Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-39-RC2, 2020

Geoscientific Model Development Discussions
Open Access

Interactive comment on "A multi-year short-range hindcast experiment for evaluating climate model moist processes from diurnal to interannual time scales" *by* Hsi-Yen Ma et al.

Anonymous Referee #2

Received and published: 5 May 2020

1 Reviewer's Summary of Manuscript

In "A multi-year short-range hindcast experiment for evaluating climate model moist processes from diurnal to interannual time scales," Ma et al. describe an experimental design for (1) differentiating model errors that arise largely due to errors in parameterized processes (fast processes) versus errors that arise largely due to errors in the model's dynamical state (slow processes); and (2) for aiding in the identification of physical processes that might need focused model development to correct errors in fast processes. The experimental design consists of using a specific atmosphere-only GCM for two sets of simulations:

Printer-friendly version

Discussion paper

- 13 years of 3-day hindcast simulations, initialized using state information from ERA-Interim (interpolated to the model grid), and

- 13 years of an Atmospheric Model Intercomparison Project (AMIP) simulation

The authors implement this experimental protocol for the Community Earth System Model (CESM) version 1.0.5, using the CAM5 physics package. They compare simulations from the hindcast and AMIP simulations with a variety of observations, including ARM observations, reanalysis, satellite, and other. The analysis focuses on four aspects of the simulations: the diurnal cycle of convection over the southern Great Plains (SGP) of the United States; diabatic processes in the Madden-Julian Oscillation; dependence of diagnosed errors on El-Nino / Southern Oscillation (ENSO); and cloud radiative effects.

Their analysis of convection over the SGP shows that CESM1.0.5 has too much high cloud cover, which the authors speculate might be due to overly-frequent triggering of convection. They also show that the model has too little mid-level cloud coverage and too much shallow cloud coverage; possible reasons for this are not provided. Their analysis of the MJO shows that the model tends to have too little precipitation in the western half of the MJO core and too much precipitation on the eastern half, even in the hindcast simulations; and they suggest that diabatic processes associated with fast physical processes is likely the cause of these errors.

The authors show that the correlation between short- and long-term errors does not depend on the ENSO state, and they argue that their experimental design is therefore robust with respect to interannual variability. They also give evidence that errors in the large-scale circulation dominate errors in cloud radiative effects diagnosed from the model, relative to errors associated with fast physical processes (e.g., cloud parameterization) alone.

They conclude by arguing that this experimental design–and specifically the CESM 1.0.5 hindcast simulation output–are useful for model development.

2 Summary of Review

The authors do a nice job of summarizing several ways in which short-duration hindcast experiments can be used to diagnose the origins of errors in long-term climate simulations. Overall, the authors make a compelling argument that this type of experimental design is useful for decomposing errors into those associated with fast physical processes and slow (circulation-related) processes.

Despite this, I have several major reservations about the current form of the current manuscript, such that I cannot recommend this manuscript for publication at this time. Most critically, the current manuscript does not make a clear case for how this manuscript presents a new experimental design that is unique relative to what the authors have previously published on. Additionally, the authors do not provide enough information for a reader to replicate their experimental design. Finally, the authors' arguments are undermined by their use of a model version that has not been officially supported or developed for 7 years.

The following sections give a detailed description of these concerns, and the final section points out some specific issues that the authors should address if they submit a revised version of the manuscript.

3 Major Concerns

3.1 Uniquness of manuscript and relevance for GMD

The manuscript is presented as describing “a multi-year hindcast experiment and its experiment procedure” and giving “examples to demonstrate how one can better utilize simulations from this experiment design.” Given that this is the authors’ explicitly-stated premise for the paper, it is not immediately clear to me how this manuscript represents a unique contribution to the literature in general, and why it is a good fit for GMD in particular. The authors have written numerous papers about ensembles of short-duration hindcasts for over 15 years, and in that time they have already made a compelling argument that this experimental design is valuable. Why is yet-another-paper demonstrating the utility of this type of experimental design needed in the literature in general? After reading the manuscript several times, it is still not clear how this manuscript is unique relative to existing literature.

With respect to GMD in particular, there are three types of manuscript (wrt https://www.geoscientific-model-development.net/about/manuscript_types.html) that this could plausibly fit as:

1. Methods for assessment of models,
2. Model experiment description papers, or
3. Model evaluation papers.

If it is a manuscript of type #1, I believe the main contribution here would be “novel ways of comparing model results with observational data.” The authors have written extensively about comparing hindcasts with ARM-SGP data (this was one of the earliest uses of CAPT, as far as I can tell), so the 1st example analyzed doesn’t seem sufficiently novel. The authors' composite analysis of MJO in hindcasts does seem unique in the literature, so this could considered a novel way of comparing models and observations. However, I would expect this analysis to be emphasized much more prominently in the paper if it was considered the main, novel contribution of the paper; instead, it is one of three examples that the authors run through, and the authors only devote three paragraphs to the topic. Finally, the analysis of diabatic processes associated with ENSO could be unique, but it wasn't clear to me what the authors were trying to show with this analysis, aside from showing that the correspondence between errors in short- and long-duration simulations is robust with respect to ENSO phase.

The way the manuscript is written, it seems that the authors are targeting manuscript type #2 "Model experiment description papers", in which case I would expect that the main contribution of the paper is "descriptions of standard experiments for a particular type of model," and "discussion of why particular choices were made in the experiment design and sample model output". The manuscript instead seems to describe a rather vague experimental protocol, consisting of short-duration hindcasts paired with AMIP simulations. The authors also describe their specific implementation of these hindcasts. It also isn't clear how the experimental design presented here differs substantially from other, seemingly similar experimental protocols described in the literature: e.g., Transpose-AMIP II (Williams et al., 2013), and the hindcast approach described by Ma et al. (2015). If it does differ, the authors should explicitly compare and contrast their proposed experimental protocol with those already described in the literature. In particular, Ma et al. (2015)–which this manuscript indicates the hindcast procedure is based on–states that "We also hope to provide guidance for those performing transpose-AMIP/CAPT simulations with their own climate models for model development or error diagnosis purposes." That sounds strikingly similar to a "model experiment description paper," which again raises the unanswered question of how this manuscript represents a unique contribution to the literature. The authors should also follow the GMD guidelines for this particular manuscript type, including giving the experimental protocol a name and version number in the manuscript title.

In its current form, I don't think this manuscript is appropriate to be considered a paper of type #3. The authors' analysis focuses on an outdated version of CESM that is no longer supported and is decreasingly used. Further, the current manuscript doesn't refer to any of the (extensive) existing literature evaluating CESM1, so there is currently no way to determine how/why this paper is unique relative to other model evaluation papers. (All of that said, I strongly suspect the authors weren't targeting this manuscript type, in which case such a literature review of CESM1 would not be needed.)

Williams, K.D., A. Bodas-Salcedo, M. Déqué, S. Fermepin, B. Medeiros, M. Watanabe, C. Jakob, S.A. Klein, C.A. Senior, and D.L. Williamson, 2013: The Transpose-AMIP II Experiment and Its Application to the Understanding of Southern Ocean Cloud Biases in Climate Models. J. Climate, 26, 3258–3274, https://doi.org/10.1175/JCLI-D-12-00429.1

Ma, H.-Y., C. C. Chuang, S. A. Klein, M.-H. Lo, Y. Zhang, S. Xie, X. Zheng, P.-L. Ma, Y. Zhang, and T. J. Phillips (2015), An improved hindcast approach for evaluation and diagnosis of physical processes in global climate models, J. Adv. Model. Earth Syst., 7, 1810–1827, doi:10.1002/ 2015MS000490.

3.2 Repeatability of experiment

Assuming that the authors are targeting this manuscript as a "Model experiment description paper", the manuscript does not appear to provide adequate detail for an external group to be able to replicate this experimental protocol a constrained way. More detail would need to be provided for the initial condition generation procedure, the nudging procedure, for the land-surface spinup procedure, and greenhouse gas and solar forcing (do these follow the AMIP protocol?). It also seems like it would be useful to guide other modeling groups on what variables should be saved as output: otherwise, how would experiments be intercompared?

Specifically, a number of questions come to mind that could impact the results from other modeling groups implementing this experiment:

- What method(s) should be used to horizontally and vertically remap the ERA-Interim data onto the model grid?
- Do any adjustments need to be made to the initial conditions to avoid spurious gravity waves associated with differences in topography between ERA-Interim and the given model?
- When running 'nudging' simulations to generate non-state-variable ICs:
  - What nudging method is used (this is provided in Ma et al. (2015), but it should be included here for completeness)?
  - How are the ERA-Interim data interpolated from the ERA-Interim times to the model's current time (nearest-neighbor, linear, spline, other?)
  - Is the nudging simulation run in an identical configuration to the AMIP run, with the exception of the nudging term?
  - Why aren't the land-surface conditions from the nudging simulation used for the land-surface initial condition in the hindcasts?
- When running the offline land-surface spinup simulation:
  - How are surface enthalpy and moisture fluxes calculated? (readers shouldn't have to read through old CESM documentation to figure out what the CESM-community-specific lingo for 'offline' land surface simulation refers to)
  - How should chemical fluxes be handled if needed by the land-surface model (e.g., C/N in the case of CLM5 in prognostic Carbon/Nitrogen mode)?

- Why is the N. Viovy dataset used for forcing the offline land-surface spinup simulation?
- How are data from these observational datasets interpolated to the model grid (nearest neighbor, linear, conservative remapping, other)?

The authors should provide the code that they use to generate the initial conditions, since there are undoubtedly numerous other questions about implementation of this experiment that would arise when external groups attempt to implement this protocol.

3.3 Old model version

The authors state that "model developers can achieve additional useful understanding of the underlying problems in model physics by conducting a multi-year hindcast experiment." However, this statement is undermined by the author's use of CESM 1.0.5, which stopped being supported and developed by NCAR years ago. The authors state "Although newer version [sic] of the CAM and CLM is now available (CAM6/CLM5), similar systematic errors associated with moist processes remain present in the latest model version," but they state this without any reference to manuscripts that support this statement. Further, CESM2 contains numerous upgrades to key parameterizations associated with moist processes: in particular the adoption of CLUBB, MG2 prognostic microphysics, and a retuning of the convection parameterization to "increase the sensitivity to convection inhibition". Because of all these changes, I don't see how the results from this dataset could be used to inform model development in CESM2, which is the only version of CESM under active development.

If this is indeed intended to be a "Model experiment description paper," then this point is somewhat less relevant. However, the authors should be more forthcoming about these caveats and the utility of the dataset produced as part of this paper. This paper would also be much more impactful if the authors made some specific comments about what (if anything) would need to be done to implement this experimental design for CESM2.

4 Specific, minor issues

- pg 1, line 21: “associated parameterized” –> “associated with parameterized”
- pg 3, line 79: “Section 4 present” –> “Section 4 presents”
- pg 4, line 104: “output at model timestep” –> “output for every model timestep” (?)
- pg 5, lines 156-158: this is one of several theories for the MJO (e.g., see Yang and Ingersoll, 2011), so these feedback processes may not all be necessary
- pg 6, line 165: what is Q1?
- pg 7, line 220: “the response SST anomalies is much superior” –> “the response to SST anomalies is much superior”
- pg 7, line 221: “the result of poor circulation” –> “the result of errors in circulation” (‘poor circulation’ is usually reserved for describing anatomical difficulties with blood flow)
- pg 8, line 253: “the annualy cloud error metrics” –> “the annually-averaged cloud error metrics” (?)
- pg 9, lines 276-277: “These comparisons identify. . . ” this statement only makes sense to include if the authors repeat the experiments with an actively-developed model version.

Yang, D. and A.P. Ingersoll, 2011: Testing the Hypothesis that the MJO is a Mixed Rossby–Gravity Wave Packet. J. Atmos. Sci., 68, 226–239, https://doi.org/10.1175/2010JAS3563.1

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2020-39, 2020.