Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2020-39-RC3, 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 #3

Received and published: 9 June 2020

Review of GMD-2020-39

A multi-year short-range hindcast experiment for evaluating climate model moist processes from diurnal to interannual time scales

General comments:

The authors describe a multi-year hindcast experiment, in which short (3-day) initialised forecasts are performed each day with a climate model. The aim of the experiment is to understand the systematic errors associated with moist processes, which typically appear within the first few days of a model simulation. By performing short forecasts daily for 16 years, the authors are able to generate robust statistics of these systematic

Printer-friendly version

Discussion paper

errors. This allows them to understand whether the errors depend on the phase of natural variability, or to aggregate statistics across all forecasts to minimise the effects of that natural variability. The authors provide three examples of the types of analysis that could be performed with this hindcast set: the diurnal cycle of clouds over the central United States, the propagation of the Madden-Julian Oscillation through the Maritime Continent and the response of tropical precipitation and circulation to the El Nino-Southern Oscillation.

While I believe that the multi-year hindcast experiment is a useful framework to understand model systematic errors, it is not clear that this manuscript advances either the experiment protocol itself or the methods used to analyse those experiments. The authors have performed and analysed similar experiments in the past, as have other groups. The analysis and conclusions presented here are brief, mainly descriptive, and occasionally erroneous. The authors do not consistently show how the errors diagnosed in the short-range hindcasts differ from those diagnosed from AMIP experiments, which is important to demonstrate the value of their framework and to understand the AMIP errors. The interpretation of MJO propagation in such short simulations is also problematic. Finally, the authors motivate their experiment framework by invoking process-level improvements in models, but it is not clear how the analysis shown here would directly inform specific, targeted model development efforts to reduce these biases. I expand on these points below.

Major comments:

1. This manuscript is a GMD "model experiment description paper", which are "descriptions of standard experiments for a particular type of model". This manuscript is missing many of the GMD criteria for such a paper, including a name for the experiment, a version number for the protocol, and a version number for the boundary and initial conditions. The description of the experiment design in the paper also falls short of expectations; it lacks much of the detail that one would find in the description for a MIP, for instance, which seems to be the GMD gold standard. In particular, the following

details are not clear:

a. Do all initial conditions come from the nudged simulations, or do some fields come directly from ERA-Interim? The current description at L86 is not clear on this. b. What fields are nudged in the nudging simulations? c. What reanalysis and observation datasets were used force the offline land simulation? d. Is land initialised to a climatology from the offline simulations, or to the value on a particular day? If the latter, is the value taken from the last spinup cycle of the land model? e. Do greenhouse gases and aerosols follow the AMIP specification?

Also, I note that the offline land simulation data are not published, according to the citation in the manuscript, but GMD specifies that all boundary conditions must be published under version control, as part of the paper.

2. The novel contribution of the manuscript to the design and analysis of short-range hindcast experiments is not clear. The authors have conducted similar experiments in the past through the CAPT project (e.g., Ma et al., 2015); other similar experiment designs exist, such as Transpose-AMIP. This is far from the first paper to propose such an approach, which I think the authors would acknowledge. What is new and innovative about this experiment design? How does this experiment design enable new understanding of the development of systematic errors related to moist processes, beyond that which could be achieved through existing protocols?

3. At various points in the paper (e.g., L129, L155, L277), the authors motivate their analysis of short-range hindcast simulations by invoking parameterisation development or parameterisation evaluation. Yet, it is difficult to see how a parameterisation developer could gain useful insight from the simulations the authors present. If I were developing a convection parameterisation, I do not know how I would be able to apply the authors' analysis to improve that parameterisation. The authors do not test the sensitivity to the choice of model parameterisation, let alone to choices of particular model parameters. To make such a test would require repeating the full hindcast set,

potentially several times to test the sensitivity to various choices. I would expect that this would be costly, both in computational and human resources. Is there actionable information for parameterisation improvement here? Can the authors point to specific parameterisation developments that were made in response to their previous work with short-range hindcasts?

4. The authors state that their aim is to understand which of the errors seen in AMIP simulations are due to local-scale errors in moist processes, and which are due to errors in larger-scale or remote processes (e.g., L274-275). However, for the cloud regimes analysis (section 3.1) and the MJO propagation analysis (section 3.2), the authors analyse only the short-range hindcasts, without reference to the AMIP simulation. Thus, it is not possible to understand the relative contributions of moist-process errors. I suggest adding similar analysis for the AMIP simulation, as the authors have done for the ENSO analysis (section 3.3). Otherwise, the value of the short-range hindcasts over the AMIP simulations is not particularly clear.

5. The MJO analysis seems incomplete and problematic. In particular:

a. The authors show diabatic heating (Q1) profiles from the model, but do not compare them against observations (e.g., satellite-derived heating profiles). Yet, they make statements that the model heating is "very weak" (L186) and "not restricted to low levels". This suggests a bias, but the reader cannot judge the bias as there is no truth against which to compare!

b. The authors discuss MJO "propagation" through the Maritime Continent (L195), but in such short (3-day) hindcasts there is no "propagation" as such. The model is constantly reinitialised from the nudged simulations, so no MJO event could ever propagation across the Maritime Continent in a single hindcast. Likewise, the authors discuss a lack of "pre-conditioning" and "gradual transition from shallow to deep convection" in the model (L189). These are concepts often used in free-running climate simulations; the degree to which they apply to such short hindcasts is not clear. There can be no "pre-conditioning" or "gradual" transition in the span of a single hindcast.

c. Related to the above, it may be useful to diagnose how well the MJO propagates in the nudged simulation used as initial conditions. I believe comparisons of the nudged simulation against the hindcasts may provide more insight into how much of the lack of MJO propagation is due to errors in moist processes in the hindcasts.

d. The authors mention errors in "interactions with the diurnal cycle of convection over the Maritime Continent" (L200), related to the MJO propagation errors, but provide no evidence or analysis of these.

6. For the ENSO analysis (section 3.3), it is not clear why the authors analyse the ENSO regressions globally, given that the authors are interested in the fast (1-3 day) response to ENSO SST anomalies. The far-field response (e.g., over the Indian Ocean or the Atlantic) takes a few weeks to develop, at least. Thus, I suspect that the errors in these regions are not errors in the response to Pacific SST anomalies in the ENSO region, but rather are errors in response to the local circulation and SST, which may or be not be directly related to ENSO. Perhaps I am missing something, but the interpretation of the analysis here seems to be less straightforward than the authors suggest. Also, related to the ENSO section:

a. L210: The authors say that there are statistics in Table 2, but I cannot see any! Where are the pattern statistics that the authors mention?

b. At several places, (e.g., L216, L220), the authors state the the hindcasts are "superior", or have "better agreement" with observations, when compared to the AMIP simulations. This agreement is not obvious, particularly in such small figures. It needs to be quantified statistically.

7. In Figure 8, why are the OLR pattern correlations much lower than for the other variables?

8. L247: The authors suggest that robust model errors can be identified "from only one

year of hindcasts with enough ensemble members." How many is enough? Can the authors estimate the number of ensemble members required from their simulations?

9. L249: "short simulations will be effective at reducing moist process errors" - Simulations alone do not reduce errors! Only dedicated model development efforts can reduce these errors. How does this framework, and the results the authors show, contribute to this effort? See also comment 1 above.

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

Printer-friendly version

Discussion paper