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 under-studied 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.

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 °×5 ° and 2 °×2.5 ° 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.)

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.
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 $\sqrt{\frac{\sum_{i=1}^{N}(x_{sim}-x_{obs})^2}{N}}$. 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”.

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

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 HOx radicals.
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.
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 °×5 °and 2 °×2.5 °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.
9. Table 1. The caption is hard to understand. And as suggested above, check the calculation of the simulation error.

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.