Review of "Specified dynamics scheme impacts on wave-mean flow dynamics, convection, and tracer transport in CESM2 (WACCM6)" by Davis et al.

In this manuscript, entitled "Specified dynamics scheme impacts on wave-mean flow dynamics, convection, and tracer transport in CESM2 (WACCM6)", the authors present a rigorous examination of the sensitivities of various aspects of the transport and dynamical circulation in CESM2 to nudging the large-scale flow. A systematic exploration of how errors depend on both meteorology frequency as well as nudging timescale is presented. As this manuscript presents an unparalleled level of information, detailing the behavior of the nudging framework within CESM2, it represents an important study that is certainly deserving of publication. I commend the authors for performing this type of evaluation, which is somewhat lacking in the literature and, as illustrated by the authors, riddled with complexities and nuances. There are, however, a few ways in which the manuscript can be improved to both clarify the applicability of these results to other (namely reanalysis) fields as well as to focus the discussion on the major findings in order to communicate more clearly the first order results. As such, my recommendation is "major revisions", but I emphasize to the authors that this study has strong potential and will constitute a valuable contribution to the field, provided that these revisions are incorporated.

Major Comment 1:

While I agree with the authors that using meteorology produced from the same underlying GCM is, indeed, the cleanest way to assess the performance of the nudging scheme, one wonders how the major conclusions from the study change (if they do) upon nudging to reanalysis fields. In particular, one of the central challenges with nudging is striking the "correct" balance between sufficiently constraining to the observed (reanalysis) fields, while also not doing too much harm in the actual act of nudging (i.e. producing dynamical inconsistencies in the flow that may further degrade the transport characteristics of the simulation). Therefore, while it is certainly worthwhile to evaluate the performance of the nudging scheme within the (parameter and flow) space of the parent GCM, ultimately what matters in the end is how that scheme operates in the combined "GCM-reanalysis" space. In other words, one needs information (not provided in the current manuscript) about the underlying meteorological biases in the underlying GCM (CESM2). How confident are the authors that a scheme that reproduces CESM2 fields also reproduces other (reanalysis) fields, given what might be quite large model biases? The authors have shown that the behavior of the nudging scheme does not necessarily exhibit nice convergence properties (for example — the minimization of the convective mass flux errors at a 12 hour nudging timescale, with increased errors at both shorter and longer nudging timescales (lines 220-222)). If the nudging scheme presents such complicated behavior, when constrained with its own fields, one can imagine the behavior might become still more complex when applied to reanalysis fields, potentially leading to major differences in the author' conclusions.

To this end, I think a necessary addition to the manuscript is the addition of a few simulations wherein CESM2 is nudged to MERRA-2 (or any reanalysis product of the authors' choosing) using a subset of the nudging timescale parameter combinations that are explored in the original set of experiments. I understand that these caveats are mentioned in lines 431-438 but I am not

convinced that this is sufficient to address this issue. I am not suggesting an exhaustive set of runs but, rather, two or three simulations that demonstrate that the main findings of the study also hold when nudging to reanalysis fields. Since CESM2-SD is already set up to nudge to reanalysis fields, I cannot imagine that this is an unreasonable ask.

Major Comment 2:

I commend the authors for performing a quite exhaustive examination of the errors in various circulation and transport diagnostics within the context of the simulations considered in this study. At the same time, however, the complexity of the results renders the manuscript very descriptive, and it can be difficult for the reader to extract the key results from some of the more secondary points. In particular, the separate discussions of the ozone and carbon monoxide errors (Sections 7 and 8) are quite long and nuanced. Is there not a way to combine into a single section that is prefaced by a paragraph highlighting the common features among these two constituents (with respect to their response to nudging), followed by a discussion of each regarding the specific characteristics of each field? At present, the number of details presented in the manuscript renders it a bit hard to follow and the authors should better emphasize throughout the main "take-away" messages (which are currently reserved only for the conclusions).

Minor Comments:

1. Line 155: Shouldn't the stratospheric vs. tropospheric averages be somewhat latitude-dependent, at least enough to distinguish between first-order differences in tropopause height? In particular, my concern is that in the extratropics 200 hPa is already well within the lower stratosphere and yet still quite tropospheric in the tropics. Why not use something a bit more physically based? (i.e. 300 hPa for latitudes > 40S/N and 100 hPa for equatorward latitudes)?

2. Figure 4: This is a very important figure, and it is nice to see this result documented so well. It could be worth noting in the text when describing this result that similar behavior was observed in Figure 3b,d in the study indicated below, albeit for only two nudging timescales (and using CAM, nudged to MERRA). Consistent with the results presented in this study, that figure shows that the convective mass fluxes in the 5hr nudged simulation differed substantially more than the fluxes from the 50 hour nudged simulation, relative to MERRA. Perhaps it is worth highlighting this consistency between the two studies.

Orbe C., D. W. Waugh, H. Yang, J. F. Lamarque, D. Kinnison, and S. Tilmes (2016), Tropospheric Transport Differences Between Models Using the Same Large-Scale Meteorological Fields, *Geophysical Research Letters,* 44(2), 1068-1078.

3. Line 165: "Negative values of the SAME indicate that the field is weaker in magnitude than it is in the reference simulation - up to and including opposite in sign". I do not necessarily agree with this description, and I think the issue is just one of unclear wording. For example, consider

that x_ref=-2 and x_sd=3. Then, according to (5), SAME<0. And, yet, the field is *greater* in (absolute) magnitude than it is in the reference simulation. Perhaps the “up to an including opposite in sign” just needs to be clarified that it refers explicitly to x_sd.

4. Line 401: Is conclusion 1(b) true? Don’t the errors in convective mass fluxes *increase* at nudging timescales shorter and longer than 12 hr?