

Interactive comment on “An adaptive method for speeding up the numerical integration of chemical mechanisms in atmospheric chemistry models: application to GEOS-Chem version 12.0.0” by Lu Shen et al.

Anonymous Referee #1

Received and published: 11 November 2019

The integration of the chemical differential equations in a chemistry transport model is a significant computational burden. Both climate and air quality forecasting models spend a significant fraction of their computation marching forwards through time, solving these differential equations. Developments in speeding up this code have been not been forthcoming over the last decades and it has only really been the availability of more, faster CPU cores that has allowed us to run increasingly complex chemistry. Algorithmic developments to speed up the integration of this chemistry are to be welcomed.

C1

This paper outlines a method for running different subsets of the differential equations in different geographical regions, so speeding up the solutions. On the one hand, this is a fairly obvious thing to do. We don't need to be integrating all of the chemistry of isoprene over the middle of the Pacific in the same way as we do over the middle of the Amazon. However, attempting this adaptive chemistry has not been readily taken up by the community. This paper attempts a first realistic attempt at such an implementation.

In general, the paper describes a novel new technique with a potentially sound methodology which could be extremely useful to the atmospheric chemistry transport modelling community. At this point though, I am not convinced that the algorithm is working as intended, or if it is, the explanations given in the text are satisfactory for the reader to understand what is going on. If either of these outcomes can be corrected the paper should be published.

1) My major concern about the methodology is the split of species into the different blocks. The methodology for doing this is explained but there is little interpretation of the results. In many cases, the blocks seem to have lumped together some fairly random sets of species and this appears to have been hidden away in the SI.

Block 1: “Aromatics” (Benzene and Toluene but the chemically almost identical XYLE is in Block 7?) but it also contains CH₂I₂. What is the advantage of using the integrator for these together? Does this mean that the aromatic chemistry is also being solved with the integrator in the middle of the oceans where the CH₂I₂ is important? Block 2: “Organic nitrates” This contains some organic nitrates and the N atom. These organic compounds are only really at appreciable concentrations at the surface whereas the N atoms are only really applicable in the upper stratosphere? Block 3: “Isoprene” This seems to contain some isoprene chemistry but also HFCS which seems surprising. Block 6: “Halocarbons” seems to be again mainly isoprene species to me, but it also contains HFC and CFC species which would only be important in the stratosphere.

Is there an explanation for this? Are the blocks in the SI correct? I might be missing an

C2

important concept here. But it has not been explained. To me, this feels like an error has occurred somewhere either in the species list given in the SI or in the algorithm.

Given the random nature of the annealing algorithm, is this set of species blocks always the same one? What degree of variation is present when running this algorithm multiple times?

There needs to be more work done to explain why these blocks are the best ones to use given the variation seen in the species type in each block. I realise this is a result of the optimization algorithm but the situation at the moment appears to be that an algorithm has told us that this is the result and we are going to leave it at that. To my mind, the species within these blocks do not appear to have the properties you would expect given what is trying to be achieved. I'm happy to be convinced otherwise but the text at the moment does not achieve that and it is not possible therefore to be confident that the algorithms are working appropriately.

2) The Supplementary information figures should be contained within the main text. The paper is fairly short and some of the figures are central to understanding the methodology. Figure S2 could be removed by putting the species list into Table 1. The other figures are small enough to be included in the main body without overwhelming the reader.

3) I found the structure of the beginning of the paper a bit confused. We have an Introduction; a section on the chemical operator which includes a very brief description of the GEOS-Chem model and the KPP system. We then have a section on the algorithm being described. The material about the chemical operator should be moved into the introduction as this basically supports the introductory text about the chemical integration. The model description should go into a separate section. 4) Page 1 Line 30. The number of reactions thought in play in atmospheric chemistry is significantly more than the "hundreds" described in the text. The MCM has 10s of thousands and mechanisms produced by GECKO-A produces millions. Hundreds are used in the sim-

C3

plified mechanisms for atmospheric chemistry transport models. The text should be clarified here.

5) Page 2 lines 55. There are now some other approaches to speed up the chemical integration using "Machine Learning" approaches they could be cited here.

6) Page 6 line 155 the reference to Santillana is 210 rather than 2010.

7) Page 6 line 166. Do the authors mean 'fast blocks' rather than ,fast blocks,?

8) Figure 1. When asking the reader to "see text" can this be more specific? What does the shaded area represent? The SD between what, the monthly values?

9) Figure 2. Although using this approach does provide some information it would be useful to split the dataset in another way. Could there be a figure which shows a map of the world indicating whether each block is switched on at that location. This need only be done at the surface for 0 GMT and 12 GMT but it would give some confidence that the approach is working. isoprene block should only be on over continental regions etc. It is very hard to get this level of information from the figures as presented. Why is the value of delta of 100 used in this figure and 500 used in other figures?

10) Figure 4. Can the figure caption give more information here? What actually is being compared? Is this the RRMS in the monthly mean fields, or in the hourly values averaged to a monthly mean? Is this all of the species in the Jacobian?

11) It's not obvious that the code for the annealing algorithm is included in the repository. I've had a look but can't find it.

12) Conclusions. a. It would be useful to discuss whether this algorithm could be used within the adjoint framework for data assimilations, inversion studies? b. The authors discuss the suitability of this approach to minor mechanistic changes. However, if the algorithm is to be useful it needs to be sustainable within the software lifecycle of the chemistry transport model. Could this be spelt out in more detail? Presumably, if a new species was added the training algorithm (which species into which block and

C4

how many blocks etc) would need to the re-run with new data, but a small change in species lifetime would not lead to a re-running. It would be useful to have the conditions which are required for the training to be updated to be described.

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2019-279>, 2019.