BNL-216090-2020-JAAM

The ARM Data-oriented Metrics and Diagnostics Package for Climate Models

C. Zhang, W. Lin

To be published in "Bulletin of the American Meteorological Society"

June 2020

Environmental and Climate Sciences Department

Brookhaven National Laboratory

U.S. Department of Energy

USDOE Office of Science (SC), Biological and Environmental Research (BER) (SC-23)

DISCLAIMER

The ARM Data-oriented Metrics and Diagnostics Package for Climate Models

- A New Tool for Evaluating Climate Models with Field Data

C. Zhang, S. Xie, C. Tao, S. Tang, T. Emmenegger, J. D. Neelin, K. A. Schiro, W. Lin and Z. Shaheen

Affiliations:

Zhang, Xie, Tao, Tang, Shaheen – Lawrence Livermore National Laboratory, Livermore, California;

Emmenegger, Neelin – University of California, Los Angeles, California;

Schiro – Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California;

Lin – Brookhaven National Laboratory, Upton, New York;

**Now at Google LLC.*

Corresponding author: *Chengzhu Zhang, Atmospheric, Earth and Energy Division (L-103), Lawrence Livermore National Laboratory, Livermore, CA 94550.*

E-mail: zhang40@llnl.gov

To be submitted to *BAMS*

Category: *NOWCAST*

Section: *Product Announcements*

ABSTRACT. The U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) program produces ground-based long-term contiguous unique measurements for atmospheric state, precipitation, turbulent fluxes, radiation, aerosol, cloud and the land surface, which are collected at its multiple Climate Research Facilities. These comprehensive datasets have been widely used to calibrate climate models and are proven to be invaluable for climate model development and improvement. This article introduces an evaluation package to facilitate the use of ground-based ARM measurements in climate model evaluation.

The ARM data-oriented metrics and diagnostics package (ARM-DIAGS) includes both ARM observational datasets and a Python-based analysis toolkit for computation and visualization. The observational datasets are compiled from multiple ARM data products and specifically tailored for use in climate model evaluation. In addition, ARM-DIAGS also includes simulation data from models participating the Coupled Model Inter-comparison Project (CMIP), which will allow climate-modeling groups to compare a new, candidate version of their model to existing CMIP models. The analysis toolkit is designed to make the metrics and diagnostics quickly available to the model developers.

1. Introduction

A set of standard metrics and diagnostics provides an effective way for climate modeling centers to routinely assess their model performance and judge the improvement of model simulations from new parameterizations. In the past, climate model developers have often relied on satellite remote sensing products to calibrate and tune their models. Satellite data sets provide a great global coverage, however, it is difficult to apply satellite data in some process studies due to their poor temporal resolutions. Therefore, utilizing detailed high-frequency ground-based measurements for a comprehensive collection of quantities can be a complementary test in model evaluation.

Over the past three decades, the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) program has established several permanent research sites and deployed a number of ARM Mobile Facilities (AMF) in diverse climate regimes around the world to collect long-term continuous field measurements of clouds, aerosols, and radiation and their associated large-scale environments. These detailed field observations have provided a unique observational basis specifically for understanding cloud and precipitation related processes and evaluating and improving their representations in climate models. However, ARM data have not been extensively utilized in current model development workflows. With the growing interest in the climate modeling community in developing process-oriented metrics and diagnostics to aid parameterization developments (Maloney et al. 2019), the high-frequency process-oriented ARM observations should play a more important role in the future metrics and diagnostics development.

In this article, we introduce the recently developed ARM data-oriented metrics and diagnostics package (ARM-DIAGS) for the global climate community to facilitate the use of ARM field data in climate model evaluation. The focus is on ARM unique observations on clouds and aerosols, as

well as process-oriented diagnostics that are particularly aimed to improve the representation of cloud and precipitation related processes in climate models, such as those included in the Coupled Model Inter-comparison Project (CMIP). The package is available publicly with the hope that it can serve as an easy entry point for climate modelers to compare their models with ARM data and supplemented CMIP datasets.

2. Overview of the ARM data-oriented metrics and diagnostics package

The ARM-DIAGS development closely follows the CMIP protocol to efficiently distribute ARM metrics and diagnostics package along with other metrics packages to the CMIP community and other climate modeling centers. For this purpose, the diagnostic toolkit is built with the Python programming language and utilizes Python libraries for scientific analysis (such as NumPy and matplotlib). Additional Python packages developed by DOE (i.e., the Community Data Analysis Tools (CDAT), https://cdat.llnl.gov/) are also used. Four components are currently included in the ARM-DIAGS: 1) a Python-based analysis program; 2) an ARM-based collection of mean and diurnal and seasonal cycle climatologies as well as high time frequency data for process-oriented diagnostics; 3) a database of simulation data from models contributed to CMIP project; and 4) relevant technical documentations for ARM-DIAGS.

The observations used to assess model performance primarily rely on the ARM Best Estimate (ARMBE) data products (Xie et al. 2010) and other ARM value-added products (VAPs) (https://www.arm.gov/capabilities/vaps), which are available for all the ARM permanent research sites and some ARM mobile facilities. These data often rely on measurements at the ARM Central Facility (CF) locations (i.e., single point measurements). To improve model-observation comparison, the ARM long-term continuous forcing data (Xie et al. 2004), which represents an

average over a Global Climate Model (GCM) grid box, is also used when it is available. For cloud properties such as cloud liquid and ice water contents, the ARM Cloud Retrieval Ensemble Data (ACRED) (Zhao et al. 2012) is used. The detailed information about ARM data used in the ARM-DIAGS package is listed in Tables 1a and 1b. The observational data product consists of hourly averaged, diurnal cycle, monthly means or climatological summaries of the measured quantities, with variable names, units and vertical dimensions remapped to CMIP convention. They are currently available for the Southern Great Plains (SGP) site (Table 1a) as well as the North Slope of Alaska (NSA) Barrow site and the Tropical Western Pacific (TWP) Manus, Nauru, and Darwin sites (Table 1b). Other than the ARM observations, ARM-DIAGS also includes simulation data from models participating CMIP project, which will allow climate-modeling groups to compare a new candidate version of their model to existing CMIP models. A full list of metrics and diagnostics are as follows, with a subset demonstrated in Section 3 of this article.

- A set of basic metrics tables: mean, mean bias, correlation and root mean square error based on annual cycle of each variable.
- Line plots and Taylor Diagrams (Taylor, 2001) for annual cycle variability of each variable.
- Contour and vertical profiles of annual cycle and diurnal cycle of cloud fraction.
- Line and harmonic dial plots (Covey et al. 2016) of diurnal cycle of precipitation.
- Probability Density Function plots of precipitation rate (Pendergrass et al. 2014)
- Convection onset metrics showing statistical relationship between precipitation rate and column water vapor (Schiro et al. 2016)

3. Facilitating use of ARM data in climate model evaluation

Diagnosis of summertime warm bias. The data and diagnostics provided through ARM-DIAGS have been used for studying the systematic warm bias in surface temperature found among the climate models in summertime at mid-latitude continent including the ARM SGP site (Zhang et al. 2018). The biases are consistent with both overestimated surface shortwave radiation and underestimated evaporative fraction, which contribute to the warm bias as illustrated in Figure 1. These diagnostics provide an integrated picture with detailed field observations to identify possible model deficiency in representing cloud, radiation, and land properties, as well as their interactions.

Diurnal cycle of cloud fraction: This daily cycle could serve as a critical test of the models' representation of the physical processes controlling cloud life cycle. One unique product from ARM is cloud vertical profile measurements derived from an integration of multiple active remote sensors, including Millimeter wavelength cloud radars, laser ceilometers and Micropulse Lidars (Active Remote Sensing of Clouds product, ARSCL). Figure 2 shows a comparison between observed and simulated diurnal cycle of cloud vertical structure over the ARM midlatitude and tropical sites (i.e., SGP and Manus), where prominent climatological diurnal cycle of clouds is present. Over the SGP site, a lack of shallow to deep cloud transition during summertime (June-July-August) is shown in E3SM. This is a common model bias which is related to model deep convection that is triggered too easily and does not allow low clouds to build-up. The Manus site exhibits a strong diurnal cycle, with a maximum in low cloud fraction occurring at early local noon and followed by a maximum in high cloud hours later. Similarly, the model in general underestimated the lower cloud and overestimated high cloud, which is also lack of diurnal variability.

Diurnal cycle of precipitation. Diurnal cycle of precipitation is often served as a benchmark for climate models. The diurnal cycle diagnostics in ARM-DIAGS, which compare the precipitation intensity and its peak time, have been utilized by the Energy Exascale Earth System Model (E3SM) development team to assess the performance of a newly developed convection triggering mechanism (Xie et al. 2019). Figure 3 shows that all climate models including the default E3SM are not able to capture the observed nocturnal peak which is often associated with the eastward propagation of mesoscale convective systems. A recently developed convective triggering function, which incorporates an empirical dynamic constraint and allows elevated convection to be captured, started to pick up the early morning peak time, although the intensity is still too weak. These diagnostics are useful to repeat continually, especially when new features in convection parameterizations are implemented.

Precipitation distribution. The probability density function analysis for daily mean precipitation at the SGP site during June-July-August is shown in Figure 4. This example illustrates that models tend to underestimate heavy rainfall (>10 mm per day) occurrences which contribute more to the total precipitation amount than lighter rainfall occurrence. The overlaying result from GPCP (Global Precipitation Climatology Project One-Degree Daily Precipitation Data Set) also confirms this systematic model bias.

Convection onset metrics. Convection onset metrics allow users to compare diagnostics for the behavior of deep convection from ARM observations to model output. The statistics quantify robust relationships between precipitation, column water vapor (CWV) and temperature. This includes the sharp increase or "pickup" in conditional-average precipitation rate above a critical CWV value seen in Figure 5a, which is easily identifiable for short time averages at tropical ARM

sites. The pickup represents the onset of conditional instability yielding strong convective precipitation (Schiro et al. 2016) and is also seen in the probability of precipitation (Figure 5b). The probability density of CWV and the contribution from precipitating points (Figure 5c) have a drop in probability density at high CWV corresponding to the regime with high precipitation loss above critical.

The statistics discussed here can distinguish between models' convective parametrizations and their qualitative characteristics are robust to space and time averaging (Kuo et al. 2018). An example of model comparison is given in Figure 5. An important diagnostic in the model evaluation of convection onset concerns the critical CWV value where the precipitation pickup begins. Many models exhibit a pickup at lower CWV than observations (Kuo et al. 2020), as seen in Figure 5a for E3SM. This mismatch persists even when temperature dependence (not shown but will be included in a feature release of ARM-DIAGS) is included by binning by the saturation water vapor.

4. Summary and future work

The ARM metrics and diagnostics package is designed and developed to facilitate the use of ARM ground-based in-situ measurements in climate model evaluation. Metrics and diagnostics evaluating the simulated atmospheric and cloud fields are generated by running a Python program in a simple software environment based on CDAT. The v2.0 ARM-DIAGS's analysis codes are currently publicly available through GitHub (https://github.com/ARM-DOE/arm-gcm-diagnostics) under the ARM User Facility project space. Analysis data include ARM observational datasets and the reference CMIP5 AMIP data can be downloaded through the ARM archive (https://www.arm.gov/capabilities/vaps/adcme-123). For now, the default requirement for the

input model is that the data are under CMIP conventions. Anyone interested in applying ARM-DIAGS to a specific model should contact the development team via our GitHub page for specific configurations for a model run.

Future work includes extending the ARM-DIAGS to the ARM Eastern North Atlantic (ENA) site (a new fixed site) and ARM AMF sites. CMIP6 data will be included as it becomes available. The diagnostics suite will be continuously improved with close collaboration with scientists in the field. Ongoing work includes incorporation of the recently developed ARM cloud radar simulator (Y. Zhang et al. 2018) into ARM-DIAGS to improve the comparison between model clouds and ARM cloud radar observations, as well as adding temperature dependence to convection onset statistics. In addition, utilizing other sources of observations, such as those retrieved from satellites, as supplementary data, can help address issues associated with observation uncertainty and data resolution.

Acknowledgement: This research was supported by the DOE Atmospheric Radiation Measurement (ARM) program and performed under the auspices of the U. S. Department of Energy by Lawrence Livermore National Laboratory under contract No. DE-AC52-07NA27344. IM release: LLNL-ABS-789643. Work at UCLA was supported by U.S. Department of Energy Grant DE-SC0011074 and subcontract B634021 and National Science Foundation AGS-1540518, AGS-1936810. We acknowledge the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison to provide coordinating support and lead development of software infrastructure in partnership with the Global Organization for Earth System Science Portals for making CMIP data available.

For Further Reading:

Bond, D. (2005), Soil Water and Temperature System (SWATS) Handbook, ARM Technical Report TR-063, U.S. Department of Energy,Washington, D. C.

Cook, D. R. (2011a). Energy Balance Bowen Ratio Station (EBBR) Handbook (No. DOE/SC-ARM/TR-037). DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States).

Cook, D. R. (2011b). Eddy Correlation Flux Measurement System (ECOR) Handbook (No. DOE/SC-ARM/TR-052). DOE Office of Science Atmospheric Radiation Measurement (ARM) Program (United States).

Covey, C., Gleckler, P. J., Doutriaux, C., Williams, D. N., Dai, A., Fasullo, J., ... & Berg, A. (2016). Metrics for the diurnal cycle of precipitation: Toward routine benchmarks for climate models. *Journal of Climate*, *29*(12), 4461-4471.

Clothiaux, E. E., T. P. Ackerman, G. G. Mace, K. P. Mo- ran, R. T. Marchand, M. Miller, and B. E. Martner, 2000: 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.

Chandra, A. S., C. Zhang, S. A. Klein, and H.-Y. Ma (2015), Low-cloud characteristics over the tropical western Pacific from ARM observations and CAM5 simulations, J. Geophys. Res. Atmos., 120, 8953–8970, doi:10.1002/2015JD023369.

Knootz, et al. (2013), Aerosol Optical Depth Value-Added Product, DOE/SC-ARM/TR-129, USDOE Office of Science, Washington, D. C.

Kuo, Y. H., Schiro, K. A., & Neelin, J. D. (2018). Convective transition statistics over tropical oceans for climate model diagnostics: Observational baseline. *J. Atmos. Sci.*, *75*, 1553-1570.

Kuo, Y. H., Neelin, J. D., Chen, C. C., Chen, W.T , Donner, L. J. , Gettelman, A , Jiang, X , Kuo, K. T., Maloney, E. , Mechoso, C. R. , Ming, Y., Schiro, K. A., Seman, C. J. , Wu, C. M., & Zhao, M. (2020). Convective transition statistics over tropical oceans for climate model 2 diagnostics: GCM evaluation, *J. Atmos. Sci.*, *77*, 379-403.

Long, C. N. and Y. Shi, (2006): The QCRad Value Added Product: Surface Radiation Measurement Quality Control Testing, Including Climatologically Configurable Limits, Atmospheric Radiation Measurement Program Technical Report, ARM TR-074, 69 pp.

Long, C. N., and Y. Shi, (2008): An Automated Quality Assessment and Control Algorithm for Surface Radiation Measurements, TOASJ, 2, 23-37, doi: 10.2174/1874282300802010023.

Maloney, E. D., Gettelman, A., Ming, Y., Neelin, J. D., Barrie, D., Mariotti, A., ... & Annamalai, H. (2019). Process-Oriented Evaluation of Climate and Weather Forecasting Models. *Bulletin of the American Meteorological Society*, *100*(9), 1665-1686.

Pendergrass, A. G., & Hartmann, D. L. (2014). Two modes of change of the distribution of rain. Journal of Climate, 27(22), 8357-8371.

Schiro, K. A., Neelin, J. D., Adams, D. K., & Lintner, B. R. (2016). Deep convection and column water vapor over tropical land versus tropical ocean: A comparison between the Amazon and the tropical western Pacific. Journal of the Atmospheric Sciences, 73(10), 4043-4063.

Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res., 106(D7), 7183– 7192, doi:10.1029/2000JD900719.

Xie, S., Wang, Y. C., Lin, W., Ma, H. Y., Tang, Q., Tang, S., ... & Zhang, M. (2019). Improved Diurnal Cycle of Precipitation in E3SM with a Revised Convective Triggering Function. Journal of Advances in Modeling Earth Systems.

Xie, S., McCoy, R. B., Klein, S. A., Cederwall, R. T., Wiscombe, W. J., Jensen, M. P., ... & Mather, J. H. (2010). Clouds and more: ARM climate modeling best estimate data: A new data product for climate studies. Bulletin of the American Meteorological Society, 91(1), 13-20.

Xie, S. C., R. T. Cederwall, and M. H. Zhang (2004), Developing long-term single-column model/cloud system-resolving model forcing data using numerical weather prediction products constrained by surface and top of the atmosphere observations, Journal of Geophysical Research-Atmospheres, 109(D1), doi: 10.1029/2003jd004045.

Zhang, C., and Xie, S., (2017). ARM Data-Oriented Metrics and Diagnostics Package for Climate Model Evaluation Value-Added Product. Ed. by Robert Stafford, ARM Climate Research Facility. DOE/SC-ARM-TR-202.

Zhang, C., Xie, S., Klein, S. A., Ma, H. Y., Tang, S., Van Weverberg, K., ... & Petch, J. (2018). CAUSES: diagnosis of the summertime warm bias in CMIP5 climate models at the ARM southern great plains site. Journal of Geophysical Research: Atmospheres, 123(6), 2968-2992.

Zhang, Y., Xie, S., Klein, S. A., Marchand, R., Kollias, P., Clothiaux, E. E., ... & Tang, S. (2018). The ARM Cloud Radar Simulator for Global Climate Models: Bridging Field Data and Climate Models. *Bulletin of the American Meteorological Society*, *99*(1), 21-26.

Zhao, C., Xie, S., Klein, S. A., Protat, A., Shupe, M. D., McFarlane, S. A., ... & Hogan, R. J. (2012). Toward understanding of differences in current cloud retrievals of ARM ground- based measurements. *Journal of Geophysical Research: Atmospheres*, *117*(D10).

Figure Captions Lists:

Figure 1: Annual cycle of monthly mean of (a) surface air temperature, (b) precipitation, (c) surface air relative humidity, (d) surface downward shortwave radiative flux, (e) surface sensible flux, and (f) surface latent heat flux over the ARM SGP domain (averaged over 35-38°N, 99-96°W) from ARM observation averaged over 1999 to 2011 (red line with error bars representing one standard deviation of inter-annual variability) and CMIP5 simulations averaged over 1979 to 2008 (grey lines for individual CMIP5 models and black line for multi-model mean). JJA mean values are shown in the legend. Plots are modified from Zhang et al. (2018).

Figure 2: Climatological composite diurnal cycle of clouds from observed (left panel) and simulated by E3SM (right panel). JJA mean at SGP (row 1); Annual mean at Manus (row 2).

Figure 3: Left: black dots are ARM observation. Curves are the first harmonics: grey for CMIP5 model AMIP type of runs. Color curves are from DOE's E3SM Atmosphere Model (EAM v1) with a standard control run and a run using newly developed convection triggers (Detailed experiment description can be found in (Xie et. al., 2019). Right: mapping precipitation peak time and amplitude from the first harmonics to polar coordinate.

Figure 4: Daily-Mean precipitation frequency (a) and precipitation amount (b) as a function of precipitation rate using observations from ARM (blue line) and GPCP (red line) compared with CMIP5 AMIP simulations shown as gray lines. The black line represents the multi-model mean. The precipitation bin arrangement follows method described by Pendergrass et al. (2014).

Figure 5: (a) Precipitation conditionally averaged on column water vapor (CWV) for observations indicated by color-coded dots (ARMBE precipitation and gap filled MWRRET radiometer CWV) and E3SM model output (black) over Manus Island. (b) as in (a) but for precipitation probability. (c) The PDFs of CWV for observations (dark blue) and model (black) and of the contribution to this from points with precipitation exceeding 0.5 mm/hr for observations (light blue) and model (grey).

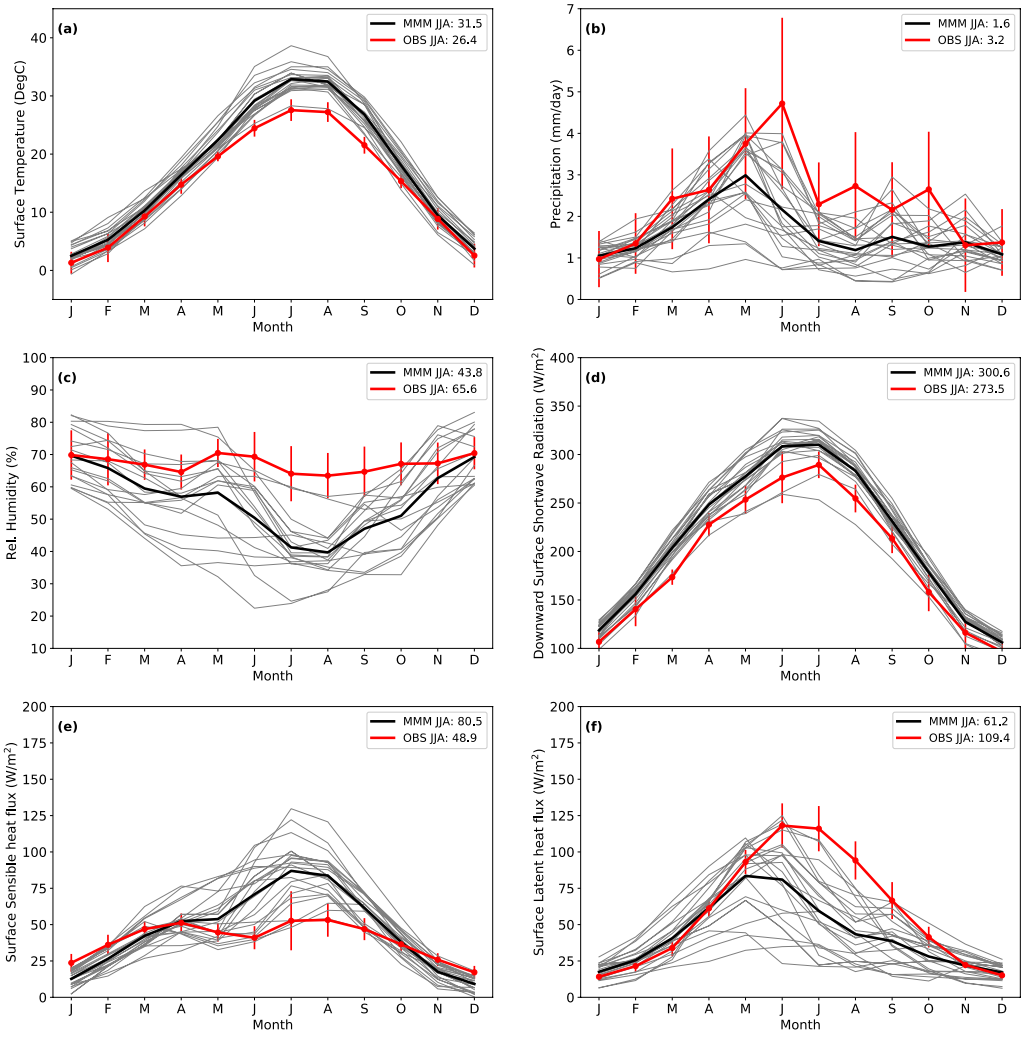
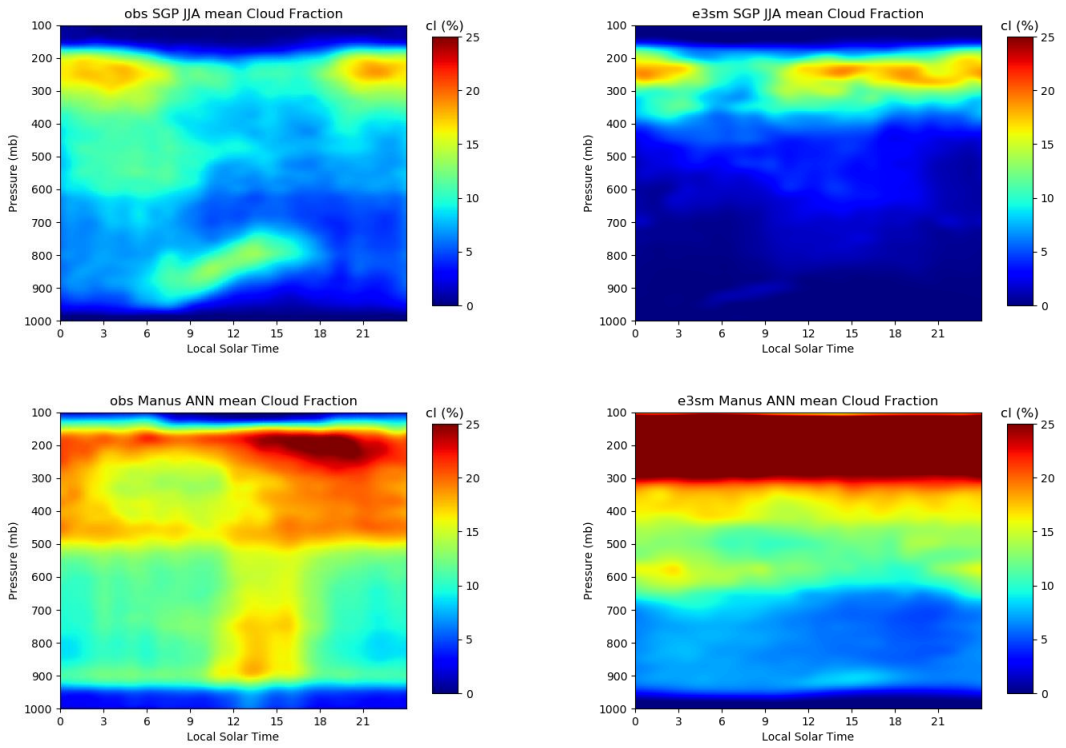


Figure 1: Annual cycle of monthly mean of (a) surface air temperature, (b) precipitation, (c) surface air relative humidity, (d) surface downward shortwave radiative flux, (e) surface sensible flux, and (f) surface latent heat flux over the ARM SGP domain (averaged over 35-38°N, 99-96°W) from ARM observation averaged over 1999 to 2011 (red line with error bars representing one standard deviation of inter-annual variability) and CMIP5 simulations averaged over 1979 to 2008 (grey lines for individual CMIP5 models and black line for multi-model mean). JJA mean values are shown in the legend. Plots are modified from Zhang et al. (2018).



329 Figure 2: Climatological composite diurnal cycle of clouds from observed (left panel) and
330 simulated by E3SM (right panel). JJA mean at SGP (row 1); Annual mean at Manus (row 2).

331

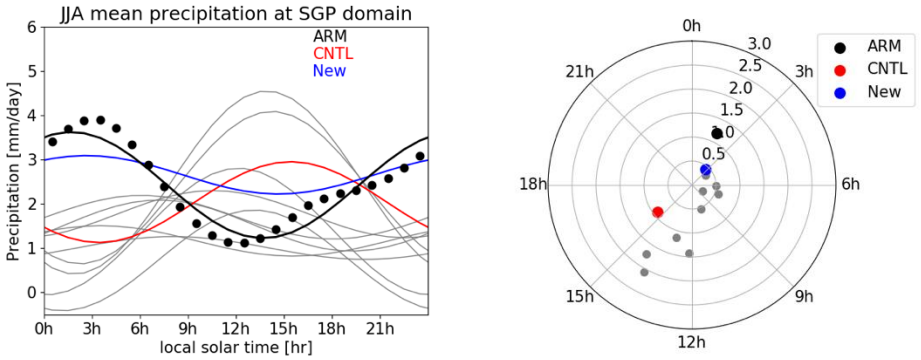
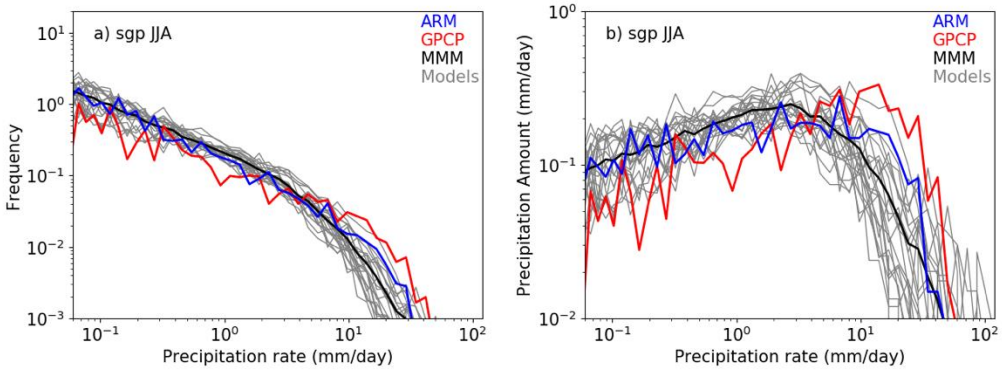


Figure 3: Left: black dots are ARM observation. Curves are the first harmonics: grey for CMIP5 model AMIP type of runs. Color curves are from DOE's E3SM Atmosphere Model (EAM v1) with a standard control run and a run using newly developed convection triggers (Detailed experiment description can be found in (Xie et. al., 2019). Right: mapping precipitation peak time and amplitude from the first harmonics to polar coordinate.



341

342 Figure 4: Daily-Mean precipitation frequency (a) and precipitation amount (b) as a function of
343 precipitation rate using observations from ARM (blue line) and GPCP (red line) compared with
344 CMIP5 AMIP simulations shown as gray lines. The black line represents the multi-model mean.
345 The precipitation bin arrangement follows method described by Pendergrass et al. (2014).

346

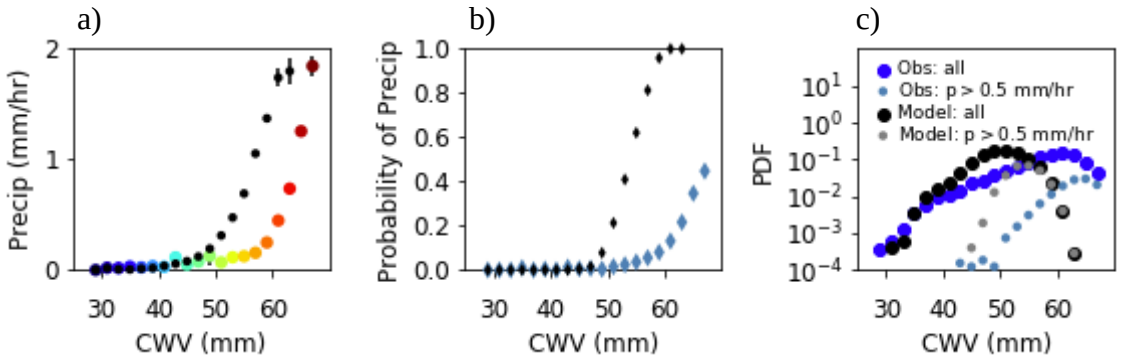


Figure 5: (a) Precipitation conditionally averaged on column water vapor (CWV) for observations indicated by color-coded dots (ARMBE precipitation and gap filled MWRRET radiometer CWV) and E3SM model output (black) over Manus Island. (b) as in (a) but for precipitation probability. (c) The PDFs of CWV for observations (dark blue) and model (black) and of the contribution to this from points with precipitation exceeding 0.5 mm/hr for observations (light blue) and model (grey).

355 Table 1a

Quantities	M Data Products	Data Source/ Instruments	Time resolution	Spatial info
Surface Screen-Level Temperature/ Humidity	ARM Continuous forcing dataset	Surface Meteorological Observation System (SMOS), Oklahoma and Kansas mesonet stations (OKM and KAM)[Xie et al. 2004]	mon, day, hr	sgp domain averaged
Temperature/Humidity profile/wind speed/large scale tendencies	Same as above	NOAA/ NCEP Rapid Update Cycle (RUC) analysis data [Xie et al. 2004]	mon, day, hr	sgp domain averaged
Surface Precipitation	Same as above	Arkansas-Red Basin River Forecast Center (ABRFC) Nexrad radar precipitation estimates w/ rain gauge	mon, day, hr	sgp domain averaged
Precipitable Water	Same as above	Microwave Radiometer (MWR) water liquid & vapor along line of sight (LOS) path (MWRLOS)	mon, day, hr	sgp domain averaged
Surface All Sky Radiative Fluxes	Same as above	Data Quality Assessment for ARM Radiation Data (QCRAD) [Long and Shi, 2006, 2008]	mon, day, hr	sgp domain averaged
Aerosol Optical Depth 550nm	MFRSRAOD1MICH	Multifilter Rotating Shadowband Radiometer (MFRSR) [Knootz et al.,2013]	mon	Averaged over sgp Site C1 and E13
Surface Latent/Sensible Heat	BAEBBR	Best-Estimate Fluxes From EBBR Measurements and Bulk Aerodynamics Calculations (BAEBBR) [Cook, 2011a]	mon	sgp domain averaged
	QCECOR	Quality Controlled Eddy Correlation Flux Measurement [Cook, 2011b]	mon	sgp domain averaged
Surface Soil Moisture Content (10 cm)	SWATS	Soil Water and Temperature System (SWATS)) [Bond, 2005]	mon	sgp domain averaged
Cloud Fraction	ARSCL	Active Remote Sensing of Clouds [Clothiaux et al, 2001]	mon, day, hr	sgp Site C1
Ice Water Content/Liquid Water Content	ACRED	ARM Cloud Retrieval Ensemble Dataset [Zhao et al. 2012]	mon, day, hr	sgp Site C1

356
357

358

359 Table 1b

Quantities	ARM Data Products	Data Source/ Instruments	Time resolution	Spatial info
Surface Screen-Level Temperature/ Humidity	ARMBE-ATM	ARM-standard meteorological instrumentation at the surface [Xie et al. 2010]	mon	twp C1; nsa C1
Surface Precipitation	ARMBE-ATM	Same as above	mon, hr	twp C1; nsa C1
Precipitable Water	ARMBE-ATM	Microwave Radiometers Retrievals (MWRRET) [Xie et al. 2010]	mon, hr	twp C1; nsa C1
Surface Radiative Fluxes	ARMBE-CLD	Data Quality Assessment for ARM Radiation Data (QCRAD) [Long and Shi, 2006, 2008]	mon	twp C1; nsa C1
Cloud Fraction	ARSCL	Active Remote Sensing of Clouds [Clothiaux et al, 2001]	mon, hr	twp C1; nsa C1

360

361 Table 1: Observed quantities selected in the evaluation package, including the quantity names,

362 the data sources, and the temporal and spatial information of the data. (a) for SGP and (b) for

363 NSA and TWP sites.