**Anonymous Referee #3**

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 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.

***Response to reviewer:***
***We thank the reviewer for all the comments. Those certainly helped to improve our manuscript.***

***Regarding the new uniqueness of the current experimental design, this is our first time documenting this suite of multi-years (1997-2012, 16 years) hindcasts and proposing on how one can better utilizing these hindcasts on both mean and variability studies. Our intention is not to promote the hindcast technique itself since climate hindcast experiment approach has become a widely used method in the Transpose AMIP project and other climate model hindcast studies as we mentioned in the introduction. This is why we submitted this manuscript as a "model experiment description paper". In our earlier hindcast studies (e.g., Xie et al. 2012; Ma et al. 2013 J. Climate), we analyzed the mean biases from two years of short-range hindcasts (May 2008 to April 2010) during Years of Tropical Convection (YOTC). In Ma et al. (2015 JAMES) paper, we proposed a refined hindcast approach in improving the initial atmospheric aerosol profiles and initial land conditions. Based on the refined procedure, we proposed a ''Core'' integration (one-year long) with the refined procedure for a simple, easily repeatable test that allows model developers to rapidly compute appropriate metrics for assessing the impacts of various parameterization changes on the fidelity of cloud-associated processes with***

***available observations. In this manuscript, we applied the refined initialization strategy in Ma et al. (2015 JAMES) and performed this "new" suite of multi-year short-range hindcasts. One can now use this suite of hindcasts to conduct more studies (like the examples listed in this manuscript) which cannot be achieved from single or two years of short-range hindcasts. This is a different purpose from what we proposed in Ma et al. (2015 JAMES). We have strengthened this point in the abstract:***

***"These analyses can only be done through this multi-year hindcast approach to establish robust statistics of the processes under well-controlled large-scale environment because these phenomena are either interannual climate variability or only happen a few times in a given year (e.g. MJO, or cloud regime types)"***

***and in the introduction over Lines 68-73:***

***"This experiment provides an new opportunity to address several modeling issues associated with moist processes, which cannot be achieved from previous short Transpose AMIP II hindcasts (Williams et al. 2013), or one or two years of short-range hindcasts that we conducted in the past (Xie et al. 2012, Ma et al, 2013, Ma et al. 2015). This is because these phenomena are either interannual climate variability or only happen a few times in a given year, and thus we need multi-years to robustly quantify the errors associated with these phenomena."***

***Please see below our point-by-point responses to your comments.***

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:

***Response to reviewer:***
***We submitted this manuscript as a "model experiment description paper". We did not submit this as a "MIP" type protocol although we plan to use to this experiment design as one of the experiment categories in the Diurnal Cycle of Precipitation (DCP, https://portal.nersc.gov/cfs/capt/diurnal/) model intercomparison project under the Global Energy and Water cycle Exchanges (GEWEX) Global Atmospheric System Studies (GASS).***

***From the GMD Manuscript types (https://www.geoscientific-model-development.net/about/manuscript_types.html#item4):***

***"For model experiment description papers, similar version control criteria apply as to model description papers: the experiment protocol should be given a version number; a data availability paragraph must be included in the manuscript; boundary conditions should be given***

***a version number and uploaded or made otherwise available; a data availability paragraph must be included in the manuscript; and links to the GMD paper should be included on the experiment website. Since the primary purpose of these papers is to make experiments accessible to the community, all input data required to perform the experiments must be made publicly available"***

***We added CESM1 to the revised manuscript title "A multi-year short-range hindcast experiment with CESM1 for evaluating climate model moist processes from diurnal to interannual time scales". We provided the model version number (CESM1_0_5) and experiment configuration (FC5) in Section 2.1. We added more information regarding how initial conditions were generated in the revised manuscript. The information of boundary conditions and other code information are already described in Code and data availability section after the Summary.***
***We also searched similar model experiment description papers published on GMD webpage. Here is a list of a few manuscript links. We also followed how they provided their model and experiment information.***

***https://gmd.copernicus.org/articles/11/3865/2018/***
***https://gmd.copernicus.org/articles/10/2833/2017/gmd-10-2833-2017.pdf***
***https://gmd.copernicus.org/articles/10/1665/2017/gmd-10-1665-2017.pdf***

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.
***Response to reviewer:***
***In Section 2.1, we did state in our initial submission from L86 that we applied the "horizontal velocities", "temperature", "specific humidity" and "surface pressure" from the ERA-Interim Reanalysis (Dee et al. 2011) for the initial atmospheric states. A nudging simulation with CAM5/CLM4 was also performed to acquire "other necessary variables" (e.g., cloud and aerosol fields), which are not available from the ERA-Interim Reanalysis for the atmospheric initial conditions. It will be helpful if the reviewer can explicitly point out which part is not clear.***

b. What fields are nudged in the nudging simulations?
***Response to reviewer:***
***We now included more information about this in the revised manuscript over Lines 98-100: "A nudging simulation (horizontal velocities nudging only following Zhang et al. 2014) with CAM5/CLM4 was also performed to acquire other necessary variables (e.g., cloud and aerosol fields)".***

c. What reanalysis and observation datasets were used force the offline land simulation?
***Response to reviewer:***
***The reanalysis and observation datasets used to force the offline land simulation is the CRUNCEP dataset. It is one of the options in driving the offline CLM simulations. It is provided by NCAR. The reason we chose it is because the forcing data covers the most recent years. This dataset is widely used in the CLM community. Unfortunately, we cannot find a reference for this dataset. In the revised manuscript, we revised the sentence to "Land initial conditions are taken from an offline land model simulation (I2000 compset) forced by***

***reanalysis and observations including precipitation, surface winds, and surface radiative fluxes (CRUNCEP, N. Viovy 2013, unpublished data) rather than coupled it to an active atmospheric model" over Lines 102-104 in Section 2.1.***

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?
***Response to reviewer:***
***The land was initialized to the values on a particular calendar day from the offline simulations. We further clarified this in the revised manuscript "The offline land model simulation started from 1990 to 2012 and we performed five cycles (1990 to 2012) for the offline simulation to allow proper spin-up of the land conditions. After that, we continued the offline land model simulation to the desire starting date and use the land model restart file (.r file) as the land initial condition" over Lines 104-107.***

e. Do greenhouse gases and aerosols follow the AMIP specification?
***Response to reviewer:***
***The model configuration of greenhouse gases and aerosols is the same for the nudging, hindcast and AMIP experiments. In the revised manuscript, we revised the sentence to "We also conducted a 16-year long AMIP simulation with the same model for the same period. In this AMIP simulation, the state of the atmosphere evolves freely without constraints. Note that the nudging simulation mentioned above has the same model configuration as the AMIP simulation with the exception of the nudging terms" over Lines 115-117.***

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.
***Response to reviewer:***
***As we stated earlier, the reanalysis and observation datasets used to force the offline land simulation is the CRUNCEP dataset. It is one of the options in driving the offline CLM simulations. It is provided by NCAR and the reason we chose it is because the forcing data covers to the most recent years. This dataset is widely used in the CLM community. Unfortunately, we cannot find a refence for this dataset.***

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?
***Response to reviewer:***
***Regarding the new uniqueness of the current experiment, this is our first time documenting this suite of multi-years (1997-2012, 16 years) hindcasts and proposing on how one can better utilizing these hindcasts on both mean and variability studies. Our intention is not to promote the hindcast technique itself since climate hindcast experiment approach has become a widely used method in the Transpose AMIP project and other climate model hindcast studies as we***

***mentioned in the introduction. This is why we submitted this manuscript as a "model experiment description paper". In our earlier hindcast studies (e.g., Xie et al. 2012; Ma et al. 2013 J. Climate), we analyzed the mean biases from two years of short-range hindcasts (May 2008 to April 2010) during Years of Tropical Convection (YOTC). In Ma et al. (2015 JAMES) paper, we proposed a refined hindcast approach in improving the initial atmospheric aerosol profiles and initial land conditions. Based on the refined procedure, we proposed a ''Core'' integration (one-year long) with the refined procedure for a simple, easily repeatable test that allows model developers to rapidly compute appropriate metrics for assessing the impacts of various parameterization changes on the fidelity of cloud-associated processes with available observations.***

***In this manuscript, we applied the refined initialization strategy in Ma et al. (2015 JAMES) and performed this "new" suite of multi-year short-range hindcasts. One can now use this suite of hindcasts to conduct more studies and analyses (like the examples listed in this manuscript) which cannot be achieved from single or two years of short-range hindcasts. For example, there won't be enough cases to make robust statistical composites for the cloud regimes over the DOE ARM SGP site with only one single year of short range hindcasts. This is the same for the MJO phase composites. Further, one cannot answer the question of whether systematic errors of moist processes show significant interannual variation in the mean state biases, or the question of whether or not errors in the response of precipitation and surface fluxes to SST anomalies associated with ENSO can be attributed to parameterization errors or whether errors in the circulation response to SST anomalies also contribute, with only single or two years of short-range hindcasts. All these analyses require a suite of multi-year hindcasts and this is the uniqueness of this suite of multi-year hindcasts.***

***We have strengthened this point in the abstract:***

***"These analyses can only be done through this multi-year hindcast approach to establish robust statistics of the processes under well-controlled large-scale environment because these phenomena are either interannual climate variability or only happen a few times in a given year (e.g. MJO, or cloud regime types)"***

***and in the introduction over Lines 68-73:***

***"This experiment provides an new opportunity to address several modeling issues associated with moist processes, which cannot be achieved from previous short Transpose AMIP II hindcasts (Williams et al. 2013), or one or two years of short-range hindcasts that we conducted in the past (Xie et al. 2012, Ma et al, 2013, Ma et al. 2015). This is because these phenomena are either interannual climate variability or only happen a few times in a given year, and thus we need multi-years to robustly quantify the errors associated with these phenomena."***

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?

***Response to reviewer:***

***The purpose of this is manuscript is a "model experiment description paper", not a model evaluation paper. We only provided examples of possible studies on how to utilize this suite of multi-year hindcasts. We did not provide vigorous evaluation of individual model issues in this manuscript since that is beyond the scope of this manuscript as a model experiment description paper. Nevertheless, some information from the current example does provide insights on parameterization issues. For example, the cloud regime analysis suggests that the model shallow convection scheme is not able to simulate the shallow convection regime well. For afternoon deep convective cloud regime, the model cannot simulate the transition from shallow to deep convective clouds. The deep convection clearly starts too early. These issues all point to specific model parameterizations as we mentioned in the manuscript. The contribution from large-scale state is much smaller because the large-scale state is still very close to the reanalysis at Day 2 hindcasts.***

***Regarding repeatability of the entire experiment, the first author alone performed another suite of 50-day long hindcasts initialized every day starting at 00Z from 1997 to 2012 with CAM5 within a two-week period. The process can be faster with a better strategy on storing the output because the disk quota was constantly reached. Since each hindcast is independent and can be completed very fast from less than half an hour to two hours depending on the speed of the computing system, one can easily bundle as many hindcasts as possible into one job submission. We now have a section in the revised manuscript to describe this process (Section 2.2 Strategy on performing the multi-year hindcasts on a high performance computing system).***

***For sensitivity tests to various parameter choices for a specific scheme or parameterization, one does not have to perform this suite of multi-year hindcast experiment. Instead, one can perform "core experiment" (i.e., series of short-range hindcast over one-year period), as we proposed in Ma et al. (2015 JAMES). These two types of experiments: "multi-year hindcast experiment" and "core experiment" have different purposes in studying model biases associated with cloud processes. Further, one can perform a set of hindcasts just for the set of key dates with the phenomena of interest (e.g. days with sallow cumulus at ARM SGP, or phase 3 of various MJOs).***

***We added this sentence: "For sensitivity tests to various parameter choices for a specific scheme or parameterization, it is not necessary to perform this suite of multi-year hindcast experiment once the issue has been identified. Instead, one could perform a "core experiment" (i.e., series of short-range hindcast over one-year period) as we proposed in Ma et al. (2015), or perform a set of hindcasts just for the set of key dates with the phenomena of interest (e.g.***

***days with sallow cumulus at ARM SGP, or phase 3 of various MJOs, which we will introduce in the following text)" in the revised manuscript over Lines 148-153 to further clarify this.***

***Regarding our previous work in guiding parameterization development, we listed many references in the introduction already. For specific example, in Zheng et al. (2017), we identified an issue of a cloudy planetary boundary layer oscillation related to interaction between CLUBB scheme and MG2 microphysics scheme in CAM5. The issue is later fixed in the CAM5/CAM6 and the U.S. DOE E3SM development. There are also many unpublished studies that our hindcast approach was used to guide the development of a scheme or diagnose parameterization errors. Specifically, we used short range hindcasts in testing the candidate convection schemes during the CAM6 and E3SM model development.***

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.

***Response to reviewer:***

***Comparing hindcasts to AMIP simulations to identify errors in parameterizations or large-scale sate is certainly beneficial. However, in some targeted studies, such as field campaign or cloud regime/MJO phase in the present manuscript, the comparison to AMIP simulations may not be possible or necessary. For cloud regimes and MJO propagation analyses, the motivation is to better isolate the bias contribution from parameterizations with a well-controlled large-scale state. As we stated in the manuscript and our response to comment 3, the analysis of cloud regimes certainly provides information on the issues of convection schemes in the model.***

***Regarding making the same AMIP composites for the comparison, it is not possible due to the definition of the cloud regimes. For the shallow convection regimes, these days are defined as precipitation rate = 0 mm day$^{-1}$ at all hours of the day, and shallow cumulus clouds are identified by Berg and Kassianov (2008), who first selected cumulus clouds based on fine temporal resolution ARSCL data at ARM SGP and then manually scrutinized cloud images taken by the Total Sky Imager (available online at http://www.arm.gov/ instruments/tsi) to eliminate low cloud types other than shallow cumulus. For MJO phase composites, one can certainly follow the definition in Wheeler and Hendon (2004). However, the days for MJO phase composites from the AMIP simulation will not correspond to the actual observation days. Also, the large-scale state in the AMIP simulation is not necessarily similar to that in the observations/reanalysis. One can obtain information about the bias correspondence between long-term climate bias and short-term hindcast bias. However, this is not what we intend to show for the MJO phase composite studies.***

***References:***

***Berg, L. K., and Kassianoy, E. I.: Temporal variability of fair- weather cumulus statistics at the ACRF SGP site, J. Climate, 21, 3344–3358, 2008.***
***Wheeler, M. C., and Hendon, H. H.: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction, Mon. Weather Rev., 132, 1917–1932, 2004.***

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

***Response to reviewer:***
***As we stated earlier, the purpose of this is manuscript is a "model experiment description paper", not a model evaluation paper. We only provided examples of possible studies to utilize this suite of multi-year hindcasts. We are not vigorously evaluating individual model issues mentioned in this study.***

***Regarding "problematic" issue, please see our response below in 5b.***

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!
***Response to reviewer:***
***Over Line 186. The sentence in the manuscript is:***

***"During Phases 2 and 3 when the MJO is over the Indian Ocean, the anomalous Q1 profiles reveal that the magnitude of shallow heating is very weak (<0.4 K day$^{-1}$) to the east over the region of suppressed convection between 100°E and 120°E in Phase 2 and the heating is not restricted to low levels between 120°E and 150°E in Phase 3."***

***Here, we were just describing the feature from the model simulations.***

***After this sentence, we stated:***

***"Instead, there is an anomalous heating associated with deep convection in Phase 3, which is not evident in the observations as indicated from previous studies (e.g., Jiang et al. 2011). This suggests that the model fails to simulate the pre-conditioning moistening processes and the gradual transition from shallow to deep convection as MJO propagates."***

***We didn't show observed heating profiles here is because the actual bias magnitudes of heating profiles are not the focus. Rather, we wanted to highlight the absence of low-level shallow convection heating in the model by Day 3. The existence of shallow convection ahead of the core of deep convection is a well-known MJO feature (e.g., Johnson et al. 1999; Kikuchi and Takayabu 2004; Chen and Del Genio 2009; Tromeur and Rossow 2010; Powell and Houze 2013; Xu and Rutledge 2014). Instead, the model shows heating profiles associated with deep convection. The absence of low-level shallow convection heating is very obvious, and we think we don't really need observation to prove what has been published in many previous literatures.***

***In the revise manuscript, we further highlight this feature that we want to focus. The revised sentence is: "Instead, there is an anomalous heating associated with deep convection in Phase 3, which is not evident in the observations as indicated from many previous studies (e.g., Figure 5a of Jiang et al. 2011). This suggests that the model fails to simulate the pre-conditioning moistening processes by shallow convection and the gradual transition from shallow to deep convection as MJO propagates by Day 3".***

***Reference:***
***Chen, Y. H. and A. D. Del Genio, 2009: Evaluation of tropical cloud regimes in observations and a general circulation model. Climate Dyn., 32, doi:10.1007/S00382-008-0386-6, 355-369.***
***Jiang, X., Waliser, D. E., Olson, W. S., Tao, W.-K., L'Ecuyer, T. S., Shige, S., Li, K.-F., Yung, Y. L., Lang, S., and Takayabu, Y. N.: Vertical diabatic heating structure of the MJO: Intercomparison between recent reanalyses and TRMM estimates, Mon. Wea. Rev., 139, 3208-3223, 2011.***
***Johnson, R. H., T. M. Rickenbach, S. A. Rutledge, P. E. Ciesielski, and W. H. Schubert, 1999: Trimodal Characteristics of Tropical Convection. J. Clim., 12, 2397-2418.***
***Kikuchi, K. and Y. N. Takayabu, 2004: The development of organized convection associated with the MJO during TOGA COARE IOP: Trimodal characteristics. Geophys. Res. Lett., 31.***
***Powell, S. W. and R. A. Houze, 2013: The cloud population and onset of the Madden-Julian Oscillation over the Indian Ocean during DYNAMO-AMIE. Journal of Geophysical Research: Atmospheres, 118, 10.1002/2013JD020421, 2013JD020421.***
***Tromeur, E. and W. B. Rossow, 2010: Interaction of Tropical Deep Convection with the Large-Scale Circulation in the MJO. J. Clim., 23, Doi 10.1175/2009jcli3240.1, 1837-1853.***
***Xu, W. and S. A. Rutledge, 2014: Convective Characteristics of the Madden–Julian Oscillation over the Central Indian Ocean Observed by Shipborne Radar during DYNAMO. J. Atmos. Sci., 71, 10.1175/JAS-D- 13-0372.1, 2859-2877.***

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.

***Response to reviewer:***

***In our analysis, we were not examining just one single hindcast. As we stated in Section 2.1 of the manuscript, "We concatenated each hindcast from 24-48 (48-72) hours lead time to form a pseudo Day 2 (Day 3) time series of 16-year duration from 1997 to 2012". Therefore, we examined the Day 2 hindcast time series over this 16-year period like an AMIP simulation. Granted, the Day 2 pseudo time series is not a continuous time series like the AMIP simulation. Nevertheless, we are examining the features of MJO phase composites rather than discussing propagation of a single MJO event in a single hindcast. Similarly, the "pre-***

***conditioning" and "gradual transition from shallow to deep convection" are discussed in those phase composites rather than in a specific MJO event 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.

***Response to reviewer:***

***We agree that the nudged simulation will be useful and may provide useful information when compared with the hindcasts. We added this information "The proposed experiment and evaluation method also complement the existing ways of climate model evaluation, such as performing GCM simulations in the AMIP, or nudging mode. Comparison among the multi-year hindcasts, AMIP and nudging simulations may provide more insights into these issues mentioned above" as a recommendation in the revised manuscript in the summary section over Lines 339-341.***

***Note that this manuscript is a "model experiment description paper", not a model evaluation paper. Therefore, further investigation with the nudging simulation for MJO propagation is beyond the scope of this manuscript.***

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.

***Response to reviewer:***

***As we mentioned above, this manuscript is a "model experiment description paper", not a model evaluation paper. Therefore, diagnosing the errors in "interactions with the diurnal cycle of convection over the Maritime Continent" (L200), related to the MJO propagation errors, will require further investigation as a separate study.***

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.

***Response to reviewer:***

***The motivation is to "gain insights into whether or not errors in the response of these fields (precipitation, surface radiative and heat fluxes, as well as zonal wind stress) to SST anomalies can be attributed to parameterization errors or whether errors in the circulation response to SST anomalies also contribute", as we stated in the beginning of Section 3.3.1. In AMIP simulations, it is difficult to disentangle errors in the response of these fields to SST anomalies from large-scale state or model parameterizations due to feedback processes. With the multi-year hindcasts, the large-scale state at Day 2 remains close to the initial state, therefore, errors in the response of these fields to SST anomalies can mainly attribute to***

***parameterization errors. Remote cloud errors are the response to the local circulation and SST anomalies, and that some of those anomalies may not be driven by ENSO. However, when one compares the result of the multi-year hindcasts to the AMIP run, it is only the circulation anomalies (correlated to ENSO) which may not be reproduced in the AMIP run, so that comparison of the cloud results between the hindcast and AMIP run provides information about the role of the errors in the circulation anomalies.***

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?
***Response to reviewer:***
***We apologize for this mistake. The table (now Table 3) in now in the revised manuscript.***

b. At several places, (e.g., L216, L220), the authors state 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.
***Response to reviewer:***
***With the Table 3, the overall spatial correlation coefficients and RMSEs are indeed superior in the hindcasts than in the AMIP quantitatively.***

7. In Figure 8, why are the OLR pattern correlations much lower than for the other variables?
***Response to reviewer:***
***We think that the possible reason for lower OLR pattern correlations is because OLR performance is mainly associated with the performance of convection, radiation and cloud micro- and macro-physics schemes in the model. Therefore, errors in all these schemes can all contribute to the errors in the OLR simulations. However, this requires further investigation and this is beyond the scope of the current manuscript.***

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?
***Response to reviewer:***
***Based on our previous study in Ma et al. (2014, JCLI), we found that ensemble members larger than 15 may be enough for identifying robust model errors associated with cloud processes. We added this information in the revision over Lines 298-300: "one may identify robust model errors in the mean state from only one year of hindcasts with enough ensemble members (with ensemble members greater than 15, Ma et al. 2014)".***

***Reference:***
***Ma, H. Y., Xie, S., Klein, S. A., Williams, K. D., Boyle, J. S., Bony, S., Douville, H., Fermepin, S., Medeiros, B., Tyteca, S., and Watanabe, M.: On the correspondence between mean forecast errors and climate errors in CMIP5 models, J. Climate, 27, 1781–1798, 2014.***

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.

***Response to reviewer:***

***We believe this was a mistake of using the wrong word in the initial submission. We have changed the word from "reducing" to "identifying".***