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 referees for their careful reading of the manuscript and the valuable comments. This document is organized as follows: the Referee'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 manuscript with tracked changes.

Referee 2

This paper present some clearly very well thought out and well executed methodologies that are shown to be effective at minimising computational costs for solving a complex chemical mechanism in the GEOS-Chem model. These tool are likely to have substantial benefits for the air quality and Earth System modelling communities. Unfortunately, the writing of the paper is confusing at times, understandable given the complexity of the subject matter but it does distracts from messages it is trying to convey. With a few small changes to the structure and presentation, this will be an excellent paper well suited for publication in GMD.

Major Comments

The biggest issue I find with the paper is the it goes into the nitty gritty of how it partitions different species/reactions into different categories (sections 2.2-2.5) before it explains the overall design philosophy and what these different categories are used for (in section 3.1). The result is that on first reading, I got very confused reading through section 2 and was only able to make sense of it on second reading. This could be improved substatially if either Section 2.1 were expanded upon to give an overview of the whole design philosophy and what each of the the different subcategories (fast, slow species, unimportant reactions etc.) are going to be used for before they are described in detail, or a new Section 2.2. is added giving an overview of all of the developments. Some of this description is in the first paragraph of Section 3.1. Also section 3.1 repeats a lot of the text currently in section 2.1 about the model description and is inapropriate here, therefore section 3.1 should also be editted to avoid repition. In short, Section 2 should contain all of the descriptions and definitions for the model and the adaptive algorithm for the chemical operator. Section 3 should focus on testing, evaluating and optimising the algorhithm in the 3D model. If these two sections were more clearly defined, the paper would be much easier to follow.

Response. Thanks. We have made the following changes to improve the presentation quality.

1. We now give an overview of the whole design philosophy in the last paragraph of the introduction part (as also suggested by the first referee).
2. We have moved part of the text in the first paragraph of Section 3.1 to Section 2.1 and 2.2.
3. We have moved Text S1 back to Section 2 to give a better description of our method.
4. We have moved Figure S1, S2 and S6 to the main text.

There are also some inconsistancies in language and definitions - I found the use of "slow species" and "slow reactions" (each of which use different threshold definitions) particularly confusing. It would help to be consistent and use "unimportant reactions" for those <10 molecules cm-3 s-1. I have specific comments below to help with this.

Response. Thanks. We now use 'slow reactions' throughout the paper because it is more accurate than 'unimportant reactions'.

In terms of the statistical analysis (Section 2.6) it is important to have a measure of bias as well as error - I would generally be more concerned about a species that has an error of 1% and bias of ~1% (which would imply a consistant error in one direction), than one which has an error of 1% and a bias of ~0% (which would implpy more random error). Looking at Figure S8, the errors in surface Ozone seem to be biased in one direction. That is potentially concerning if the bias is enough to significantly affect tropospheric ozone burden - as ozone is a radiatively important species this could have consequences in an Earth System model. Please also include a normalised measure of bias in Section 2.6.

Response. Thanks for this good point. We now define a relative abundance metric in the paper.

Line 465. A second metric to evaluate our adaptive chemical mechanism is the relative difference of atmospheric abundances for all species compared to the standard simulation. This tests for accumulating bias over long simulation periods.

The tropospheric ozone doesn't show accumulated bias in one direction. As seen from Figure 7, its bias has some seasonality. And now we have a new figure to show the relative difference in atmospheric abundances of all species.

Line 719. The median relative difference in atmospheric abundances among all species remains at 0% over this 3-year period; the relative differences for key species like ozone, OH, sulfate and NO2 also remain at 0% and are within ±10% for >99% of the other species (Figure S7).

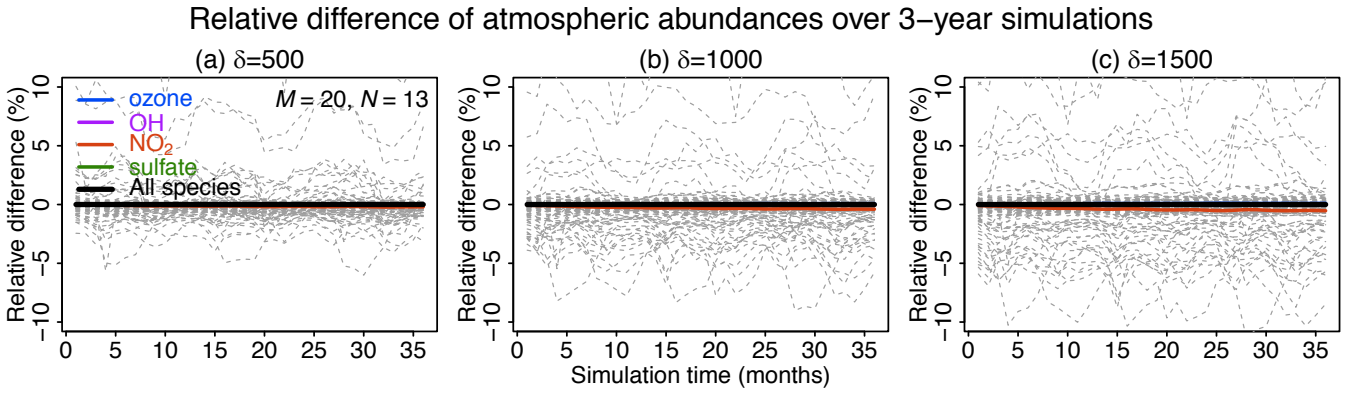

Fig. S7. Relative difference of total atmospheric abundances using adaptive reduced chemistry mechanism algorithm over a three-year GEOS-Chem simulation. Results are shown for the median relative difference across all species in the mechanism and more specifically for ozone, OH, NO_2, and sulfate. The grey dashed lines are for all 228 species.

Finally, there are a number of figures in the supplement which would have been useful in the main paper and I did not understand why they were not in the main paper. I would recomend moving at least figures S1, S2, S5, S6, S7 and S8 into the main paper.

Response. Thanks. After considering both referee's suggestions, we now move Figure S1, S2 and S6 back to the main text.

Specific comments

ln 26. Ambigous "it" in: "because it exerts strong forcing and feedbacks...". change to "because chemical and aerosol species extert strong forcings and feedbacks..."

Response. Done, thanks.

*ln 49. Change "guide us build" to "guide us **to** build".*
Response. Done, thanks.

Ln 88. here and elsewhere, please be consistent and use the term "unimportant" instead of "slow" to describe those reactions that are removed with fluxes <10 molecules cm-3 s-1, otherwise it is easily confused with the "slow species".
Response. Thanks. But we think 'slow reactions' are more accurate than 'unimportant reactions', so we use it throughout the paper.

Section 2.3. If I am to understand this correctly, when species are defined as being "slow" and/or "long-lived", they are not "removed" from the mechanism per se*, rather they are solved using the analytical approach instead of the 4-th order Rosenbrock solver. Given this, I think you need to make clear here that when you say "coupled system", you mean all of the species and reactions which are solved using the Rosenbrock solver. All of the species that are "removed" or "excluded" from the coupled system, as you say later, are instead solved using the simple analytical approach.*
Response. Thanks for correcting this. Now we say
Line 203. We solve the fast species in their submechanism using the standard Rosenbrock solver. For the slow or long-lived species, we approximate the evolution of concentrations using an explicit analytical solution that assumes first-order loss (Santillana et al., 2010).
Line 209. **As such, we still update the concentrations of all species but in a more efficient way**.

Line 99. Please can you define "distance" qualitively here before the quantitative definition.
Response. Now we say
Line 212. We construct coherent subsets ('blocks') of the species in the mechanism species based on their linkages through the mechanism reactions. This is done objectively by defining the species distances in the mechanism using graph theory. **In general, two species should have shorter distances if they appear in the same reaction more times and have similar products in the mechanism**.

Line 112. The term "distance" is overloaded in the text to mean multiple different things. You have defined a new term "Euclidian distance" |Di-Dj|, which is a scaler, rather than the previous "distance" Di which is a vector. This Euclidian distance is then modified to cluster species together (I would call this something like "clustered distance"). Please make clear that it is this "clustered distance" that is stored in the matrix and used to calculate the cost function Z (eq 4).
Response. Thanks, we prefer to name it as 'modified distances'. Now we say this in the text.
Line 298. We store these modified distances of all pairs in a 228x228 matrix.

*Line 113. When applying the 50% factor, is this applied iteratively? i.e. if two species are each others closest pair, will their Euclidian distance be reduced by a factor 0.5*0.5=0.25, as iterating through each species, or is the 50% factor only applied once? Please be clear which approach you are using.*
Response. This scaling will be applied only once.
Line 297. Second, for each species *i*, we will decrease its distance with the 5 species that have highest similarity with it by 50% and **this scaling is applied once for one pair of species**.

The 50% factor and 5 closest neighbours both seem quite arbitrary, but I think I can see how this approach will cause clustering of species into close families. What was the reason for the use of these two values, and did you test other values?
Response. We indeed tested other numbers and the results are in general consistent.
Line 297. Using 10 highest-similarity species instead of 5 and decreasing distances by 30% or 70% does not change the results.

Line 135. Would benefit from Figure S2 being in the main text here. Looking at Figure S2, I think you can make a fair cost-benefit argument that M=20, N=13 is well optimised as it is on the bottom-left of the 30% contour. It is clear from the contour lines that you get diminishing returns of the fraction of species if you were to increase M and N, hence selecting on the 30% contour seems reasonable. By selecting the bottom-left part of the 30% contour, you are minising both N and M.
Response. Thanks, we now move Figure S2 back to the main text. Now we say this in text.

Line 528. In order to make the code manageable, we choose to use $M = 20$ resulting in an optimal value $N = 13$ at which only 30% of the species need to be treated as fast in the global tropospheric and stratospheric domain (Figure 3).

Basically, you can be more rigorous in the justification for why you used M=20, N=13 than simply saying you "choose" them.
Response. Now we say this in the main text.
Line 528. In order to make the code manageable, we choose to use $M = 20$ and an optimal value $N = 13$ at which only 30% of the species need to be treated as fast in the global tropospheric and stratospheric domain.

Line 137-139. Please move this text to the top of section 2.5, as this describes the training set used to derive the submechanisms and blocks.
Response. Done, thanks.

Line 196. Please again change "slow reactions" to "unimportant reactions". Please clarrify - are the unimportant reactions removed from each of the submechanisms using the original training data, or are they removed on the fly depending on the concentrations of species in each grid cell at each timestep?
Response. Thanks. We prefer to name it 'slow reactions' and we have made it clear that the slow reactions removed in each submechanism are pre-defined and the same.
Line 524. The slow reactions removed in each submechanism are pre-defined (see Section 2.2 for more details).

Removing the reactions from each of the submechanisms in advance seems like the more efficient approach to me. However, there is a risk that the approach becomes inconsistant if used in different time periods with different chemical conditions to the training data. For example, there will be risk in using submechanisms derived with present day training data in preindustrial conditions. This is relevant for application in Earth System models.
Response. Thanks for raising this point and we now discuss it in the text.
Line 200. Here we remove these slow reactions in each submechanism based on present-day atmospheric chemistry environment and it should be re-evaluated if this method is applied in other periods (e.g. pre-industrial times) when the atmospheric conditions could be very different from our present-day one.

line 229. The full mechanism is by definition not a submechanism. Say that it is the 21st "chemical regime".
Response. Done, thanks.

Line 236. slow reactions -> unimportant reactions
Response. Now we use 'slow reactions' throughout this paper.

Figure 1. The line in panel a shows the fraction of species removed from the coupled mechanism. *Call the slow reactions unimportant reactions.*
Response. Now we use 'slow reactions' throughout this paper.

Figure 3. There are 21 chemical regimes, made up of M=20 submechanisms plus the whole mechanism.
Response. Now we use 'slow reactions' throughout this paper.

Figure 4. Number of chemical regimes is 21 (M+1), not 20.
Response. Done.

Supplementary material

line 12. think Z_2 *should be called* f *or* f(M,N) *instead. Use a different term to* D *for the gridcells, because* D *is already used for the distance vectors.*
Response. Fixed and we now use P instead of D.

line 18-20. Unclear how this algorithm works, don't know what is meant by "temperature" here.
Response. Now we add a reference to describe how this temperature parameter works.

figure S5. presumably chemical regimes should go up to 21, not 20? Also, its the 21st regime that has 100% of the mechanism, regime 20 has 90% according to Figure 3.
Response. We have updated this figure. Please check Figure S3 for more details.

Figure S6. Please include panels showing biases for the key species. If any have consistant/growing biases, that should be commented on (especially concerning for ozone). What value of δ did you use here?
Response. Thanks for this good point. We now define a relative abundance metric in the paper.

Line 465. A second metric to evaluate our adaptive chemical mechanism is the relative difference of atmospheric abundances for all species compared to the standard simulation. This tests for accumulating bias over long simulation periods.

And now we have a new figure to show the relative difference in atmospheric abundances of all species.
Line 719. The median relative difference in atmospheric abundances among all species remains at 0% over this 3-year period; the relative differences for key species like ozone, OH, sulfate and NO2 also remain at 0% and are within ±10% for >99% of the other species (Figure S7).

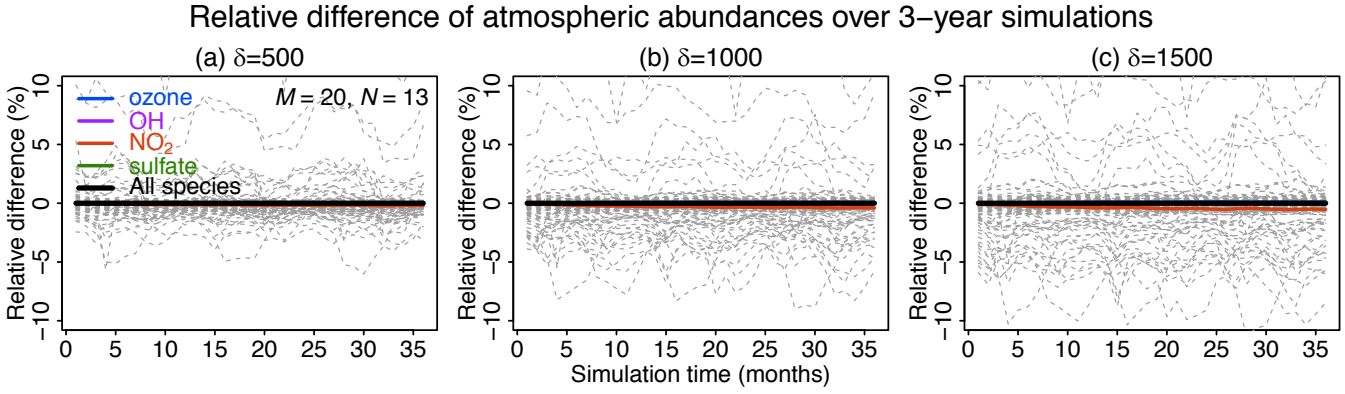

Fig. S7. Relative difference of total atmospheric abundances using adaptive reduced chemistry mechanism algorithm over a three-year GEOS-Chem simulation. Results are shown for the median relative difference across all species in the mechanism and more specifically for ozone, OH, NO_2, and sulfate. The grey dashed lines are for all 228 species.