Review of "Description and evaluation of tropospheric chemistry and aerosols in the Community Earth System Model (CESM1.2) (GMD-2014-236)", by S. Tilmes et al.

This manuscript presents an overview assessment of the performance of tropospheric chemistry within the CESM1.2. The presentation makes extensive comparison to observation for validation and assessment of model results. In the Abstract, the authors state that the system is "well suited as tools for atmospheric-chemistry modeling studies in the troposphere and lower stratosphere, whether with internally derived "free running" (FR) meteorology, or "specified dynamics" (SD)" and follows with "The main focus of this paper is to compare the performance of these configurations against observations from surface, aircraft, and satellite, as well as understand the origin of the identified differences." Unfortunately, the paper doesn't seem to sufficiently address the second part of the statement well enough to support the first.

It is recognized that a truly thorough evaluation of a system as complex as this is difficult and time consuming, but the content of this paper is a pretty extensive information dump that for the most part lacks anything more than speculative and unsupported guesses regarding the results, their main drivers and why they are what they are relative to observation.

- Why and what is the impact of the new reactions added?
- Can the impacts of clouds, dynamics, aerosols, etc. be more explicitly evaluated to address the inter-model differences and differences between model and observation?

It seems that this paper should go in one of two directions:

1) It should be extensively revised to include a more explicit statistical analysis of the results and their comparison to observation. This would be perhaps be best done with a smaller set of observations that highlight some of the main driving factors that impact results: e.g. tropospheric O_3 is often a clear signal of vertical transport. Methane lifetimes reflect OH abundance and chemistry, but this pits one of the slowest reacting species as an indicator of one of the fastest reacting. Is it possible to more directly assess their interaction?
2) The paper should be retracted and resubmitted as two papers: One as a technical description paper with a lot more technical discussion and evaluation. And a second that evaluates the model's performance against observation, including a proper statistical analysis.

The authors have clearly done a lot of work preparing and testing the model system, compiling observations, and developing tools for comparison. They are well regarded for their effort, but this paper, unfortunately, is sub-par. I add, lastly, that with the density of acronyms it is somewhat difficult to read.

Please consider my comments below for more detail.

P8878, L1: The intro starts off rather abruptly. Please include a general introduction to the CESM.

P8879, L27: With fixed SST, what impact might this have on truly evaluating the interactions between chem., physics, and dynamics in the results?

P8880 – Since the system uses chemistry coupled to the modal aerosol scheme, comparison to results from Long et al. (2013, GMD; 2014, ACP) with more explicit multiphase chem would perhaps be valuable to this MS.

P. 8882, L15-17: Small particles DO impact het. chem. Why do the authors believe they don't? This would only be the case if the specific chemical mechanism used specifically neglected the chemistry associated with smaller particles.

P. 8888, L 20-24: At the resolution used, is CAM even able to resolve STE well enough to permit diagnostic analysis of the impact of stratospheric chemistry and exchange on the results?

P8889, L10: SAD is used here but defined on p. 8890

P8891, L13-15: This statement is self-contradicting. It appears you're using the term SAD in place of the more appropriate term surface-to-volume ratio. Perhaps this should be changed.

P8892 and in general: Since this is a chemistry paper, more or less, it is important that the appropriate symbols and terminology are used. Is SO_2 meant to refer to SO_2 gas? If so, it should be $SO_2(g)$. SO_4, as presented, suggests some non-ionic sulfur oxide radical. If it is meant to be particulate or aqueous SO_4^{2-}, it should be presented as such.

P8892, L17-20: Unfortunately, for a publication seeking to present a proper analysis of a chemistry model system, the terms "reasonable agreement" and "agrees well with" are insufficient. This work should be a proper analysis based on observational and modeled statistics. Avoiding this type of analysis completely undermines the utility and ability to extract meaningful information from this as a scientific work. The work that the authors have done to date compiling data and creating unbelievably useful tools for making comparisons and analysis possible should be leveraged to do at least something to step beyond the current state. IF, on the other hand, it is the authors intention to perform a more complete analysis for presentation in another manuscript for submission elsewhere – which may be the more suitable way forward – then this should be stated and much of this analysis should be removed from the manuscript.

P8896, L6, L8; P8897, L1; P8904, L9: Invoking cloud impacts on O_3 seems speculative here. Further analysis is both possible and necessary.

P8893, L28 – P8894, L1: If the AOD bias is the result of too much seasalt and/or dust, wouldn't this mean that the winds may be too high? This would be a straightforward analysis, wouldn't it? Also, given that AOD is based on the MIRAGE system, wouldn't a sensitivity to RH also be just as likely to affect AOD?

P8904, L3-4: This is improper English.

Table 1 (and most of the rest of the data shown): What are the statistics? Are they means? If so, what about std. dev. If they're medians, what about range? In any case, are the statistics appropriate? For example, if the optical depth data are means, it should be pointed out that AOD is not normally distributed and therefore means are meaningless (no pun intended). In fact, most of the quantities presented in this table and elsewhere are known to not only NOT be normally distributed, but the distribution and resulting statistics are known to differ with model horizontal resolution. In most cases, there are clear modes within the distributions that can be attributed to individual geographic regions, features, or locations in the atmosphere. As such, changes in one area may appear to affect the whole from a statistical standpoint, when in fact there is no change at all elsewhere. It is possible that any analysis or comparison based on these data are insupportable.