportional to the product of reactant concentrations raised to their orders.” to illustrate the compatibility of separation method to other forms of reaction rate expressions. (Line 174–176)

Comment 4: Lines 253-256: I don't understand this sentence. Is there a separate solver to predict the total concentrations, and then the ratios between the summed tagged concentrations and independently solved total concentrations used to scale each tagged concentration?

Response: Thanks for the comment and sorry for the misunderstanding. Yes, this might cause some ambiguity. Total concentrations are not calculated by a separate solver; rather, the original EBI calculates them prior to solving for source-tagged species. Photochemical species (e.g., NO, NO₂, O₃, O₃P) are calculated using the analytical solution subroutines mentioned earlier. Once these total concentrations are determined (as in 4a-4d), the source-tagged species are then solved (e.g., in 5a & 5b). For instance, solving for source-tagged NO and NO₂ requires the concentration of O₃, which appears on the right-hand side of 5a & 5b. Notably, the total concentrations of NO, NO₂, O₃, and O₃P (from 4a-4d) are determined by the aforementioned explicit method. Therefore, these total concentrations are a prerequisite for predicting source-tagged NO and NO₂. The purpose of this readjustment is to avoid numerical errors arising from the inversion of ill-conditioned matrices (6a-6c), a common issue in source apportionment calculations due to significant differences in the magnitudes of tagged species concentrations. Additional description has been added in the introduction "In the EBI solver, four family groups of strongly coupled species, are excluded from the general equation (2): (1) NO, NO₂, O₃ and O(³P), (2) OH, HO₂, HONO and HNO₄, (3) peroxyacetyl radical (CH₃C(O)OO·, or C₂O₃) and peroxyacetyl nitrate (PAN), and (4) NO₃ and N₂O₅. For these species, analytical solutions of four sets of non-linear algebraic equations (Eqs. 9-12 in Hertel et al., 1993) are applied to determine their concentrations at time t instead of using the P and L terms with equation (2). The detailed mathematical procedures involved in obtaining these analytical solutions are listed in Appendices A.1-A.4 of Hertel et al. (1993)." (Line 130–137)

Comment 5: Line 260: The use of "final" here is ambiguous. Would it be more accurate to say "current iteration"? The use of t in the subscript already implies final

in terms of time (though the authors refer to this as current timestep, which also seems appropriate). But from what I understood from the results in 3.1, this must be applied at every iteration up to convergence, not just an adjustment applied post convergence to the solution to make the “final” solution $\sim 20\%$ (or $1-\alpha$) more like the previous timestep? This would be important to make clearer in a revision.

Response: Thanks for the comment and sorry for the misunderstanding. Yes, this might cause ambiguity. Regarding the term "final," which refers to the concentration value being updated in the solution at the end of an iteration, it should be changed to "updated." (Line 281) The modification using values from the previous iteration is introduced to avoid the non-converging oscillations characteristic of fixed-point iteration. Furthermore, the description in equation 11 are inaccurate. The under-relaxation method updates the solution vector at the end of an iteration based on values from the previous iteration and the predicted values in the current iteration, not values from old and current time steps. Consequently, all instances of "time step" should be replaced with "iteration," including in Eq. (11). And the word "final" changed to "updated", and corrections are made in line 279-285 as well as in Eq. (11).

Comment 6: Fig 2 and table 2: This is a very interesting result, and a revision could go in more depth on these results, which would be a major addition to the study. I noticed the dynamic under-relaxation factor is not fully monotonic, is there a reason for that? More generally, how did the authors obtain the values for alpha shown in table 2? Is there any explanation why the solution in later *iterations* should weight the previous *timestep* more heavily? Finally, do the authors have any explanation (or intuition) for why a dynamic under-relaxation factor converges faster than a constant alpha?

The dynamic under-relaxation factor is not fully monotonic, is there a reason for that? More generally, how did the authors obtain the values for alpha shown in table 2? Is there any explanation why the solution in later iterations should weight the previous timestep more heavily? Finally, do the authors have any explanation (or intuition) for why a dynamic under-relaxation factor converges faster than a constant alpha?

Response: Thanks for the valuable comment. Yes, an in-depth discussion should be added. The under-relaxation approach is commonly employed in commercial Computational Fluid Dynamics (CFD) codes to mitigate numerical instability in highly non-linear systems. For under-relaxation of scalar variables such as concentration or energy, a recommended α value is 0.8, consistent with the optimal fixed α shown in Fig. 2. However, a dynamic α generally offers better performance due to its greater flexibility. Analysis of predicted solutions without under-relaxation reveals that in the initial iterations, the solution vector rapidly shifts towards the vicinity of the true solution before oscillating around it until convergence is achieved. Therefore, to accelerate convergence, an aggressive selection of α (larger value) is beneficial in the early iterations to quickly propel the solution vector towards this near-optimal space. In later iterations, a more conservative selection of α (smaller value, weighting the previous iteration more heavily) is preferred to update the solution vector slowly towards the true solution, gradually eliminating residual errors and effectively damping out slight overshoots. This initial "kick" provided by the larger initial α explains why a dynamic α often converges faster than a fixed α , which lacks this initial boost. This behavior is analogous to a damped oscillatory system, where critical or slightly under-damped systems achieve equilibrium faster than over-damped or highly under-damped ones, which are either too slow or prone to instability.

The α values in Table 2 were derived from the aforementioned concepts and fine-tuned through several numerical experiments. The non-monotonic behavior of α is attributed to the fact that for most test cases, convergence was achieved within 10 iterations. In the less frequent cases which require more iterations, a slightly larger adjustment step could accelerate the approach to the true solution. The α value of 0.79 for iterations greater than 16 represents a compromise between rapid movement and effective damping of overshoots. And in-depth discussion has been added to the relevant places. (Line 398–409)

Minor Comments:

1. Line 25 and the copied phrase in the conclusion: improved -> improving
Response: Thanks for the comment, “improved” has been changed to “improving”.
(Line 24)
2. RS1: To distinguish source from stoich, I would suggest using superscripts for tags or operators (NO)_{S1} + (NO₃)_{S1}
Response: Thank you for your suggestion. Subscript for source tag in RS1 changed to superscript. (Line 51)
3. RS2: Would a different example with tagged products be better here, to illustrate the relevance? I know this is just an example, but why have different reactions if RO₂_R and RCHO products are the same species for every reaction? This approach is more useful than just for estimating different sinks of reactants.
Response: Thank you for your suggestion. A different example with TERP + NO₃ has replaced the previous one. (Line 70)
4. Line 59: Is this “also” as in additionally, or is it specifically for the reason in the previous sentence? If also refers to species quadratic scaling as well as reactions, maybe this point could be made right before the N₂O₅ example.
Response: Thanks for the comment. Also means increase in both number of reactions and tagged species, a clarification has been made before “also” in the manuscript. The sentence in manuscript has been modified to: “The necessity to deal with N₂O₅, which can be generated from NO₂ and NO₃ from different sources, is handled with double-source-tagged species N₂O_{5,sij}. In addition to a potential quadratic scaling of the number of reactions, the number of N₂O₅ species also increases quadratically with the number of explicit sources, leading to near-quadratic growth of the overall number of species when the number of types to track gets higher.” (Line 56-61)
5. Line 94: Jacobin -> Jacobian
Response: Thanks for the comment and sorry for the mistake. Jacobin has been changed to Jacobian. (Line 94)
6. Line 121: I would suggest reminding the reader here of $\Delta C_t^{(m+1)}$ here, to avoid ambiguity between changes in C over time versus changes in C over iteration.

Response: Thank you for your suggestion. C has been changed to $C_{i,t}$ for a new iteration. (Line 126)

7. Line 127: The use of explicit here is confusing (see the major comment). The authors can disambiguate in many ways, e.g. if explicit solutions mean closed-form algebraic expressions arising from pseudo steady-state assumptions, say that. In the context of iterative solvers, explicit often refers to forward Euler methods versus implicit backward Euler methods.

Response: Thanks for the comment, “explicit” has been changed to “analytical” (Line 134)

8. Line 212: unknowns -> unknowns

Response: Thanks for the comment and sorry for the mistake. The “unknowns” has been changed to “unknowns”. (Line 231)

9. Line 237: quadratic -> quadratic

Response: Thanks for the comment and sorry for the mistake. The “quadratic” has been changed to “quadratic”. (Line 257)

10. Table 1: Maybe some point in the text explain V2-S is V2 with SMVGEAR?

Response: Thanks for the comment. V2-S has been explained with additional words “# V2-S: V2 reaction mechanism with SMVGEAR solver” underneath the Table 1. (Line 333)

11. Line 324-325: “For most of the species, a relative tolerance of $1e-3$ is used”. What are the exceptions, and why do they differ?

Response: Thanks for the comment. The convergence criteria are $1e-3$ for most of species that are in differential forms, however some pseudo steady state species such as O_3P and O_1D , the convergence criteria are 1.0, indicate convergence check not applicable. Explanation has been added “For most of the species, a relative tolerance of 1×10^{-3} is used. Exceptions include pseudo-steady-state species like O_3P and O_1D , for which a tolerance of 1.0 is applied, indicating that a convergence check is not applicable.” (Line 353–355)

12. Line 359: “for more than three folds” was confusing at first. Not sure if it refers to the 74% figure comparing total time to V2-S, or some other comparison in Table 1.

Maybe this could be a little clearer where this conclusion is coming from. I might also suggest wording such as “by more than threefold” or “by a factor of 3 or more”.

Response: Thanks for the comment, “for more than three folds” has been changed to “by a factor of 3 or more”. (Line 390)

13. Line 402: “than” could be substituted with “compared to” or something similar.

Response: Thanks for the comment. The word “than” has been substituted with “compared to”. Also, this part has been rewritten for in-depth discussion of the errors. (Line 454)

14. Line 432: This sentence might work better without “zone”

Response: Thanks for the comment. We had deleted the word “zone”. (Line 494)

15. Line 450: “needs only 11 times of the computation time” Is this the first timing comparison to the non-source-oriented mechanism in the manuscript? This result would be good to report before the conclusions, though fine to reiterate it in the conclusion. For a frame of reference, it would be good to compare the other simulations to the untagged mechanism computation time, at least V1, and ideally V2-S as well. If including this result, why not include it in Table 1?

Response: Thanks for the comment. The phrase, "Test cases based on the Texas Air Quality Study 2006," was mistakenly placed in the conclusion and has since been moved to the introduction, where it is now properly referenced. Additional data regarding the timing of untagged mechanism with SMVGEAR solver have been added into Table 1. (Line 328-333)