

Anonymous Referee #1

This study presents an innovative approach by conducting a series of short-range hindcasts based on a climate model that can be used to expose deficiencies in model parameterizations that are responsible for biases in climate simulations from the same model. Justification of this approach includes that model biases in climate mean and variability in long-term simulations start to emerge in the first several days of hindcasts by this model, which is thought to result from model parameterized processes since the large-scale state is still very close to the observations specified in the hindcast initial conditions. This study also proposed to conduct these short-range hindcasts for multiple years to achieve stable statistics when examining model deficiencies and associated processes, although it is shown that a systematic association between model biases in short-range hindcasts and long-term simulations do not significantly change in different individual years. Three examples are further illustrated for applications of this approach to understand model biases in simulating the diurnal cycle of warm season precipitation over central U.S., the tropical MJO, and the local and remote influences by ENSO on the interannual time-scale. This approach along with the long-term hindcast datasets based on CAM5 produced from this project, can be used in many other studies for both tropical and extratropical climate phenomena, thus are valuable for the climate research community. In addition to identify model parameterization deficiencies as mentioned in this study, this approach can also provide a very useful avenue to diagnose and understand critical processes regulating various climate and weather phenomena by taking advantage of detailed model output with largely realistic representation of these phenomena in hindcasts at day-2. This can also be stressed in the manuscript. The paper is generally very clearly written although there is room for further improvement. I recommend this paper for publication after some minor revisions as listed below.

Response to reviewer:

We thank the reviewer for all the comments. Those certainly helped to improve our manuscript. We have stressed the point raised by the review regarding “This experiment can also provide a very useful avenue to diagnose and understand critical processes regulating various climate and weather phenomena by taking advantage of detailed model output with a largely realistic representation of the large-scale state in hindcasts” in the revised manuscript over Lines 339-341 in the Summary Section.

Please see our point-by-point responses to the minor moments below in bold.

Minor comments:

Line 12: suggest to change “3-day long hindcasts every day” to “3-day hindcasts initialized every day . . .”. Also many grammatical errors need to be corrected throughout the manuscript, which I may not list all of them in the following.

Response to reviewer:

The sentence is revised as suggested and we have carefully checked the grammatical errors throughout the entire manuscript.

Line 14: suggest to delete “propagation”, since MJO amplitude is also involved.

Response to reviewer:

Propagation deleted as suggested.

Line 14: also not sure about “the responses of moist processes to sea surface temperature anomalies. . .” here. Why “moist processes” since net heat fluxes and wind stresses are also shown in Fig. 7?

Response to reviewer:

We changed the sentence to “the response of precipitation, surface radiative and heat fluxes, as well as zonal wind stress to sea surface temperature anomalies associated with the El Niño-Southern Oscillation - ...” to better matched what is shown in Figure 7.

Line 21: again, in addition to “parameterized moist processes”, could other parameterizations also related, for example, boundary layer and radiation parameterizations?

Response to reviewer:

Boundary layer and radiation parameterizations are certainly related based on our previous studies and experience. We did not mention these two in the abstract because we didn’t show any results directly linked to these two schemes. In the revised manuscript, we added the information “Although we only showed examples relevant to moist processes, other processes related to planetary boundary layer or radiation schemes can also be examined through this suite of experiments.” in the summary section over Lines 337-339.

Line 52: “AMIP” first appears here, but is defined in Line 99. Also, suggest to remove “of years”.

Response to reviewer:

AMIP is now defined when it first appears in the introduction, and “of years” is removed as suggested.

Line 57: CAM5 is mentioned here, but CESMv1 is mentioned in abstract and other places. Better be consistent.

Response to reviewer:

We now changed it to CESM1 to be consistent across the manuscript.

Line 66: suggest to delete “from diurnal to interannual time scales” in this line.

Response to reviewer:

Deleted as suggested.

Line 75: variation(s)

Response to reviewer:

Revised.

Line 79: present(s)

Response to reviewer:

Revised.

Line 96: of the 16-year duration

Response to reviewer:

Revised.

Line 120-125: would be better if how these regimes are defined can be briefly explained.

Response to reviewer:

The cloud regimes are now defined and briefly explained following Zhang and Klein (2010) in the revised manuscript over Lines 160-169.

Line 135: cloud regime(s)

Response to reviewer:

Revised.

Line 147: parameterization deficiencies

Response to reviewer:

Revised.

Line 150: suggest change “scheme developer” to “model developer”

Response to reviewer:

Revised as suggested.

Line 153: the intraseasonal variability

Response to reviewer:

Revised.

Line 166: present(s)

Response to reviewer:

Revised.

Line 170-171: “the intensity of OLR decreases” is a little confusing since active convection corresponds to smaller OLR. Not sure if this means convection is enhanced or weakened.

Response to reviewer:

The sentence is revised to “the intensity of convection decreases (with OLR anomalies increasing) after the core of MJO crosses over the Maritime Continent” to avoid the confusion.

Figure captions 3,6: CAPT are used but not defined.

Response to reviewer:

We now removed CAPT from the figure captions.

The presentation of Figures 4,5 can be improved, particularly Fig. 5. It may be not needed to use both precipitation and OLR since both of them represent MJO convection. At their present forms, these figures look very busy and a little difficult to see some detailed features particularly in wind vectors.

Response to reviewer:

We have removed OLR contours in Figure 5 to make the figure less busy in the revised manuscript.

Lines 179-180: Seems to me, the wet biases are largely evident over regions around the Maritime Continent, which may also relate to model biases in representing the regional diurnal cycle.

Response to reviewer:

We agree that the wet biases may also relate to the biases in the diurnal cycle of the Maritime Continent. We now revised the sentence to “Further, there is a persistent dry bias over Borneo and part of Sumatra and wet bias around the Maritime Continent for all the phases indicating a possible local effect of diurnal cycle of convection.” in the revised manuscript over Lines 228-229.

Line 220: the circulation anomalies are discussed here, but temperature and moisture fields are also used to constrain model in the initial conditions. “the response SST anomalies is” needs to be fixed

Response to reviewer:

We revised the sentence to “The large-scale state is well constrained in the hindcasts and the response of those fields to SST anomalies in Figure 7 is much superior”.

Line 231: “larger” root mean square errors?

Response to reviewer:

“larger” is added in the revised manuscript.

Line 254: Possibly to briefly mention these “cloud measures”.

Response to reviewer:

We now added the following sentences “ E_{TCA} measures the error in total cloud amount, and E_{CTP-l} measures the errors in the frequency of optically intermediate and thick clouds at high, middle, and low-levels of the atmosphere. E_{SW} and E_{LW} measures the errors in the impacts on top-of-atmosphere shortwave and longwave radiation in the same cloud-top pressure and optical depth categories used for E_{CTP-l} , respectively” in the revised manuscript.

Line 284-285: “Indeed, GCMs usually perform more poorly for . . .”. Not quite follow the purpose of this statement.

Response to reviewer:

We have removed this sentence in the revised manuscript.

Line 297: suggest change to “to understand the processes that control the diurnal and sub-diurnal variations of . . .”

Response to reviewer:

This sentence is revised as suggested.

Line 301: CESM1 is mentioned here, and CAM5 in the model introduction part.

Response to reviewer:

We now use CESM1 across the manuscript.

Anonymous Referee #2

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:

- 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.

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 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 (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.”

For our initial submission, we did not include detailed information on the initialization procedure since this was already described in Ma et al. (2015 JAMES). We did not want to duplicate the information because the initialization technique is not the main focus of the manuscript. In this revised version of the manuscript, we provided more information on how to conduct the multi-year hindcasts including the initialization information. Recently, we made our scripts and hindcast procedure documentation available to the public on GitHub (<https://github.com/PCMDI/CAPT>). We also include this information in the revised manuscript over Lines 107-109.

Regarding the concern of using CESM1/CAM5/CLM4 as our base model, we understand this model is not currently supported by NCAR anymore. However, simulations from this version are still used by many people including students in universities because the performance of this model is still very good. For model development and diagnosis purpose, people still compare model performance between the latest version and older version to understand the impacts of parameterization changes to the older version. For example, CAM5 results were still being mentioned in the recent CESM workshop in June 2020. As we are preparing another suite of multi-year short-range hindcasts with DOE E3SM/EAM v1, we will compare our results from E3SM to CESM1. The E3SM EAM v1 was originally branching from CAM5.3, which has very similar model performance to CAM5. Therefore, the comparison between the two suites of hindcasts will provides us information on the impacts of parameterization and other changes on cloud process performance. Also note that E3SM/EAM v1 has a new set of atmospheric physical parameterizations that are very similar to CAM6, the latest CAM.

Finally, this suite of hindcasts was completed back in 2016 and we have been thinking of more ways on how to better utilize these hindcasts other than what we had planned for. Therefore, it took us a while to prepare this manuscript because we want to provide the community more ways on how to use this type of hindcast experiment.

For our responses to other comments, please see our point-by-point responses below in bold.

3 Major Concerns

3.1 Uniqueness 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.

Response to reviewer:

Please see our response earlier regarding the concern of uniqueness of manuscript and relevance for GMD.

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 be 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.

Response to reviewer:

This manuscript is not under the methods for assessment of models.

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.

Response to reviewer:

As we responded earlier, we submitted this manuscript as a “model experiment description paper”. We aren’t sure if this information was directly provided to the reviewers or not. It is shown on the GMD manuscript page (<https://www.geosci-model-dev-discuss.net/gmd-2020-39/#discussion>). As we have explained the uniqueness of this suite of multi-year hindcast experiment and why this is different from we have published in earlier studies including Ma et al. (2015 JAMES), please refer to the response above.

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”

With the description above, it is not clear whether the version number should be on the manuscript title. Nevertheless, we added CESM1 to the manuscript title. The title is now “A multi-year short-range hindcast experiment with CESM1 for evaluating climate model moist processes from diurnal to interannual time scales”. We also 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 described in Code and data availability section after the Summary.

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.)

Response to reviewer:

This manuscript is not a model evaluation type paper.

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?

Response to reviewer:

For our initial submission, we did not include detailed information on the initialization procedure since this was already described in Ma et al. (2015 JAMES). We did not want to duplicate the information because the initialization technique is not the main focus of the manuscript and modeling groups who have the capability to conduct hindcast studies already have their own initialization strategy. Nevertheless, in the revised manuscript, we provided more information on how to conduct the multi-year hindcasts including the initialization generation procedure, the nudging procedure, land-surface spin-up procedure. The greenhouse gas and solar forcing is based on the setup of the CESM1 FC5 compset, which corresponds to the year 2000 level for all the simulation period. This is because the CMIP5 forcing data does not go beyond 2005. We also don't want the interannual variations in the solar and greenhouse gas forcing to affect our results so that we can better identify possible causes of model biases associated with model parameterizations. Recently, we made our scripts and hindcast procedure documentation available to the public on GitHub (<https://github.com/PCMDI/CAPT>). We also include this information in the revised manuscript over Lines 107-109.

Regarding what variables should be saved in the experiment, this really depends on the subjects of the studies. For example, we participated in the MJO diabatic heating process model intercomparison project and we also organized the US summertime warm bias model intercomparison project. There is an overlap of output variables but there are also variables, which are unique to MJO studies, such as the diabatic heating profiles and all the state variable budget terms. While in the warm bias study, we requested more variables associated with clouds and surface processes. Nevertheless, we do have a list of output variable available to others over https://portal.neresc.gov/archive/home/h/hyma/www/CAPT/CAPT_Long/CAPT_Long_output_cesm1_0_5_v5.pdf. This information is under the Code and data availability section.

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?

Response to reviewer:

The goal is to preserve the state variable information from the reanalysis as much as possible after the interpolation. In the present study, we applied the bilinear interpolation with ESMF regriding utility for the horizontal remap for all the state variables. For vertical remap, we follow the procedure used at ECMWF when initializing model with foreign analysis: Quadratic interpolation is used for temperature remap, linear interpolation is used for specific humidity, and a combination of linear and quadratic interpolation is used for zonal and meridional winds. The vertical interpolation scripts were originally constructed following the ECMWF procedure by one of our colleagues who has retired back in 2013. We didn't have the reference listed here because we can't find the link to the reference anymore after an update of ECMWF webpage.

Note that based on the experience from the past model intercomparison projects that we have organized or participated in, there were generally no guidelines on what method(s) should be used to do horizontal or vertical remap. This is because each model group, especially for operational centers, already have a set of tools to do this.

- 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?

Response to reviewer:

We applied a spatial smoothing for the state variables. This is based on Gerrity and McPherson (1970).

Gerrity, J. P. and R. D. McPherson, 1970: Noise analysis of a limited-area fine-mesh prediction model. ESSA Technical Memoranda, WBTM NMC 46, PB-191-188. 81pp.

We also adjusted the surface pressure associated with differences in topography between ERA-Interim and CAM5 using hydrostatic approximation.

Note that the above information is available in the documentation over our GitHub repository (<https://github.com/PCMDI/CAPT>).

- 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)?

Response to reviewer:

A U, V only nudging was used in the nudging run. We follow the recommendation from Zhang et al. (2014 ACP) and also made a few test runs to reach the same conclusion in Ma et al. (2015 JAMES). This information is included in the revised manuscript over Lines 98-101.

Reference:

Zhang, K., Wan, H., Liu, X., Ghan, S. J., Kooperman., G. J., Ma, P.-L., Rasch, P. J., Neubauer, D., and Lohmann, U.: Technical note: On the use of nudging for aerosol–climate model intercomparison studies, Atmos. Chem. Phys., 14, 8631–8645, 2014.

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.

- – How are the ERA-Interim data interpolated from the ERA-Interim times to the model’s current time (nearest-neighbor, linear, spline, other?)

Response to reviewer:

In the nudging simulation, the reanalysis data are linear interpolated between two time steps. Note that the nudging and interpolation codes were created by a NCAR software engineer and we didn’t make any more changes.

- – Is the nudging simulation run in an identical configuration to the AMIP run, with the exception of the nudging term?

Response to reviewer:

That is correct. We also added this information in the revised manuscript over Lines 116-117.

- – Why aren’t the land-surface conditions from the nudging simulation used for the land-surface initial condition in the hindcasts?

Response to reviewer:

This is because in a nudging simulation, biased precipitation, winds, and surface fluxes are allowed to pass to the land model. This will cause larger biases in the simulated soil moisture and temperature. This was discussed in Ma et al. (2015 JAMES). Therefore, we did not mention this in the present manuscript since the focus is on the multi-year hindcast experiment, not the initialization procedure.

- 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)

Response to Reviewer:

These are two different questions regarding what is an offline simulation, and how are surface enthalpy and moisture fluxes calculated in CLM.

For an offline land model simulation, we stated in Section 2.1 that it is a land model simulation forced by reanalysis and observations including precipitation, surface winds, and surface radiative fluxes. We added another sentence after the original sentence to make this clear. This is now “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 (N. Viovy 2013, unpublished data) rather than coupled it to an active atmospheric model” over Lines 102-104.

Regarding the calculation of the surface enthalpy and moisture fluxes, this is really not the focus of this manuscript. For us (atmospheric scientists), we also need to read through the CESM documentation to know the exact equations to calculate these fluxes since we are not land model developers. In short, the surface fluxes are based on Monin-Obukhov Similarity Theory. Without copying the entire chapter from the CLM4 documentation to here, we refer the reviewer to see Section 5 Momentum, Sensible Heat, and Latent Heat Fluxes of the CLM4 technical note on how to calculate these fluxes.

http://www.cesm.ucar.edu/models/ccsm4.0/clm/CLM4_Tech_Note.pdf

- – 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)?

Response to Reviewer:

This is an open question since we do not turn on the C/N option in all of our hindcasts. One could perform the same offline land model simulation with C/N turned on although the spin-up time may require much longer time (more cycles).

- – Why is the N. Viovy dataset used for forcing the offline land-surface spinup simulation?

Response to Reviewer:

We selected this option (CRUNCEP) in CLM because they offer the most recent years of atmospheric forcing data at that time. Also, the land state from offline simulation is superior than that from a nudging simulation (Ma et al. 2015 JAMES).

- – How are data from these observational datasets interpolated to the model grid (nearest neighbor, linear, conservative remapping, other)?

Response to Reviewer:

We used the default option in the model setup when running the offline land model simulation. The default method is bilinear interpolation according to CESM1.0 documentation:

<http://www.cesm.ucar.edu/models/cesm1.0/clm/models/land/clm/doc/UsersGuide/x7895.html>

- 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.

Response to Reviewer:

We have now made our scripts and hindcast procedure documentation available to the public on GitHub (<https://github.com/PCMDI/CAPT>). We also included this information in the revised manuscript over Lines 107-109.

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.

Response to Reviewer:

This manuscript is indeed a “Model experiment description paper”. We have provided our response to this concern in the beginning of our responses.

4 Specific, minor issues

pg 1, line 21: “associated parameterized” → “associated with parameterized”

Response to reviewer:

Revised.

pg 3, line 79: “Section 4 present” → “Section 4 presents”

Response to reviewer:

Revised.

pg 4, line 104: “output at model timestep” → “output for every model timestep” (?)

Response to reviewer:

Revised

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

Response to reviewer:

We revised the sentence to “These feedback processes may contribute to better MJO simulations if they are well represented in the GCMs”, so the sentence does not seem to suggest that these feedback processes are all necessary for realistic MJO simulations.

pg 6, line 165: what is Q1?

Response to reviewer:

We added the following short paragraph for the information of Q1 in the revised manuscript:

“The diabatic heating rate or apparent heating of large-scale motion system (Q1) consists of the heating due to radiation, the release of latent heat by net condensation, and vertical convergence of the vertical eddy transport of sensible heat (Yanai et al. 1973). In CESM1/CAM5, Q1 can be calculated through summing up all the tendency terms with all the diabatic processes.”

Reference:

Yanai, M., Esbensen, S., and Chu, J.–H.: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets, J. Atmos. Sci., 30, 611–627, 1973.

pg 7, line 220: “the response SST anomalies is much superior” → “the response to SST anomalies is much superior”

Response to reviewer:

We revised the sentence to “the response of those fields to SST anomalies in Figure 7”.

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)

Response to reviewer:

Revised as suggested.

pg 8, line 253: “the annually cloud error metrics” → “the annually-averaged cloud error metrics” (?)

Response to reviewer:

Revised as suggested.

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.

Response to reviewer:

Yes, we do plan to perform multi-year hindcasts with an actively-developed model version (E3SM) and compare with the CESM1-CAM5 results. We added “We will also compare the results from E3SM to CESM1 to understand the impact of parameterization and model changes to the performance of moist processes since the atmospheric component of E3SM was originally branching from CAM5.3, which has very similar performance as CAM5 (Xie et al. 2018, Rasch et al. 2019). Note that E3SM version 1 has a new set of atmospheric physical parameterizations that are very similar to CAM6, the latest CAM.” in the revised manuscript over Lines 353-356.

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>

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.neresc.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 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 shallow 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 shallow 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 state 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 reinitialized from the nudged simulations, so no MJO event could ever propagate 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) is 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”.

A multi-year short-range hindcast experiment with CESM1 for evaluating climate model moist processes from diurnal to interannual time scales

Hsi-Yen Ma¹, Chen Zhou², Yunyan Zhang¹, Stephen A. Klein¹, Mark D. Zelinka¹, Xue Zheng¹,
5 Shaocheng Xie¹, Wei-Ting Chen³, and Chien-Ming Wu³

¹Lawrence Livermore National Laboratory, Livermore, California, USA

²School of Atmospheric Sciences, Nanjing University, Nanjing, China

³Department of Atmospheric Sciences, National Taiwan University, Taipei, Taiwan

Correspondence to: Hsi-Yen Ma (ma21@llnl.gov)

10 **Abstract.** We present a multi-year short-range hindcast experiment and its experimental design for better evaluation of both
the mean state and variability of atmospheric moist processes in climate models from diurnal to interannual time scales and
facilitate model development. We use the Community Earth System Model version 1 as the base model and performed a suite
of 3-day hindcasts initialized every day starting at 00Z from 1997 to 2012. Three processes – the diurnal cycle of clouds during
different cloud regimes over the Central U.S., precipitation and diabatic heating associated with the Madden-Julian Oscillation
15 (MJO), and the response of precipitation, surface radiative and heat fluxes, as well as zonal wind stress to sea surface
temperature anomalies associated with the El Niño-Southern Oscillation – are evaluated as examples to demonstrate how one
can better utilize simulations from this experiment to gain insights into model errors and their connection to physical
parameterizations or large-scale state. This is achieved by comparing the hindcasts with corresponding long-term observations
for periods based on different phenomena. These analyses can only be done through this multi-year hindcast approach to
20 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). Furthermore,
comparison of hindcasts to the typical simulations in climate mode with the same model allows one to infer what portion of a
model’s climate error directly comes from fast errors in the parameterizations of moist processes. As demonstrated here, model
biases in the mean state and variability associated with parameterized moist processes usually develop within a few days, and
25 manifest within weeks to affect the simulations of large-scale circulation and ultimately the climate mean state and variability.
Therefore, model developers can achieve additional useful understanding of the underlying problems in model physics by
conducting a multi-year hindcast experiment.

Deleted: procedure

Deleted: ng

Deleted: to

Deleted: d

Deleted: 3-day long hindcasts every day

Deleted: propagation

Deleted: moist processes

Deleted: design

30

40 1 Introduction

The representation of moist processes - clouds, convection, precipitation and the associated radiative perturbations - and their interactions with the large-scale circulation in Global Climate Models (GCMs) or Earth System Models (ESMs) remains one of the grand challenges for the modeling community (Bony et al. 2015). Aside from using high resolution cloud modeling (cloud-resolving models or large-eddy simulations) to study detailed cloud dynamics and physics and their interactions with large-scale environment, significant progress on process-level understanding of moist processes has also been achieved recently through the application of a climate model hindcast approach (Phillips et al. 2004; Williams et al. 2013) to gain insights relevant to the improvement of parameterizations in climate models. This progress has been made either through individual modeling studies (e.g., Xie et al. 2004, 2008; Klein et al. 2006; Barton et al. 2012, 2014; Medeiros et al. 2012; Hannah and Maloney 2014; Chandra et al. 2015; Van Weverberg et al. 2015; Zheng et al. 2016; 2017; Qin et al. 2018; [Chen et al. 2019](#); [Zhang et al. 2020](#)), or through coordinated model intercomparison projects (Lin et al. 2012; Williams et al. 2013; Ma et al. 2014; Klingaman et al. 2015; Xavier et al. 2015; Morcrette et al. 2018; Ma et al. 2018; Van Weverberg et al. 2018).

Deleted: ¶

Earlier studies with climate model hindcast experiments usually focused on a relatively short period of time, such as those spanning Intensive Observation Periods or field campaigns. However, determining the robust aspects of certain cloud processes may not be achieved through these short simulations. Recent work in climate model studies has demonstrated the benefit of using multiple years of short-range hindcasts. For example, O'Brien et al. (2016) presented a direct comparison between observed and simulated weather events across multiple resolutions through the analysis of 5-day long hindcasts performed every day during a five-year period. This hindcast modeling framework allows them to assess the degree to which increased resolution improves the fidelity of extreme events in the model. Further, Phillips et al. (2017) studied the land-atmosphere coupling over the U.S. Southern Great Plains (SGP) using a suite of 16 years of short-range hindcasts as well as a free-running [Atmospheric Model Intercomparison Project \(AMIP, Gates 1992\)](#) simulation for the same period, in comparison with the U.S. Department of Energy Atmospheric Radiation Measurement (ARM) observations. Although the surface climate state of the hindcasts deviates less from the observations in contrast to the AMIP simulation, they further identify that the model surface characteristics (e.g., vegetation cover) or physical parameterizations involving land-atmosphere coupling are more important factors than the performance of surface climate state in controlling the coupling behaviors. Chen et al. (2019) assessed precipitation biases in the [Community Earth System Model version 1 \(CESM1\)](#) during the abrupt onset of the South China Sea summer monsoon, a key precursor of the overall East Asia summer monsoon. A multi-year hindcast approach was utilized to obtain the well-constrained synoptic-scale horizontal circulation each year during the onset period. Their results highlighted the need for an appropriate representation of land-ocean-convection interactions over coastal areas in order to improve the simulation of monsoon onset.

Deleted: ; Chen et al. 2019

Deleted: of years

Deleted: Community Atmosphere Model, version 5 (CAM5)

The above examples indicate the benefit of using multiple years of short-range hindcasts for robust process-level modeling studies in comparison with long-term observations. Also, evaluating ensemble short-range hindcasts with the same climate model can complement the traditional way of conducting AMIP-type model evaluation. In the present paper, we present a multi-year short-range hindcast experiment and its experimental design for better evaluating both the mean state and variability of atmospheric moist processes in climate models to facilitate model development using the CESM1 as the base model. 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. We will demonstrate the unique value of diagnosing systematic model errors from diurnal to interannual time scales with this suite of multi-year short-range hindcasts paired with long-term observations, such as from various satellites or from major field programs like the U.S. Department of Energy ARM Program. Process-level understanding can be achieved by comparing hindcasts with observations for periods based on the phenomena of interest rather than the climatological mean state. Three processes – the diurnal cycle of clouds during different cloud regimes at the ARM SGP Site, precipitation and diabatic heating associated with the Madden-Julian Oscillation (MJO), and moist processes response to sea surface temperature (SST) anomalies associated with the El Niño-Southern Oscillation (ENSO) – are evaluated as examples to gain insights into model errors and their connection to physical parameterizations. We also demonstrate that systematic errors in the mean state of moist processes over the global scale are very robust and do not show significant interannual variations in either error magnitudes or patterns over large spatial domains. Our focus of this manuscript is to document this multi-year experiment as a model description paper and provide examples on how to better utilize this suite of hindcasts. The remainder of this manuscript is organized into three sections. Section 2 describes the hindcast experiment design, experiments performed and validation datasets. Section 3 presents three examples of how we can better utilize this suite of multi-year short-range hindcasts to evaluate the variability of moist processes over various time scales. Section 4 presents a summary.

2 Model experiments and validation data

2.1 Model and experiment design

All simulations were conducted with the CESM1 using the active atmospheric and land model components (version cesm1_0_5, FC5 compset, Neale et al. 2012). The atmospheric model component is the CAM5 with the finite volume dynamical core at a horizontal resolution of 0.9° latitude by 1.25° longitude and 30 vertical levels. The land model component is the Community Land Model version 4.0 (CLM4) with the same horizontal resolution. The ocean and sea ice components are prescribed with the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation weekly SSTs and sea ice (Reynolds et al. 2002).

Deleted: propose

Deleted: protocol

Deleted: from diurnal to interannual time scales

Deleted:

Deleted: CAM5

Deleted:

Deleted: using

Moved (insertion) [1]

Deleted: Both experiments

The hindcast procedure is based on Ma et al. (2015). In summary, 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 (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), which are not available from the ERA-Interim Reanalysis for the atmospheric initial conditions. The nudging simulation started from January 1st, 1996 and stopped at December 31st, 2012 with a 6 h relaxation time scale. Land initial conditions are taken from an offline land model simulation (I 2000 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. 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. The carbon and nitrogen cycles in this particular CLM4 setup are not active. 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. We have published a documentation and relevant scripts in generating initial conditions on GitHub (<https://github.com/PCMDI/CAPT>), which includes the initialization generation procedure, the nudging procedure, and land-surface spinup procedure. The multi-year hindcast experiment is a suite of 3-day long hindcasts starting at 00Z every day for the years of 1997 to 2012 (Figure 1) using the initial conditions obtained from the procedure described above. 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. Day 1 data are not analyzed to minimize the impact of model spin-up (Ma et al. 2013; 2014).

Deleted: .

Deleted: w

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. Also, the greenhouse gas and solar forcing is based on the setup of the CESM1 FC5 compset, which corresponds to the year 2000 level for all the simulation period. This is because the CMIP5 forcing data does not go beyond 2005. We also want to exclude the impact of the interannual variations in the solar and greenhouse gas forcings to our simulations so that we can better identify possible causes of model biases associated with parameterizations. To compare with high-temporal frequency observations collected at the ARM permanent sites as well as at various major field campaign locations within the simulation period, we have additionally generated output for every model timestep (30 minutes interval) in addition to output for the entire global domain. Figure 2 and Table 1 identify their geographical locations and output grids.

Deleted: s

Deleted: Atmospheric Model Intercomparison Project (AMIP, Gates 1992)...

Moved up [1]: Both experiments are prescribed with the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation weekly SSTs and sea ice (Reynolds et al. 2002).

Deleted: at

2.2 Strategy on performing the multi-year hindcasts on a high performance computing system

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 high performance computing (HPC) system, one can easily bundle as many hindcasts as possible into one job submission on a HPC system. This is the concept of submitting multiple parallel jobs simultaneously in a single batch script.

160 Most HPC systems nowadays have this capability and even encourage people to submit large jobs with discount on the charge of computer hours. For example, a single hindcast takes one hour to complete with 5 computer nodes. One can then request 1,825 nodes for performing short-range hindcasts for one year period in a single job submission. The queue time may require a longer wait but the entire hindcasts will finish within an hour. One can submit multiple jobs in the queue for multiple years of hindcasts. One issue to keep in mind is the storage space for model output. There is usually a disk quota for scratch space on a HPC system. Therefore, a script or code to save model output to a long-term storage system, such as a High Performance Storage System (HPSS), may be necessary to prevent reaching the scratch disk quota while the model is running.

2.3 Comparison datasets

Daily global observational precipitation is adopted from the Global Precipitation Climatology Project Version 1.2 (GPCP, Adler et al. 2003). Absorbed shortwave flux at top of atmosphere (SWAbs), outgoing longwave radiation (OLR) as well as net surface shortwave and longwave fluxes are obtained from Clouds and the Earth's Radiant Energy System (CERES) Energy 170 Balanced And Filled (EBAF) observations (Loeb et al. 2009; Kato et al. 2013, Edition 2.8). Total cloud fraction is from the International Satellite Cloud Climatology Project (ISCCP) D2 dataset (Rossow et al. 1999). Global winds and surface turbulent heat fluxes are from the ERA-Interim Reanalysis. Vertical profiles of cloud fraction at the ARM SGP site are from the ARM Best Estimate (ARMBE; Xie et al. 2010) Active Remote Sensing of Clouds data (ARSCL; Clothiaux et al. 2000, 2001). The available time period of the dataset is listed in Table 2. We interpolated all the datasets onto model's grid for better comparison.

175 3 Example analysis

Our goal here is to demonstrate the usefulness of the multi-year hindcasts in providing a different perspective on several long-standing moist processes errors in GCMs through three examples. Note that identifying causes for individual model issue requires further investigation and is beyond the current scope of this model experiment description paper. For sensitivity tests to various parameter choices for a specific scheme or parameterization, it is not necessary to perform this suite of multi-year 180 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 shallow cumulus at ARM SGP, or phase 3 of various MJOs, which we will introduce in the following text).

3.1 Cloud regimes at the ARM SGP site

185 One common application of hindcasts for model evaluation is during major field campaigns where intensive observations are available at very high temporal scale. However, field campaigns are usually confined to a certain short period and cannot determine the robust aspects of certain cloud processes, which are available only from long-term monitoring as provided by satellites or permanent ground-based sites. From over ten years of cloud radar observations at the ARM SGP site, Zhang and Klein (2010) computed the diurnal cycle of cloud vertical structure for four distinct cloud regimes: daytime clear sky, daytime

Deleted: 2

Deleted: CAM5's

shallow convection, afternoon deep convection and nighttime convection (Figure 3 a-d). They are defined in as: (1) daytime clear sky day: precipitation rate = 0 mm day⁻¹ at all hours of the day and cloud fraction ≤ 5% at all levels between 0800 and 1600 local standard time (LST), (2) daytime shallow convection day: 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, (3) afternoon deep convection day: the diurnal maximum hourly precipitation rate ≥ 1 mm day⁻¹ and occurs between 1500 and 2000 LST and is at least twice more than the precipitation rate at any other hour of the day outside of 1500-2000 LST) and (4) nighttime deep convection day: the diurnal maximum hourly precipitation rate ≥ 1 mm day⁻¹ and occurs between 0000 and 0700 LST. Each composite consists of somewhere between 79 and 229 days spanning the warm seasons for a 10-year period. We also created the same model composites (Figure 3 e-h) from the model grid box nearest the SGP site (Figure 2), for the exact same days from the Day 2 hindcasts. We use the model cloud fraction for this comparison because the variables for using a radar simulator (Zhang et al. 2018; 2019) were not saved at the time the hindcasts were done. This analysis, which cannot be achieved from the usual AMIP simulations, is a more precise way of model parameterization evaluation because it minimizes the impact of erroneous large-scale states on the clouds. This is because the atmospheric large-scale state is closer to observations during each diurnal cycle of the hindcasts than it is in the AMIP simulation. Furthermore, multi-year hindcasts provide a sufficient number of events to make a meaningful comparison with observations so that conclusions from such studies are more statistically robust.

Formatted: Superscript

In Figure 3, the model overestimates high clouds regardless of cloud regimes, even for clear sky condition. For clear sky condition, the model also shows middle- and low-level clouds. One possible explanation is that the deep convection scheme in the model is triggered whenever the convective available potential energy (CAPE) is larger than 70 J kg⁻¹. During the daytime in the warm seasons, CAPE is usually larger than the threshold and deep convection is easily triggered, resulting in the transport of water vapor and detrainment of cloud condensates. For the shallow convection regime, the model overestimates middle-level clouds by ~4-6% but underestimates shallow clouds by ~10%. 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 from around 11 local time rather than 15 in the afternoon. Also, the model underestimates both shallow and middle-level cloud fraction by ~10%. The model completely misses the nighttime convection regime, and only shows some deep convection starting around noon. The too early afternoon convection and the lack of nocturnal convection over land are common model problems as reported in previous studies (e.g., Dai 2006; Jiang et al. 2006; Covey et al. 2016). The missing nocturnal precipitation is likely related to the incapability of the model to capture elevated convection that often occurs at night at Southern Great Plains (Xie et al. 2019). We do realize that there are already small errors in the Day 2 large-scale state and they can also contribute to the errors in the simulated cloud fields. Nevertheless, their impact is still much smaller compared to the errors due to parameterization deficiencies in convection.

Deleted: as easily

Deleted: s

230 With multi-year hindcasts and long-term cloud observations to build up robust statistics, these comparisons help identify specific cloud regime deficiencies under very similar large-scale meteorological conditions, and model developers can further focus on improving specific processes represented in the cloud and convection parameterizations.

Deleted: scheme

3.2 Model biases associated with MJO

235 The MJO (Madden and Julian 1971, 1972) is the dominant mode of intraseasonal variability in the tropics. MJO has significant impacts on the global water cycle as it can interact with many weather and climate phenomena (Zhang 2013). Nevertheless, contemporary GCMs still simulate poor MJO including its weak amplitude and lack of eastward propagation (Jiang et al.2015, Ahn et al. 2017). Recent studies suggest that the instability and propagation of the MJO are regulated by various feedback processes including cloud-radiation and wind-evaporation feedbacks (Sobel and Maloney 2012, 2013; Adams and Kim 2016; Ciesielski et al. 2017). These feedback processes may contribute to better MJO simulations if they are well represented in the GCMs. A particularly relevant process responsible for the eastward propagation of the MJO is the “pre-conditioning” process consisting of low-level moistening and shallow convective heating structure at the eastern edge of MJO deep convection (e.g., Jiang et al. 2011; Johnson and Ciesielski 2013; Powell and Houze 2013; Xu and Rutledge 2014). This process destabilizes the environment encouraging subsequent development of deep convection.

Deleted: -

Deleted: need to be well represented in GCMs in order to produce realistic

245 As each MJO event is unique from one another, one can take advantage of the multi-year hindcasts to composite precipitation, winds and diabatic heating profiles based on observed MJO phases with the focus on identifying robust model biases associated with the MJO. The diabatic heating rate or apparent heating of large-scale motion system (Q1) consists of the heating due to radiation, the release of latent heat by net condensation, and vertical convergence of the vertical eddy transport of sensible heat (Yanai et al. 1973). In CESM1/CAM5, Q1 can be calculated through summing up all the tendency terms of all the diabatic processes. Figure 4 presents the observed composites of November to April 20-100 day band-pass filtered NOAA Interpolated Outgoing Longwave Radiation (OLR) anomalies and horizontal wind anomalies at 850mb from ERA-Interim, as a function of the eight phases of the MJO (Wheeler and Hendon 2004). The observed MJO shows a core of deep convection (center of negative OLR anomalies) over the Indian Ocean around 80°E associated with low-level convergence in winds during Phase 2. The core of deep convection slowly propagates eastward, and the intensity of convection decreases (with OLR anomalies increasing) after the core of MJO crosses over the Maritime Continent and reaches the central Pacific (Phases 6-8). Figure 5 shows composites of November to April precipitation and horizontal wind biases from Day 3 hindcasts as a function of the eight phases of the MJO. We find that there is a dry bias in Day 3 hindcasts over the core of deep convection (center of negative OLR anomalies in Figure 4) associated with MJO, and a wet bias to the east over the region of suppressed convection (center of positive OLR anomalies in Figure 4) for all the phases as the MJO moves eastward. The dry bias is largest over Indian Ocean during Phase 2 with magnitude ~ -6 mm day⁻¹, and the wet bias is largest over western Pacific during Phase 8 with magnitude ~5-6 mm day⁻¹. The dry bias is usually attributed to the lack of organized convection in the model (Moncrieff et al.

Deleted: (Q1, Yanai et al. 1973)

Deleted: de

Deleted: es

Deleted: , as well as 20-100 day band-pass filtered NOAA Interpolated OLR anomalies,

Deleted: The observed OLR anomalies are superimposed to highlight the location of the MJO at each phase.

2017), and the wet bias is consistent with the too frequently triggered deep convection scheme even under suppressed large-scale condition. Further, there is a persistent dry bias over Borneo and part of Sumatra [and wet bias around the Maritime Continent](#) for all the phases indicating a possible local effect of diurnal cycle of convection. The dry bias is more significant during Phase 4 and 5 as the MJO crosses over the Maritime Continent. The 850 mb winds show a biased low-level convergence near the Equator consistent with the excessive precipitation bias to the east over the region of suppressed convection.

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 ($<0.4 \text{ K day}^{-1}$) 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. 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 [in Phase 3 of Day 3 hindcast composites](#).

Figure 6 presents Hovmöller diagrams (longitude versus time in lag days) of rainfall anomalies along an equatorial band based on the lag correlation over an Indian Ocean box for both GPCP and Day 3 hindcasts. The model shows an eastward propagation of precipitation anomalies associated with the MJO. However, the slightly lower correlation coefficients in the hindcasts (-0.3 vs ~ 0.4 in observation) east of 105°E is consistent with the weakening of MJO eastward propagation particularly over the Maritime Continent. Previous studies have identified Maritime Continent as a “barrier” for MJO propagation and most MJO episodes in the models fail to propagate across it due to several possible reasons, such as interactions of convection, clouds, surface fluxes and local circulation within a diurnal cycle, land-sea contrast, terrain effect, or east-west low-level moisture gradient (Hagos et al. 2016; Jiang 2017; Zhang and Ling 2017). Although further diagnosis is required to fully understand the underlying problems for MJO propagation, our analysis with multi-year hindcasts indicates issues in the early development of biases in the shallow and deep convections associated with the MJO, as well as their interactions with the diurnal cycle of convection over the Maritime Continent.

3.3 Interannual variations of moist processes

3.3.1 Variations of moist processes associated with ENSO

Being the leading mode of interannual variability in the tropics and extra-tropics, ENSO has significant impact on both regional and global temperature, circulation, and moist processes through teleconnections. To gain insights into whether or not errors in the response of these fields to SST anomalies can be attributed to parameterization errors or whether errors in the circulation response to SST anomalies also contribute, one can further contrast the multi-year hindcasts with the behavior of a companion AMIP simulation with the same boundary conditions (SST and sea ice). [This question cannot be addressed before with one-year short-range hindcasts as we proposed in Ma et al. \(2015\)](#). To this end, we first selected several fields to compute their monthly anomalies and then regressed these anomalous fields onto the Nino 3.4 index. Figure 7 shows the regression maps of

Deleted: as MJO propagates

305 precipitation, SWAbs, surface net flux (from atmosphere to the surface), and the surface zonal wind stress from observations,
Day 2 hindcasts and the AMIP simulation (pattern statistics are shown in Table 3). The motivation for selecting these fields is
because the tropical response of precipitation represents the atmospheric diabatic heating that forces circulation anomalies. On
the other hand, surface radiation, turbulent heat fluxes, and wind stress provide critical heat and momentum forcings for SST
anomalies and govern the ENSO behavior. The performance of these fields from an uncoupled atmospheric GCM is considered
310 to be highly relevant for evaluation when it couples to an ocean model (Sun et al. 2006; Guilyardi et al. 2009).

Deleted: 2

The responses of these fields from Day 2 hindcasts show a better agreement both in the spatial patterns and magnitude with
observations compared to the AMIP response (right column in Figure 7). This is especially evident for precipitation, absorbed
shortwave flux and zonal wind stress over the Western North Pacific, South Pacific Convergence Zone and Indian Ocean. The
315 remote teleconnections may be chaotic or poorly done by the model, causing a poor simulation in the AMIP mode. The large-
scale state is well constrained in the hindcasts and the response of those fields to SST anomalies in Figure 7 is much superior.
This shows that remote errors are mostly the result of errors in circulation on long-term time scales although the errors in
circulation may be caused by model physics in the first place and deteriorates through feedback processes with time. This is
evident as there are still biases in the hindcasts indicating problems from parameterizations in representing those response to
320 SST changes even over local Nino 3.4 region.

Deleted: circulation anomalies

Deleted: are

Deleted: poor

Deleted: poor

Surface net flux and zonal wind stress also show a greater change between hindcast and AMIP response compared to
precipitation and SWAbs. It is reasonable for the latter two moist processes to show less changes as they are fast processes
and the biases associated with model parameterizations usually develop within a few days of model integration (Xie et al.
325 2012; Ma et al. 2014). It is also reasonable for zonal wind stress to show greater change as the low-level winds are well
constrained for the hindcasts. For surface net heat flux, the errors are contributed from various flux terms including radiation
and turbulent heat fluxes, which are affected by both model physics and dynamics. Therefore, the net heat flux shows the
lowest spatial correlation and larger root mean square errors in both hindcasts and the AMIP simulation compared to other
fields.

330 3.3.2 Robustness of systematic errors

One question raised from earlier studies (Xie et al. 2012 and Ma et al. 2014) of the correspondence between short- and long-
time-scale errors is whether systematic errors of moist processes show significant interannual variation in the mean state biases.
Figure 8 shows the pattern statistics between errors from the individual annual means in the hindcasts or AMIP simulation,
and errors in the 16-year mean of the AMIP simulation (the reference fields) for precipitation, total cloud fraction (from the
335 ISCCP cloud simulator), SWAbs, and OLR. Compared to the long-term mean errors in the AMIP simulations, annual mean
errors of the individual years for these fields show very similar magnitude in correlation and the normalized spatial standard
deviation from the hindcasts at either time lag. This is also the case for individual AMIP years although the correlations and

standard deviations show slightly larger spread. Compared to Day 2 hindcasts, the correlations and standard deviations from Day 3 hindcasts are closer to those from the AMIP simulations indicating the bias growth toward the AMIP bias with hindcast lead time. We further find that the magnitude of correlations for annual mean errors between individual hindcast years and the long-term AMIP simulations are not sensitive to the ENSO phase in a given year for these fields. This is also the case if seasonal means are compared (Figures not shown here). These results suggest that mean errors in the moist processes are very robust and do not show significant interannual variations. Indeed, averaging the hindcast errors over many years (indicated by “2” or “3” in Figure 8) only slightly improves the agreement with the AMIP reference field. Thus, 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). A similar conclusion with multiple years of short AMIP-type simulations was also suggested by Wan et al. (2014). These results suggest that relatively short simulations will be effective at identifying the systematic moist process errors of a very high-resolution climate model which is too expensive to regularly perform multi-year simulations.

Deleted: can

Deleted: reducing

It is also of interest to compare the absolute magnitude of errors in individual years to that of the long-term systematic error in the AMIP simulation. To do so, we calculated the annually-averaged cloud error metrics proposed in Klein et al. (2013) in Figure 9. These metrics are scalar measures of performance in simulating the space-time distribution of several cloud measures, with better performance indicated by smaller E values. E_{TCA} measures the error in total cloud amount, and $E_{CTP-\tau}$ measures the errors in the frequency of optically intermediate and thick clouds at high, middle, and low-levels of the atmosphere. E_{sw} and E_{lw} measures the errors in the impacts on top-of-atmosphere shortwave and longwave radiation in the same cloud-top pressure and optical depth categories used for $E_{CTP-\tau}$, respectively. It is not surprising that the hindcasts show better performance in all the cloud metrics as the large-scale circulation and state are not too far from the reanalysis. This is also true for the interannual variations in global mean cloud radiative effect at the top of the atmosphere (Figure 10). We find that all the metrics and the cloud radiative effect show interannual variations indicating that the circulation and state anomalies make a significant contribution to interannual variations although these metrics or errors in the cloud radiative effect are not sensitive to ENSO phase. We further find that there is a larger difference between hindcasts and AMIP in the total cloud amount error metric (E_{TCA}) implying that errors in the large-scale circulation and state make a larger contribution to errors in E_{TCA} than cloud radiative properties (Figure 10).

4 Summary

In this study, we present a multi-year short-range hindcast experiment and its experiment design for better evaluating both the mean state and variability of atmospheric moist processes in climate models from diurnal to interannual time scales to facilitate model development. We also demonstrate that one can obtain unique understanding into robust GCM systematic moist processes errors by diagnosing these processes with corresponding observations for periods based on different phenomena. The present experiment also demonstrates that it is now feasible to systematically evaluate climate model moist processes in deterministic weather-prediction mode just as the moist processes in weather prediction models are often evaluated in analyses

Deleted: propose

Deleted: protocol

Deleted: study

or re-analyses (Jakob 1999, Yang et al. 2006). This experiment can also provide a very useful avenue to diagnose and understand critical processes regulating various climate and weather phenomena by taking advantage of detailed model output with a largely realistic representation of the large-scale state in hindcasts.

385 Three processes – the diurnal cycle of clouds during different cloud regimes at the ARM SGP Site, precipitation and diabatic heating associated with the MJO, and the response of moist processes to ENSO SST anomalies – are evaluated as examples of using this multi-year hindcast experiment to gain insights into robust model errors and their connection to physical parameterizations and large-scale state. These analyses can only be done through this multi-year hindcast experiment to establish robust statistics of the processes under well-controlled large-scale environment because these phenomena are either
390 interannual climate variability or only happen a few times in a given year (e.g. MJO, or cloud regime types). These comparisons identify specific model deficiencies that subsequent parameterization development should focus on. Results from the multi-year hindcasts also suggest that systematic errors in the mean state of moist processes are very robust and do not show significant interannual variation in error magnitude or patterns over large spatial domain. Although we only showed examples relevant to moist processes, other processes related to planetary boundary layer or radiation schemes can also be examined
395 through this suite of experiments. 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.

Deleted: c

Deleted: approach

Deleted: The proposed experiment and evaluation method also complement the existing ways of climate model evaluation, such as performing GCM simulations in the AMIP mode.

Deleted: ¶

In addition to processes indicated above, further studies on monsoon variability (e.g., South American and Asian monsoons,
400 Chen et al. 2019), land-atmosphere interactions (Phillips et al. 2017), or detailed MJO studies with longer hindcast duration (Klingaman et al. 2015), are currently being explored with these hindcasts. As demonstrated in previous studies and here,
model mean biases associated parameterized moist processes usually develop within a few days and manifest within weeks to affect the simulations of large-scale circulation and ultimately the climate mean state and variability. Therefore, model developers can achieve useful understanding of the underlying problems in model physics by conducting multiple years of
405 hindcasts as demonstrated in the present work. Although newer version of the CAM and CLM is now available (CAM6/CLM5), similar systematic errors associated with moist processes remain present in the latest model version. Therefore, it is still worthwhile to study these hindcasts and compare the results to hindcasts with newer model version. In the meantime, we also plan to conduct another suite of multi-year hindcasts with the latest DOE Exascale Energy Earth System Model (E3SM, Golaz et al. 2019). We will also compare the results from E3SM to CESM1 to understand the impact of
410 parameterization and model changes to the performance of moist processes since the atmospheric component of E3SM was originally branching form CAM5.3, which has very similar performance as CAM5 (Xie et al. 2018, Rasch et al. 2019). Note that E3SM version 1 has a new set of atmospheric physical parameterizations that are very similar to CAM6, the latest CAM.
The hindcasts will also be available to the community once available.

Deleted: Indeed, GCMs usually perform more poorly for climate variability than for mean state.

Finally, the multi-year hindcast approach presented in this study is also intended as one of the experiment protocols which will be used 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). This project is aimed to understand the processes that control the diurnal and sub-diurnal variations of precipitation over different climate regimes in observations and in models. The project will also identify the deficiencies and missing physics in current GCMs to gain insights for further improving the parameterization of convection.

Deleted: understand what processes control the diurnal and sub-diurnal ...

Code and data availability

The model code is the CESM1 (cesm1_0_5, FC5 compset, F09_F09 resolution) and is available over <http://www.cesm.ucar.edu/models/cesm1.0/>. All model necessary input files are available over <https://svn-ccsm-inputdata.cgd.ucar.edu/trunk/inputdata/>. The boundary conditions of SST and sea ice data are available over <https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html>. The simulations is available online through the NERSC Science Gateways (https://portal.nersc.gov/archive/home/h/hyma/www/CAPT/CAPT_Long). An detailed documentation for this experiment and variable list is over https://portal.nersc.gov/archive/home/h/hyma/www/CAPT/CAPT_Long/CAPT_Long_output_cesm1_0_5_v5.pdf.

Author contribution

HM designed, performed, analyzed the experiments, and wrote the first draft of the manuscript. SK and SX contributed to the design of the experiments and to the interpretation of the results. CZ, YZ, MZ, and XZ contributed to the analysis of experiments, and the interpretation of the results. WC, and CW contributed to the interpretation of the results. All coauthors contributed to the manuscript text.

Acknowledgments

This study was funded by the U.S. Department of Energy (DOE) Regional and Global Model Analysis program area, and the Atmospheric System Research and Atmospheric Radiation Measurement Programs. This work was performed under the auspices of the U.S. DOE by LLNL under contract DE-AC52-07NA27344.

References

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., and Nelkin, E.: The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979-present), *J. Hydrometeorol.*, 4, 1147-1167, 2003.

- Ahn, M.-S., Kim, D., Sperber, K. R., Kang, I.-S., Maloney, E., Waliser, D., and Hendon, H.: MJO simulation in CMIP5 climate models: MJO skill metrics and process-oriented diagnosis, *Clim. Dyn.*, <https://doi.org/10.1007/s00382-017-3558-4>, 1-23, 2017.
- 455 Barton, N. P., Klein, S. A., Boyle, J. S., and Zhang, Y.: Arctic synoptic regimes: Comparing domain-wide Arctic cloud observations with CAM4 and CAM5 during similar dynamics, *J. Geophys. Res.*, 117, D15205, <https://doi.org/10.1029/2012JD017589>, 2012.
- Barton, N. P., Klein, S. A., and Boyle, J. S.: On the contribution of longwave radiation to global climate model biases in Arctic lower tropospheric stability, *J. Climate*, 27, 7250-7269, <https://doi.org/10.1175/JCLI-D-14-00126.1>, 2014.
- 460 Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma, A. P., Sobel, A. H., Watanabe, M., and Webb, M. J.: Clouds, circulation and climate sensitivity, *Nat. Geosci.*, 8, 261–268. <https://doi.org/10.1038/NGEO2398>, 2015.
- [Berg, L. K., and Kassianov, E. I.: Temporal variability of fair- weather cumulus statistics at the ACRF SGP site, *J. Climate*, 21, 3344–3358, 2008.](https://doi.org/10.1175/JCLI-D-14-00126.1)
- 465 Ciesielski, P. E., Johnson, R. H., Jiang, X., Zhang, Y., and Xie, S.: Relationships between radiation, clouds, and convection during DYNAMO, *J. Geophys. Res. Atmos.*, 122, 2529–2548, <https://doi.org/10.1002/2016JD025965>, 2017.
- Chen, W.-T., Wu, C.-M., and Ma, H.-Y.: Evaluating the bias of South China Sea summer monsoon precipitation associated with fast physical processes using climate model hindcast approach, *J. Climate*, 32, 4491–4507, <https://doi.org/10.1175/JCLI-D-18-0660.1>, 2019.
- 470 Clothiaux, E. E., Ackerman, T. P., Mace, G. G., Moran, K. P., Marchand, R. T., Miller, M., and Martner, B. E.: Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites, *J. Appl. Meteor.*, 39, 645–665, 2000.
- Clothiaux, E. E., Miller, M. A., Perez, R. C., Turner, D. D., Moran, K. P., Martner, B. E., Ackerman, T. P., Mace, G. G., Marchand, R. T., Widener, K. B., Rodriguez, D. J., Uttal, T., Mather, J. H., Flynn, C. J., Gaustad, K. L., and Ermold, B.: The
- 475 ARM millimeter wave cloud radars (MMCRs) and the active remote sensing of clouds (ARSCL) value added product (VAP), U.S. Department of Energy Tech. Memo. ARM VAP-002.1, 56 pp, 2001.
- Chandra, A. S., Zhang, C., Klein, S. A., and Ma, H.-Y.: Low-cloud characteristics over the tropical western Pacific from ARM observations and CAM5 simulations, *J. Geophys. Res. Atmos.*, 120, <https://doi.org/10.1002/2015JD023369>, 2015.
- Covey, C., Gleckler, P. J., Doutriaux, C., Williams, D. N., Dai, A., Fasullo, J., Trenberth, K., and Berg, A.: Metrics for diurnal
- 480 cycle of precipitation: Toward routine benchmarks for climate models, *J. of Climate*, 29, 4461-4471, 2016.
- Dai, A.: Precipitation characteristics in eighteen coupled climate models, *J. of Climate*, 19, 4605-4630, 2006.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kallberg, P., Köhler, M., Matricardi, M.,
- 485 McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and

- Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, 137, 553–597, 2011.
- Gates, W. L.: AMIP: The Atmospheric Model Intercomparison Project, *Bull. Amer. Meteor. Soc.*, 73, 1962–1970, 1992.
- Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., et al: The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, 11, <https://doi.org/10.1029/2018MS001603>, 2019.
- Guilyardi, E., Wittenberg, A., Fedorov, A., Collins, M., Wang, C., Capotondi, A., van Oldenborgh, G. J., and Stockdale, T.: Understanding El Niño in ocean-atmosphere general circulation models, *Bull. Amer. Met. Soc.*, 90, 325–340, 2009.
- Hagos, S. M., Zhang, C., Feng, Z., Burleyson, C. D., DeMott, C., Kerns, B., Benedict, J., and Martini, M.: The impact of the diurnal cycle on the propagation of Madden-Julian Oscillation convection across the Maritime Continent, *J. Adv. Model. Earth Syst.*, 8, 1552-1564. <https://doi.org/10.1002/2016MS000725>, 2016.
- Hannah, W. M., and Maloney, E. D.: The moist static energy budget in NCAR CAM5 hindcasts during DYNAMO, *J. Adv. Model. Earth Syst.*, 6(2), 420–440, 2014.
- Jakob, C.: Cloud cover in the ECMWF reanalysis, *J. Climate*, 12, 947–959, 1999.
- 500 Jiang, X., Lau, N.-C., and Klein, S. A.: Role of eastward propagating convection systems in the diurnal cycle and seasonal mean of summertime rainfall over the U.S. Great Plains, *Geophys. Res. Lett.*, 33, L19809, <https://doi.org/10.1029/2006GL027022>, 2006.
- 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.
- Jiang, X., Waliser, D. E., Xavier, P. K., Petch, J., Klingaman, N. P., Woolnough, S. J., Guan, B., Bellon, G., Crueger, T., DeMott, C., Hannay, C., Lin, H., Hu, W., Kim, D., Lappen, C.-L., Lu, M.-M., Ma, H.-Y., Miyakawa, T., Ridout, J. A., Schubert, S. D., Scinocca, J., Seo, K.-H., Shindo, E., Song, X., Stan, C., Tseng, W.-L., Wang, W., Wu, T., Wu, X., Wyser, K., Zhang, G. J., Zhu, H.: Vertical structure and physical processes of the Madden-Julian oscillation: Exploring key model physics in climate simulations, *J. Geophys. Res. Atmos.*, 120, 4718-4748, <https://doi.org/10.1002/2014JD022375>, 2015.
- 510 Jiang, X.: Key processes for the eastward propagation of the Madden-Julian Oscillation based on multimodel simulations, *J. Geophys. Res. Atmos.*, <https://doi.org/10.1002/2016JD025955>, 2017.
- Johnson, R. H. and P. E. Ciesielski (2013), Structure and Properties of Madden–Julian Oscillations Deduced from DYNAMO Sounding Arrays. *J. Atmos. Sci.*, 70, 3157-3179, <https://doi.org/10.1175/JAS-D-13-065.1>, 2013.
- 515 Kato, S., Loeb, N. G., Fred, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., Yu, L., Weller, R. A.: Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances, *J. of Climate*, 26(9), 2719-2740. <http://dx.doi.org/10.1175/JCLI-D-12-00436.1>, 2013.
- Klingaman, N. P., Woolnough, S. J., Jiang, X., Waliser, D., Xavier, P. K., Petch, J., Caian, M., Hannay, C., Kim, D., Ma, H.-Y., Merryfield, W. J., Miyakawa, T., Pritchard, M., Ridout, J. A., Roehrig, R., Shindo, E., Vitart, F., Wang, H., Cavanaugh,

- 520 N. R., Mapes, B. E., Shelly, A., and Zhang, G. J.: Vertical structure and physical processes of the Madden–Julian oscillation: Linking hindcast fidelity to simulated diabatic heating and moistening, *J. Geophys. Res.-Atmos.*, 120, 4690–4717, <https://doi.org/10.1002/2014JD022374>, 2015.
- Klein, S. A., Jiang, X., Boyle, J. S., Malyshev, S., and Xie, S.: Diagnosis of the summertime warm and dry bias over the U.S. Southern Great Plains in the GFDL climate model using a weather forecasting approach. *Geophys. Res. Lett.*, 33, L18805, <https://doi.org/10.1029/2006GL027567>, 2006.
- 525 Klein, S. A., Zhang, Y., Zelinka, M. D., Pincus, R., Boyle, J. S., and Gleckler, P. J.: Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator, *J. Geophys. Res.*, 118, <https://doi.org/10.1002/jgrd.50141>, 2013.
- Lin, Y., Donner, L. J., Petch, J., Bechtold, P., Boyle, J. S., Klein, S. A., Komori, T., Wapler, K., Willett, M., Xie, X., Zhao, M., Xie, S., McFarlane, S. A., and Schumacher, C.: TWP-ICE global atmospheric model intercomparison: convection
- 530 responsiveness and resolution impact, *J. Geophys. Res.*, <https://doi.org/10.1029/2011JD017018>, 2012.
- Loeb, G. N., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., Manalo-Smith, N., and Wong, T.: Toward optimal closure of the Earth’s top-of-atmosphere radiation budget, *J. Climate*, 22, 748–766, <https://doi.org/10.1175/2008JCLI2637.1>, 2009.
- Ma, H.-Y., Xie, S., Boyle, J. S., Klein, S. A., and Zhang, Y.: Metrics and diagnostics for precipitation-related processes in
- 535 climate model short-range hindcasts, *J. Climate*, 26, 1516–1534, 2013.
- Ma, H. Y., Xie, S., Klein, S. A., Williams, K. D., Boyle, J. S., Bony, S., Douville, H., Fermin, 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.
- Ma, H.-Y., Chuang, C. C., Klein, S. A., Lo, M.-H., Zhang, Y., Xie, S., Zheng, X., Ma, P.-L., Zhang, Y., and Phillips, T. J.: An
- 540 improved hindcast approach for evaluation and diagnosis of physical processes in global climate models, *J. Adv. Model. Earth Syst.*, 7, 1810–1827, <https://doi.org/10.1002/2015MS000490>, 2015.
- Ma, H. Y., Klein, S. A., Xie, S., Zhang, C., Tang, S., Tang, Q., Morcrette, C. J., Van Weverberg, K., Petch, J., Ahlgrimm, M., Berg, L. K., Cheruy, F., Cole, J., Forbes, R., Gustafson Jr., W. I., Huang, M., Liu, Y., Merryfield, W., Qian, Y., Roehrig, R., and Wang, Y.-C.: CAUSES: On the role of surface energy budget errors to the warm surface air temperature error over the
- 545 Cen- tral United States, *J. Geophys. Res.-Atmos.*, 123, 2888–2909, <https://doi.org/10.1002/2017JD027194>, 2018.
- Madden, R. A., and Julian, P. R.: Detection of a 40- 50 day oscillation in the zonal wind in the tropical Pacific, *J. Atmos. Sci.*, 28, 702–708, 1971.
- Madden, R. A., and Julian, P. R.: Description of global-scale circulation cells in the tropics with a 40-50 day period, *J. Atmos. Sci.*, 29, 1109–1123, 1972.
- 550 Medeiros, B., Williamson, D. L., Hannay, C., Olson, J. G.: Southeast Pacific stratocumulus in the Community Atmosphere Model, *J. Climate*, 25, 6175–6192, 2012.
- Moncrieff, M. W., Liu, C., and Bogenschutz, P.: Simulation, modeling, and dynamically based parameterization of organized tropical convection for global climate models, *J. Atmos. Sci.*, 74, 1363–1380, 2017.

Deleted: -

- 555 Morcrette, C. J., Van Weverberg, K., Ma, H. Y., Ahlgrim, M., Bazile, E., Berg, L. K., Cheng, A., Cheruy, F., Cole, J., Forbes, R., and Gustafson Jr., W. I.: Introduction to CAUSES: Description of weather and climate models and their near- surface temperature errors in 5 day hindcasts near the South- ern Great Plains, *J. Geophys. Res.-Atmos.*, 123, 2655–2683, <https://doi.org/10.1002/2017JD027199>, 2018.
- Neale, R. B., Chen, C. C., Gettelman, A., Lauritzen, P. H., Park, S., Williamson, D. L., Conley, A. J., Garcia, R., Kinnison, 560 D., Lamarque, J. F., and Marsh, D.: Description of the NCAR community atmosphere model (CAM 5.0), NCAR Tech. Note NCAR/TN-486+ STR, 2010.
- O'Brien, T. A., Collins, W. D., Kashinath, K., Rübél, O., Byna, S., Gu, J., Krishnan, H., and Ullrich, P. A.: Resolution dependence of precipitation statistical fidelity in hindcast simulations, *J. Adv. Model. Earth Syst.*, 8, 976–990, <https://doi.org/10.1002/2016MS000671>, 2016.
- 565 Phillips, T. J., Potter, G. L., Williamson, D. L., Cederwall, R. T., Boyle, J. S., Fiorino, M., Hnilo, J. J., Olson, J. G., Xie, S., and Yio, J. J.: Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bull. Amer. Meteor. Soc.*, 85, 1903–1915, 2004.
- Phillips, T. J., Klein, S. A., Ma, H.-Y., Tang, Q., Xie, S., Williams, I. N., Santanello, J. A., Cook, D. R., and Torn, M. S.: Using ARM Observations to Evaluate Climate Model Simulations of Land-Atmosphere Coupling on the U.S. Southern Great Plains, 570 *J. Geophys. Res., Atmos.*, 122, 11,524–11,548. <https://doi.org/10.1002/2017JD027141>, 2017.
- Powell, S. W. and Houze, R. A.: The cloud population and onset of the Madden-Julian Oscillation over the Indian Ocean during DYNAMO-AMIE, *J. Geophys. Res., Atmos.*, 118, <https://doi.org/10.1002/2013JD020421>, 2013.
- [Rasch, P. J., Xie, S., Ma, P., Lin, W., et al.: An Overview of the Atmospheric Component of the Energy Exascale Earth System Model. *Journal of Advances in Modeling Earth Systems*, DOI:10.1029/2019MS001629, 2019.](https://doi.org/10.1029/2019MS001629)
- 575 Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., and Wang, W.: An improved in situ and satellite SST analysis for climate, *J. Climate*, 15, 1609-1625, 2002.
- Rossow, W. B., and Schiffer, R. A.: Advances in Understanding Clouds from ISCCP, *Bull. Amer. Meteor. Soc.*, 80, 2261–2288, 1999.
- Sun, D.-Z., Thang, T., Covey, C., Klein, S. A., Collins, W. D., Hack, J. J., Kiehl, J. T., Meehl, G. A., Held, I. M., and Suarez, 580 M.: Radiative and dynamical feedbacks over the equatorial cold tongue: Results from nine atmospheric GCMs, *J. Climate*, 19, 4059–4074, 2006.
- Van Weverberg, K., Morcrette, C. J., Ma, H.-Y., Klein, S. A. and Petch, J. C.: Using regime analysis to identify the contribution of clouds to surface temperature errors in weather and climate models, *Q. J. R. Meteorol. Soc.*, <https://doi.org/10.1002/qj.2603>, 2015.
- 585 Van Weverberg, K., Morcrette, C. J., Petch, J., Klein, S. A., Ma, H. Y., Zhang, C., Xie, S., Tang, Q., Gustafson Jr., W. I., Qian, Y., and Berg, L. K.: CAUSES: Attribution of surface radia- tion biases in NWP and climate models near the U.S. South- ern Great Plains, *J. Geophys. Res.-Atmos.*, 123, 3612–3644, <https://doi.org/10.1002/2017JD027188>, 2018.

- Xavier, P. K., Petch, J. C., Klingaman, N. P., Woolnough, S. J., Jiang, X., Waliser, D. E., Caian, M., Cole, J., Hagos, S. M., Han-nay, C., and Kim, D.: Vertical structure and physical processes of the Madden-Julian Oscillation: Biases and uncertainties at short range, *J. Geophys. Res.-Atmos.*, 120, 4749–4763, <https://doi.org/10.1002/2014JD022718>, 2015.
- Xu, W., and Rutledge, S. A.: Convective characteristics of the Madden–Julian Oscillation over the Central Indian Ocean observed by shipborne radar during DYNAMO, *J. Atmos. Sci.*, 71, 2859–2877, 2014.
- Wan, H., Rasch, P. J., Zhang, K., Qian, Y., Yan, H., and Zhao, C.: Short ensembles: An efficient method for discerning climate-relevant sensitivities in atmospheric general circulation models, *Geosci. Model Dev.*, 7(5), 1961–1977, <https://doi.org/10.5194/gmd-7-1961-2014>, 2014.
- 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.
- Williams, K. D., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., Jakob, C., Klein, S. A., Senior, C. A., and Williamson, D. L.: The Transpose-AMIP II Experiment and Its Application to the Understanding of Southern Ocean Cloud Biases in Climate Models, *J. Climate*, 26, 3258–3274, 2013.
- Xie, S., Zhang, M. H., Boyle, J. S., Cederwall, R. T., Potter, G. L., and Lin, W. Y.: Impact of a revised convective triggering mechanism on CAM2 model simulations: results from short-range weather forecasts, *J. Geophys. Res.*, 109, D14102, <https://doi.org/10.1029/2004JD004692>, 2004.
- Xie, S., Boyle, J. S., Klein, S. A., Liu, X. and Ghan, S.: Simulations of Arctic mixed-phase clouds in forecasts with CAM3 and AM2 for M-PACE, *J. Geophys. Res.*, 113, D04211, <https://doi.org/10.1029/2007JD009225>, 2008.
- Xie, S. C., McCoy, R. B., Klein, S. A., Cederwall, R. T., Wiscombe, W. J., Clothiaux, E. E., Gaustad, K. L., Golaz, J. C., Hall, S. D., Jensen, M. P., Johnson, K. L., Lin, Y. L., Long, C. N., Mather, J. H., McCord, R. A., McFarlane, S. A., Palanisamy, G., Shi, Y., and Turner, D. D. D.: Arm Climate Modeling Best Estimate Data a New Data Product for Climate Studies, *B. Am. Meteorol. Soc.*, 91, 13–20, <https://doi.org/10.1175/2009bams2891.1>, 2010.
- Xie, S., Ma, H.-Y., Boyle, J. S., Klein, S. A., and Zhang, Y.: On the correspondence between short- and long- timescale systematic errors in CAM4/CAM5 for the Years of Tropical Convection, *J. Climate*, 25, 7937–7955, 2012.
- [Xie, S., Lin, W., Rasch, P. J., Ma, P.-L., Neale, R., Larson, V. E., et al.: Understanding cloud and convective characteristics in version 1 of the E3SM atmosphere model. *Journal of Advances in Modeling Earth Systems*, 10, 2618–2644. <https://doi.org/10.1029/2018MS001350>, 2018.](https://doi.org/10.1029/2018MS001350)
- [Xie, S., Wang, Y.-C., Lin, W., Ma, H.-Y., Tang, Q., Tang, S., et al.: Improved Diurnal Cycle of Precipitation in E3SM with a Revised Convective Triggering Function. *Journal of Advances in Modeling Earth Systems*, 11. <https://doi.org/10.1029/2019MS001702>, 2019.](https://doi.org/10.1029/2019MS001702)
- Yang, F., Pan, H., Krueger, S. K., Moorthi, S., and Lord, S. J.: Evaluation of the NCEP Global Forecast System at the ARM SGP site, *Mon. Wea. Rev.*, 134, 3668–3690, 2006.
- Yanai, M., Esbensen, S., and Chu, J.-H.: Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets, *J. Atmos. Sci.*, 30, 611–627, 1973.

Zhang, C.: Madden-Julian Oscillation: Bridging weather and climate, *Bull. Am. Meteorol. Soc.*, 94, 1849-1870, <https://doi.org/10.1175/bams-d-12-00026.1>, 2013.

Zhang, C. and Ling, J.: Barrier effect of the Indo-Pacific Maritime Continent on the MJO: Perspectives from tracking MJO precipitation, *J. Climate*, 30, 3439–3459, 2017.

Zhang, K., Wan, H., Liu, X., Ghan, S. J., Kooperman, G. J., Ma, P.-L., Rasch, P. J., Neubauer, D., and Lohmann, U.: Technical note: On the use of nudging for aerosol–climate model intercomparison studies, *Atmos. Chem. Phys.*, 14, 8631–8645, 2014.

Zhang, M. Xie, S., Liu, X., Lin, W., Zhang K., Ma, H.-Y., Zheng, X., Zhang, Y.: Toward Understanding the Simulated Phase Partitioning of Arctic Single-Layer Mixed-Phase Clouds in E3SM. *Earth and Space Science. In Press.* doi: 10.1002/essoar.10502164.1, 2020

Zhang, Y., and Klein, S. A.: Mechanisms affecting the transition from shallow to deep convection over land: Inferences from observations of the diurnal cycle collected at the ARM Southern Great Plains Site, *J. Atmos. Sci.*, 67, 2943–2959, 2010.

Zhang, Y., Xie, S., Klein, S. A., Marchand, R., Kollias, P., Clothiaux, E. E., Lin, W., Johnson, K., Swales, D., Bodas-Salcedo, A., Tang, S., Haynes, J. M., Collis, S., Jensen, M., Bharadwaj, N., Hardin, J., and Isom, B.: The ARM cloud radar simulator for global climate models: Bridging field data and climate models, *Bull. Amer. Met. Soc.*, 99, 21–26, <https://doi.org/10.1175/BAMS-D-16-0258.1>, 2018.

Zhang, Y., Xie, S., Lin, W., Klein, S. A., Zelinka, M. D., Ma, P.-L., Rasch, P. J., Qian, Y., Tang, Q., and Ma, H.-Y.: Evaluation of clouds in version 1 of E3SM Atmosphere Model with satellite simulators, *J. Adv. Model. Earth Syst.*, 11, <https://doi.org/10.1029/2018MS001562>, 2019.

Zheng, X., Klein, S. A., Ma, H.-Y., Bogenschutz, P., Gettelman, A. and Larson, V. E.: Assessment of marine boundary layer cloud simulations in the Community Atmosphere Model with Cloud Layers Unified By Binormals and updated microphysics scheme based on ARM observations from the Azores, *J. Geophys. Res.-Atm.*, 121, 8472–8492, <https://doi.org/10.1002/2016JD025274>, 2016.

Zheng, X., Klein, S. A., Ma, H.-Y., Caldwell, P. M., Larson, V. E., Gettelman, A., and Bogenschutz, P.: A cloudy planetary boundary layer oscillation arising from the coupling of turbulence with precipitation in climate simulations, *J. Adv. Model. Earth Syst.*, 9, 1973–1993, <https://doi.org/10.1002/2017MS000993>, 2017.

650

655

Deleted: ¶

Table 1: Locations of model patch/site output associated with major field campaigns or DOE ARM sites.

Locations	Longitude	Latitude	grids
1.Niamey-1	357E-359E	14N-17N	10
2.Niamey-2	0-9E	5N-18N	120
3.DYNAMO	70E-83E	10S-9N	231
4.China-Shouxian	114E-119E	31N-34N	25
5.Darwin	129E-133E	14S-10S	20
6.Manus	146E-149E	4S-1S	12
7.Nauru	166E-169E	2S-1N	15
8.SHEBA	190E-220E	74N-78N	125
9.MAGIC-1	201E-214E	16N-25N	121
10.MAGIC-2	207E-232E	24N-30N	147
11.MAGIC-3	221E-243E	29N-36N	162
12.NSA	202E-206E	70N-73N	16
13.CARE	238E-240E	37N-39N	9
14.SGP	261E-264E	35N-38N	12
15.Vocals	272E-291E	23S-17S	112
16.Amazonia	296E-301E	13S-9S	25
17.Manaus	298E-302E	5S-1S	25
18.Azores-Graciosa	329E-334E	38N-41N	20
19.Barbados	291E-301E	10N-20N	108

- Deleted: ¶ ... [1]
- Formatted ... [2]
- Formatted ... [3]
- Formatted ... [4]
- Formatted ... [5]
- Formatted ... [6]
- Formatted ... [7]
- Formatted ... [8]
- Formatted ... [9]
- Formatted ... [10]
- Formatted ... [11]
- Formatted ... [12]
- Formatted ... [13]
- Formatted ... [14]
- Formatted ... [15]
- Formatted ... [16]
- Formatted ... [17]
- Formatted ... [18]
- Formatted ... [19]
- Formatted ... [20]
- Formatted ... [21]
- Formatted ... [22]
- Formatted ... [23]
- Formatted ... [24]
- Formatted ... [25]
- Formatted ... [26]
- Formatted ... [27]
- Formatted ... [28]
- Formatted ... [29]
- Formatted ... [30]
- Formatted ... [31]
- Formatted ... [32]
- Formatted ... [33]
- Formatted ... [34]
- Formatted ... [35]
- Formatted ... [36]
- Formatted ... [37]
- Formatted ... [38]
- Formatted ... [39]
- Formatted ... [40]
- Formatted ... [41]
- Formatted ... [42]
- Formatted ... [43]
- Formatted ... [44]
- Formatted ... [45]
- Formatted ... [46]
- Formatted ... [47]
- Formatted ... [48]
- Formatted ... [49]
- Formatted ... [50]
- Formatted ... [51]
- Formatted ... [52]
- Formatted ... [53]
- Formatted ... [54]

660

665

685 **Table 2: List of observation datasets.**

Formatted: Font: 10 pt

Observations	Analyzed period	References
GPCP Precipitation V1.2	1997-2012	Adler et al. 2003
CERES EBAF Radiation Edition 2.8	2000-2012	Loeb et al. 2009; Kato et al. 2013
ISCCP D2 Cloud	1997-2009	Rossow et al. 1999
ERA-Interim Reanalysis	1997-2012	Dee et al. 2011
ARMBE ARSCL Cloud	1997-2007	Xie et al. 2010; Clothiaux et al. 2000, 2001

Formatted: Font: 10 pt

690

695

700

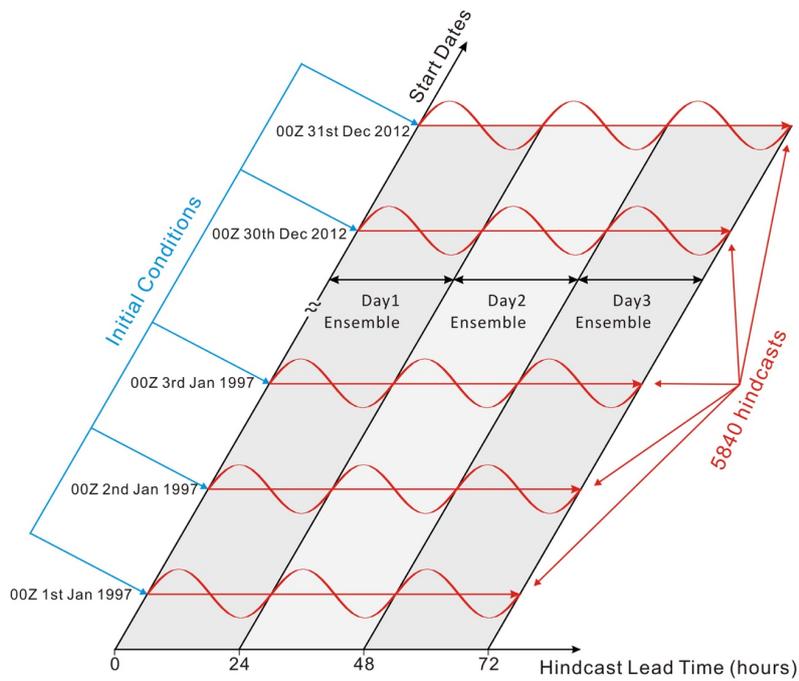
705

710 **Table 3. Pattern statistics of regression maps of selected fields onto the Niño 3.4 index between observations and model simulations (Day 2 hindcasts or AMIP).**

	<u>Spatial Correlation</u>		<u>RMSE</u>		<u>Normalized Spatial</u>	
	<u>Coefficient</u>				<u>Standard Deviation</u>	
	<u>Day 2</u>	<u>AMIP</u>	<u>Day 2</u>	<u>AMIP</u>	<u>Day 2</u>	<u>AMIP</u>
<u>Precipitation</u>	0.94	0.81	0.25	0.44	1.02	0.97
<u>JSCCP Total</u>	0.83	0.68	2.11	2.84	0.91	0.82
<u>Cloud Fraction</u>						
<u>MODIS Total</u>	0.85	0.72	2.12	2.74	1.03	0.94
<u>Cloud fraction</u>						
<u>Absorbed</u>						
<u>Shortwave</u>	0.87	0.74	2.70	3.62	1.08	1.02
<u>Radiation</u>						
<u>Outgoing</u>						
<u>Longwave</u>	0.95	0.82	1.97	3.34	1.09	0.97
<u>Radiation</u>						
<u>Net Surface</u>	0.79	0.61	4.42	6.19	0.90	0.98
<u>Fluxes</u>						
<u>TAUX</u>	0.91	0.72	3.49	5.84	0.80	0.81
<u>OMEGA500</u>	0.94	0.80	2.40	4.18	1.00	0.96
<u>U850</u>	0.99	0.88	0.14	0.40	1.08	1.08
<u>U200</u>	0.99	0.89	0.23	1.16	0.98	0.81

- Formatted: Font: Times New Roman, 10 pt
- Formatted: Font: Times New Roman, 10 pt
- Formatted: Font: 10 pt, Bold
- Formatted: Font: Times New Roman, 10 pt
- Formatted: Font: Times New Roman, 10 pt, Bold
- Formatted: Font: Bold
- Formatted: Centered
- Formatted Table
- Formatted: Centered
- Formatted: Font: Bold
- Formatted: Font: Bold
- Formatted: Centered
- Deleted: ¶
- Formatted: Font: 9 pt
- Deleted: ¶

715



720 **Figure 1: Schematic diagram for the multi-year hindcast procedure (modified from Ma et al. 2015) applied to series of three-day hindcasts from 1997 to 2012. Each hindcast is initialized with ERA Interim Reanalysis and the starting time is 00Z every day.**

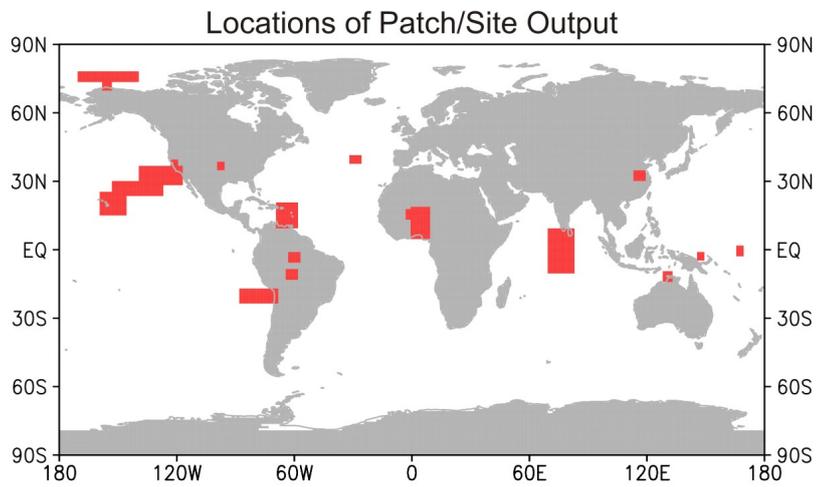


Figure 2: Locations of model patch/site output associated with major field campaigns or U.S. Department of Energy Atmospheric Radiation Measurement (ARM) sites. See Table 1 for detailed longitude and latitude information.

725

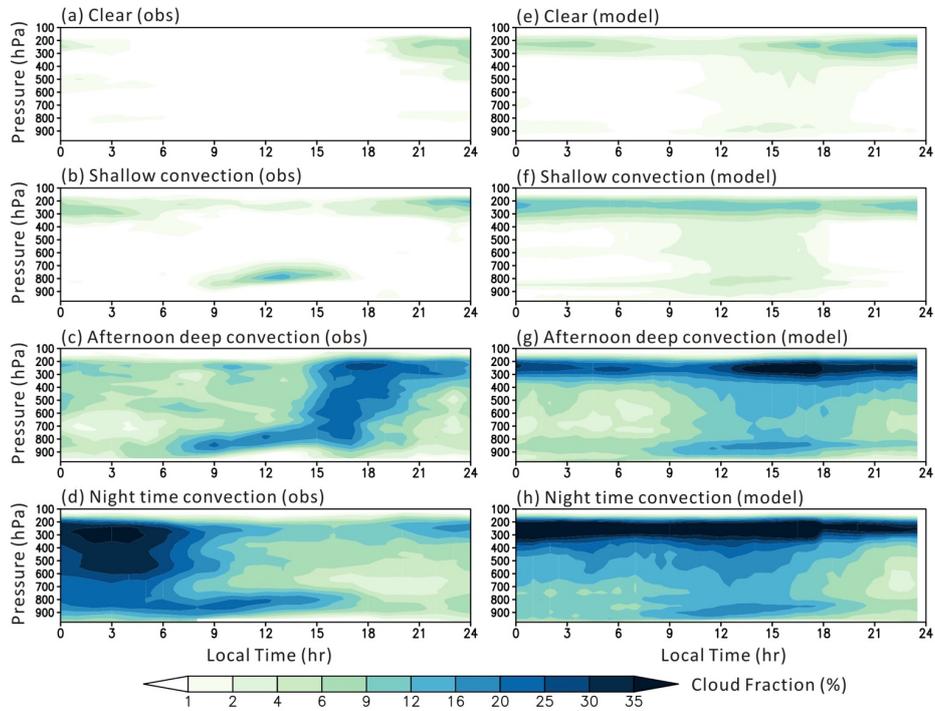
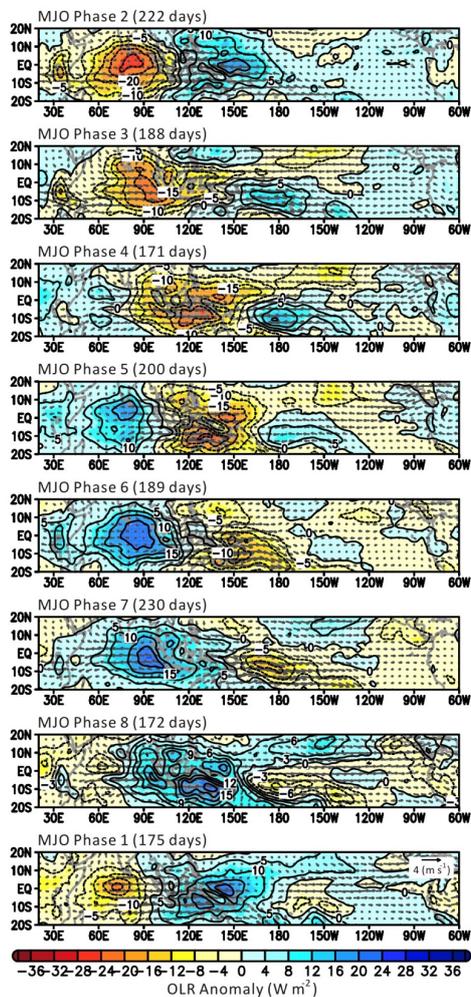


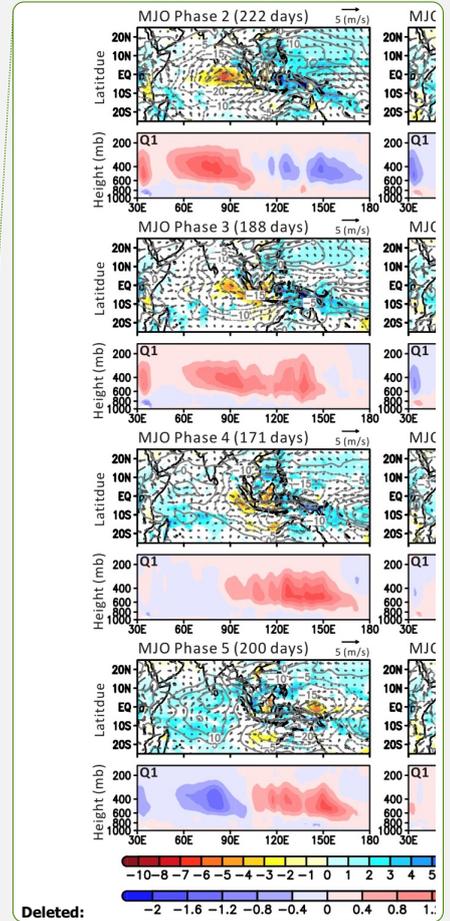
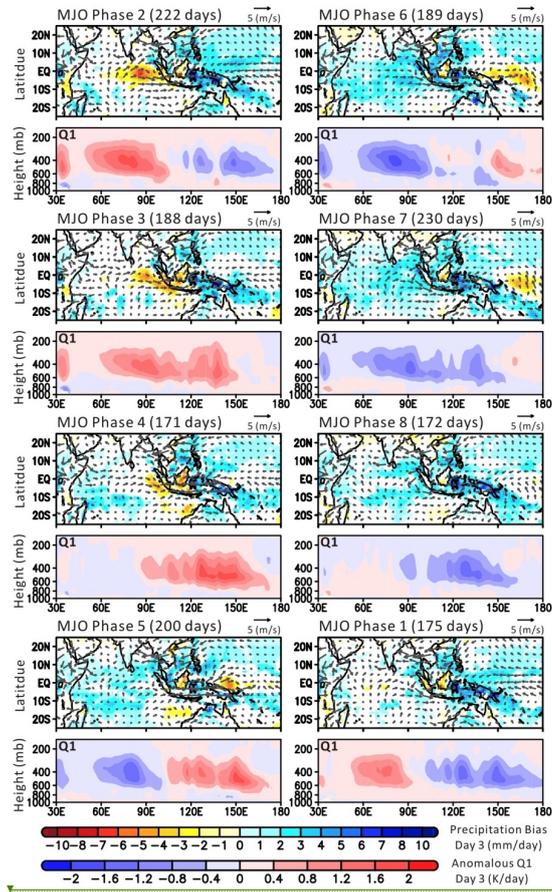
Figure 3: Diurnal cycle of cloud fraction composites (%) from May to August in the years of 1997 to 2007 from (a)–(d) ARMBE ARSCL and (e)–(h) Day 2 hindcasts for different convection regimes: (a)(e) clear-sky regime, (b)(f) fair-weather shallow cumulus regime, (c)(g) late-afternoon deep convection regime, and (d)(h) nighttime deep convection regime. Figures (a)–(d) were modified from Figure 3 in Zhang and Klein (2010).

Deleted: CAPT

730



735 Figure 4: Composites of November to April 20-100 day band-pass filtered NOAA Interpolated Outgoing Longwave Radiation anomalies (color shades and contours, $W m^{-2}$) and horizontal wind anomalies (vectors, $m s^{-1}$) at 850mb from ERA-Interim, as a function of the eight phases of the MJO (Wheeler and Hendon 2004). Number of days for composites are indicated on the top of each panel. Years of analysis are from 1997 to 2012.



Deleted: as well as the 20-100 day band-pass filtered NOAA Interpolated Outgoing Longwave Radiation anomalies (contours, $W m^{-2}$), ...

740

745

Figure 5: Composites of November to April precipitation bias (color shades, $mm day^{-1}$), 850 mb horizontal wind biases (vectors, $m s^{-1}$), and anomalous 20-100 day band-pass filtered Q1 vertical profiles (averaged over $10^{\circ}S$ to $10^{\circ}N$, $K day^{-1}$) from Day 3 hindcasts, as a function of the eight phases of the MJO (Wheeler and Hendon 2004). The number of days comprising each composite are indicated on the top of each panel. The observed precipitation and winds are from GPCP and ERA-Interim, respectively. Only precipitation biases that are significant at the 95% confidence level are shaded. The Q1 profiles are computed directly from model's tendency terms with all the diabatic processes.

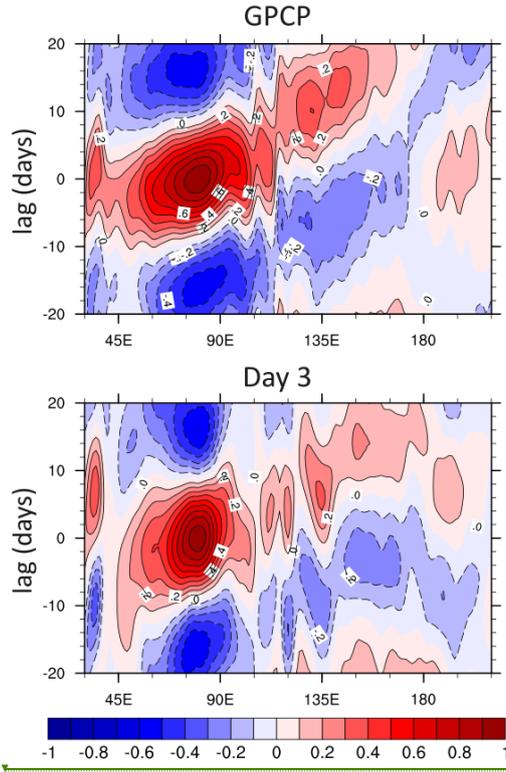
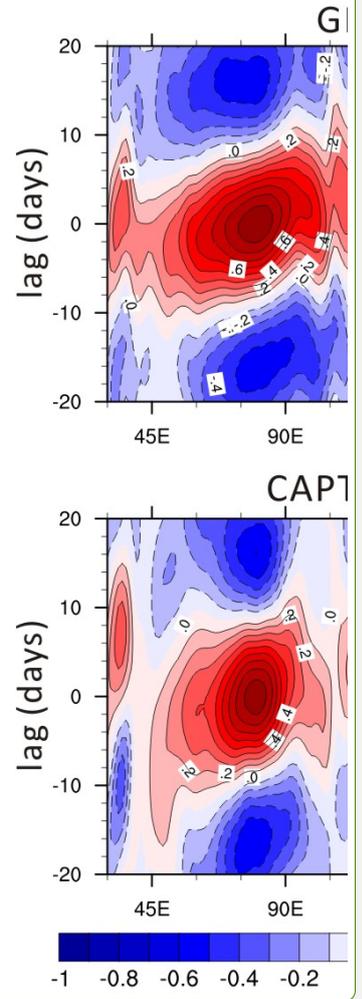


Figure 6: Time-longitude anomalous rainfall correlation Homvöller diagram along the Equator (10°S - 10°N) based on GPCP and Day3 hindcasts. The rainfall anomalies associated with the MJO are derived based on lag-correlation over an Indian Ocean box (75°-85°E; 5°S-5°N) for northern winter (November-April) of 1997 to 2012.



Deleted:
Deleted: CAP1

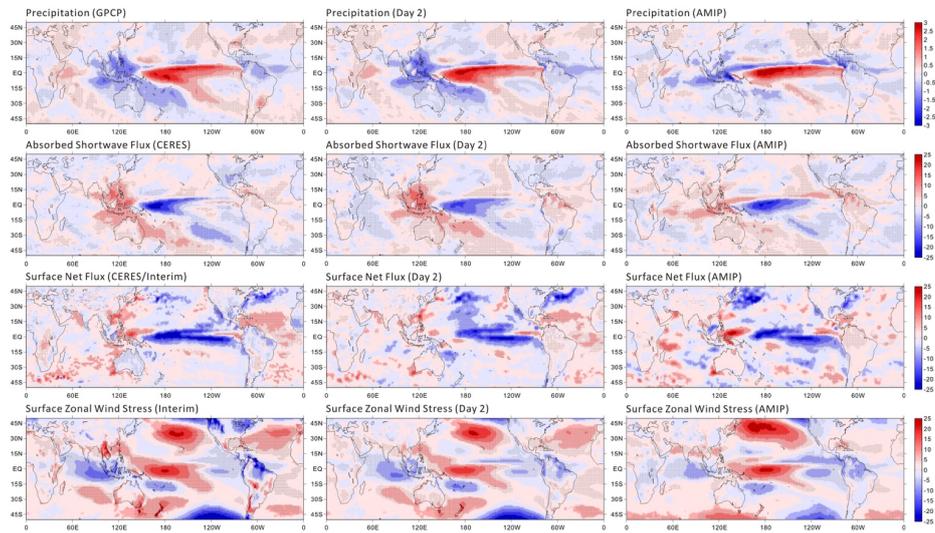
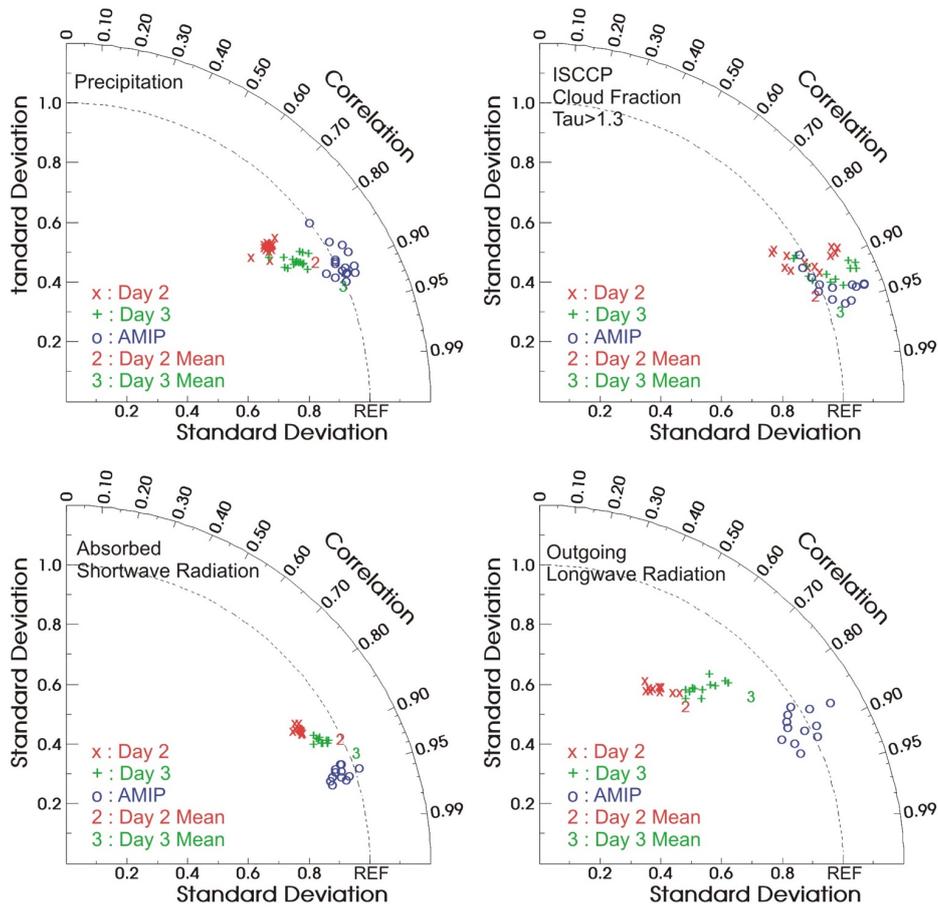
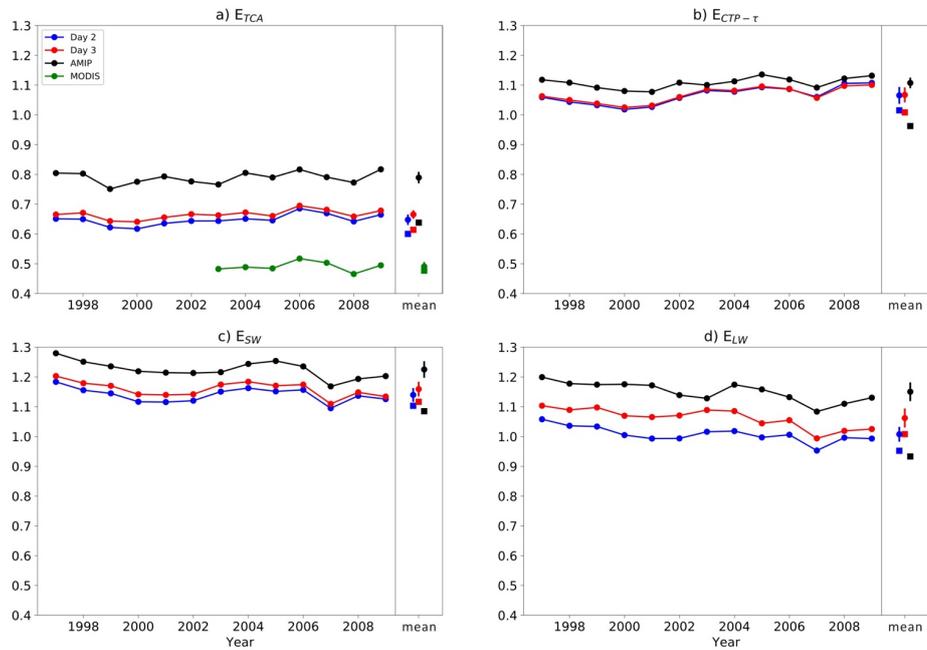


Figure 7: Regression maps of precipitation ($\text{mm day}^{-1} \text{K}^{-1}$), absorbed shortwave radiation ($\text{W m}^{-2} \text{K}^{-1}$), net surface heat flux ($\text{W m}^{-2} \text{K}^{-1}$), surface zonal wind stress ($\text{N m}^{-2} \text{K}^{-1}$) onto the Niño 3.4 index from observations (left panels), Day 2 hindcasts (middle panels), and AMIP simulation (right panels).

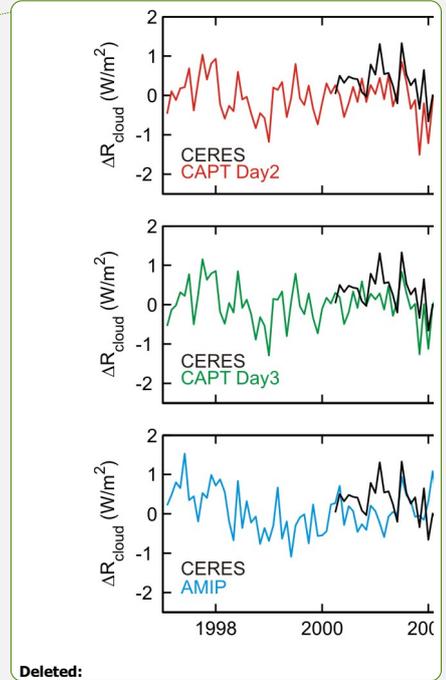
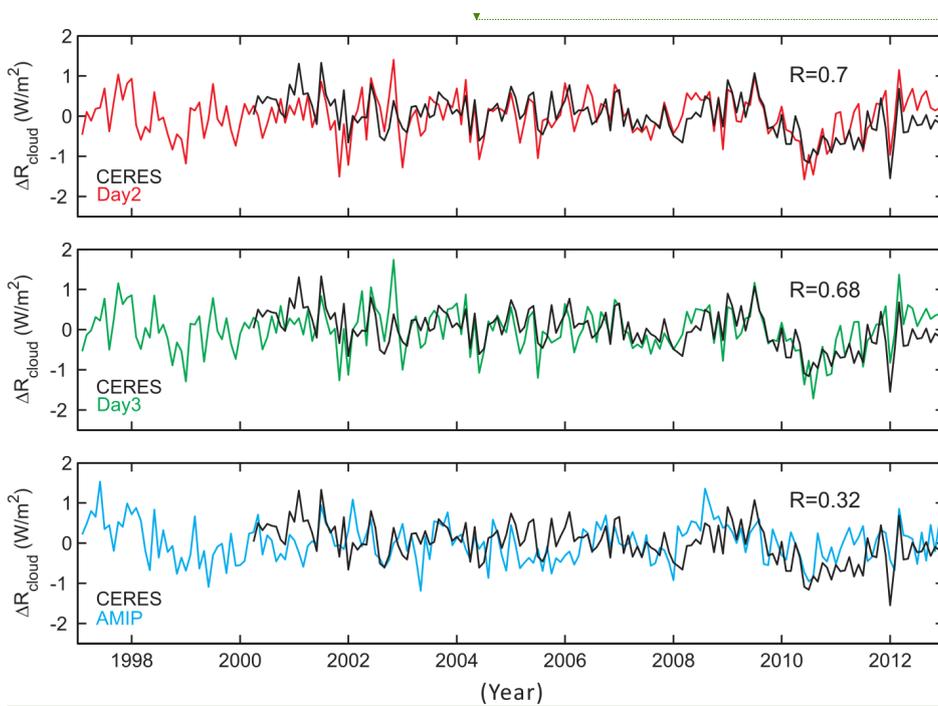
760



765 Figure 8: Spatial pattern statistics of the bias errors (with respect to observations) of annual mean precipitation, ISCCP total cloud
 fraction (with tau > 1.3), absorbed shortwave radiation, and outgoing longwave radiation from the multi-year CAM5 hindcasts and
 an AMIP simulation. These statistics are illustrated with a Taylor diagram (Taylor et al. 2001) that shows the level of agreement of
 a given field to a common reference. The reference fields (REF) are the correspondent multi-year mean bias errors from the AMIP
 simulation. Each "x", "+", or "o" represents the Day 2, Day 3 or AMIP annual mean bias for individual years between 1997 and
 770 2012 whenever the observations are available, and "2" and "3" represents the Day 2 and Day 3 hindcast mean biases averaged over
 all available years, respectively.



775 Figure 9: Cloud metrics as defined using the ISCCP simulator in Klein et al. (2013) for Day 2 and 3 hindcasts, as well as the AMIP
 780 simulation. These metrics are scalar measures of performance in simulating the space-time distribution of several cloud measures,
 with better performance indicated by smaller E values. E_{TCA} measures total cloud amount, and $E_{CTP-\tau}$ measures cloud-top pressure
 and optical depth in different categories of optically intermediate and thick clouds at high, middle, and low-levels of the atmosphere.
 E_{SW} and E_{LW} measure the impacts on top-of-atmosphere shortwave and longwave radiation in the same categories used for $E_{CTP-\tau}$,
 respectively. The markers with errorbars show the average and 1- σ interannual variation in these error metrics. The square symbols
 are the error metrics computed by comparing model and observed monthly-resolved climatological means. MODIS cloud amount
 for E_{TCA} is plotted as a measure of observational uncertainty.



785 **Figure 10:** Time series of monthly anomalies of global mean cloud radiative effect from CERES, Day 2 and 3 hindcasts, as well as
 790 from the AMIP simulation. The correlation coefficient between the model and observation is shown in the upper right portion of
 each panel.

▲ **Page 19: [1] Deleted** Microsoft Office User 7/8/20 2:27:00 PM

▼ **Page 19: [2] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [3] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [4] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [5] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [6] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [7] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [8] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [9] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [10] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [11] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [12] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [13] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [14] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [15] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [16] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Font: 10 pt

▲ **Page 19: [17] Formatted** Microsoft Office User 6/30/20 12:13:00 PM

Page 19: [18] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [19] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [20] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [21] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [22] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [23] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [24] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [25] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [26] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [27] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [28] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [29] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [30] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [31] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [32] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Page 19: [33] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [35] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [36] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [37] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [38] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [39] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [40] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [41] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [42] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [43] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [44] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [45] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [46] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [47] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [48] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [49] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [50] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

Font: 10 pt

▲
Page 19: [52] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [53] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [54] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [55] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [56] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [57] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [58] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [59] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [60] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [61] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [62] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [63] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [64] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [65] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [66] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲

Page 19: [68] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [69] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [70] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [71] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [72] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [73] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [74] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [75] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [76] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [77] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [78] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [79] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [80] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [81] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲
Page 19: [82] Formatted **Microsoft Office User** **6/30/20 12:13:00 PM**

Font: 10 pt

▲