Response to Reviewer Comments (C2459) of the manuscript, **"Development of a grid-independent GEOS-chem chemical transport model as an atmospheric chemistry module for Earth System Models"** by M.S. Long et al.

Jan. 9, 2015

We thank the reviewer for their thoughtful comments and recommendations for improving the manuscript. To the extent possible, we have addressed the reviewer's concerns. Each comment (in italics) followed by our corresponding response to that comment is listed below. Unless otherwise noted, line numbers refer to those in the original, PDF version of the discussion manuscript published online.

Reviewer 1 – C2459

It would be useful to outline details of QA procedures involved in ensuring that the "state-of-science" and other developments mentioned are safe and fit for purpose in terms of scientific and technical performance. For example:

p.7507, lines 19-29: The document refers to quick and efficient implementation of new developments. It would be useful to outline what quality assurance methods are employed to prevent new developments being taken up before they are proved beneficial and/or safe. (e.g. any automated testing? "Benchmarking" may mean different things to different people, so it would be helpful if this were briefly clarified in terms of how scientifically valid model evolution is ensured as well as technical performance.)

We now specify in the introduction that "benchmarking" refers to rigorous QA. Further QA details are given at the end of Section 2.1 and we have expanded on those in revision.

p.7508, lines 13-15: This seems to suggest new developments are immediately adopted without any QA. Presumably the impression is incorrect in which case it would be useful to discuss such procedures.

We now clarify that QA is applied to any new developments. Again, further QA details are given at the end of Section 2.1.
.

It would be interesting, if possible, to comment in more detail, in the context of the abstract and the main document on the expected scalability and performance beyond the range of processors actually tested.

The following text has been included in the manuscript starting on p. 7515, L21:

"The results further suggest that the chemistry module would remain efficient for simulations beyond the range of values tested"

Individual Issues:

p.7510, line 18: Use of the term "leverage" seems rather flowery. Words with more obviously understandable meaning such as "use", "adopt" or "employ" would aid clarity.

The sentence now reads, " much of the FORTRAN-77 code base was updated to Fortran-90".

p.7511, line 7: It seems unclear unclear what "hooks" are in this context. Are these additional interface or wrapper routines or some sort of trigger mechanism?

The text has been revised. The text at p. 7511, L7 now reads,
   "The GEOS-Chem code includes specific conditional-compilation flags to accommodate the ESMF interface and permit coupling with external data streams. These flags do not interfere with GEOS-Chem's scientific operation and are used exclusively in grid, I/O, and utility operations. There are three flags invoked as C-preprocessor statements: *ESMF_, EXTERNAL_GRID,* and *EXTERNAL_FORCING.*"

p.7511, line 10: Re the sentence: "They can remain invisible to the scientific programmer." To what end? i.e. When would that be desirable and how is it achieved?

This statement has been removed from the text.

p.7512, lines 3-8: It would help the document flow to establish this as a standard working practice earlier in the document. (See earlier specific comments about QA.)

We have added mention of QA in the introduction.

p.7512, line 21: Missing word? Existing wording seems to imply that initialization is performed at the beginning of each time step. That seems unlikely. Should this say "at the beginning of the first time step" or at the "start of the run" or something similar?

The words "time step" have been replaced with "simulation".

p.7514, line 22: Use of the term: "Scalability simulations" for clarity of meaning. Earth system simulations have been run using different resource configurations in order to establish scalability, so the term "scalability tests" or "scalability analysis" would seem more appropriate. i.e. scalability itself is not being simulated.

The line has been revised. It now reads, "Simulations used to test strong scalability of the coupled system were run…"

p.7515, line 8: Suggest the use of "Wall-clock time", "elapsed time" or two separate words "wall time" rather than "walltime".

Instances of "walltime" were replaced with "wall time".

p.7515, line 19: Suggest rewording to avoid the suggestion that scaling efficiency has truly been demonstrated for ANY number of cores.

The words, "all numbers of cores", have been replaced with "the range of cores tested."

p.7515, line 22: The word "other" is in quotes. The reason for this is unclear. Would it help to provide general examples of the scalability, performance and code structure characteristics of the important elements of these "other" components as compared with the chemistry module?

The word "other" has been removed from the text.