

Response to Referee comments # 2

The paper examines the sensitivity of chemical simulation to the operator duration in a global CTM (GEOS-Chem). By proposing a measure called “normalized error”, the paper also seeks the optimized operator duration setup that balances the accuracy and the computation. Finally, the paper compares the errors from operator splitting to those from coarser resolution and concludes that the errors from coarser resolution dominate. The topic is an important yet understudied problem in the atmospheric modeling community and thus is well suited for GMD. However, I find the paper has several problems, which I list in the following comments. The paper is publishable if the authors address these comments in a revised manuscript.

We thank the referee for these valuable comments. Our responses to referee comments are below in blue. We have modified the article based on the referee’s suggestions. Line numbers refer to the revised manuscript.

Major comments

1. Confusion of terminology. There are two different “timesteps” in a CTM. One is the duration or timesteps of the operators (e.g., chemistry, advection, etc...) resulting from the operator splitting, i.e., the focus of this paper. The other is the integration time step in an ODE or PDE solver within operators. The example of the latter is the SMVGEAR chemistry solver in GEOS-Chem, which uses variable time steps to integrate the stiff ODE system efficiently and accurately. (Even 1 min is way too long to solve the rapid chemical reactions!) Similar internal integration timesteps exist in the advection scheme to integrate PDE. The authors do not seem to understand the difference in these two types of timesteps. For example, although the focus of this paper is the “operator duration”, the authors introduced the Courant number, which is in theory more related to the “integration timestep” within the advection operator (Less than unity Courant number guarantees a stable solution from the PDE solver. The Courant criterion does not consider operator splitting at all!).

In addition, with coarse $4^{\circ}\times 5^{\circ}$ and $2^{\circ}\times 2.5^{\circ}$ resolution in GEOS-Chem, the Courant number should be almost always less than unity, even when the operator duration is 1 hour. In P. 9599 L.9, the authors claim “...until the transport timestep exceeds 30 min. Then the large Courant

number increases errors by an order of magnitude for long lived species of O₃ and CO". The statement is unlikely to be true.

I strongly suggest that the authors clearly define the "timestep" that they focus on and make a distinction between the "operator duration" and "integration timestep". (Unfortunately, there seems no widely-accepted terminology. So define them clearly from the very beginning.)

We thank the referee for suggesting ways to address this source of confusion, which also addresses the major concerns raised by the first reviewer. We now use the terms "operator duration" and "integration timestep" throughout the paper, and avoid the ambiguous term "timestep". As suggested, we removed the description of the CFL criterion in the introduction section to reduce confusion.

We eliminated the concerning phrase about Courant number on lines 252-254, "Errors increase fairly smoothly with increasing chemical and transport operator duration until the transport operator duration exceeds 30 min. Then errors increase by an order of magnitude for long lived species of O₃ and CO."

2. The authors used a reference simulation with the shortest operator duration to represent the "truth". I am fine with this method for the purpose of this paper. But I would also recommend the author to put the "errors" in the context of the model-observation errors. In this way, readers can have a good idea whether the operator-duration errors are significant compared with model uncertainties, which, I think, is a very important piece of information for readers. For example, Fig. 3 shows the relative errors are on the order of 10^{-4} or 10^{-5} (But I think these results are from a wrong calculation, see the next comment), which is very small and negligible, given the usual model-observation errors in current state-of-art models.

Helpful suggestion. We added lines 255-259, "These relative errors are comparable to typical model-observation errors of ~30% for NO_x (Boersma et al., 2008; Hudman et al., 2006) and 20 - 40% for SIA (Philip et al., 2014; Heald et al., 2012), while smaller than typical model-observation errors of ~20% for ozone (Zhang et al., 2011; Wang et al., 2009) and 10 - 20% for CO (Duncan et al., 2007; Shindell et al., 2006)."

3. Incorrect definition of statistics. P.9596 L. 11 the equation is not a RMSE normalized by mean concentration, as the author claimed in P.9596 Line 8-10. In statistics, RMSE is usually

defined as $\frac{\sqrt{\sum_{i=1}^N (x_{sim} - x_{obs})^2}}{N}$. Therefore, RMSE normalized by mean concentration, or the

coefficient of variation of the RMSE, should be $\frac{\sqrt{N} \sqrt{\sum_{i=1}^N (x_{sim} - x_{true})^2}}{\sum_{i=1}^N x_{true}}$. The author's simulation errors defined in P.9596 Line 11 will decrease with increasing number of samples, and thus is not a good statistics!

And I think that is why the numbers in Figure 3 are so small (on the order of 10^{-4} or 10^{-5}) (because you have a large number of samples with model simulation results) and is inconsistent with the order of magnitude (10^{-2} to 10^{-3}) in Figure 4.

In addition, the name of the statistics, "simulation error", does not reflect the "relative" nature of the statistic. I would suggest using "relative simulation error".

Good points. As suggested, we have revised Equation 3, Figure 4 and Table 1. We note that normalized error (or CNE; equation 4) and Figure 6 are unaffected by this change.

As suggested, we replaced "simulation error" by "relative simulation error".

Minor comments

1. P.9592 Line 21-22, "For computational convenience, production and loss terms also simulated as individual operators". This is not true. With the SMVGEAR solver, chemical production and loss are simulated at the same time within the chemistry operator.

Sentence removed.

2. P. 9593 L. 22, P.9595, L.9. Authors mentioned the nested simulation, but did not have an adequate presentation of and discussion on those results other than Fig. 5. Remove the nested simulation or add more contents.

We added an additional figure (Figure 3), and added lines 243-248, “Figure 3 shows the sensitivity of simulated species to changes in operator duration (C20T10 to C10T05) at two other horizontal resolutions (global $4^\circ \times 5^\circ$, and nested North America $0.5^\circ \times 0.67^\circ$ simulations) considered here. Spatial patterns of monthly mean ground-level concentrations, and absolute differences are similar, albeit with finer spatial heterogeneity resolved in the nested simulation. However, some resolution dependent differences do arise reflecting nonlinear feedbacks.”

3. P. 9594 L. 6. Equation (3) shows the order of operators in standard GEOS-Chem with two groups of operators, one with duration T and the other with duration C. It is easy to think of this equation when $C=T$, but I wonder in which order these operators are implemented when $C=2*T$. Please add some clarification.

We added line 122, “Transport operations are repeated twice before a chemical operation when $C = 2 \times T$.”

4. P. 9597 L.11. The author uses the Normalized Error defined in this line to find the optimal setup. First, the name Normalized Error, is not proper, because it does not reflect that the measure adjusts for the computation expense or that it includes four different species. I suggest the author to come up with a better name for the measure. Maybe CPU-time adjusted composite normalized error? Second, although it is a good idea to try to find a balance between computation (CPU time) and accuracy (simulation error), I doubt the measure is useful in practice, because the measure does not take into account the respective cost of computation and errors, and thus is in fact comparing apple to orange. I suggest the author to add a discussion on this weakness.

As recommended, we changed the terminology to CPU-time adjusted Composite Normalized Error (CNE). We added lines 199-201, “The relative value of CPU time versus simulation accuracy is subjective and depends on scientific objective. This definition of CNE gives equal weighting to the respective cost of CPU time and simulation accuracy.”

5. P. 9598 L. 6-10. The explanation for why concentrations of primary tracers increase is hard to understand. Why “more homogeneous fields have lower dry deposition and chemical loss rates”? Should the effects of longer operator duration also depend on the chemical regime? For

example, NO_x clearly decreases in Eastern US and Southern China. Also see the additional comments in the end.

We modified lines 221-222, “More homogeneous fields have lower dry deposition rates as a larger fraction is mixed aloft, and lower chemical loss rates depending on the chemical regime.”

6. P. 9598 L.11. The oxidation of CO cannot effectively decrease OH over the source region, because the reaction is slow and CO oxidation does not consume HO_x radicals.

We modified lines 222-223, “The increase in CO over source regions is partly associated with decreases in OH.”

7. I suggest the author use quotation mark with every “True” or “Truth” that refers to the reference simulation, e.g., P. 9600 L. 15.

Modified throughout the paper.

8. P. 9600 L.19-20. The author claims the error due to coarse resolution dominates over that the error due to longer duration operator duration. This claim may well be true in this study since $4^{\circ} \times 5^{\circ}$ and $2^{\circ} \times 2.5^{\circ}$ are both coarse horizontal resolution, in relative to the horizontal extent of the air pollution phenomenon. However, as the resolution goes further down to 36 km, 12 km, and 4 km and 1 km, the operator splitting errors are expected to become more important. If the author can explore this aspect using the nested simulation results or other regional CTM, the paper will be much more significant. But at least acknowledge this point in the discussion.

Good point. We revised the discussion in lines 294-303, “The simulation error for all species at $4^{\circ} \times 5^{\circ}$ resolution increases by an order of magnitude compared to $2^{\circ} \times 2.5^{\circ}$ resolution for any choice of operator duration tested here. The error in this configuration is insensitive to operator duration, and dominated by representativeness differences due to spatial structure resolved at $2^{\circ} \times 2.5^{\circ}$ resolution, but not at $4^{\circ} \times 5^{\circ}$ resolution. Nonlinear chemistry at different horizontal resolutions (e.g., Wild and Prather, 2006) also plays a role. Numerical errors due to advection processes generally exceed those from operator splitting (e.g., Prather et al., 2008; Santillana et

al., 2016). We therefore recommend prioritizing horizontal resolution over operator duration for offline CTMs using time-averaged meteorological fields as tested here. As meteorological fields used in CTMs become available at finer temporal and spatial resolution, the value of shorter operator duration should further increase.”

9. Table 1. The caption is hard to understand. And as suggested above, check the calculation of the simulation error.

We modified Table 1 statistics based on the corrected Equation 3. Table 1 caption now reads (lines 723-724): “Comparison of mean* relative simulation error versus horizontal resolution, with “truth” defined at $2^\circ \times 2.5^\circ$ horizontal resolution”

Additional comments

P. 9598 L. 6-10. In my opinion, the effect of longer operator duration depends on the order of operators and where you archive your output. In another word, the effects from the sequence “advection”, “chemistry”, and then “output” and those from the sequence “chemistry”, “advection”, and then “output” should be different, **even if the operator duration is the same**. The differences from these two setups decrease as the operator duration becomes smaller. Therefore, the effects of increased concentration of primary species in P.9598 L.6-10 may only apply to this study’s implementation and may be misleading to other model setups.

So I suggest that a more rigid and a more theoretical study on this paper’s topic (the impact of operator splitting) should measure how good the results from operators in different orders converge as a function of operator duration instead of errors relative to a reference simulation. I hope the authors can explore this aspect if they are interested.

We thank the referee for these thoughtful comments. A study of the order of operators was recently conducted by Santillana et al. (2016) using the GEOS-Chem model. Therefore we focus on operator duration here.