

Response to referee comments on “A machine learning-guided adaptive algorithm to reduce the computational cost of atmospheric chemistry in Earth System models: application to GEOS-Chem versions 12.0.0 and v12.9.1”

We thank the editor for his careful reading of the manuscript and the valuable comments. This document is organized as follows: the editor’s comments are in *italic*, our responses are in plain text, and all the revisions in the manuscript are shown in blue. **Boldface blue text** denotes text written in direct response to the Referee’s comments. **The line numbers in this document refer to the updated document with tracked changes.**

We have made these major changes.

1. We conducted an 8-year simulation from 2013 to 2020 and found that the error is still stable over this period.
2. We removed all texts claiming that our algorithm can be used in earth system models. See tracked changes in the manuscript for more details.
3. We have a new Figure 8. We show the error of all species in Figure 7, 8 and all related figures in the supplementary materials.
4. We calculate the errors using three different definitions as suggested by reviewers and the editor. See section 2.5 for more details.

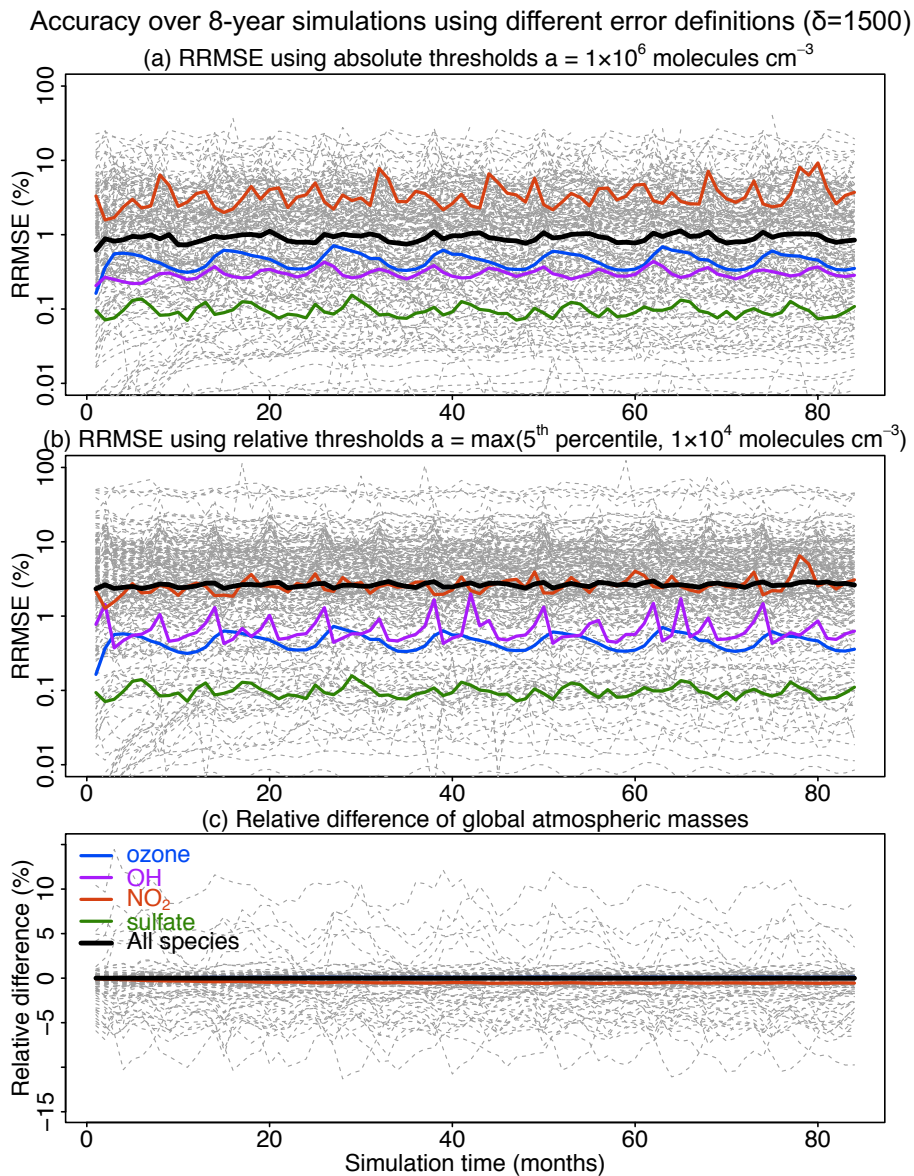


Figure 8. Accuracy of the adaptive reduced chemistry mechanism algorithm over an 8-year GEOS-Chem simulation using a threshold δ of 1500 molecules $\text{cm}^{-3} \text{ s}^{-1}$ to separate fast and slow species. (a) Same as Figure 7c but for the 8-year simulation. Here we use an absolute threshold of species concentration ($a = 1 \times 10^6 \text{ molecules cm}^{-3}$) to define the RRMSE (See Section 2.5). (b) Same as Figure 8a but using a relative threshold ($a = \max(5^{\text{th}} \text{ percentile}, 1 \times 10^4 \text{ molecules cm}^{-3})$) to define the RRMSE (See Section 2.5). (c) Relative difference of global atmospheric masses for each species. Dashed lines show results for all 228 species in the mechanism. Results are also shown for the median RRMSE across all species in the mechanism and more specifically the RRMSE for ozone, OH, NO₂, and sulfate.

Major comments

One of my main concerns is the superficial and nature of discussion you offer, often having little physical context. Such instance are:

Response to Reviewer 1 comment on Line 87: Your response still does not offer any context regarding the

selection criteria. Why 10 molecules and not 100 or 1000? How much larger is “slightly larger than 0 molecules cm⁻³ s⁻¹” you mention? What “because of numerical precisions” means here at all? I cannot accept this response as scientific.

Response. We examined different thresholds in some short-term experiments in the early stage of this project. We found that the error quickly increases if the threshold is above 10 molecules cm⁻³ s⁻¹. You are right that our previous discussion was superficial, and we provide more details here.

Line 143. In each submechanism, if a reaction is slower than 10 molecules cm⁻³ s⁻¹ for all gridboxes that select this submechanism, then the reaction is considered negligible and removed from the submechanism. **The logic is that such a slow reaction will not contribute significantly to the total species production/loss rate threshold $\delta = 500-1500$ molecules cm⁻³ s⁻¹ for the species to be included in the Jacobian matrix.**

Similarly,

Reviewer 1 on Line 243, Reviewer 2 on Line 297 (and generally comments/replies regarding the choice of thresholds): You select a new threshold without explicating why – an approach “we use now lower value and it does (or does not) change things” is vaguely related to the nature of the selection, it does not help understanding which factors are important or not. Where are the important thresholds for distances? Why testing for 10 species is now sufficient as compared to 5 species before? It appears to be an arbitrary choice.

Response. Thanks for this good point. This was done with repeated tests of different thresholds and then we determine if the optimized chemical blocks are consistent with our chemical intuition. We rewrite it here.

Line 208. Second, for each species *i*, we decrease its distance with the 5 species that have highest similarity with it by 50% and this scaling is applied only once for each species pair. We found in tests that using 10 highest-similarity species instead of 5 and decreasing distances by 30%-70% instead of 50% did not significantly change the results. The logic is that the number of species with similar chemical characteristics is usually around 5-10 and decreasing the distances among them by 30-70% can increase the probability of these species to be in the same chemical blocks after the optimization process.

Furthermore,

Reviewer 1 on Figure S6, Reviewer 2 major comment: Showing the average difference for the species is alike reporting the average body temperature for all patients in the hospital. Not only that does not help understanding the biases for the species of importance, how would you judge/set the input (e.g. weight) of every species to this average?

Response. Thanks. Now we show the error of all species in Figure 6, 8, S4, S7 and S8. Please check them for more details.

A large bias in simulating N₂O in troposphere (where it is mostly inert) will hardly affect CH₄ lifetime there, whilst a small bias in OH will do so (especially on the long-term, I will return to that below).

Response. We don't simulate the evolution of long-lived species in our standard chemistry model, following standard practice of simulations focused on tropospheric chemistry – we now clarify that this is our focus. We have also shown that the error is stable in an 8-year simulation.

Line 113. In the simulations presented here, methane, N₂O, and other long-lived halocarbons have fixed concentrations in surface air (Eastham et al., 2014; Murray, 2016) so that the longest resolved chemical modes are less than a year.

We also removed all text claiming that our algorithm can be applied in earth system models because we haven't tested it. We have also changed the title,

Title. A machine learning-guided adaptive algorithm to reduce the computational cost of integrating kinetics in global atmospheric chemistry models: application to GEOS-Chem versions 12.0.0 and 12.9.1

Fig. S7 and its vertical scale are deceptive – behind larger relative differences for some species (not indicated which) one does not check for growing biases (I spot one in the right panel for NO₂ nonetheless)

Response. Thanks for pointing this out. Now we show all errors (in log scale) of all species in figures. Please check updated Fig. 7, 8 and S7.

The stated errors that “also remain at 0%” make no sense from the numerical point of view – does this mean that you compute numerically identical result (read copy it)? So how large are deviations really? Do you mean below 0.1%? The metric “+/-10% for >99% of the other species” is also not reasonable – what if the remaining 1% of species has other important (in addition to key ones) species, e.g. reactive compounds?

Response. Sorry, our old narrative wasn't precise, and we have removed all these sentences. We now show all errors in Figure 7, 8 and the distribution of errors in Figure S4.

Line 376. The relative differences of global atmospheric masses are within 10% for >99% species (Figure 8c) and show no sign of increasing trends. The relative difference for NO₂ increases slightly from 0% to 0.4% in the first 30 months and then stabilizes at 0.4%. Key species like OH, ozone and sulfate have a relative difference smaller than 0.01% throughout this simulation period.

*I give a few hints to help adequately re-address these comments. Classical ODE integrators employ absolute and relative tolerances for estimating the errors about the solution that is obtained. This is done to constrain errors when relative scales of changes for the spectrum of variables integrated vary many orders of magnitude, and is shown to be valid using through the strict mathematical analysis (see, e.g., [1], Chapter on ODE integration). Whilst I do not require such here, a **minimum sound mathematical analysis of the order of the errors in your method should be present in the revised version.***

Response. Thanks, but we are afraid that a simplified ODE assumption may not be applied here. We mentioned in text that the Rosenbrock chemical solver is a 4th-order implicit method and the linear approximation for slow species is a 1st-order explicit method. But we feel it is difficult to further analyze the order of the errors mathematically because of operator splitting with other processes including deposition and transport.

A shortcoming of your current method, analysis and presentation of the results is also in using only absolute thresholds. Ozone and OH differ by orders of magnitude, yet you use the same absolute threshold values for these two species in the analysis. Presenting errors at the surface and at 15 km altitude (Figs. S6-7) has little sense for OH, whose tropospheric sink is 80% in the free troposphere. Which atmospheric abundances are compared in Fig. S7, are these gridboxes or domain averages? These are only a few questions, I could go on... but they already show how inconclusive and not-well-thought-through are the analysis and discussion you present.

Response. Thanks for these good points. Here we report the error using both absolute and relative thresholds. See Figure 8 for more details. And we have a new paragraph discussing this.

Line 272.

$$RRMSE_i = \sqrt{\frac{1}{Q_i} \sum_{j=1}^{Q_i} \left(\frac{n_{i,j}^{\text{reduced}} - n_{i,j}^{\text{full}}}{n_{i,j}^{\text{full}}} \right)^2} \quad (\text{where } n_{i,j}^{\text{full}} \geq a) \quad (7)$$

Absolute threshold $a = 1 \times 10^6$ molecules cm^{-3}

Relative threshold $a = \max(5^{\text{th}} \text{ percentile}, 1 \times 10^4 \text{ molecules } \text{cm}^{-3})$

Line 372. Our algorithm also shows no sign of increasing errors over an 8-year simulation. Figure 8a-b displays the RRMSE over this period by taking account of all gridboxes with species concentrations above an absolute threshold of 1×10^6 molecules cm^{-3} or a relative threshold of the 5th percentile of species concentrations across all gridboxes (see Section 2.5 for more details). In both cases, the median and the maximum RRMSE remain constant over this simulation period. The relative differences of global atmospheric masses are within 10% for >99% species (Figure 8c) and show no sign of increasing trends. The relative difference for NO₂ increases slightly from 0% to 0.4% in the first 30 months and then stabilizes at 0.4%. Key species like OH, ozone and sulfate have a relative difference smaller than 0.01% throughout this simulation period.

To recap, you need to substantially improve the analysis of errors, presentation and discussion in order to confirm that your method is accurate and convergent (w.r.t. the reference) and support your statements, especially the ones in the conclusion section. Please add a definition or explicate what exactly you imply by “chemical coherence” (better in the introduction).

Response. Thanks for pointing this out. We use chemical initiation instead.

Line 69. A major development here is to enforce that chemically connected species be grouped in the same blocks, so that the blocks are consistent with chemical intuition and can be logically modified and extended as the mechanism changes.

As of now, you show no evidence of “no error growth over multi-year global simulations” because many atmospheric compounds have lifetimes longer than 3 years (read they need to equilibrate to be consistently compared to the reference). Fig. S7 adds dubiety here, too. I suggest that you simulate at least a 30 years long period in order to corroborate that no substantial errors in CH₄, NO_x and other reactive N reservoirs and hence O₃ are building up in the optimised mechanism.

Response. A 30-year simulation is not necessary for tropospheric oxidant/aerosol chemistry, which we now clarify as our focus. Here we conduct an 8-year simulation and we find the error is stable over this time period. Please check Figure 8 for more details.

Line 372. Our algorithm also shows no sign of increasing errors over an 8-year simulation. Figure 8a-b displays the RRMSE over this period by taking account of all gridboxes with species concentrations above an absolute threshold of 1×10^6 molecules cm^{-3} or a relative threshold of the 5th percentile of species concentrations across all gridboxes (see Section 2.5 for more details). In both cases, the median and the maximum RRMSE remain constant over this simulation period.

Note that an error of a few percent in oxidising radicals may lead to largely diverging results in a few decades. For instance, using a simple spreadsheet integration you may infer that a longer by 1% (as compared to assumed average of 9 years) atmospheric CH₄ lifetime would yield CH₄ abundance larger by already 3% (above 50 ppb, or growth rate in last 5 years!) in just recent 30 years, assuming that emissions are the same in both cases. So can your optimised method really simulate 3 decades of CH₄ evolution with an accuracy of 1%?

Response. We don't simulate the evolution of long-lived species in our standard chemistry model. We now make it clear in text.

Line 114. In the simulations presented here, methane, N₂O, and other long-lived halocarbons have fixed concentrations in surface air (Eastham et al., 2014; Murray, 2016) so that the longest resolved chemical modes are less than a year.

Minor comments

Reviewer 1 on Line 212: What are the objective criteria, if you state "This is done objectively here..."

Response. We removed 'objectively' and now explain in detail how we construct coherent blocks in the following sentences.

How can "two species ... appear in the same reaction more times"? Present an example.

Response. "appear in the same reaction" was confusing. We have replaced by "appear together in multiple reactions". For example, NO and NO₂, OH and HO₂ may appear in a lot of reactions related to VOC oxidation.

Line 157. In general, two species should have shorter distances if they appear together in multiple reactions (e.g. NO and NO₂, HO and HO₂) or have similar products in the mechanism.

Reviewer 1 on Line 285: Specify exactly which index denotes product and which denotes educt, i.e. "... include both educt i and product j species".

Response. Sorry, we didn't make it clear. Either species i or j could be the reactant.

Line 164. Where T_(i, j) is the number of reactions that include both species i and j (with one as reactant and the other as product, **either i or j could be the reactant**).

Reviewer 2 on Line 465: Copy-pasted responses do not respect reviewers and editors time invested in this manuscript. I consider it a mauvais ton, especially as this happens not the first time in this submission.

Response. Sorry, I did this because reviewers have similar questions. It is not a good manner.

Reviewer 1 on Line 209: Do you "think" that there is more active iodine in the Block 3, or do you "believe" in that, or do you postulate that? Can you actually look into the model results to confirm that?

Response. We indeed looked into the model results. Please see the details here.

Percentage of gridboxes in the global tropospheric+stratospheric domain that treat this species as fast.

Block 3: BrNO₂(16.5%), IONO(9.5%), OIO(28.9%), ClOO(78.6%), OCIO(20.6%), BrCl(25.5%), HOI(35.1%), Br₂(28.2%), IONO₂(24.7%), BrNO₃(40.2%), I(51.1%), IO(50.7%), HOBr(48.6%), HOCl(34.4%), ClNO₃(30.8%), BrO(51.3%), HCl(46%), HBr(26.9%), Cl(79%), Br(51.4%), ClO(49.1%)

Block 4: AERI(0.2%), ISALA(0.1%), ISALC(0.2%), I₂O₄(1.2%), I₂O₂(4.7%), I₂O₃(0.9%), IBr(0.7%), INO(0.9%), HI(1.2%), ICl(4.4%), Cl₂O₂(2.2%), ClNO₂(2%), BrSALC(1.9%), BrSALA(2.2%), I₂(2.3%), Cl₂(2.5%)