

PROCESS-ORIENTED EVALUATION OF CLIMATE AND WEATHER FORECASTING MODELS

ERIC D. MALONEY, ANDREW GETTELMAN, YI MING, J. DAVID NEELIN, DANIEL BARRIE, ANNARITA MARIOTTI, C.-C. CHEN, DANIELLE R. B. COLEMAN, YI-HUNG KUO, BOHAR SINGH, H. ANNAMALAI, ALEXIS BERG, JAMES F. BOOTH, SUZANA J. CAMARGO, AIGUO DAI, ALEX GONZALEZ, JAN HAFNER, XIANAN JIANG, XIANWEN JING, DAEHYUN KIM, ARUN KUMAR, YUMIN MOON, CATHERINE M. NAUD, ADAM H. SOBEL, KENTAROH SUZUKI, FUCHANG WANG, JUNHONG WANG, ALLISON A. WING, XIAOBIAO XU, AND MING ZHAO

Outcomes of NOAA MAPP Model Diagnostics Task Force activities to promote process-oriented diagnosis of models to accelerate development are described.

Realistic climate and weather forecasting models grounded in sound physical principles are necessary to produce confidence in projections of future climate for the next century and predictions for days to seasons. However, global models continue to suffer from important and often common biases that impact their ability to provide reliable representations of weather and future climate. These include biases in the cold tongue and intertropical convergence zone regions (e.g., Li and Xie 2014; Grose et al. 2014), the structure of El Niño–Southern Oscillation (ENSO) sea surface temperature (SST) and precipitation anomalies (e.g., Bellenger et al. 2014; Grose et al. 2014), simulation of the Madden–Julian oscillation (MJO; Kim et al. 2014a; Hung et al. 2013; Jiang et al. 2015; Ahn et al. 2017), tropical monsoon precipitation and Indian Ocean processes (e.g., Sperber et al. 2013; Annamalai et al. 2017), the strength of the Atlantic meridional overturning circulation (AMOC; e.g., Wang et al. 2014), extratropical cyclone tracks (Zappa et al. 2013), tropical–extratropical teleconnections (e.g., Sheffield et al. 2013a,b; Henderson et al. 2017), and general interactions of clouds with the large-scale circulation (Stevens and Bony 2013), among others. Some aspects of simulations can often be improved,

but seemingly for the wrong reasons. For example, improving biases in model tropical intraseasonal variability often systematically degrades other aspects of the simulation like the mean state (Kim et al. 2011; Mapes and Neale 2011; Hannah and Maloney 2014). Model biases are rooted in imperfect parameterizations of unresolved processes.

The climate and weather forecasting communities have a long-standing and high interest in conducting process studies and applying process-oriented diagnostics (PODs) that are designed to inform parameterization improvements to address these long-standing model biases (e.g., Eyring et al. 2019). A POD characterizes a specific physical process or emergent behavior that is hypothesized to be related to the ability to simulate an observed phenomenon. An example of an observed phenomenon is the intraseasonal variability of tropical convection, as could be measured by an index or a power spectra of precipitation variance in the tropics. PODs representing the sensitivity of atmospheric convection to free-tropospheric humidity demonstrate a strong coupling between convection and moisture on daily time scales, which are also able to discern models with strong and weak intraseasonal variability (e.g., Kim

et al. 2014a). Evaluating new model configurations against observations can determine whether a particular process is well represented, ensure that models produce the right answers for the right reasons, and identify gaps in the understanding of phenomena. Process oriented “metrics” are scalar quantities that can be derived from some PODs.

A key need is incorporation of PODs into standard diagnostics packages that are applied to development versions of models, allowing the diagnosis to be repeatable across multiple model versions and generations. A significant barrier is the lack of a mechanism for getting community-developed PODs into the modeling center development process. This paper describes outcomes of activities by the National Oceanic and Atmospheric Administration (NOAA) Modeling, Analysis, Prediction, and Projections program (MAPP) Model Diagnostics Task Force (MDTF) to promote development of PODs and their application to climate and weather prediction models. A product of the first phase of the MDTF (2015–18) is the creation of select demonstrative PODs and a modeler-oriented open-source software framework that is portable, extensible, and open for contribution of PODs from the community. The framework is conceived to be compatible with and complementary to other efforts such as European Earth System Model Bias Reduction and Assessing Abrupt Climate Change (EMBRACE) project/Earth System Model eValuation Tool (ESMValTool) and Coordinated Set of Model Evaluation Capabilities (CMEC) that

use open-source software packages for multimodel evaluation. Because most other efforts have thus far largely emphasized basic performance metrics for models, the MDTF effort described here is complementary and advantageous to these other efforts as they expand their POD capabilities.

This paper is the centerpiece of an American Meteorological Society (AMS) special collection devoted to process-oriented evaluation of climate and Earth system models (https://journals.ametsoc.org/topic/process_oriented_model_diagnostics). Other articles in this collection describe the scientific basis for individual PODs. This centerpiece paper provides a summary of these individual diagnostics and others being developed by the MDTF, including development of a software framework to entrain these PODs to allow ease of use by modeling centers. The second section describes existing institutional efforts and needs, including details on the MDTF and modeling center perspectives. Existing community efforts at process-oriented diagnosis are then discussed. The fourth section provides examples of key PODs and metrics developed by the MDTF during its first three years and plans for expansion of this diagnostic set. The integrative open-source software framework to entrain these diagnostics is then described. The last section provides a summary and a path forward for PODs.

THE NOAA MAPP MDTF AND MODELING CENTER NEEDS. *Brief summary of MDTF.* In 2015 and 2018, NOAA's MAPP program solicited projects

AFFILIATIONS: MALONEY AND SINGH—Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado; GETTELMAN, CHEN, AND COLEMAN—National Center for Atmospheric Research, Boulder, Colorado; MING AND ZHAO—Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey; NEELIN AND KUO—Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, Los Angeles, California; BARRIE AND MARIOTTI—National Oceanic and Atmospheric Administration/Climate Program Office, Silver Spring, Maryland; ANNAMALAI AND HAFNER—International Pacific Research Center/Department of Oceanography, University of Hawai'i at Mānoa, Honolulu, Hawaii; BERG—Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey; BOOTH—City College of the City University of New York, New York, New York; CAMARGO—Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York; DAI AND J. WANG—Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, Albany, New York; GONZALEZ AND JIANG—Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, Los Angeles, California; JING AND SUZUKI—Atmosphere and Ocean Research Institute, The University of Tokyo, Kashiwa, Japan; KIM AND MOON—

Department of Atmospheric Sciences, University of Washington, Seattle, Washington; KUMAR—National Oceanic and Atmospheric Administration/Climate Prediction Center, College Park, Maryland; NAUD—Columbia University, New York, New York; SOBEL—Lamont-Doherty Earth Observatory, Columbia University, Palisades, and Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York; F. WANG—Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida, and Institute of Oceanography, Shanghai Jiao Tong University, Shanghai, China; WING—Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, Florida; Xu—Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, Florida

CORRESPONDING AUTHOR: Eric D. Maloney, emaloneyn@atmos.colostate.edu

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/BAMS-D-18-0042.1

In final form 23 April 2019

©2019 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

to develop PODs for model development. These funded projects and the investigators leading them ultimately constituted the NOAA MDTF. At the time of the initial proposal solicitation, the global modeling community was between cycles of the Coupled Model Intercomparison Project (CMIP), and modeling centers were actively moving on from their CMIP5-class models toward developing and testing their CMIP6-class models. Performance evaluations and analyses of CMIP5 class models were becoming less useful for informing next-generation model development activities.

NOAA's CMIP5 Task Force (2011–14) discussed the idea of expanding upon nascent POD development efforts in the field, such as the Working Group on Numerical Experimentation (WGNE) MJO Task Force's initial work on PODs (Wheeler et al. 2013; Kim et al. 2014a, 2015), and others described below. MDTF activities were designed to build on analyses performed by the CMIP5 Task Force and others by providing an opportunity for nonfederal scientists to contribute to model development activities at the NOAA Geophysical Fluid Dynamics Laboratory (GFDL) and the National Science Foundation National Center for Atmospheric Research (NCAR). The MDTF enabled nonfederal participating scientists to gain access to development versions of the next-generation models and work with modeling center staff toward developing PODs that could provide physical insight into the sources of model bias. The second phase of this activity, which began in 2018, is leveraging the CMIP6 experiments for further diagnostic development and model evaluation.

The MDTF has engaged over 50 scientists from 6 laboratories and operational centers and 15 academic institutions. During its first phase (2015–18), the MDTF was led by Eric Maloney (Colorado State University), and co-led by Yi Ming (GFDL), Andrew Gettelman (NCAR), David Neelin (UCLA), and Aiguo Dai (University at Albany). MDTF activities have included two major thrusts: 1) coordinating and supporting community development of diagnostics and metrics for a variety of physical systems and modeling and process areas targeting known model biases, and 2) designing a software framework useable at GFDL and NCAR that is flexible enough to incorporate PODs from disparate community efforts that may be written in diverse coding languages. PODs developed or under development for the first task include

- cloud microphysical processes;
- tropical and extratropical cyclones;
- ENSO teleconnections and atmospheric dynamics;

- land–atmosphere interactions;
- MJO moisture, convection, and radiative processes;
- precipitation diurnal cycle;
- AMOC;
- Arctic sea ice;
- lake-effect processes;
- North American monsoon;
- radiative forcing and cloud–circulation feedbacks; and
- temperature and precipitation extremes.

These diverse, somewhat eclectic focal areas were determined by the submitted competitive proposals that emerged successfully from the MAPP panel reviews. They also reflect key model biases that impact climate and climate variability. While modeling centers are aware of biases in all of these areas, the specific PODs developed by the MDTF to address these biases are unique to our knowledge. Continued development of these PODs is supported through MAPP proposal solicitations. This competitive solicitation model for advancing the process-oriented activity encourages a bottom-up design of the diagnostics framework and is driven by organic, mutually beneficial interactions between modeling center and academic scientists and staff as opposed to top-down engineered engagements. NOAA has provided a funding commitment to this activity going forward, and the new leadership team of the MDTF led by David Neelin of UCLA has made explicit plans to increase engagement with complementary efforts such as those at the Program for Climate Model Diagnosis and Intercomparison (PCMDI), as well as outreach to the community through sessions and presentations at conferences and through participation in MDTF teleconferences. These efforts will help navigate the political and technical challenges in pursuing our POD concept going forward.

Modeling center perspectives. MDTF activities are designed to support model development and the diagnostic workflow at major modeling centers. Centers typically have a workflow containing a package of diagnostic comparisons with model output, which enables rapid analysis of many aspects of a model run. The method typically is for a large package to be constructed to generate diagnostics that many different developers may want to look at, enabling a multivariate and multidisciplinary approach to model evaluation. One major MDTF goal is to provide an extensible mechanism for community scientists to contribute PODs that can be integrated into the workflow at modeling centers.

Traditionally, diagnostics for climate models are based on monthly mean statistics and climatologies. Increasingly, models are being analyzed in more detail against observations of specific processes, and the MDTF is approaching PODs in this spirit. The closer to a model process the observations and evaluation are, the better the ability to constrain the process and hence provide a guide to parameterization improvement. For a simple example: cloud radiative effects at the top of the atmosphere are a nonunique function of cloud microphysical properties (drop number and liquid water path). Thus, constraining radiative effects of clouds is better done in conjunction with detailed observations of cloud microphysics than with just radiative fluxes.

EXISTING PROCESS-ORIENTED DIAGNOSTIC EFFORTS. The MDTF PODs effort is inspired by, builds upon, and in many cases is complementary to prior and existing community efforts at model diagnosis. Such existing efforts that have influenced the MDTF are described here, although this list is likely not exhaustive. Individual modeling centers also have their own diagnostics suites that perform diagnosis in a similar spirit, but for individual models. The MDTF developed differently from many of the efforts described here, in that it started from an initially small group of POD developers in collaboration with two modeling centers with a focus on model improvement rather than general model evaluation, although other efforts have recently expanded emphasis on process evaluation. Community developers and modeling centers worked together from day 1 to craft a mechanism that was as flexible and useful as possible for entrainment of community PODs into center development packages.

The WGNE MJO task force. The WGNE MJO task force (www.wmo.int/pages/prog/arep/wwrp/new/MJO_Task_Force_index.html) was developed in 2010 under the Years of Tropical Convection (Waliser et al. 2012) with an explicit goal to foster improved MJO simulations in global models (Wheeler et al. 2013). One goal of the MJO task force is to promote PODs/metrics for the MJO that facilitate model improvement. Several PODs were developed including an assessment of the sensitivity of tropical convection to lower-free-tropospheric moisture (Kim et al. 2014a), normalized gross moist stability (Benedict et al. 2014; Hannah and Maloney 2014; Jiang et al. 2015), and the strength of cloud–radiative feedbacks (Kim et al. 2015). The POD work of the MJO task force was an early inspiration behind the efforts of the

NOAA MDTF, and a mutual benefit is that the framework developed by the MDTF may allow broader community dissemination of PODs that foster the improvement of MJO simulations. The MDTF held a joint meeting with the MJO Task Force at the WGNE Systematic Errors Workshop in Montreal in 2017 to initiate collaboration on diagnostic efforts.

European EMBRACE project/ESMValTool. The European Union–funded EMBRACE project has developed a package called the ESMValTool (Eyring et al. 2016a,b). This tool was originally developed from the Chemistry–Climate Model Validation Activity (CCMVal) diagnostic package (Gettelman et al. 2012). The ESMValTool is a flexible and community-oriented diagnostic framework that uses standard model files as input, similar to the MDTF tool described below, and provides a structured set of diagnostic output plots. The spirit of the tool is similar to that of the MDTF, and because the ESMValTool uses similar inputs and a similar structure, diagnostics coded for one tool [ESMValTool is largely in Python, derived from NCAR Command Language (NCL) code] should be applicable in the other. The ESMValTool is increasingly incorporating process-level information, motivated by a cited community need to bring more process-level information to bear on model evaluation (Eyring et al. 2019).

The Coordinated Set of Model Evaluation Capabilities. CMEC is an open-source package incorporating the PCMDI Metrics Package (PMP), the International Land Modeling Benchmarking Project tool (ILAMB), and the parallel Toolkit for Extreme Climate Analysis (TECA). As described in Gleckler et al. (2016), the PMP currently provides an open-source package based on Python and Ultrascale Visualization Climate Data Analysis Tools (UV-CDAT; e.g., Santos et al. 2013; Williams et al. 2016) that compares climate models to observations using a set of basic performance metrics and statistics. The PMP development team is open to working with community users to entrain more diagnostics into the package, and future releases plan to incorporate more extensive model evaluation based on emergent constraints and process-level diagnosis. ILAMB provides a framework for evaluating land surface models that includes benchmarking the realism of specific processes that allow good land surface performance (Luo et al. 2012). TECA is a parallelized software package in C++ designed to detect extreme climate events in model fields such as tropical and extratropical cyclones and atmospheric rivers (Prabhat et al. 2012). As of

late 2018, the MDTF has already entrained PCMDI into explicit discussions in teleconferences and at meetings to assure compatibility and complementarity of diagnostics efforts.

GEWEX Process Evaluation Study. The Global Energy and Water Exchanges (GEWEX) Process Evaluation Study (PROES) has been launched as a GEWEX-wide community effort that aims to advance understanding of key climate processes and their representations in weather prediction and global climate models (Stephens et al. 2015). In particular, GEWEX-PROES is intended to exploit multiple satellite observations to diagnose the processes relevant to water and energy balances and thereby to advance the models at a fundamental level. The aims of GEWEX-PROES are to better understand Earth's energy and water cycles, diagnose reasons for model bias in simulating these cycles, and facilitate improved representation of processes underlying the energy and water cycles in models (Stephens et al. 2015). Although proposed as a GEWEX-based effort, the GEWEX-PROES also seeks strong connection with other efforts in the climate study community beyond GEWEX, such as the World Climate Research Program Grand Challenges, CMIP, the Cloud Feedback Model Intercomparison Project (CFMIP), the Observations for Model Intercomparisons Project (obs4MIPs), and WGNE. GEWEX-PROES is composed of projects including three main ingredients: 1) collection of datasets that allow for process diagnosis, 2) development of tools or methodologies constructed from data that enable process evaluation, and 3) design and execution of model simulations that will be analyzed with the diagnostic methodologies applied.

CFMIP Diagnostic Codes Catalogue. The CFMIP Diagnostic Codes Catalogue is a showcase of metrics and diagnostics on cloud-related processes to evaluate their representations in global climate models (Tsushima et al. 2017). It is intended to integrate existing methodologies for diagnosing key aspects of the cloud–climate feedback developed by members of the CFMIP community. This community effort assembles the metrics and diagnostics in the form of code repositories that allow open access. The result helps facilitate use of the metrics/diagnostics by the wider climate community and also encourages additional diagnostics to be included in the catalogue as long as they are documented in peer-reviewed publications and source code is provided. Given that the effort emerges from CFMIP, the catalogue is intended to serve as a shared toolkit that enhances analysis of

output from CFMIP and CMIP6 model experiments with a particular focus on clouds. The European Union Cloud Intercomparison, Process Study and Evaluation Project (www.euclipse.eu/index.html) produced one such set of cloud diagnostics entrained into the CFMIP Diagnostic Codes Catalogue.

EXAMPLES OF PROCESS-ORIENTED DIAGNOSTICS. This section describes efforts to develop PODs for climate model evaluation by the NOAA MAPP MDTF during its first three years. References are provided where expanded science description for these diagnostics can be found. We are not to our knowledge aware that the specific PODs presented here have been previously employed by modeling centers. This diagnostic set will be supplemented with PODs developed by new investigators entrained into the task force during the 2018 solicitation, and hence is continually evolving. The PODs currently entrained into the MDTF software framework at the time of this writing are noted at the website linked in the next section. This approach thus differs from products provided by bodies such as the U.S. Climate Variability and Predictability Program (CLIVAR) MJO Working Group (Waliser et al. 2009), which provided a package more or less frozen at the time of publication. We also stress that the software framework described below was developed to easily incorporate other PODs contributed by the community in addition to those described here. For example, approaches that allow the spatial and temporal dependence of two geophysical fields to be assessed in increasingly sophisticated ways, including more robust assessments of causality, might be employed in future diagnostics contributed by the MDTF and broader community (e.g., Livina et al. 2008; Moise and Delage 2011; Gilleland et al. 2016; Abatan et al. 2018; McGraw and Barnes 2018).

Convective transition statistics. Figure 1b shows an example of PODs for the transition between nonprecipitating and precipitating regimes for the tropics, where deep convection dominates precipitation production. A basic set of diagnostics is shown for precipitation dependence on measures of the water vapor–temperature environment, evaluated at short time scales comparable to those at which parameterized convection acts (Neelin et al. 2009; Schiro et al. 2016). Observations (Kuo et al. 2018) and an example model (GFDL) are shown with, left to right, panels for precipitation conditionally averaged as a function of column water vapor (CWV) for various values of troposphere-average temperature (colors),

probability of precipitation (exceeding a threshold of 0.25 mm h^{-1}), and the probability density function (PDF) of CWV and of CWV for precipitating points. The sharp pickup of precipitation and probability of precipitation above a threshold in CWV for each temperature provides a measure of conditional instability, as it occurs in each model. In an advanced-diagnostics module of this POD, the location of the sharp pickup is identified and compared to observations for each model, and the different temperatures are collapsed onto a dependence that is very similar in observations for the pickup in conditional-average precipitation, probability of precipitation, and PDF of water vapor for precipitating points. The GFDL model provides an example that reproduces these observational measures fairly well—other models can exhibit considerable spread. Model representations of entrainment can be a significant factor in correctly obtaining the water vapor–temperature dependence of the transition, although microphysics and other aspects of the convective parameterization can also play a role (Holloway and Neelin 2009; Sahany et al. 2012, 2014; Kuo et al. 2017; Schiro et al. 2018).

MJO teleconnection biases. Henderson et al. (2017) documented reasons for MJO midlatitude teleconnection errors in CMIP5 models. Since MJO teleconnections have significant impacts on atmospheric rivers, blocking, and other extreme events in the midlatitudes, teleconnection errors in models have important implications for the subseasonal prediction

of midlatitude weather extremes (e.g., Henderson et al. 2016; Mundhenk et al. 2018; Baggett et al. 2017). Henderson et al. (2017) developed diagnostics linking teleconnection biases to biases in the position and extent of the North Pacific jet.

Figure 2 (from Henderson et al. 2017) contains two panels, each having MJO teleconnection performance during December–February on the y axis. In Fig. 2a, the x axis represents an MJO skill metric. While Fig. 2a shows a relationship between MJO skill and teleconnection performance, even models with a good MJO can have poor teleconnection performance. For only the models assessed to have a sufficiently good MJO, Fig. 2b assesses the relationship between teleconnection performance and biases in the North Pacific zonal flow. Plus signs are a measure of the total root-mean-square (RMS) error of the 250-hPa zonal flow over the region 15° – 60°N , 110°E – 120°W , and the filled circle provides a measure of the RMS error in the length of the North Pacific subtropical jet. Both measures are correlated with MJO teleconnection performance, although biases in the jet provides a somewhat better metric ($r = -0.7$ versus -0.6 for the total RMS). Subsequent analysis showed that models with a jet that extends too far east tend to have degraded teleconnection performance. Model physics appears to play a key role in the extent of the Pacific jet, as was demonstrated by Neelin et al. (2013) in diagnosing projected California precipitation changes between CMIP3 and CMIP5 models into the late twenty-first century.

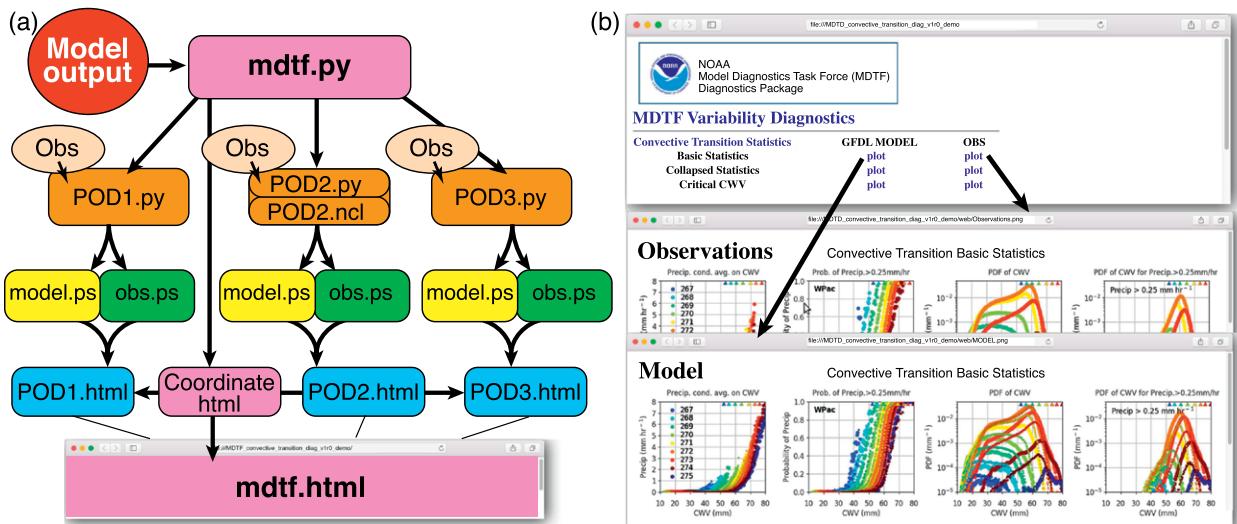


FIG. 1. Schematic of MDTF PODs framework (mdtf.py). (a) Workflow under the API. The model output is that of a candidate version that the development team wants to compare to observations under the various diagnostics. The observations for each POD are supplied in analyzed form within the module. (b) Example web page (edited for brevity and clarity) for one POD, described in the “Convective transition statistics” subsection.

MJO propagation and amplitude diagnostics. A POD for MJO propagation is motivated by findings that the horizontal advection of lower-tropospheric moisture plays a critical role in eastward propagation of the winter MJO (e.g., Maloney 2009; Kiranmayi and Maloney 2011; Sobel et al. 2014; Chikira 2014; Kim et al. 2014b; Adames and Wallace 2015; Jiang 2017; Kim et al. 2017; Jiang et al. 2018). Under this process, the spatial distribution of the winter mean lower-tropospheric moisture over the equatorial Indo-Pacific region (Fig. 3a) is critically important for moistening (drying) to the east (west) of MJO convection through advection by MJO anomalous winds. The critical role of the mean lower-tropospheric moisture pattern for MJO eastward propagation is supported based on multimodel simulations from the MJO Task Force/GEWEX Global Atmospheric System Studies (GASS) MJO model comparison project (Jiang 2017; Gonzalez and Jiang 2017). In particular, model skill in representing the 900–650-hPa mean moisture pattern over the Maritime Continent region (red rectangle in Fig. 3a) exhibits a high correlation (about 0.8) with MJO eastward propagation skill across about 25 GCM simulations (Fig. 3b).

The convective moisture adjustment time scale τ , defined by the ratio of intraseasonal perturbations of precipitable water and surface precipitation (e.g., Bretherton et al. 2004; Peters and Neelin 2006; Sobel and Maloney 2013), is selected as a metric for model MJO amplitude, which is motivated by the high anticorrelation (-0.72) between τ and MJO amplitude across multimodel simulations in Jiang et al. (2016, Fig. 3c). Parameter τ depicts how rapidly precipitation must occur to remove excess column water vapor, or alternately the efficiency of surface precipitation generation per unit column water vapor anomaly, and is highly relevant to the convective onset diagnostics described above.

AMOC structure diagnostic. The AMOC, with large temperature (T) and salinity (S) differences between the northward-flowing upper limb and southward-flowing lower limbs, is responsible for large oceanic transport of heat and freshwater, thus playing a fundamental role in establishing the mean state and the variability of the climate system. The focus on diagnosing AMOC in climate models has been

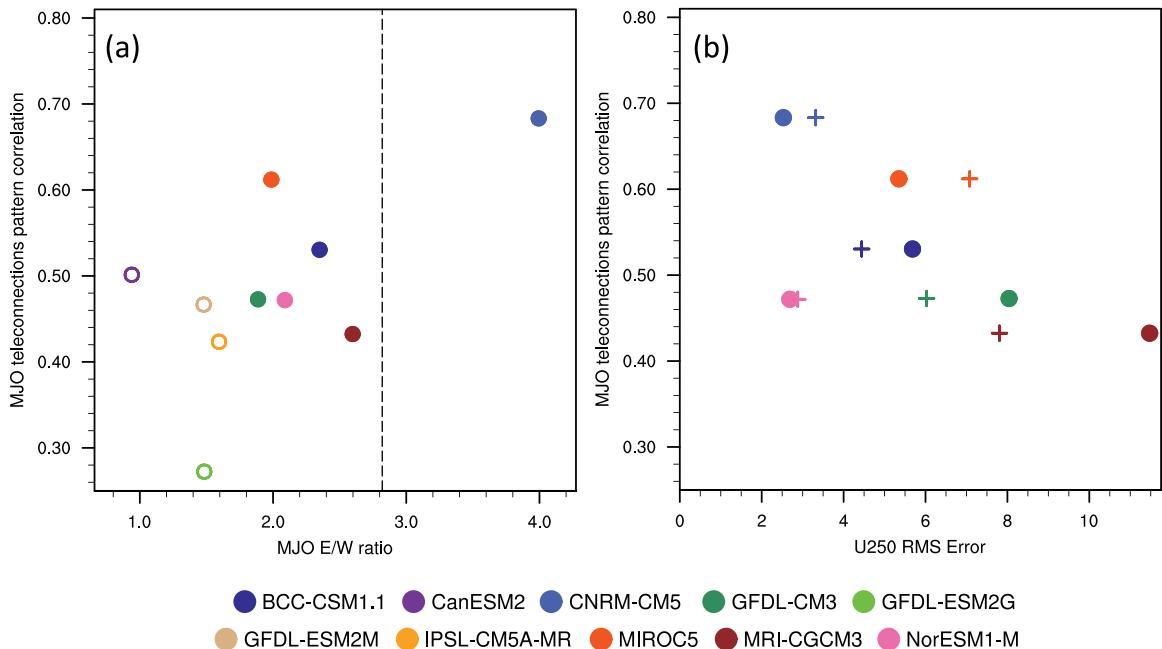


FIG. 2. From Henderson et al. (2017). December–February teleconnection performance averaged across all MJO phases (y axis) vs (a) MJO skill (MJO E/W ratio) and (b) the RMS error of the 250-hPa December–February zonal wind. The observed East/West (E/W) precipitation (GPCP) ratio is provided (dashed line, left panel). The MJO skill is derived as the ratio of eastward to westward power of equatorial precipitation in the 30–60-day, zonal wavenumber-1–3 band (e.g., Ahn et al. 2017). Teleconnection performance was determined as pattern correlation of North Pacific and North America (15°S–80°N, 130°E–60°W) MJO composite 250-hPa geopotential height anomalies between CMIP models and ERA-Interim averaged over all MJO phases. In (b), the crosses show the model 250-hPa zonal wind RMS error over the full Pacific basin, while the closed circles indicate the longitudinal RMS error of the subtropical jet. See the text and Henderson et al. (2017) for more detailed explanations.

mostly on the magnitude or the volume transport of the circulation (e.g., Cheng et al. 2013; Collins et al. 2013) and the role of water properties has been less quantified. In an AMOC structure POD, we examine the water properties of the AMOC by projecting the meridional transport on T - S space, then use the transport-weighted T and S as the characteristic T and S of the upper and lower limbs. The results show that the modeled AMOC in CMIP5 historical simulations has a smaller temperature difference between the upper and lower limbs compared to the results of a high-resolution ocean simulation that well represents the observed AMOC structure and the heat/freshwater transports (Xu et al. 2016). The model spread of time-mean heat transport among different CMIP5 simulations is significantly correlated with

the volume transport/magnitude of the AMOC, not with the temperature difference between the upper and lower limbs (Figs. 4a,b). The smaller temperature difference, however, is the main reason for a weaker multimodel mean heat transport in CMIP5 models (Fig. 4b). However, the averaged freshwater transport in CMIP5 models is similar to high-resolution simulation and observations, and the spread of freshwater transports in different CMIP5 models is significantly correlated with the salinity difference between the upper and lower AMOC limbs (Figs. 4c,d).

ENSO-precipitation diagnostics along the equatorial Pacific. Sustained research in theory, numerical modeling and observations has demonstrated that SST anomalies associated with ENSO serve as the leading

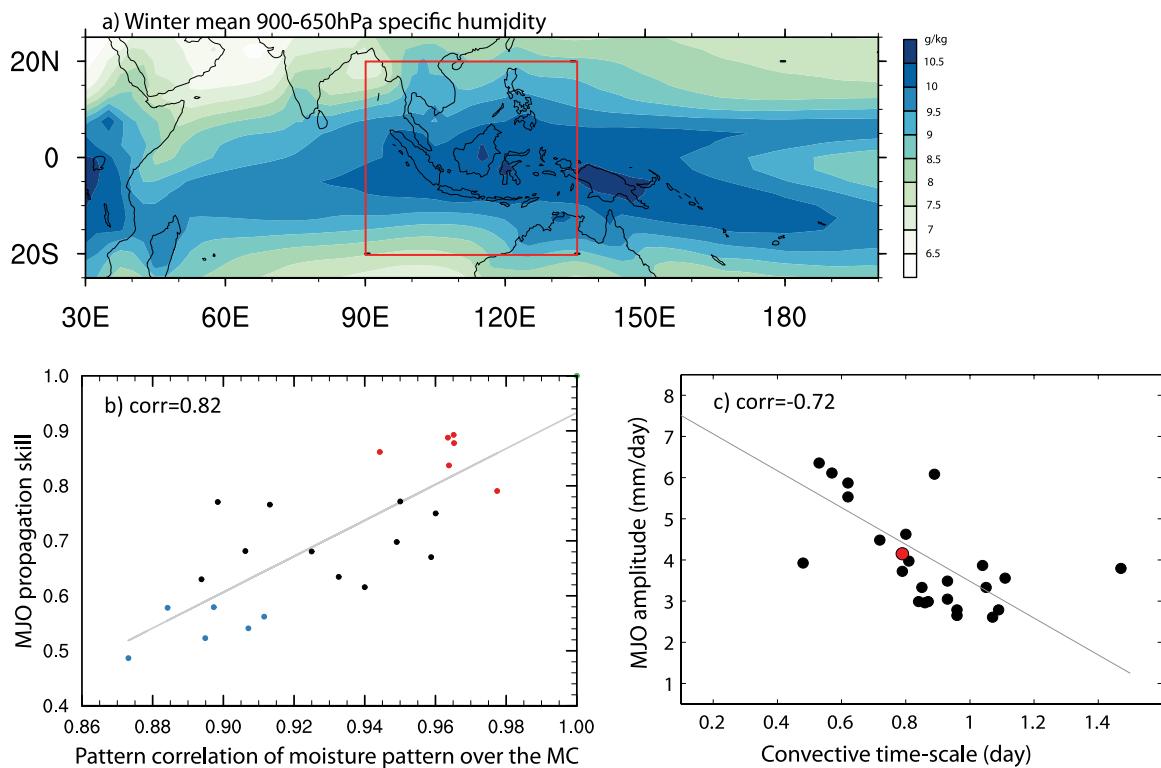


FIG. 3. (a) Winter (November–April) mean 650–900-hPa specific humidity based on ERA-Interim. **(b)** Scatterplot of model skill for eastward propagation of the MJO vs model skill of the mean 650–900-hPa moisture pattern over the Maritime Continent [red rectangle in (a)] based on multimodel simulations from the MJO Task Force/GASS project. Model MJO propagation skill is defined by the pattern correlation of Hovmöller diagrams of model simulated rainfall anomalies associated with the MJO against its observed counterpart following Jiang et al. (2015). Red (blue) dots denote good (poor) MJO models. **(c)** Scatterplot of MJO amplitude and model convective moisture adjustment time scale in models (black dots) and observations (red dots). The MJO amplitude in each model is defined by the standard deviation of 20–100-day bandpass-filtered rainfall over the Indian Ocean (10°S – 10°N , 75° – 85°E) during winter. Convective time scale in a model is defined by the ratio of precipitable water (W) anomaly to precipitation (P) anomaly associated with the MJO and derived by a regression approach. Before conducting the regression, both W and P anomalies are subject to 20–100-day filtering and averaged over the Indian Ocean (10°S – 10°N , 75° – 85°E) region. Adapted from Jiang et al. (2016) and Gonzalez and Jiang (2017).

source of predictability of seasonal to interannual climate anomalies over North America (Hoskins and Karoly 1981; Horel and Wallace 1981) and the U.S. Affiliated Pacific Islands (USAPI; Annamalai et al. 2014). Recognizing that equatorial Pacific precipitation and associated diabatic heating anomalies are fundamental to this framework, and that in regions of weak horizontal temperature gradients such as the tropical oceans, moist static energy (MSE) variations are primarily due to moisture variations and have a close association with precipitation (Neelin and Held 1987; Bretherton et al. 2006), we developed a POD based on vertically integrated MSE budget. The POD identifies leading model processes that are important in translating ENSO-related SST anomalies into precipitation anomalies. Further, to identify and quantify compensating errors in model processes, MSE variance analysis (Wing and Emanuel 2014) is also included in the POD. With a focus on

ENSO winters, this POD is applied to CMIP5 models' historical simulations and reanalysis products, and metrics are developed to assess models' fidelity in representing processes. Apart from identifying systematic errors across models (e.g., Fig. 5), the POD identifies compensating errors in individual models, and assesses progress in generations of models from the same center (Annamalai 2019, manuscript submitted to *J. Climate*).

Figure 5 shows scatterplots between simulated anomalous precipitation and net radiative flux divergence into the column F_{rad} for composites of El Niño winters over the equatorial central (Fig. 5a) and eastern (Fig. 5b) Pacific Ocean, respectively. The strong intermodel correlations in these plots suggest that systematic biases in precipitation are similarly tied to biases in F_{rad} . Annamalai (2019, manuscript submitted to *J. Climate*) note that both during El Niño and La Niña winters, F_{rad} , particularly the bias in net

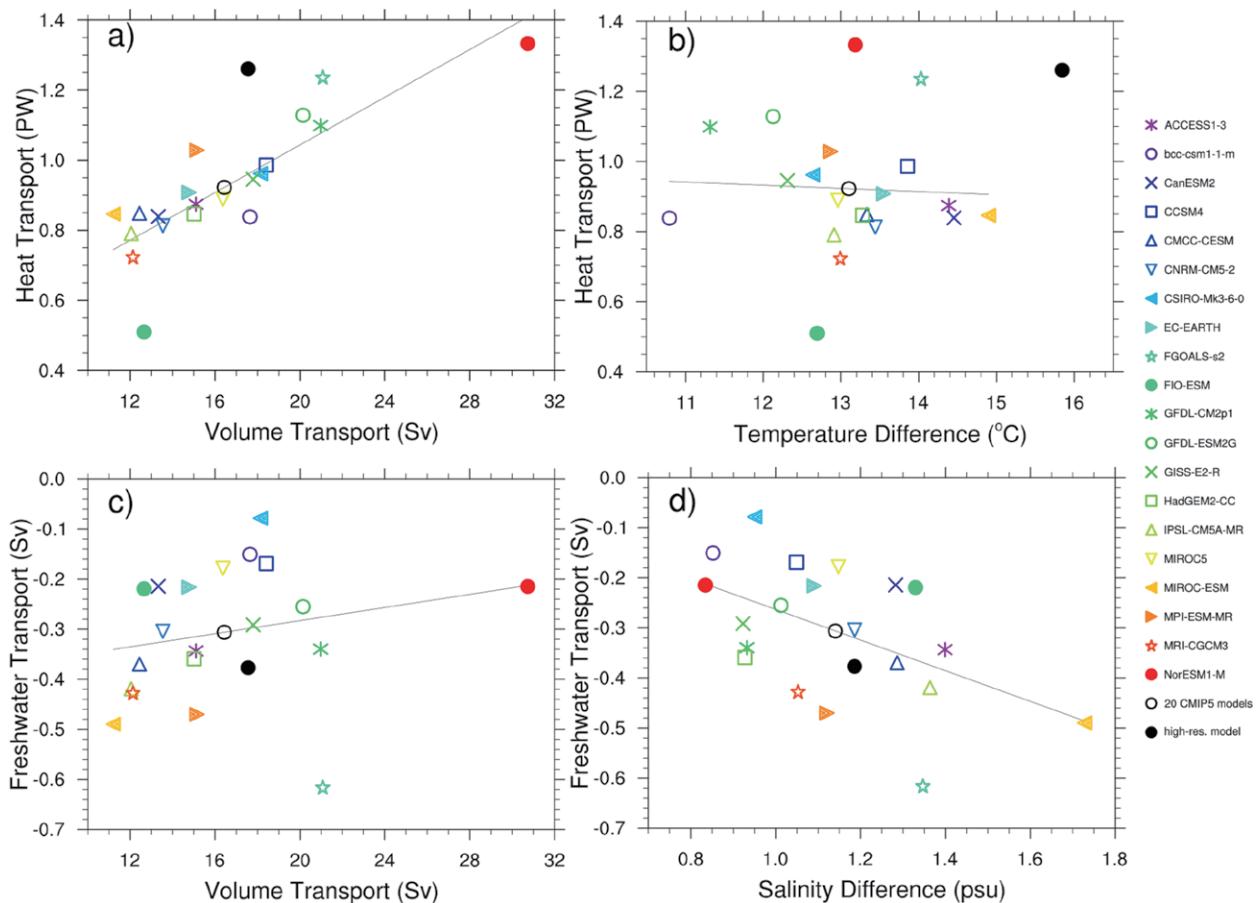


FIG. 4. Relation of the modeled oceanic heat transport with (a) volume transport and (b) temperature difference (between the upper and lower AMOC limbs) at 26°N. Relation of the modeled freshwater transport with (c) volume transport and (d) salinity differences. Colored symbols and the black circle denote results from 20 CMIP5 historical simulations and their average, respectively; black dot denotes high-resolution ocean simulation results that represent well the observed mean heat and freshwater transports and the AMOC structure at 26°N (Xu et al. 2016).

longwave (LW) component, dominates the systematic bias in the MSE budget across all models. Here, higher F_{rad} values indicate stronger cloud–radiative feedbacks that relate to perturbation of the radiative energy budget by condensate produced by convection (Stephens et al. 2008). Furthermore, systematic biases in F_{rad} are strongly linked to simulated free-troposphere moisture anomalies that in turn are strongly linked to precipitation biases (not shown).

Warm rain processes. Combined analysis of multiple satellite measurements from *CloudSat* and the Moderate Resolution Imaging Spectroradiometer (MODIS) has provided new insights into the warm rain process, a key process that governs the low-cloud radiative properties and is a major pathway through which aerosols influence clouds. Suzuki et al. (2010) proposed a methodology for combining the radar reflectivity profile from *CloudSat* (Marchand et al. 2008) and the cloud properties (optical thickness and effective radius) from MODIS (Platnick et al. 2003; Nakajima et al. 2010) to probe how the warm rain process occurs within clouds. The methodology composites the radar reflectivity profiles in the form of the probability density function normalized at each in-cloud optical depth, which is determined by vertically slicing the cloud optical thickness according to the adiabatic profile assumption. The statistics thus constructed, referred to as contoured frequency by optical depth diagram (CFODD), are

further classified according to ranges of cloud-top particle size (Fig. 6, top panels), which is another observable from MODIS, to reveal how the vertical microphysical structure of warm-topped clouds tends to transition from a nonprecipitating regime (Fig. 6a) to a precipitating regime (Fig. 6c) as a fairly monotonic function of the particle size. The statistics provide a direct insight into the coalescence process.

The methodology has been applied to output from multiple global models (Suzuki et al. 2015; Jing et al. 2017) to construct the statistics corresponding to those from satellite observations. The statistics are then compared to evaluate how the models represent the warm rain formation process against satellite observations. Examples for such a comparison with state-of-the-art global models are shown in Fig. 6 (middle and bottom panels) that indicate the models tend to produce rain too efficiently even when the cloud-top particle size is small. The behavior of the model biases identified in these statistics is further traced to formulations of model cloud microphysics, particularly the autoconversion process (Suzuki et al. 2015), implying that the CFODD statistics could serve as a clue to constrain a key uncertainty in cloud microphysics parameterization with satellite observations. This bottom-up constraint on model physics, however, tends to produce an overly negative forcing due to the aerosol indirect effect, which contradicts the top-down requirement for models to reproduce the historical temperature trend (Suzuki et al. 2013),

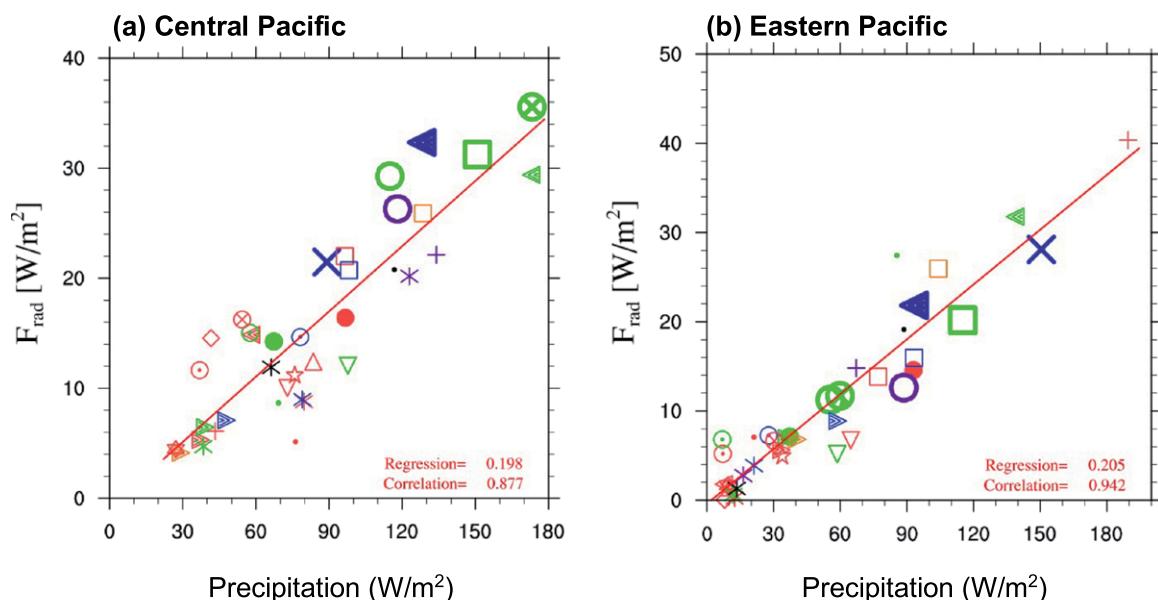


FIG. 5. Scatterplots between CMIP5 models simulated precipitation vs net radiative flux divergence F_{rad} : (a) equatorial central Pacific (10°S – 5°N , 160°E – 160°W) and (b) eastern Pacific (5°S – 5°N , 160° – 80°W). Results are for El Niño winter composites. In each panel, intermodel correlations and best-fit regressions are also provided.

implying the presence of error compensation at a fundamental level.

Tropical cyclones. The tropical cyclone (TC) POD contains a set of diagnostic codes that facilitate examination of TCs in global model simulations. When supplied with storm information (e.g., center position and intensity), this module computes azimuthal averages of dynamic and thermodynamic fields around the storm center that are helpful in identifying physical processes that lead to intermodel

differences in simulated TCs. Figure 7 shows an example output from this POD. The top two rows show radius–pressure plots of tangential and radial velocity, and relative humidity and pressure velocity, while the bottom row shows rainfall rates. The composite structures of TCs from four different GCM simulations show cyclonic tangential winds and typical secondary circulations that are made of low-level radial inflow toward the center, rising motions around the center, and upper-level radial outflow away from the center. The TC POD was used by Kim

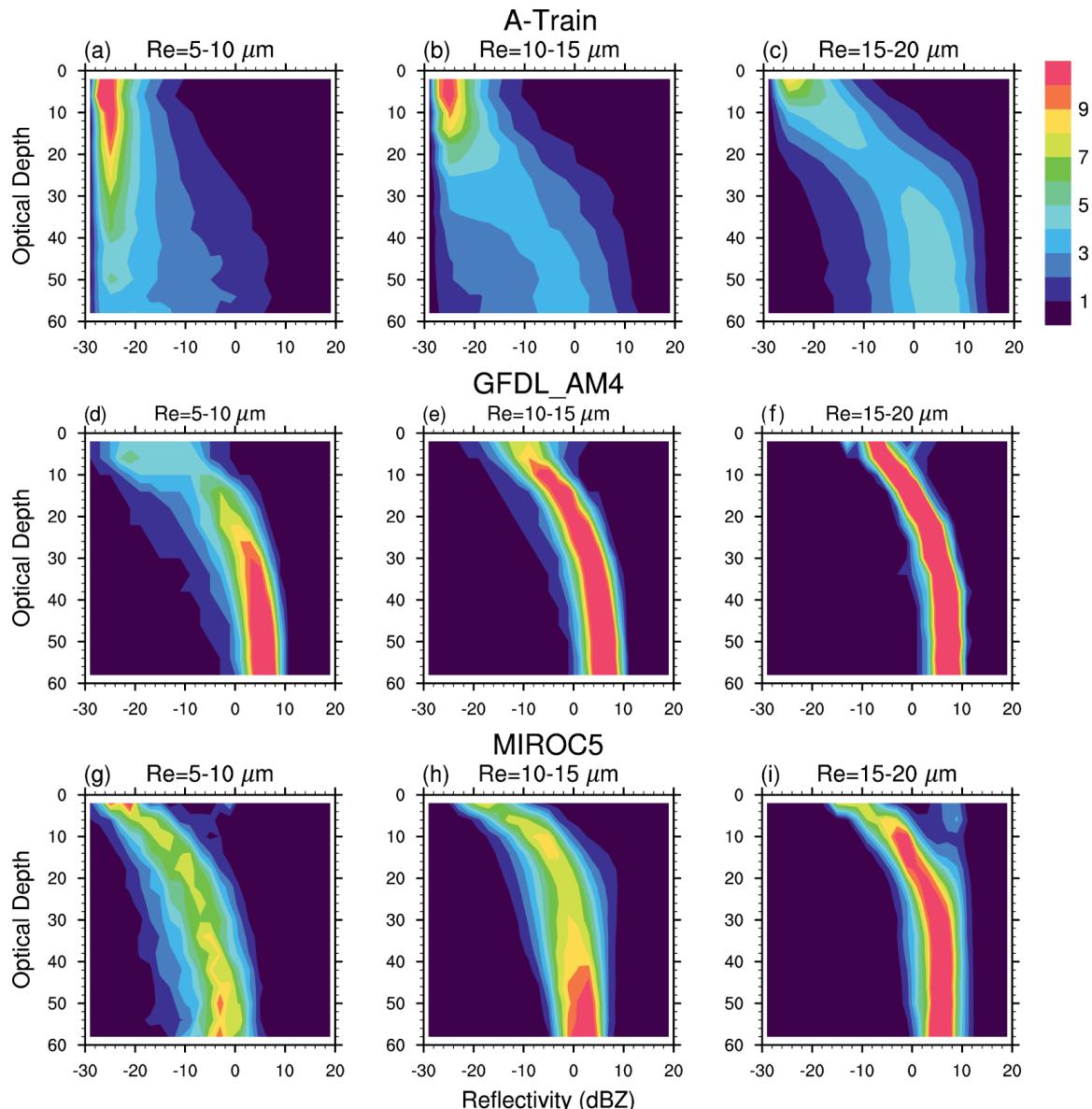


FIG. 6. The PDF (color shading; % dBZ^{-1}) of radar reflectivity (abscissa) normalized as a function of cloud optical depth increasing downward (ordinate), which is further classified according to different ranges of cloud-top particle sizes for (left) 5–10, (center) 10–15, and (right) 15–20 mm obtained from (a)–(c) A-Train satellite observations, (d)–(f) GFDL AM4, and (g)–(i) Model for Interdisciplinary Research on Climate, version 5 (MIROC5; adapted from Jing et al. 2017).

et al. (2018) to examine why the High Resolution Atmospheric Model (HiRAM) simulation produces stronger TCs than the GFDL Atmosphere Model 2.5 (AM2.5) and Forecast-Oriented Low Resolution Ocean (FLOR) simulations. A key finding in the study was that at comparable intensity, the HiRAM model produces a greater amount of precipitation near the TC center than the other models (cf. the left two panels in the bottom row of Fig. 7). The greater amount of diabatic heating associated with more rainfall in the TC inner-core region in the HiRAM model favors intensification (e.g., Schubert and Hack 1982; Nolan et al. 2007). Moon et al. (2019, manuscript submitted to *J. Climate*) applied the TC POD to further examine intermodel spread among eight different global model simulations with different resolutions and physics.

In the second set of TC PODs, a framework based on the column-integrated MSE variance budget, which was originally developed to study convective organization in cloud-resolving model simulations (Wing and Emanuel 2014), has been adapted for climate model

simulations of TCs. This POD focuses on the relative role of feedback processes associated with tropical cyclogenesis by computing the product of MSE anomalies from the mean of a 10° box surrounding a TC and anomalous sources and sinks of MSE. Figure 8 shows an example of this POD, for the same GCM simulations and composites as used in Fig. 7. The first row shows the squared MSE anomalies, and the bottom two rows show two of the terms in the MSE variance budget—the radiative and surface flux feedbacks. While the feedbacks are generally positive and thus act to amplify MSE anomalies and favor development of the TC, they tend to be stronger in the models with more intense TCs. This strength disparity indicates that the representation of the interaction of spatially varying radiative cooling and surface fluxes with the developing TC is partially responsible for intermodel spread in TC simulation. Wing et al. (2019) applied this POD to six different global model simulations.

Soil moisture control on evapotranspiration. Soil moisture–atmosphere interactions are a key factor

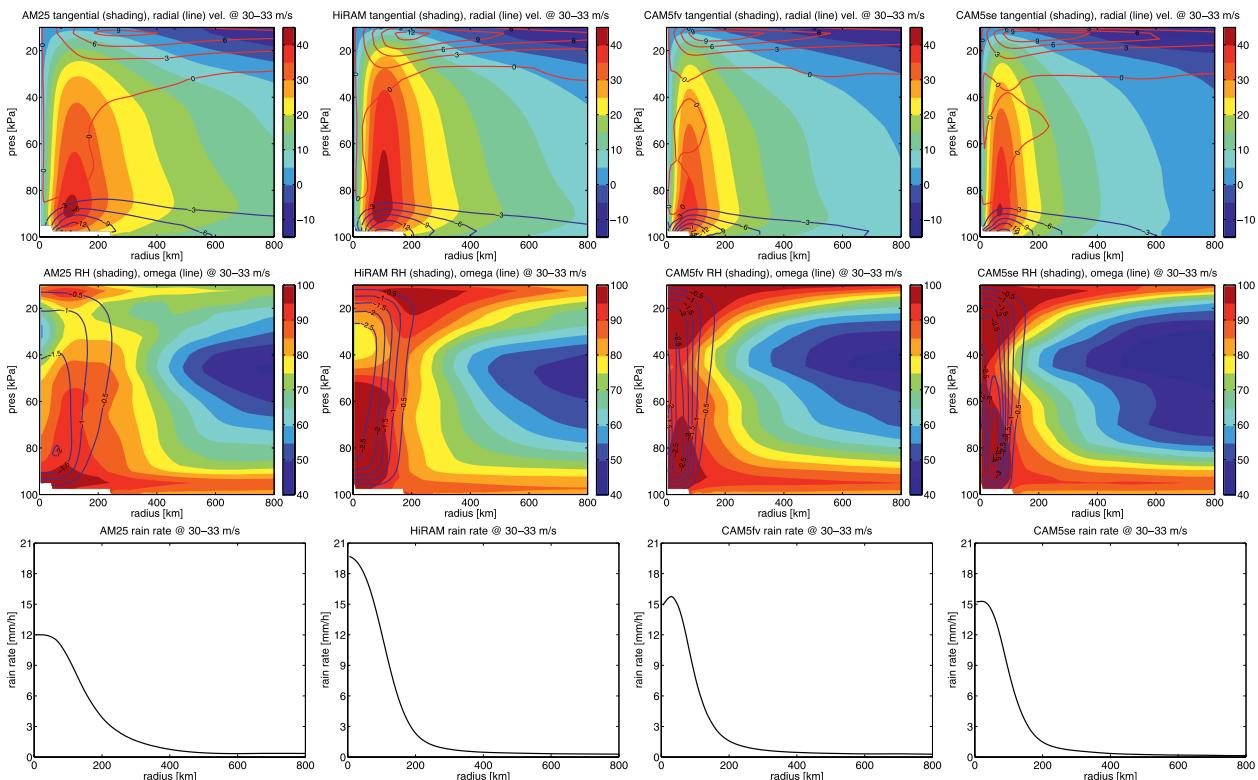


Fig. 7. (top),(middle) Azimuthally averaged radius–pressure plots of (top) tangential velocity (shading) and radial velocity (lines) and (middle) relative humidity (shading) and pressure velocity (lines). **(bottom)** Azimuthally averaged rain rate. All panels show the composites when the simulated tropical cyclones have a similar intensity, in this case between 30 and 33 $m s^{-1}$. Positive (negative) values are plotted in red (blue) contour lines. Four different GCM simulations [GFDL AM2.5 and High Resolution Atmospheric Model (HiRAM) models at 0.5° resolution and NCAR Community Atmosphere Model version 5–Finite Volume (CAM5-FV) and CAM5-Spectral Element (CAM-SE) models at 0.25° resolution] are used.

modulating surface climate over land. Soil moisture variations are forced by the atmosphere; in turn, they regulate surface water and energy fluxes [e.g., evapotranspiration (ET)], and thus feedback onto near-surface surface climate (e.g., Seneviratne et al. 2010). One of the PODs focuses on the so-called “terrestrial leg” of this coupling, that is, the dependence of ET on soil moisture. Models exhibit significant uncertainties in the representation of this relationship (Guo et al. 2006; Dirmeyer et al. 2006; Berg and Sheffield 2018), which strongly influences summertime warming projections.

The hydrological and radiative controls on ET were assessed with a first-order diagnosis consisting of the correlations at the interannual time scale between summertime-mean values of surface (top 10 cm) soil moisture (SM) and incoming solar radiation (R_{Sds}), respectively, with ET (Berg and Sheffield 2018). Regions of positive SM–ET correlations in

Fig. 9 indicate soil moisture–limited regions, where soil moisture variability controls ET variability—generally in drier summer midlatitude regions. The value of the correlation indicates how strongly SM controls ET. Conversely, negative values indicate that ET variations drive variations in soil moisture levels: this occurs in the tropics and high latitudes, where available soil moisture is sufficient and the limiting factor for ET becomes atmospheric evaporative demand. This is consistent with the positive R_{Sds} –ET correlations in the same regions.

Figure 9 shows that the CMIP5 multimodel mean qualitatively reproduces the climatological pattern from the European Centre for Medium-Range Weather Forecasts (ECMWF) interim reanalysis (ERA-Interim). However, large uncertainties exist between models in the detailed spatial pattern and amplitude of the soil moisture control on ET. Overall, model uncertainty in SM–ET coupling tends to be

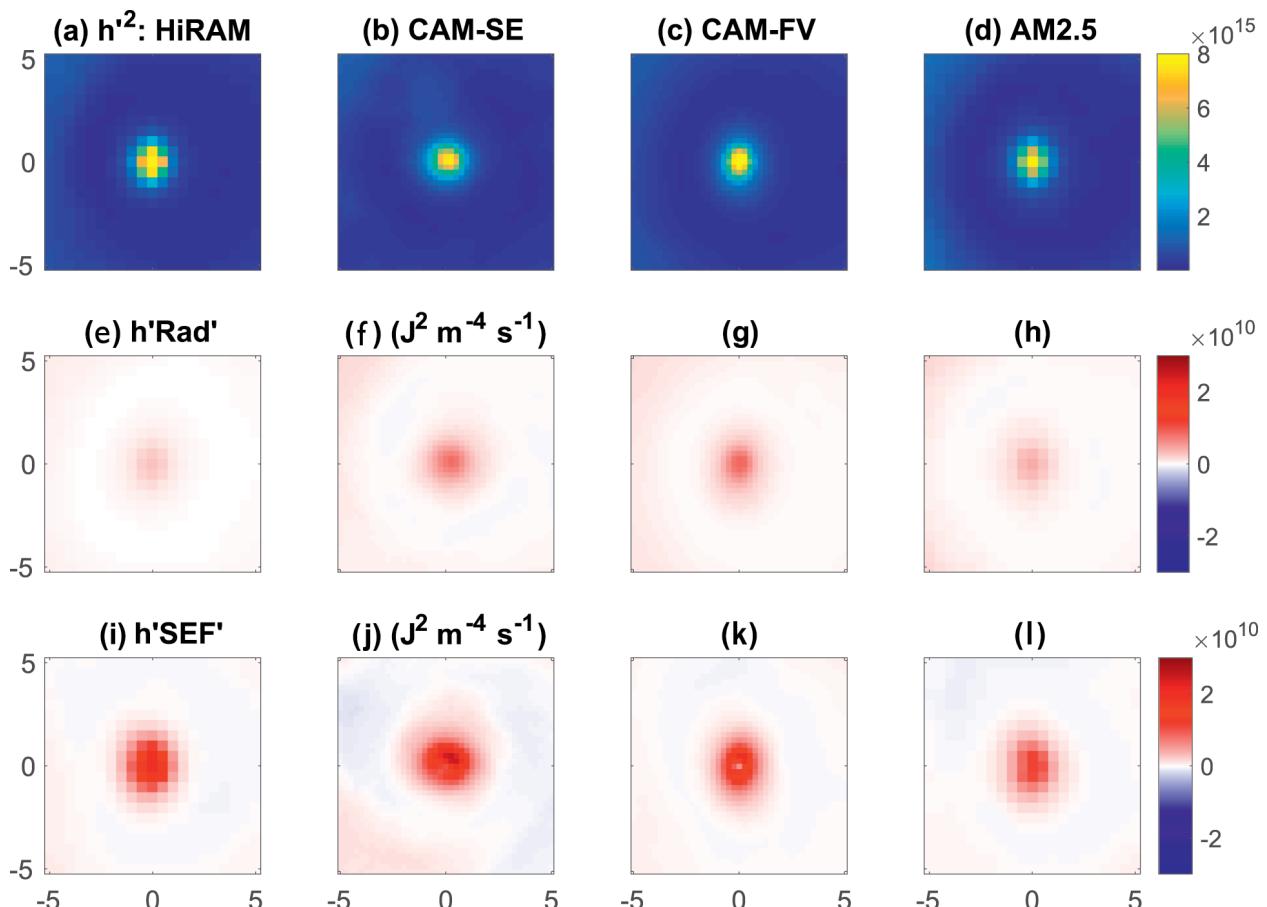


FIG. 8. (top) Squared anomalies of column-integrated MSE ($J^2 m^{-4}$), where anomalies are from the mean of a $10^\circ \times 10^\circ$ box surrounding a TC. (middle) Radiative and (bottom) surface flux feedback terms in the MSE variance budget, plotted as a function of latitude and longitude relative to the TC center. All panels show the composites when the simulated tropical cyclones have a similar intensity, in this case between 30 and 33 $m s^{-1}$. Four different GCM simulations are shown: (d),(h),(l) GFDL AM2.5 and (a),(e),(i) HiRAM models at 0.5° resolution and NCAR (c),(g),(k) CAM5-FV and (b),(f),(j) CAM5-SE models at 0.25° resolution.

greatest on the outer margins of regions of positive coupling, extending into regions of energy-limited ET. The complementarity across space between hydrological and radiative controls on ET extends across models: models that are less soil moisture limited are more radiation limited, and vice versa (not shown). In regions of greatest model spread, up to half of the intermodel variance in SM–ET coupling is explained by model differences in model precipitation; the remaining spread may be related to further differences in rainfall characteristics such as intraseasonal distribution, but differences are also likely to stem from differences in model treatment of land hydrology, including differences in the simulation of vegetation and the representation of soil water stress.

Midlatitude cyclones, fronts, and storm tracks. One focus of the task force was extratropical cyclones, which generate precipitation, winds, and clouds in the midlatitudes. GCMs need to capture the dynamics and thermodynamic properties of both the individual cyclone events and their accumulated behavior. Eulerian storm-track analysis revealed that model sea surface temperature biases impact the surface storm tracks and precipitation near ocean western boundary currents (Booth et al. 2017; Small et al. 2019). Targeted analyses of features in cyclones

and/or their fronts were carried out using Lagrangian tracking algorithms and compositing. These metrics facilitated process-oriented analyses of satellite observations of clouds that lead to 1) explanations for relationships between stability and cloud cover (Naud et al. 2016), and 2) pin-pointing the synoptic locations and conditions where biases in GCM clouds occur (Fig. 10). Task force efforts on cyclone-centered precipitation led to 1) a satellite-based benchmarks (Naud et al. 2019), and 2) results showing GCMs represent cyclone total precipitation as well as reanalysis, but the models have markedly different levels of contributions from their convection parameterizations (Booth et al. 2018). The Lagrangian metrics require 6-hourly, three-dimensional data, some of which are not standard in the CMIP archive.

Diurnal cycle as a test bed. Diurnal variations are large in near-surface temperature, pressure, winds, energy fluxes, precipitation, and other fields, especially over land during the warm season. These variations are linked to many land surface and atmospheric processes; therefore, they can be used as a test bed for diagnosing and evaluating weather and climate models (Dai and Trenberth 2004). One POD is the diurnal cycle in surface temperature and related fields. Analyses of surface air temperature in the GFDL AM4

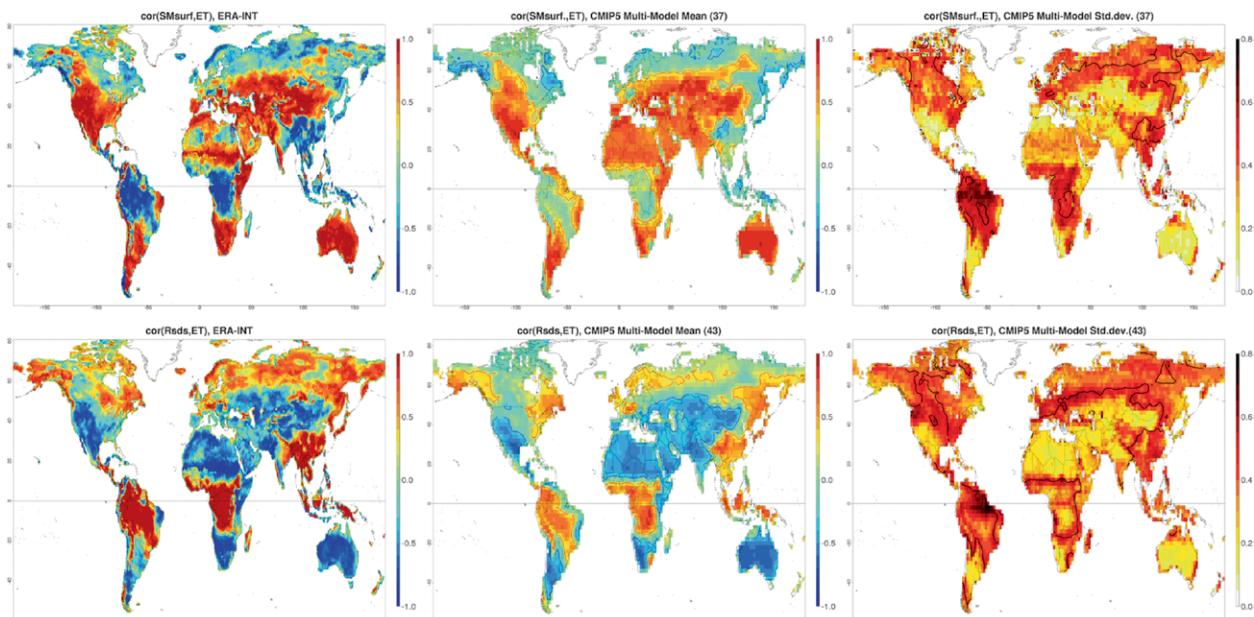


FIG. 9. (top) Correlation between summertime-mean surface SM and ET (left) in ERA-Interim, (center) in CMIP5 models, and (right) the standard deviation across CMIP5 models. (bottom) As in the top row, but for the correlation between ET and Rds. Summertime is defined as June–August in the Northern Hemisphere and December–February in the Southern Hemisphere. Correlations are over 1979–2014 for ERA-Interim and 1950–2005 for CMIP5 models, using outputs from the historical simulations. SM and ET outputs were available for 37 models, and for 43 models for Rds and ET.

(Zhao et al. 2018a,b) revealed some systematic biases in the daily minimum (T_{min}) and maximum (T_{max}) air temperature, despite the relatively small biases in the daily mean (T_{mean}) temperature over many land areas (Lu 2018). For example, T_{max} in the AM4 showed a cold bias of $1^{\circ}\text{--}4^{\circ}\text{C}$ over most land areas during all seasons, while T_{min} in the AM4 showed small to slight warm biases when compared with station observations, resulting in greatly reduced diurnal temperature range (DTR) in the model (Lu 2018).

Analyses of surface energy fluxes (Lu 2018) revealed many biases, including higher surface albedo, higher downward shortwave radiation, and weaker surface winds than ERA-Interim (Dee et al. 2011).

However, large uncertainties in existing surface energy flux data made it difficult to precisely quantify the model biases in these fields. Furthermore, the inconsistent definitions of 2-m air temperature in the model and in observations further complicated the evaluation because the T_{min} , T_{max} , and DTR vary with the height of the measurement above the ground (Fig. 11). In the AM4, the reference for the 2-m air temperature is close to the displacement height (rather than the ground, which is the reference for the 2-m air temperature from weather stations), which is about two-thirds of the canopy height (a function of vegetation types). Thus, the 2-m air temperature from the AM4 is likely at a higher level than the 2-m air

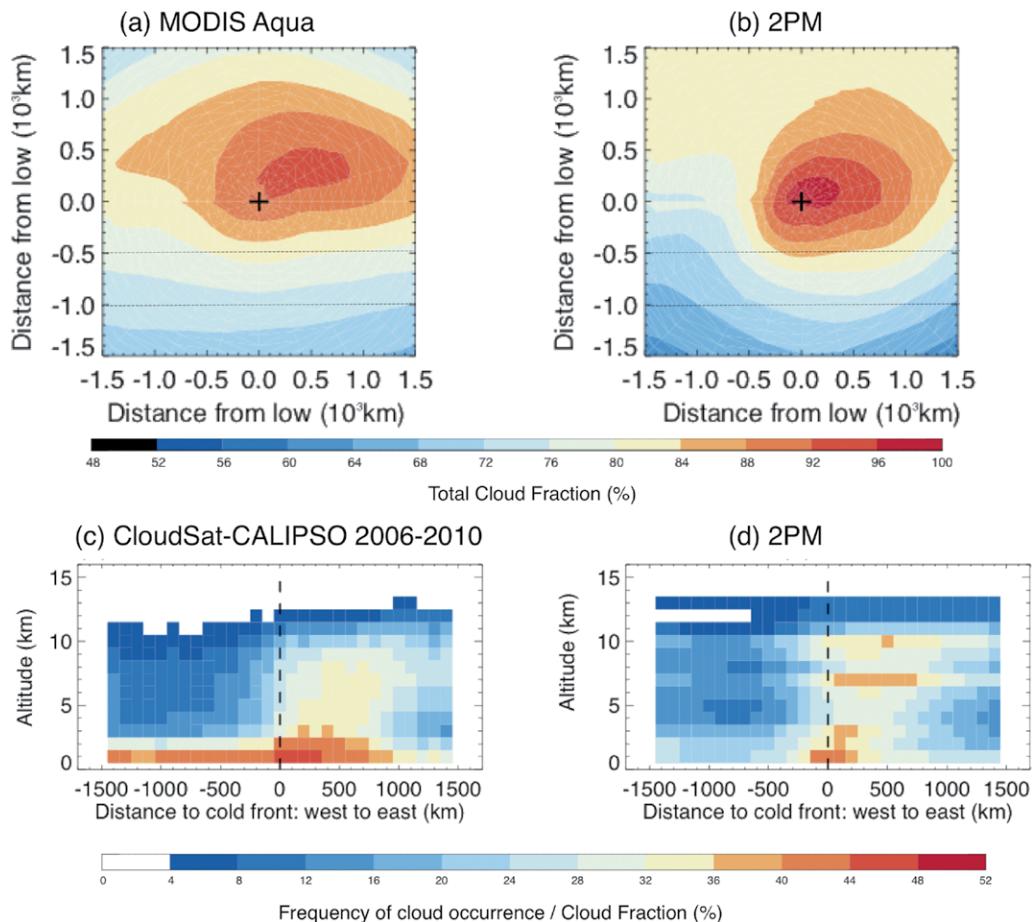


FIG. 10. Composites of (top) cloud fraction for cyclone-centered map view and (bottom) cold-front-centered transects. Observational data: (a) MODIS Aqua and (c) CloudSat–Cloud–Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO). Model data: GFDL pre-CMIP6 model that uses a two-plume convection scheme (i.e., 1400 local time). Composites are generated using cyclones or fronts from 5 years of observations or model data, within the $30^{\circ}\text{--}60^{\circ}$ latitude band in both hemispheres. The plus sign in the top row represents the center of the cyclones. The region between the black dashed lines in the top row is the approximate location for transects used in frontal analysis. The black dashed line in the bottom row indicates location of the cold front based on 850-hPa potential temperature and wind gradient data [see Naud et al. (2014) and Naud et al. (2016) for compositing methodology details].

temperature from station measurements, especially over forests. Since Tmax decreases with height (due to solar heating on the ground) while Tmin increases with height (due to radiative cooling of the ground; Fig. 11), the higher reference in the AM4 could contribute to and help partially explain the systematic cold bias in Tmax and the warm bias in Tmin, as well as the smaller DTR in AM4. However, the difference between the model and station temperatures is larger than that between the observed 2- and 9-m temperatures at many of the stations shown in Fig. 11, suggesting that other factors besides the higher reference height likely played a role.

THE NOAA MDTF PROCESS-ORIENTED DIAGNOSTICS FRAMEWORK.

As alluded to above, a product of the NOAA MAPP MDTF is an

evolving software framework to aid application of the PODs described above to the model development process. Extensive documentation of the current state of the framework including a developer's guide, a "getting started" guide, standardized POD-specific documentation, sample .html output, and the code itself is available at www.cesm.ucar.edu/working_groups/Atmosphere/mdtf-diagnostics-package/. The framework has been developed as a Python code that integrates modules with PODs provided by contributing teams. While the Python framework is useful for modeling centers, it is important to emphasize that it is primarily a vehicle to facilitate adoption of the intellectual content—a center with its own diagnostics framework could easily adapt any part into its own interface and workflow. The PODs themselves follow an applications programming interface

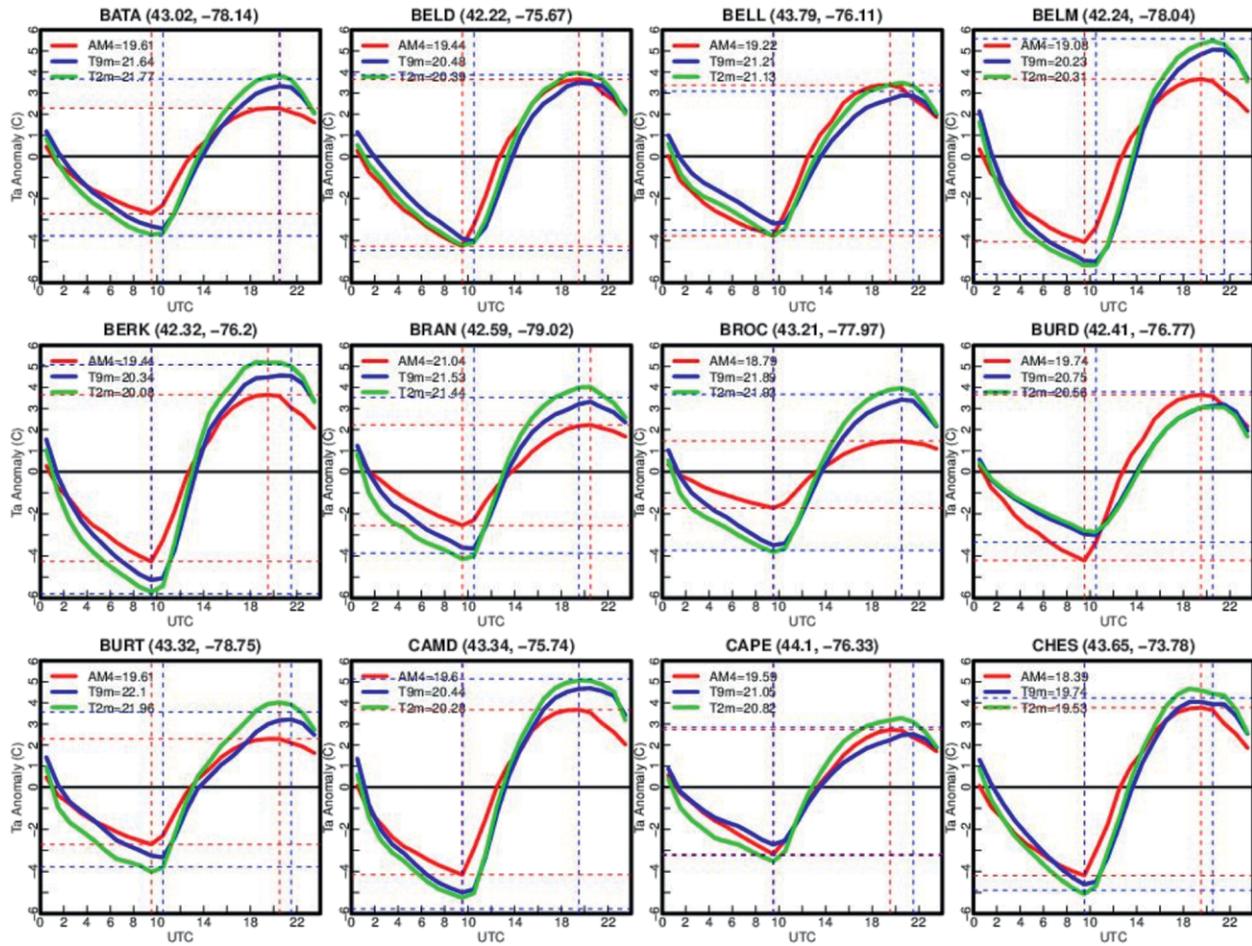


FIG. 11. The observed mean diurnal cycle of surface air temperatures at 2 (green) and 9 m (blue) above the ground during July of 2016 and 2017 at 12 New York Mesonet stations, compared with the GFDL AM4 model-simulated multiyear mean 2-m air temperature (red) for the grid box containing the station. The daily mean value is shown in the line legend. The 2-m height in the AM4 is 2 m above a reference height, which is close to the displacement height ($\sim 2/3$ of the canopy height) in the AM4, rather than 2 m above the ground as for weather stations.

(API) that specifies how the modules interact with the output from the candidate model version that is being diagnosed. We stress that the API developed here differs from packages like Grid Analysis and Display System (GrADS) or Giovanni that provide general tools that can be used for a broad set of visualization, data analysis, or data manipulation tasks. Rather, the functionality of this API provides a vehicle for organizing specific PODs that have been contributed by developers into a consistent package that runs on Climate and Forecast (CF) formatted model output or raw model output with extensions, and is designed to be used by modeling centers in their development workflows. This API also differs from those like Model Evaluation Tools (MET) at NCAR that provide a general set of tools for model verification (<https://ral.ucar.edu/solutions/products/model-evaluation-tools-met>), although it is being designed to be flexible enough to conceivably interface with such packages in the future.

Figure 1a illustrates both the Python framework and the API. Key features include the following:

- A Python script sets up paths, variable names, etc. for the model data to be analyzed. It calls PODs contributed by various groups; these yield plots, and each group provides the observational comparison for its own POD.
- The output plots are then composed into a web page, with subpages that permit easy comparison of the candidate model and observations.
- The PODs must be open source, but need not be based on Python; they just need to be callable from Python (e.g., POD2.ncl in the schematic are entrained in the framework using a simple Python wrapper).
- PODs are repeatable in modeling center workflow, and focused on model improvement. Any group can test a POD to submit, contributing to the library of POD.
- The PODs are independent, so that one can be added without reference to any other, making the MDTF framework extensible and amenable to parallel development.

As described above, Fig. 1b shows an edited example of a web page from a particular POD, illustrating how the comparison appears between observations and the model output analyzed by the POD. The details of the format vary according to the POD, but each provides the developer with model-to-observation comparison for a process of interest. The diagnostic set also provides a work set of examples

addressing different processes, each with their unique requirements and approach.

It is also important to acknowledge that not all PODs fit conveniently into the Python framework. Some require specialized output or large datasets that would not routinely be provided, or must interact with other software at the center, such as cyclone tracking routines. Such PODs will be provided separately, or in preprocessed form, with instructions for adoption. Nonetheless these diagnostics form part of the same intellectual framework. An updated status of PODs implemented into the framework can be found at the web link above.

SUMMARY AND PATH FORWARD. This article described the outcomes of the NOAA MAPP MDTF activities to promote development of PODs and their application to climate and weather prediction models. These activities include development of an open-source framework that is portable, community extensible, and usable to aid application of PODs to the model development process. Moving forward, a renewed MDTF that began its term in the fall of 2018 plans expansion, refinement, and steps to increase the diagnostic utility of the framework. Development and entrainment of additional PODs will be an ongoing activity. For example, there is a need for standardized basin-scale heat uptake and sea level change PODs. PODs for feedback mechanisms in regional hydroclimate extremes including cloud feedbacks will be developed, complemented by parameter-perturbation experiments with models. Diagnostics will be brought into the framework for processes affecting temperature and precipitation distribution tails, including advanced convective and moist-static energy diagnostics. Collaborations will continue with GFDL and NCAR model development teams and an expanded number of other centers to refine PODs to increase their range and usability for model development teams. A particular interest is expanding the POD suite for use with weather forecasting models, which is conceptually attractive given common physical roots of climate and forecasting models and the shared imperative to reduce biases in both types of models.

The MDTF plans to develop protocols to optimize application of the diagnostic framework to CMIP6 model simulations. The API already uses standard CF model formats and variable names used for CMIP6 output, thus the package will read CMIP6 output. Further development includes developing tools to assist modelers in navigating trade-offs among multiple observational constraints and expanding functionality

to enable ensembles to be examined. The aim is to emphasize those aspects where the multimodel ensemble provides information about processes that tend to be ill constrained, and thus should be targeted for close scrutiny against observations (i.e., there must be added value beyond simply comparing a development version to existing models). Approaches considered will range from simply placing the candidate model within a multimodel plot of process-oriented metrics to new means of assessing parameter perturbation experiments systematically against observations.

Given the numerous community efforts related to process-oriented model diagnosis described above, greater coordination among these efforts would provide efficiencies, optimize science and technical approaches, and foster the greatest benefit to the climate and modeling communities. To further this goal, the MDTF has been proactive in forging connections to other efforts, for example fostering stronger links to PCMDI to leverage community data standards and enhance coordination of metrics and diagnostics development across agencies. The project will provide complementary process diagnosis to PCMDI capabilities that are expected to provide routine performance evaluation of all CMIP6 Diagnostic, Evaluation and Characterization of Klima (DECK) and Historical simulations. As MAPP PODs crystalize via experience at GFDL and NCAR, it is expected that some will be entrained into the broader community-based efforts, including possible collaborations with other modeling groups contributing to CMIP. Hence, the NOAA MDTF effort will benefit from experience such as PCMDI's working with the broader modeling community, and in particular its support of the developing standards and protocols.

ACKNOWLEDGMENTS. The PIs acknowledge discussions with Peter Gleckler, and constructive comments of three reviewers, that enabled substantial improvements of this manuscript. We would also like to thank Joyce Meyerson for graphical assistance and Fiaz Ahmed and Baird Langenbrunner for discussions and testing. This work was supported by the NOAA Climate Program Office MAPP Program as part of the MDTF, and by grants NA15OAR4310099, and NA18OAR4310280. The statements, findings, conclusions, and recommendations do not necessarily reflect the views of NOAA or the Department of Commerce.

REFERENCES

- Abatan, A. A., and Coauthors, 2018: Statistics of multi-year droughts from the Method for Object-Based Diagnostic Evaluation (MODE). *Int. J. Climatol.*, **38**, 3405–3420, <https://doi.org/10.1002/joc.5512>.
- Adames, Á. F., and J. M. Wallace, 2015: Three-dimensional structure and evolution of the moisture field in the MJO. *J. Atmos. Sci.*, **72**, 3733–3754, <https://doi.org/10.1175/JAS-D-15-0003.1>.
- Ahn, M.-S., D. Kim, K. Sperber, E. Maloney, D. Waliser, and H. Hendon, 2017: MJO simulation in CMIP5 climate models: MJO skill metrics and process-oriented diagnosis. *Climate Dyn.*, **49**, 4023–4045, <https://doi.org/10.1007/s00382-017-3558-4>.
- Annamalai, H., J. Hafner, A. Kumar, and H. Wang, 2014: A framework for dynamical seasonal prediction of precipitation over the Pacific islands. *J. Climate*, **27**, 3272–3297, <https://doi.org/10.1175/JCLI-D-13-00379.1>.
- , B. Taguchi, J. P. McCreary, M. Nagura, and T. Miyama, 2017: Systematic errors in South Asian monsoon simulation: Importance of equatorial Indian Ocean processes. *J. Climate*, **30**, 8159–8178, <https://doi.org/10.1175/JCLI-D-16-0573.1>.
- Baggett, C. F., E. A. Barnes, E. D. Maloney, and B. D. Mundhenk, 2017: Advancing atmospheric river forecasts into subseasonal timescales. *Geophys. Res. Lett.*, **44**, 7528–7536, <https://doi.org/10.1002/2017GL074434>.
- Bellenger, H., E. Guilyardi, J. Leloup, M. Lengaigne, and J. Vialard, 2014: ENSO representation in climate models: From CMIP3 to CMIP5. *Climate Dyn.*, **42**, 1999–2018, <https://doi.org/10.1007/s00382-013-1783-z>.
- Benedict, J. J., E. D. Maloney, A. H. Sobel, and D. M. Frierson, 2014: Gross moist stability and MJO simulation skill in three full-physics GCMs. *J. Atmos. Sci.*, **71**, 3327–3349, <https://doi.org/10.1175/JAS-D-13-0240.1>.
- Berg, A., and J. Sheffield, 2018: Soil moisture-evapotranspiration coupling in CMIP5 models: relationship with simulated climate and projections. *J. Climate*, **31**, 4865–4878.
- Booth, J. F., Y.-K. Kwon, S. Ko, J. Small, and R. Madsek, 2017: Spatial patterns and intensity of the surface storm tracks in CMIP5 models. *J. Climate*, **30**, 4965–4981, <https://doi.org/10.1175/JCLI-D-16-0228.1>.
- , C. M. Naud, and J. Willison, 2018: Evaluation of extratropical cyclone precipitation in the North Atlantic basin: An analysis of ERA-Interim, WRF, and two CMIP5 models. *J. Climate*, **31**, 2345–2360, <https://doi.org/10.1175/JCLI-D-17-0308.1>.
- Bretherton, C. S., M. E. Peters, and L. E. Back, 2004: Relationships between water vapor path and precipitation over the tropical oceans. *J. Climate*, **17**, 1517–1528, [https://doi.org/10.1175/1520-0442\(2004\)017<1517:RBWVPA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<1517:RBWVPA>2.0.CO;2).

- , P. N. Blossey, and M. E. Peters, 2006: Interpretation of simple and cloud-resolving simulations of moist convection–radiation interaction with a mock-Walker circulation. *Theor. Comput. Fluid. Dyn.*, **20**, 421–442, <https://doi.org/10.1007/s00162-006-0029-7>.
- Cheng, W., J. C. H. Chiang, and D. Zhang, 2013: Atlantic meridional overturning circulation (AMOC) in CMIP5 models: RCP and historical simulations. *J. Climate*, **26**, 7187–7197, <https://doi.org/10.1175/JCLI-D-12-00496.1>.
- Chikira, M., 2014: Eastward-propagating intraseasonal oscillation represented by Chikira–Sugiyama cumulus parameterization. Part II: Understanding moisture variation under weak temperature gradient balance. *J. Atmos. Sci.*, **71**, 615–639, <https://doi.org/10.1175/JAS-D-13-038.1>.
- Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136.
- Dai, A., and K. E. Trenberth, 2004: The diurnal cycle and its depiction in the Community Climate System Model. *J. Climate*, **17**, 930–951, [https://doi.org/10.1175/1520-0442\(2004\)017<0930:TDCAID>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<0930:TDCAID>2.0.CO;2).
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- Dirmeyer, P. A., R. D. Koster, and Z. Guo, 2006: Do global models properly represent the feedback between land and atmosphere? *J. Hydrometeor.*, **7**, 1177–1198, <https://doi.org/10.1175/JHM532.1>.
- Eyring, V., and Coauthors, 2016a: ESMValTool (v1.0) – A community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geosci. Model Dev.*, **9**, 1747–1802, <https://doi.org/10.5194/gmd-9-1747-2016>.
- , and Coauthors, 2016b: Towards improved and more routine Earth system model evaluation in CMIP. *Earth Syst. Dyn.*, **7**, 813–830, <https://doi.org/10.5194/esd-7-813-2016>.
- , and Coauthors, 2019: The multiple and growing benefits of confronting climate models with observations. *Nat. Climate Change*, **9**, 102–110, <https://doi.org/10.1038/s41558-018-0355-y>.
- Gettelman, A., and Coauthors, 2012: A community diagnostic tool for chemistry climate model validation. *Geosci. Model Dev.*, **5**, 1061–1073, <https://doi.org/10.5194/gmd-5-1061-2012>.
- Gilleland, E., M. Bukovsky, C. L. Williams, S. McGinnis, C. M. Ammann, B. G. Brown, and L. O. Mearns, 2016: Evaluating NARCCAP model performance for frequencies of severe-storm environments. *Adv. Stat. Climatol. Meteor. Oceanogr.*, **2**, 137–153, <https://doi.org/10.5194/ascmo-2-137-2016>.
- Gleckler, P. J., C. Doutriaux, P. J. Durack, K. E. Taylor, Y. Zhang, D. N. Williams, E. Mason, and J. Servonnat, 2016: A more powerful reality test for climate models. *Eos, Trans. Amer. Geophys. Union*, **97**, <https://doi.org/10.1029/2016EO051663>.
- Gonzalez, A. O., and X. Jiang, 2017: Winter mean lower tropospheric moisture over the Maritime Continent as a climate model diagnostic metric for the propagation of the Madden-Julian oscillation. *Geophys. Res. Lett.*, **44**, 2588–2596, <https://doi.org/10.1002/2016GL072430>.
- Grose, M. R., and Coauthors, 2014: Assessment of the CMIP5 global climate model simulations of the western tropical Pacific climate system and comparison to CMIP3. *Int. J. Climatol.*, **34**, 3382–3399, <https://doi.org/10.1002/joc.3916>.
- Guo, P. A., and Coauthors, 2006: GLACE: The Global Land–Atmosphere Coupling Experiment. Part II: Analysis. *J. Hydrometeor.*, **7**, 611–625, <https://doi.org/10.1175/JHM511.1>.
- Hannah, W. M., and E. D. Maloney, 2014: The moist static energy budget in NCAR CAM5 hindcasts during DYNAMO. *J. Adv. Model. Earth Syst.*, **6**, 420–440, <https://doi.org/10.1002/2013MS000272>.
- Henderson, S. A., E. D. Maloney, and E. A. Barnes, 2016: The influence of the Madden–Julian oscillation on Northern Hemisphere winter blocking. *J. Climate*, **29**, 4597–4616, <https://doi.org/10.1175/JCLI-D-15-0502.1>.
- Henderson, S. A., E. D. Maloney, and S.-W. Son, 2017: Madden–Julian oscillation teleconnections: The impact of the basic state and MJO representation in general circulation models. *J. Climate*, **30**, 4567–4587, <https://doi.org/10.1175/JCLI-D-16-0789.1>.
- Holloway, C. E., and J. D. Neelin, 2009: Moisture vertical structure, column water vapor, and tropical deep convection. *J. Atmos. Sci.*, **66**, 1665–1683, <https://doi.org/10.1175/2008JAS2806.1>.
- Horel, J., and J. M. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.*, **109**, 813–828, [https://doi.org/10.1175/1520-0493\(1981\)109<0813:PSAPAW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2).
- Hoskins, B. J., and D. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196, [https://doi.org/10.1175/1520-0469\(1981\)038<1179:TSLROA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1981)038<1179:TSLROA>2.0.CO;2).
- Hung, M.-P., J.-L. Lin, W. Wang, D. Kim, T. Shinoda, and S. J. Weaver, 2013: MJO and convectively coupled

- equatorial waves simulated by CMIP5 climate models. *J. Climate*, **26**, 6185–6214, <https://doi.org/10.1175/JCLI-D-12-00541.1>.
- Jiang, X., 2017: Key processes for the eastward propagation of the Madden-Julian Oscillation based on multimodel simulations. *J. Geophys. Res. Atmos.*, **122**, 755–770, <https://doi.org/10.1002/2016JD025955>.
- , and Coauthors, 2015: 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>.
- , M. Zhao, E. D. Maloney, and D. E. Waliser, 2016: Convective moisture adjustment time scale as a key factor in regulating model amplitude of the Madden-Julian Oscillation. *Geophys. Res. Lett.*, **43**, 10412–10419, <https://doi.org/10.1002/2016GL070898>.
- , Á. F. Adames, M. Zhao, D. Waliser, and E. Maloney, 2018: A unified moisture moist framework for seasonality of MJO propagation. *J. Climate*, **31**, 4215–4224, <https://doi.org/10.1175/JCLI-D-17-0671.1>.
- Jing, X., K. Suzuki, H. Guo, D. Goto, T. Ogura, T. Koshiro, and J. Mulmenstadt, 2017: A multi-model study on warm precipitation biases in global models compared to satellite observations. *J. Geophys. Res. Atmos.*, **122**, 11 806–11 824, <https://doi.org/10.1002/2017JD027310>.
- Kim, D., A. H. Sobel, E. D. Maloney, D. M. W. Frierson, and I.-S. Kang, 2011: A systematic relationship between intraseasonal variability and mean state bias. *J. Climate*, **24**, 5506–5520, <https://doi.org/10.1175/2011JCLI4177.1>.
- , and Coauthors, 2014a: Process-oriented MJO simulation diagnostic: Moisture sensitivity of simulated convection. *J. Climate*, **27**, 5379–5395, <https://doi.org/10.1175/JCLI-D-13-00497.1>.
- , J.-S. Kug, and A. H. Sobel, 2014b: Propagating versus nonpropagating Madden-Julian oscillation events. *J. Climate*, **27**, 111–125, <https://doi.org/10.1175/JCLI-D-13-00084.1>.
- , M.-S. Ahn, I.-S. Kang, and A. D. Del Genio, 2015: Role of longwave cloud-radiation feedback in the simulation of the Madden-Julian oscillation. *J. Climate*, **28**, 6979–6994, <https://doi.org/10.1175/JCLI-D-14-00767.1>.
- , H. Kim, and M.-I. Lee, 2017: Why does the MJO detour the Maritime Continent during austral summer? *Geophys. Res. Lett.*, **44**, 2579–2587, <https://doi.org/10.1002/2017GL072643>.
- , and Coauthors, 2018: Process-oriented diagnosis of tropical cyclones in high-resolution GCMs. *J. Climate*, **31**, 1685–1702, <https://doi.org/10.1175/JCLI-D-17-0269.1>.
- Kiranmayi, L., and E. D. Maloney, 2011: Intraseasonal moist static energy budget in reanalysis data. *J. Geophys. Res.*, **116**, D21117, <https://doi.org/10.1029/2011JD016031>.
- Kuo, Y.-H., J. D. Neelin, and C. R. Mechoso, 2017: Tropical convective transition statistics and causality in the water vapor-precipitation relation. *J. Atmos. Sci.*, **74**, 915–931, <https://doi.org/10.1175/JAS-D-16-0182.1>.
- , K. A. Schiro, and J. D. Neelin, 2018: Convective transition statistics over tropical oceans for climate model diagnostics: Observational baseline. *J. Atmos. Sci.*, **75**, 1553–1570, <https://doi.org/10.1175/JAS-D-17-0287.1>.
- Li, G., and S.-P. Xie, 2014: Tropical biases in CMIP5 multi-model ensemble: The excessive equatorial Pacific cold tongue and double ITCZ problems. *J. Climate*, **27**, 1765–1780, <https://doi.org/10.1175/JCLI-D-13-00337.1>.
- Livina, V., N. Edwards, S. Goswami, and T. Lenton, 2008: A wavelet-coefficient score for comparison of two-dimensional climatic-data fields. *Quart. J. Roy. Meteor. Soc.*, **134**, 941–955, <https://doi.org/10.1002/qj.261>.
- Lu, Y., 2018: Surface air temperature biases and other associated biases in GFDL AM4.0 model. M.S. thesis, Department of Atmospheric and Environmental Sciences, University at Albany, State University of New York, 60 pp.
- Luo, Y. Q., and Coauthors, 2012: A framework for benchmarking land models. *Biogeosciences*, **9**, 3857–3874, <https://doi.org/10.5194/bg-9-3857-2012>.
- Maloney, E. D., 2009: The moist static energy budget of a composite tropical intraseasonal oscillation in a climate model. *J. Climate*, **22**, 711–729, <https://doi.org/10.1175/2008JCLI2542.1>.
- Mapes, B., and R. Neale, 2011: Parameterizing convective organization to escape the entrainment dilemma. *J. Adv. Model. Earth Syst.*, **3**, M06004, <https://doi.org/10.1029/2011MS000042>.
- Marchand, R., G. Mace, T. Ackerman, and G. Stephens, 2008: Hydrometeor detection using CloudSat—An Earth-orbiting 94-GHz cloud radar. *J. Atmos. Oceanic Technol.*, **25**, 519–533, <https://doi.org/10.1175/2007JTECHA1006.1>.
- McGraw, M. C., and E. A. Barnes, 2018: Memory matters: A case for Granger causality in climate variability studies. *J. Climate*, **31**, 3289–3300, <https://doi.org/10.1175/JCLI-D-17-0334.1>.
- Moise, A. F., and F. P. Delage, 2011: New climate model metrics based on object-orientated pattern matching of rainfall. *J. Geophys. Res.*, **116**, D12108, <https://doi.org/10.1029/2010JD015318>.
- Mundhenk, B. D., E. A. Barnes, E. D. Maloney, and C. F. Baggett, 2018: Skillful subseasonal prediction of

- atmospheric river activity based on the Madden-Julian oscillation and the Quasi-Biennial oscillation.. *ndp Climate Atmos. Sci.*, **1**, 2017, <https://doi.org/10.1038/s41612-017-0008-2>.
- Nakajima, T. Y., K. Suzuki, and G. L. Stephens, 2010: Droplet growth in warm water clouds observed by the A-Train. Part II: A multisensor view. *J. Atmos. Sci.*, **67**, 1897–1907, <https://doi.org/10.1175/2010JAS3276.1>.
- Naud, C. M., J. F. Booth, and A. D. Del Genio, 2014: Evaluation of ERA-Interim and MERRA cloudiness in the Southern Ocean. *J. Climate*, **27**, 2109–2124, <https://doi.org/10.1175/JCLI-D-13-00432.1>.
- , —, and A. D. Del Genio, 2016: The relationship between boundary layer stability and cloud cover in the post-cold-frontal region. *J. Climate*, **29**, 8129–8149, <https://doi.org/10.1175/JCLI-D-15-0700.1>.
- , J. Jeyaratnam, L. J. Donner, C. J. Seman, M. Zhao, H. Guo, and Y. Ming, 2019: Extratropical cyclone clouds in the GFDL climate model: Diagnosing biases and the associated causes. *J. Climate*, <https://doi.org/10.1175/JCLI-D-19-0421.1>, in press.
- Neelin, J. D., and I. M. Held, 1987: Modeling tropical convergence based on the moist static energy budget. *Mon. Wea. Rev.*, **115**, 3–12, [https://doi.org/10.1175/1520-0493\(1987\)115<0003:MTCBOT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0003:MTCBOT>2.0.CO;2).
- , O. Peters, and K. Hales, 2009: The transition to strong convection. *J. Atmos. Sci.*, **66**, 2367–2384, <https://doi.org/10.1175/2009JAS2962.1>.
- , B. Langenbrunner, J. E. Meyerson, A. Hall, and N. Berg, 2013: California winter precipitation change under global warming in the Coupled Model Intercomparison Project 5 ensemble. *J. Climate*, **26**, 6238–6256, <https://doi.org/10.1175/JCLI-D-12-00514.1>.
- Nolan, D. S., Y. Moon, and D. P. Stern, 2007: Tropical cyclone intensification from asymmetric convection: Energetics and efficiency. *J. Atmos. Sci.*, **64**, 3377–3405, <https://doi.org/10.1175/JAS3988.1>.
- Peters, O., and J. D. Neelin, 2006: Critical phenomena in atmospheric precipitation. *Nat. Phys.*, **2**, 393–396, <https://doi.org/10.1038/nphys314>.
- Platnick, S., M. King, S. Ackerman, W. Menzel, B. Baum, J. Riedi, and R. Frey, 2003: The MODIS cloud products: Algorithms and examples from Terra. *IEEE Trans. Geosci. Remote Sens.*, **41**, 459–473, <https://doi.org/10.1109/TGRS.2002.808301>.
- Prabhat, O. R., S. Byna, K. Wu, F. Li, M. Wehner, and W. Bethel, 2012: TECA: A parallel toolkit for extreme climate analysis. *Procedia Comput. Sci.*, **9**, 866–876, <https://doi.org/10.1016/j.procs.2012.04.093>.
- Sahany, S., J. D. Neelin, K. Hales, and R. B. Neale, 2012: Temperature-moisture dependence of the deep convective transition as a constraint on entrainment in climate models. *J. Atmos. Sci.*, **69**, 1340–1358, <https://doi.org/10.1175/JAS-D-11-0164.1>.
- , —, —, and —, 2014: Deep convective transition characteristics in the Community Climate System Model and changes under global warming. *J. Climate*, **27**, 9214–9232, <https://doi.org/10.1175/JCLI-D-13-00747.1>.
- Santos, E., J. Paco, Y. Wei, S. Liu, B. Cook, D. Williams, and C. Silva, 2013: UV-CDAT: Analyzing climate datasets from a user's perspective. *Comput. Sci. Eng.*, **15**, 94–103, <https://doi.org/10.1109/MCSE.2013.15>.
- Schiro, K. A., J. D. Neelin, D. K. Adams, and B. R. Lintner, 2016: Deep convection and column water vapor over tropical land versus tropical ocean: A comparison between the Amazon and the tropical western Pacific. *J. Atmos. Sci.*, **73**, 4043–4063, <https://doi.org/10.1175/JAS-D-16-0119.1>.
- , F. Ahmed, S. E. Giangrande, and J. D. Neelin, 2018: GoAmazon2014/5 campaign points to deep-inflow approach to deep convection across scales. *Proc. Natl. Acad. Sci. USA*, **115**, 4577–4582, <https://doi.org/10.1073/pnas.1719842115>.
- Schubert, W. H., and J. J. Hack, 1982: Inertial stability and tropical cyclone development. *J. Atmos. Sci.*, **39**, 1687–1697, [https://doi.org/10.1175/1520-0469\(1982\)039<1687:ISATCD>2.0.CO;2](https://doi.org/10.1175/1520-0469(1982)039<1687:ISATCD>2.0.CO;2).
- Seneviratne, S. I., T. Corti, E. L. Davin, M. Hirschi, E. B. Jaeger, I. Lehner, B. Orlowsky, and A. J. Teuling, 2010: Investigating soil moisture–climate interactions in a changing climate: A review. *Earth Sci. Rev.*, **99**, 125–161, <https://doi.org/10.1016/j.earscirev.2010.02.004>.
- Sheffield, J., and Coauthors, 2013a: North American climate in CMIP5 experiments. Part I: Evaluation of 20th century continental and regional climatology. *J. Climate*, **26**, 9209–9245, <https://doi.org/10.1175/JCLI-D-12-00592.1>.
- , and Coauthors, 2013b: North American climate in CMIP5 experiments. Part II: Evaluation of 20th century intra-seasonal to decadal variability. *J. Climate*, **26**, 9247–9290, <https://doi.org/10.1175/JCLI-D-12-00593.1>.
- Small, J., R. Msadek, Y.-O. Kwon, J. F. Booth, and C. Zarzycki, 2019: Atmosphere surface storm track response to resolved ocean mesoscale in two sets of global climate model experiments. *Climate Dyn.*, **52**, 2067–2089, <https://doi.org/10.1007/s00382-018-4237-9>.
- Sobel, A., and E. Maloney, 2013: Moisture modes and the eastward propagation of the MJO. *J. Atmos. Sci.*, **70**, 187–192, <https://doi.org/10.1175/JAS-D-12-0189.1>.
- , S. Wang, and D. Kim, 2014: Moist static energy budget of the MJO during DYNAMO. *J. Atmos. Sci.*, **71**, 4276–4291, <https://doi.org/10.1175/JAS-D-14-0052.1>.

- Sperber, K. R., H. Annamalai, I.-S. Kang, A. Kitoh, A. Moise, A. Turner, B. Wang, and T. Zhou, 2013: The Asian Summer Monsoon: An intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Climate Dyn.*, **41**, 2711–2744, <https://doi.org/10.1007/s00382-012-1607-6>.
- Stephens, G. L., S. Van den Heever, and L.A. Pakula, 2008: Radiative convective feedback in idealized states of radiative-convective equilibrium. *J. Atmos. Sci.*, **65**, 3899–3916.
- Stephens, G. L., C. Jacob, and G. Tselioudis, 2015: The GEWEX Process Evaluation Study: GEWEX-PROES. *GEWEX News*, Vol. 27, No.4, International GEWEX Project Office, Silver Spring, MD, 4–5, www.gewex.org/gewex-content/files_mf/1464112934Nov2015.pdf.
- Stevens, B., and S. Bony, 2013: What are climate models missing. *Science*, **340**, 1053–1054, <https://doi.org/10.1126/science.1237554>.
- Suzuki, K., T. Nakajima, and G. L. Stephens, 2010: Particle growth and drop collection efficiency of warm clouds as inferred from joint *CloudSat* and MODIS observations. *J. Atmos. Sci.*, **67**, 3019–3032, <https://doi.org/10.1175/2010JAS3463.1>.
- , J.-C. Golaz, and G. L. Stephens, 2013: Evaluating cloud tuning in a climate model with satellite observations. *Geophys. Res. Lett.*, **40**, 4464–4468, <https://doi.org/10.1002/grl.50874>.
- , G. Stephens, A. Bodas-Salcedo, M. Wang, J.-C. Golaz, T. Yokohata, and T. Koshiro, 2015: Evaluation of the warm rain formation process in global models with satellite observations. *J. Atmos. Sci.*, **72**, 3996–4014, <https://doi.org/10.1175/JAS-D-14-0265.1>.
- Tsushima, Y., and Coauthors, 2017: The Cloud Feedback Model Intercomparison Project (CFMIP) Diagnostic Codes Catalogue – Metrics, diagnostics and methodologies to evaluate, understand and improve the representation of clouds and cloud feedbacks in climate models. *Geosci. Model Dev.*, **10**, 4285–4305, <https://doi.org/10.5194/gmd-10-4285-2017>.
- Waliser, D., and Coauthors, 2009: MJO simulation diagnostics. *J. Climate*, **22**, 3006–3030, <https://doi.org/10.1175/2008JCLI2731.1>.
- , and Coauthors, 2012: The “year” of tropical convection (May 2008 to April 2010): Climate variability and weather highlights. *Bull. Amer. Meteor. Soc.*, **93**, 1189–1218, <https://doi.org/10.1175/2011BAMS3095.1>.
- Wang, C., L. Zhang, S.-K. Lee, L. Wu, and C. R. Mechoso, 2014: A global perspective on CMIP5 climate model biases. *Nat. Climate Change*, **4**, 201–205, <https://doi.org/10.1038/nclimate2118>.
- Wheeler, M. W., E. D. Maloney, and the MJO Task Force, 2013: Madden-Julian Oscillation (MJO) Task Force: A joint effort of the climate and weather communities. *CLIVAR Exchanges*, No. 61, International CLIVAR Project Office, Southampton, United Kingdom, 9–12, www.clivar.org/sites/default/files/documents/Exchanges61.pdf.
- Williams, D. N., and Coauthors, 2016: UV-CDAT v2.4.0. Zenodo, <https://doi.org/10.5281/zenodo.45136>.
- Wing, A. A., and K. A. Emanuel, 2014: Physical mechanisms controlling self-aggregation of convection in idealized numerical modeling simulations. *J. Adv. Model. Earth Syst.*, **6**, 59–74, <https://doi.org/10.1002/2013MS000269>.
- , and Coauthors, 2019: Moist static energy budget analysis of tropical cyclone formation and intensification in high-resolution climate models. *J. Climate*, <https://doi.org/10.1175/JCLI-D-18-0599.1>, in press.
- Xu, X., P. B. Rhines, and E. P. Chassignet, 2016: Temperature-salinity structure of the North Atlantic circulation and associated heat and freshwater transports. *J. Climate*, **29**, 7723–7742, <https://doi.org/10.1175/JCLI-D-15-0798.1>.
- Zappa, G., L. C. Shaffrey, and K. I. Hodges, 2013: The ability of CMIP5 models to simulate North Atlantic extratropical cyclones. *J. Climate*, **26**, 5379–5396, <https://doi.org/10.1175/JCLI-D-12-00501.1>.
- Zhao, M., and Coauthors, 2018a: The GFDL global atmosphere and land model AM4.0/LM4.0: 1. Simulation characteristics with prescribed SSTs. *J. Adv. Model. Earth Syst.*, **10**, 691–734, <https://doi.org/10.1002/2017MS001208>.
- , and Coauthors, 2018b: The GFDL global atmosphere and land model AM4.0/LM4.0: 2. Model description, sensitivity studies, and tuning strategies. *J. Adv. Model. Earth Syst.*, **10**, 735–769, <https://doi.org/10.1002/2017MS001209>.