

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25

**A Framework for Process-Oriented Evaluation of  
Climate and Weather Forecasting Models**

The NOAA MAPP Model Diagnostics Task Force

Eric D. Maloney<sup>1</sup>, Andrew Gettelman<sup>2</sup>, Yi Ming<sup>3</sup>, J. David Neelin<sup>4</sup>, Daniel Barrie<sup>5</sup>, Annarita Mariotti<sup>5</sup>, C.-C. Chen<sup>2</sup>, Yi-Hung Kuo<sup>4</sup>, Bohar Singh<sup>1</sup>, H. Annamalai<sup>6</sup>, Alexis Berg<sup>7</sup>, James F. Booth<sup>8</sup>, Suzana J. Camargo<sup>9</sup>, Aiguo Dai<sup>10</sup>, Alex Gonzalez<sup>11</sup>, Jan Hafner<sup>6</sup>, Xianan Jiang<sup>11</sup>, Xianwen Jing<sup>12</sup>, Daehyun Kim<sup>13</sup>, Arun Kumar<sup>14</sup>, Yumin Moon<sup>13</sup>, Catherine M. Naud<sup>15</sup>, Adam H. Sobel<sup>9,16</sup>, Kentaroh Suzuki<sup>12</sup>, Fuchang Wang<sup>17,18</sup>, Junhong Wang<sup>10</sup>, Allison A. Wing<sup>19</sup>, Xiaobiao Xu<sup>17</sup>, Ming Zhou<sup>3</sup>

*Bulletin of the American Meteorological Society*

Submitted

1. *Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado*
2. *National Center for Atmospheric Research, Boulder, Colorado*

- 26 3. *Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey*
- 27 4. *Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles,*  
28 *Los Angeles California*
- 29 5. *National Oceanic and Atmospheric Administration, Climate Program Office, Silver Spring,*  
30 *Maryland*
- 31 6. *International Pacific Research Center/Department of Oceanography, University of Hawaii*  
32 *at Manoa, Honolulu, Hawaii*
- 33 7. *Department of Civil and Environmental Engineering, Princeton University, Princeton, New*  
34 *Jersey*
- 35 8. *City University of New York, City College, New York, New York*
- 36 9. *Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York*
- 37 10. *Department of Atmospheric and Environmental Sciences, University at Albany, SUNY,*  
38 *Albany, New York*
- 39 11. *Joint Institute for Regional Earth System Science & Engineering, University of California,*  
40 *Los Angeles, California*
- 41 12. *Atmosphere and Ocean Research Institute, University of Tokyo, Kashiwa, Japan*
- 42 13. *Department of Atmospheric Sciences, University of Washington, Seattle, Washington*
- 43 14. *Climate Prediction Center, National Oceanic and Atmospheric Administration, College*  
44 *Park, Maryland.*
- 45 15. *Columbia University, New York, New York*
- 46 16. *Department of Applied Physics and Applied Mathematics, Columbia University, New York,*  
47 *New York*

48 *17. Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee,*  
49 *Florida*

50 *18. Institute of Oceanography, Shanghai Jiao Tong University, Shanghai, China*

51 *19. Department of Earth, Ocean, and Atmospheric Science, Florida State University,*  
52 *Tallahassee, Florida*

53

54 Corresponding Author Address: Eric D. Maloney, Department of Atmospheric Science,  
55 Colorado State University, 1371 Campus Delivery, Fort Collins, CO 80523-1371. Email:  
56 [emaloney@atmos.colostate.edu](mailto:emaloney@atmos.colostate.edu)

57

*Capsule*

58 A process-oriented diagnostics framework has been developed by the NOAA MAPP Model

59 Diagnostics Task Force to aid model improvement.

60

61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82

*Abstract*

Realistic climate and weather prediction models are necessary to produce confidence in projections of future climate over many decades and weather forecasts for days to seasons. These models must be physically justified and validated for multiple weather and climate processes. A key need is greater incorporation of process-oriented diagnostics into standard diagnostic packages that can be applied during the model development process, allowing the application of diagnostics to be repeatable across multiple model versions and used as a benchmark for model improvement. A "process-oriented diagnostic" characterizes a specific physical process or emergent behavior that is related to the ability to simulate an observed phenomenon. The Model Diagnostics Task Force (MDTF) under the NOAA Climate Program Office (CPO) Modeling, Analysis, Predictions and Projections (MAPP) program has created an open-source diagnostic package that is portable, community extensible and usable to aid application of process-oriented diagnostics to the model development process. This paper will first describe the MDTF and modeling center perspectives on the need for expanded process-oriented diagnosis of models. Then, the basic framework of the applied programming interface developed by the MDTF that underlies this diagnostics package will be presented. Initial process-oriented diagnostics contributed to this package by the MDTF are also described. The paper concludes with a discussion of the path forward for the MDTF effort and for community process-oriented diagnosis in general.

## 83 INTRODUCTION

84 Realistic climate and weather forecasting models with formulations grounded in sound  
85 physical principles are necessary to produce confidence in projections of future climate for the  
86 next century and weather forecasts for days to seasons. However, global models continue to  
87 suffer from important and often common biases that may impact their ability to provide reliable  
88 representations of weather and future climate. These include biases in the cold tongue and  
89 Intertropical Convergence Zone regions (e.g. Li and Xie 2014; Grose et al. 2014), the structure  
90 of El Niño-Southern Oscillation (ENSO) sea surface temperature (SST) and precipitation  
91 anomalies (e.g. Bellinger et al. 2014; Grose et al. 2014), simulation of the Madden-Julian  
92 oscillation (MJO; Kim et al. 2014; Hung et al. 2013; Jiang et al. 2015; Ahn et al. 2017), the  
93 strength of the Atlantic meridional overturning circulation (AMOC, e.g. Wang et al. 2014),  
94 extratropical cyclone tracks (Zappa et al. 2013), tropical-extratropical teleconnections (e.g.  
95 Sheffield et al. 2013a,b; Henderson et al. 2017), and general interactions of clouds with the  
96 large-scale circulation (Stevens and Bony 2013), among others. Some aspects of model  
97 simulations can often be improved, but seemingly for the wrong reasons. For example,  
98 improving biases in model tropical intraseasonal variability often systematically degrades other  
99 aspects of the simulation like the mean state (Kim et al. 2011; Mapes and Neale 2011; Hannah  
100 and Maloney 2014). Model biases are rooted in imperfect representation in parameterizations of  
101 unresolved processes.

102 The climate and weather forecasting communities have a long-standing and high interest  
103 in conducting process studies and applying process-oriented diagnoses that are designed to  
104 inform parameterization improvements to address these long-standing model biases (e.g. Eyring  
105 et al. 2018). A "process-oriented diagnostic" characterizes a specific physical process or

106 emergent behavior that is hypothesized to be related to the ability to simulate an observed  
107 phenomenon. An example of an observed phenomenon is the intraseasonal variability of tropical  
108 convection, as could be measured by an index or a power spectra of precipitation variance in the  
109 tropics. Process-oriented diagnostics representing the sensitivity of atmospheric convection to  
110 free-tropospheric humidity have been developed that demonstrate a strong coupling between  
111 convection and moisture on daily timescales, which are also able to discern models with strong  
112 and weak intraseasonal variability (e.g. Kim et al. 2014). Process-oriented diagnostics can be  
113 used to compare models to observations in a way that focuses model improvement on specific  
114 processes. Evaluating the new model configurations against these observations can give insight  
115 into whether a particular process is being well represented, help to ensure that models produce  
116 the right answers for the right reasons, expose potential cancellation of errors, and identify gaps  
117 in the understanding of phenomena. Process oriented “metrics” are scalar quantities that can be  
118 derived from some diagnostics.

119         A key need is incorporation of process-oriented diagnostics into standard diagnostics  
120 packages that are applied to development versions of the models, allowing the application of  
121 diagnostics to be repeatable across multiple model versions and throughout model generations  
122 and be used for rapid analysis of intermediate model versions during the development process. A  
123 significant barrier is the lack of a mechanism for getting community-developed diagnostics into  
124 the modeling center development process. This paper describes an effort by the National Oceanic  
125 and Atmospheric Administration (NOAA) Modeling, Analysis, Prediction, and Projections  
126 program (MAPP) Model Diagnostics Task Force (MDTF) to create a diagnostic package that is  
127 portable, extensible, and open for contribution from the community. The diagnostics developed  
128 here will also ideally be compatible with other diagnostic efforts such as European Earth System

129 Model Bias Reduction and Assessing Abrupt Climate Change (EMBRACE) project/Earth  
130 System Model eValuation Tool (ESMValTool) and Coordinated set of Model Evaluation  
131 Capabilities (CMEC) described below that use open source software frameworks for multimodel  
132 evaluation. Because most other efforts have mainly emphasized basic performance metrics for  
133 models to this stage, the effort described here will be highly complementary and advantageous to  
134 these other efforts as they seek to expand their process-oriented diagnostic capabilities.

135 This paper is the centerpiece of an American Meteorological Society (AMS) special  
136 collection spanning several AMS journals devoted to process-oriented evaluation of climate and  
137 Earth system models. Other articles in this collection describe the scientific basis for individual  
138 diagnostics, while this centerpiece paper provides a synthesis and overview of the broader  
139 diagnostics package, with a shorter summary of individual diagnostics. The second section  
140 describes existing institutional efforts and needs, including details on the MDTF and modeling  
141 center perspectives. Existing and developing efforts at process-oriented diagnosis are then  
142 discussed. The integrative process-oriented diagnostics framework and associated flow of the  
143 MDTF package are described in the fourth section. The fifth section provides examples of key  
144 process-oriented diagnostics and metrics developed by the MDTF. The last section provides a  
145 summary and a path forward for process-oriented diagnostics.

146

## 147 **THE NOAA MAPP MDTF AND MODELING CENTER NEEDS**

### 148 *Brief summary of MDTF*

149 In 2015 and 2018, NOAA's MAPP program solicited projects to develop process-  
150 oriented diagnostics for model development. These funded projects and the investigators leading  
151 them ultimately constituted the NOAA MDTF. At the time of the initial proposal solicitation, the

152 global modeling community was between cycles of the Coupled Model Intercomparison Project  
153 (CMIP), and modeling centers were actively moving on from their CMIP5 class models toward  
154 developing and testing their CMIP6-class models. Performance evaluations and analyses of  
155 CMIP5 class models were becoming less useful for informing next-generation model  
156 development activities.

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

170 The MDTF has engaged over 50 scientists from 6 laboratories and operational centers  
171 and 15 academic institutions. Initially, the MDTF was led by Eric Maloney (Colorado State  
172 University), and co-led by Yi Ming (GFDL), Andrew Gettelman (NCAR), David Neelin  
173 (UCLA), and Aiguo Dai (University at Albany) during its first phase (2015-2018). MDTF  
174 activities have included two major thrusts: (1) designing a software package useable at GFDL

175 and NCAR and flexible enough to incorporate diagnostics from disparate community efforts that  
176 may be written in diverse coding languages; and (2) coordinating and supporting community  
177 development of diagnostics and metrics for a variety of physical systems and modeling and  
178 process areas targeting known model biases including:

- 179 ● Cloud microphysical processes
- 180 ● Tropical and extratropical cyclones
- 181 ● ENSO teleconnections and atmospheric dynamics
- 182 ● Land-atmosphere interactions
- 183 ● MJO moisture, convection, and radiative processes
- 184 ● Precipitation diurnal cycle
- 185 ● AMOC
- 186 ● Arctic sea ice
- 187 ● Lake effect processes
- 188 ● North American Monsoon
- 189 ● Radiative forcing and cloud-circulation feedbacks
- 190 ● Temperature and precipitation extremes

191

192 These diverse, somewhat eclectic focal areas were determined by the submitted  
193 competitive proposals that emerged successfully from the MAPP panel reviews. They also  
194 reflect key model biases that impact climate and climate variability. Continued development of  
195 the software platform and diagnostics is supported through MAPP proposal solicitations. This  
196 competitive solicitation model for advancing the process-oriented activity encourages a bottom-  
197 up design of the diagnostics package and is driven by organic, mutually beneficial interactions

198 between modeling center and academic scientists and staff as opposed to top-down engineered  
199 engagements.

200

201 *Modeling center perspectives*

202 MDTF activities are designed to support model development and the diagnostic  
203 workflow at major modeling centers. One major MDTF goal is to provide a mechanism for  
204 diagnostic development that enables modeling centers to use these diagnostics in their  
205 development cycle. Centers typically have a workflow containing a package of diagnostic  
206 comparisons with model output. This enables rapid analysis of many aspects of a model run. The  
207 method typically is for a large package to be constructed to generate diagnostics that many  
208 different developers may want to look at, enabling a multi-variate and multi-disciplinary  
209 approach to model evaluation. The philosophy behind the MDTF diagnostic package is to  
210 provide an extensible framework for outside scientists to contribute diagnostics for a community  
211 package that can be used by the community, but is also integrated into the workflow at modeling  
212 centers.

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

221

222 **EXISTING PROCESS-ORIENTED DIAGNOSTIC EFFORTS**

223         The MDTF process-oriented diagnostics effort is inspired by, builds upon, and in many  
224 cases is complementary to prior and existing community efforts at model diagnosis. Such  
225 existing efforts that have influenced the MDTF are described here, although this list is likely not  
226 inclusive of all existing community efforts. Individual modeling centers also have their own  
227 diagnostics suites that perform diagnosis in a similar spirit, but for individual models.

228

229 *The WGNE MJO Task Force*

230         The WGNE MJO task force  
231 ([www.wmo.int/pages/prog/arep/wwrp/new/MJO\\_Task\\_Force\\_index.html](http://www.wmo.int/pages/prog/arep/wwrp/new/MJO_Task_Force_index.html)) was developed in  
232 2010 under the Year of Tropical Convection with an explicit goal to “facilitate improvements in  
233 the representation of the MJO in weather and climate models in order to increase the predictive  
234 skill of the MJO and related weather and climate phenomena” (Wheeler et al. 2013). While the  
235 task force has several terms of reference, of particular relevance is the goal to development and  
236 promote process-oriented diagnostics/metrics for the MJO that facilitate model improvement. To  
237 further this goal, the MJO task force has developed several diagnostics including an assessment  
238 of the sensitivity of tropical convection to lower free tropospheric moisture (Kim et al. 2014),  
239 normalized gross moist stability (Benedict et al. 2014; Hannah and Maloney 2014; Jiang et al.  
240 2015), and the strength of cloud-radiative feedbacks (Kim et al. 2015). The process-oriented  
241 diagnostics efforts of the MJO task force were an early inspiration behind the efforts of the  
242 NOAA MDTF, and a mutual benefit is that the framework developed by the MDTF may allow

243 broader community dissemination for diagnostics such as those developed by the MJO task  
244 force.

245

#### 246 *European EMBRACE project/ESMValTool*

247 The European Union funded EMBRACE project has developed a package called the  
248 ESMValTool (Eyring et al 2016a,b). This tool was originally developed from the Chemistry-  
249 Climate Model Validation Activity (CCMVal) diagnostic package (Gettelman et al 2012). The  
250 ESMValTool is a flexible and community oriented diagnostic framework that uses standard  
251 model files as input, similar to the MDTF tool described below, and provides a structured set of  
252 diagnostic output plots. The spirit of the tool is similar to that of the MDTF, and because the  
253 ESMValTool uses similar inputs and a similar structure, diagnostics coded for one tool  
254 (ESMValTool is largely in python, derived from NCL code) should be applicable in the other.  
255 The ESMValTool is increasingly incorporating process-level information.

256

#### 257 *The Coordinated set of Model Evaluation Capabilities*

258 The Coordinated set of Model Evaluation Capabilities (CMEC) is an open source package  
259 incorporating the Program for Climate Model Diagnosis and Intercomparison (PCMDI) Metrics  
260 Package (PMP), the International Land Modeling Benchmarking Project tool (ILAMB), and the  
261 parallel toolkit for extreme climate analysis (TECA). As described in Gleckler et al. (2016), the  
262 PMP currently provides an open source package based on Python and Ultrascale Visualization  
263 Climate Data Analysis Tools (UV-CDAT; e.g. Santos et al. 2013; Williams et al. 2016) that  
264 compares climate models to observations using a set of basic performance metrics and statistics.  
265 The PMP development team is open to working with community users to entrain more

266 diagnostics into the package, and future releases plan to incorporate more extensive model  
267 evaluation based on emergent constraints and process-level diagnosis. ILAMB provides a  
268 framework for evaluating land surface models that includes benchmarking the realism of specific  
269 processes that allow good land surface performance (Luo et al. 2012). It is recognized in the  
270 ILAMB effort that model improvement can only come by providing specific information on the  
271 fidelity of processes acting at a range of time and space scales. TECA is a parallelized software  
272 package in C++ designed to detect extreme climate events in model fields such as tropical and  
273 extratropical cyclones and atmospheric rivers (Prabhat et al. 2012).

274

#### 275 *WCRP Grand Challenges*

276 The World Climate Research Program (WCRP) has developed “Grand Challenges” to focus  
277 near-term efforts in modeling, climate analysis, observations, and general research  
278 (<https://www.wcrp-climate.org/grand-challenges/grand-challenges-overview>) . These challenges  
279 are designed to overcome particularly nettlesome and compelling barriers in climate science that  
280 require sustained focus, and that lend themselves to substantial progress on a 5-10 year  
281 timeframe. Current areas of emphasis include ice melt, carbon cycle feedbacks, weather and  
282 climate extremes, water resources, regional sea level changes and impacts on coasts, near term  
283 climate prediction, and clouds, circulation, and climate sensitivity. All of these areas require  
284 improved process understanding and process-level evaluation of models. Integrative activities  
285 such as the Global Energy and Water Exchanges (GEWEX) project Global Atmospheric System  
286 Studies (GASS) Panel help to accelerate model improvement through development of improved  
287 model parameterization schemes through process research  
288 (<https://www.gewex.org/panels/global-atmospheric-system-studies-panel/>).

289

290 *GEWEX Process Evaluation Study*

291 The GEWEX Process Evaluation Study (PROES) has been launched as a GEWEX-wide  
292 community effort that aims to advance understanding of key climate processes and their  
293 representations in weather prediction and global climate models (Stephens et al. 2015). In  
294 particular, GEWEX-PROES is intended to exploit multiple satellite observations to diagnose the  
295 processes relevant to water and energy balances and thereby to advance the models at a  
296 fundamental process level. The goals of GEWEX-PROES are (i) “to provide a better  
297 understanding of the mechanisms involved in energy and water exchanges on Earth”, (ii) “to  
298 diagnose the sources of major model shortcomings”, and (iii) “to use the knowledge gained in  
299 new treatments of energy and water exchange processes in models” (Stephens et al. 2015).  
300 Although proposed as a GEWEX-based effort, the GEWEX-PROES also seeks strong  
301 connection with other efforts in the climate study community beyond GEWEX, such as the  
302 WCRP Grand Challenges, CMIP, the Cloud Feedback Model Intercomparison Project (CFMIP),  
303 the Observations for Model Intercomparisons Project (obs4MIPs), and WGNE. GEWEX-  
304 PROES is comprised of projects including three main ingredients: (i) collection of data sets that  
305 allow for process diagnostics, (ii) development of diagnostic tools or methodologies constructed  
306 from data that enable process evaluation, and (iii) design and execution of model simulations that  
307 will be analyzed with the diagnostic methodologies applied.

308

309 *CFMIP Diagnostic Codes Catalogue*

310 The CFMIP Diagnostic Codes Catalogue is a showcase of metrics and diagnostics on cloud-  
311 related processes to evaluate their representations in global climate models (Tsushima et al.

312 2017). It is intended to integrate existing methodologies for diagnosing key aspects of the cloud-  
313 climate feedback developed by members of the CFMIP community. This community effort  
314 assembles the metrics and diagnostics in the form of their code repositories that allow open  
315 access. In doing so, this helps facilitate use of the metrics/diagnostics by the wider climate  
316 community and also encourages additional diagnostics to be included in the catalogue as long as  
317 they are documented in peer-reviewed publications and source code is provided. Given that the  
318 effort emerges from CFMIP, the catalogue is intended to serve as a shared toolkit that enhances  
319 analysis of output from CFMIP and CMIP6 model experiments with a particular focus on clouds.

320

#### 321 **THE NOAA MDTF PROCESS-ORIENTED DIAGNOSTICS FRAMEWORK**

322 As alluded to above, a central activity of the NOAA MAPP MDTF is development of a  
323 framework to aid application of process-oriented diagnostics to the model development process.

324 More information and the code itself is available at:

325 [http://www.cesm.ucar.edu/working\\_groups/Atmosphere/mdtf-diagnostics-package/](http://www.cesm.ucar.edu/working_groups/Atmosphere/mdtf-diagnostics-package/). The

326 framework has been developed as a Python code that integrates modules with Process Oriented

327 Diagnostics (PODs) provided by contributing teams. While the Python framework is useful for

328 modeling centers, it is important to emphasize that this is primarily a vehicle to facilitate

329 adoption of the intellectual content — a center with its own diagnostics framework could easily

330 adapt any part into its own interface and workflow. The PODs themselves follow an applications

331 programming interface (API) that specifies how the modules interact with the output from the

332 candidate model version that is being diagnosed. Figure 1 illustrates both the Python framework

333 and the API. Key features include:

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

346 Figure 1b shows an edited example of a webpage from a particular POD, illustrating how  
347 the comparison appears between observations and the model output analyzed by the POD. The  
348 details of the format vary according to the POD, but each provides the developer with model-to-  
349 observation comparison for a process of interest. The set of PODs also provides a work set of  
350 examples of process-oriented diagnostics addressing different processes, each with their unique  
351 requirements and approach.

352 It is also important to acknowledge that not all diagnostics fit conveniently into the  
353 Python framework. Some require specialized output or large data sets that would not routinely be  
354 provided, or must interact with other software at the center, such as cyclone tracking routines.  
355 Such PODs will be provided separately, or in preprocessed form, with instructions for adoption.  
356 Nonetheless these diagnostics form part of the same intellectual framework.

357

## 358 **EXAMPLES OF PROCESS-ORIENTED DIAGNOSTICS MODULES**

359 This section describes the suite of PODs developed by the NOAA MAPP MDTF for climate  
360 model evaluation. These diagnostics are currently implemented or in the process of being  
361 implemented in the MDTF framework. References are provided where expanded documentation  
362 for these diagnostics can be found elsewhere in this special collection or other locations.

363 Development of the PODs framework is an ongoing activity within the NOAA MAPP MDTF,  
364 and hence the status of the framework is continually evolving. For example, the suite of  
365 diagnostics being added to the framework is being continually augmented by the MDTF and  
366 others. We stress that the PODs framework is flexible enough to easily incorporate other  
367 diagnostics contributed by the community.

368

### 369 *Convective transition statistics*

370 Figure 1b shows an example of PODs for the transition between non-precipitating and  
371 precipitating regimes for the tropics, where deep convection dominates precipitation production.  
372 A basic set of diagnostics is shown for precipitation dependence on measures of the water vapor-  
373 temperature environment, evaluated at short timescales comparable to those at which  
374 parameterized convection acts (Neelin et al. 2009; Schiro et al. 2016). Observations (Kuo et al.  
375 2018) and an example model (GFDL) are shown with, left to right, panels for precipitation  
376 conditionally averaged as a function of column water vapor (CWV) for various values of  
377 troposphere-average temperature (colors), probability of precipitation (exceeding a threshold of  
378 0.25 mm/hour), and the probability density function (PDF) of CWV and of CWV for  
379 precipitating points. The sharp pickup of precipitation and probability of precipitation above a

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

391

#### 392 *MJO teleconnection biases*

393 Henderson et al (2017) documented reasons for MJO midlatitude teleconnection errors in  
394 CMIP5 models. Since MJO teleconnections have significant impacts on atmospheric rivers,  
395 blocking, and other extreme events in the midlatitudes, teleconnection errors in models have  
396 important implications for the subseasonal prediction of midlatitude weather extremes (e.g.  
397 Henderson et al. 2016; Mundhenk et al. 2018; Baggett et al. 2017). In addition to extended  
398 analyses of stationary wavenumber biases and use of a linear baroclinic model to diagnose CMIP  
399 model biases, Henderson et al. (2017) developed diagnostics linking teleconnection biases to  
400 biases in the position and extent of the North Pacific jet.

401 Figure 2 from Henderson et al. (2017) contains two panels, each having MJO teleconnection  
402 performance during December-February on the y-axis. In Figure 2a, the x-axis represents an

403 MJO skill metric. While Figure 2a shows a relationship between MJO skill and teleconnection  
404 performance, even models with a good MJO can have poor teleconnection performance. For only  
405 the models assessed to have a sufficiently good MJO, Figure 2b assesses the relationship  
406 between teleconnection performance and biases in the North Pacific zonal flow. Plus signs are a  
407 measure of the total root mean squared (RMS) error of the 250 hPa zonal flow over the region  
408 15°N – 60°N, 110°E – 120°W, and the filled circle provides a measure of the RMS error in the  
409 length of the North Pacific subtropical jet. Both measures are correlated with MJO  
410 teleconnection performance, although biases in the jet provides a somewhat better metric ( $r = -0.7$   
411 versus  $-0.6$  for the total RMS). Subsequent analysis showed that models with a jet that extends  
412 too far east tend to have degraded teleconnection performance. Model physics appears to play a  
413 key role in the extent of the Pacific jet, as was demonstrated by Neelin et al. (2016) in diagnosing  
414 projected California precipitation changes between CMIP3 and CMIP5 models into the late 21st  
415 Century.

416

#### 417 *MJO propagation and amplitude diagnostics*

418 A diagnostic for MJO propagation is motivated by findings that the horizontal advection  
419 of column moist static energy, or equivalently the lower-tropospheric moisture, plays a critical  
420 role in driving the eastward propagation of the winter MJO (e.g., Maloney 2009; Kiranmayi and  
421 Maloney 2011; Sobel et al. 2014; Chikira 2014; Kim et al. 2014; Adames and Wallace 2015;  
422 Jiang 2017; Kim et al. 2017). Under this process, the spatial distribution of the winter mean  
423 lower-tropospheric moisture distribution over the equatorial Indo-Pacific region (Figure 3a) is  
424 critically important for moistening (drying) to the east (west) of the MJO convection through  
425 advection by MJO anomalous winds. The critical role of the mean lower-tropospheric moisture

426 pattern for the MJO eastward propagation is indeed confirmed based on multi-model simulations  
427 from the MJO Task Force / GEWEX GASS MJO model comparison project (Jiang 2017;  
428 Gonzalez and Jiang 2017). In particular, model skill in representing the 900-650hPa mean  
429 moisture pattern over the Maritime Continent region (red rectangle in Fig. 3a) exhibits a high  
430 correlation (about 0.8) with model MJO eastward propagation skill in about 25 GCM simulations  
431 (Figure 3b).

432         The convective moisture adjustment time scale ( $\tau$ ) in a model, defined by the ratio of  
433 intraseasonal perturbations of precipitable water and surface precipitation (e.g., Bretherton et al.  
434 2004; Peters and Neelin 2006; Sobel and Maloney 2013), is selected as a diagnostic metric for  
435 model MJO amplitude, which is motivated by the high anti-correlation (-0.72) between  $\tau$  and  
436 model MJO amplitude across multi-model simulations as indicated in Jiang et al. (2016, Fig. 3c).  
437 The convective moisture adjustment time scale depicts how rapidly precipitation must occur to  
438 remove excess column water vapor, or alternately the efficiency of surface precipitation  
439 generation per unit column water vapor anomaly, and is highly relevant to the convective onset  
440 diagnostics described above.

441

#### 442 *AMOC structure diagnostic*

443 The AMOC, with large temperature and salinity differences between the northward-flowing upper  
444 and southward-flowing lower limbs, is responsible for a large oceanic transport of heat and fresh  
445 water, thus playing a fundamental role in establishing the mean state and the variability of the  
446 climate system. The focus on diagnosing AMOC in climate models (e.g., CMIP5) has been mostly  
447 on the magnitude or the volume transport of the circulation (e.g., Cheng et al., 2013; Collins et al.,  
448 2013) and the role of water properties has been less quantified. In an AMOC structure diagnostic,

449 we examine the water properties of the AMOC by projecting the meridional transport on  
450 temperature-salinity (T-S) space, then use the transport-weighted T and S as the characteristic  
451 temperature and salinity of the upper and lower limbs. The results show that the modeled AMOC  
452 in CMIP5 historical simulations has a smaller temperature difference between the upper and lower  
453 limbs, compared to the results of a high-resolution ocean simulation that represents well the  
454 observed AMOC structure and the heat/freshwater transports (Xu et al., 2016). The model spread  
455 of time-mean heat transport among different CMIP5 simulations is significantly correlated with  
456 the volume transport/magnitude of the AMOC, not with the temperature difference between the  
457 upper and lower limbs (Figure 4a-b). The smaller temperature difference, however, is the main  
458 reason for a weaker heat transport in CMIP5 models (Figure 4b). However, the averaged  
459 freshwater transport in CMIP5 models is similar to high resolution simulation and observations,  
460 and the spread of freshwater transports in different CMIP5 models is significantly correlated with  
461 the salinity difference between the upper and lower AMOC limbs, not the volume transport of the  
462 AMOC (Figure 4c-d).

463

#### 464 *ENSO-precipitation diagnostics along the equatorial Pacific*

465 Sustained research in theory, numerical modeling and observations has demonstrated that  
466 SST anomalies associated with ENSO serve the leading source of predictability of seasonal to  
467 interannual climate anomalies over North America (Hoskins and Karoly 1981; Horel and Wallace  
468 1981) and the U.S. Affiliated Pacific Islands (USAPI; Annamalai et al. 2014). Recognizing that  
469 equatorial Pacific precipitation and associated diabatic heating anomalies are fundamental to this  
470 framework, and that in regions of weak horizontal temperature gradients such as the tropical  
471 oceans, moist static energy (MSE) variations are primarily due to moisture variations and have a

472 close association with precipitation (Neelin and Held 1987; Bretherton et al. 2006), we developed  
473 a POD based on vertically integrated MSE budget. The package identifies leading model processes  
474 that are important in translating ENSO-related SST anomalies into precipitation anomalies.  
475 Further, to identify and quantify compensating errors in model processes, MSE variance analysis  
476 (Wing and Emanuel 2014) is also included in the package. With a focus on ENSO winters, this  
477 POD is applied to CMIP5 models' historical simulations, and reanalysis products, and metrics are  
478 developed to assess models' fidelity in representing processes. Apart from identifying systematic  
479 errors across models (e.g., Fig. 5), the POD identifies compensating errors in individual models,  
480 and provides progress (or lack thereof) in generations of models from the same modeling group  
481 (see Annamalai 2018).

482 Figure 5 shows scatter plots between simulated anomalous precipitation and net radiative  
483 flux divergence into the column ( $F_{\text{rad}}$ ) for composite of El Niño winters over the equatorial central  
484 (Fig. 5a) and eastern (Fig. 5b) Pacific Ocean, respectively. The strong correlation in these plots  
485 suggest that systematic biases in precipitation are tied to biases in  $F_{\text{rad}}$ . Annamalai (2018) note  
486 that both during El Niño and La Nina winters,  $F_{\text{rad}}$ , particularly the bias in net longwave (LW)  
487 component, dominates the systematic bias in the MSE budget across all models. Here, higher  $F_{\text{rad}}$   
488 values indicate stronger cloud-radiative feedbacks that relate to perturbation of the radiative energy  
489 budget by condensate produced by convection (Stephens et al. 2008). Furthermore, systematic  
490 biases in  $F_{\text{rad}}$  are strongly linked to simulated free troposphere moisture anomalies that in turn are  
491 strongly linked to precipitation biases (not shown).

492

493 *Warm rain processes*

494 Combined analysis of multiple satellite measurements from CloudSat and the Moderate  
495 Resolution Imaging Spectroradiometer (MODIS) has provided new insights into the warm rain  
496 process, a key process that governs the low-cloud radiative properties and is a major pathway  
497 through which aerosols influence clouds. Suzuki et al. (2010) proposed a methodology for  
498 combining the radar reflectivity profile from CloudSat (Marchand et al. 2008) and the cloud  
499 properties (optical thickness and effective radius) from MODIS (Platnick et al. 2003; Nakajima et  
500 al. 2010) to probe how the warm rain process occurs within clouds. The methodology composites  
501 the radar reflectivity profiles in the form of the probability density function normalized at each in-  
502 cloud optical depth, which is determined by vertically slicing the cloud optical thickness according  
503 to the adiabatic profile assumption. The statistics thus constructed, referred to as Contoured  
504 Frequency by Optical Depth Diagram (CFODD), are further classified according to ranges of  
505 cloud-top particle size (Figure 6 top panels), which is another observable from MODIS, to reveal  
506 how the vertical microphysical structure of warm-topped clouds tends to transition from non-  
507 precipitating regime (Figure 6a) to precipitating regime (Figure 6c) as a fairly monotonic function  
508 of the particle size. The statistics provides a direct insight into the coalescence process.

509 The methodology has been applied to output from multiple global models (Suzuki et al. 2015;  
510 Jing et al. 2017) to construct the statistics corresponding to those from satellite observations. The  
511 statistics are then compared to evaluate how the models represent the warm rain formation process  
512 against satellite observations. Examples for such a comparison with state-of-the-art global models  
513 are shown in Figure 6 (middle and bottom panels) that indicate the models tend to produce rain  
514 too efficiently even when the cloud-top particle size is small. The behavior of the model biases  
515 identified in these statistics is further traced to formulations of model cloud microphysics,  
516 particularly the auto-conversion process (Suzuki et al. 2015), implying that the CFODD statistics

517 could serve as a clue to constrain a key uncertainty in cloud microphysics parameterization with  
518 satellite observations. This “bottom-up” constraint on model physics, however, tends to produce  
519 an overly negative forcing due to the aerosol indirect effect, which contradicts the “top-down”  
520 requirement for models to reproduce the historical temperature trend (Suzuki et al. 2013), implying  
521 the presence of error compensation at a fundamental level.

## 522 *Tropical cyclones*

523 The tropical cyclone (TC) diagnostics package contains a set of diagnostic codes that  
524 facilitate examination of TCs in global model simulations. When supplied with storm information  
525 (e.g., center position and intensity), this module computes azimuthal averages of dynamic and  
526 thermodynamic fields around the storm center that are helpful in identifying physical processes  
527 that lead to inter-model differences in simulated TCs. Figure 7 shows an example output from this  
528 package. The top two rows show radius-pressure plots of tangential and radial velocity, and  
529 relative humidity and pressure velocity, while the bottom row shows rainfall rates. The composite  
530 structures of TCs from four different GCM simulations show cyclonic tangential winds and “in-  
531 up-out” secondary circulations that are made of low-level radial inflow toward the center, rising  
532 motions around the center, and upper-level radial outflow away from the center – which are  
533 typically associated with observed TCs. The TC diagnostic package was used by Kim et al. (2018)  
534 to examine why the HiRAM simulation produces stronger TCs than the AM2.5 and Forecast-  
535 Oriented Low Resolution Ocean (FLOR) simulations. A key finding in the study was that at  
536 comparable intensity, the HiRAM model produces a greater amount of precipitation near the TC  
537 center than the other models (cf. the left two panels in the bottom row of Figure 7). The greater  
538 amount of diabatic heating associated with more rainfall in the TC inner-core region is associated  
539 with more favorable conditions for TC intensification (e.g., Schubert and Hack 1982; Nolan et al.

540 2007), and this allowed for stronger TCs in the HiRAM model. Moon et al. (2018) applied the TC  
541 diagnostic package to further examine inter-model spread among eight different global model  
542 simulations with different resolutions and physics.

543 In the second set of TC diagnostics, a framework based on the column-integrated MSE  
544 variance budget, which was originally developed to study convective organization in cloud-  
545 resolving model simulations (Wing and Emanuel 2014), has been adapted for climate model  
546 simulations of TCs. This diagnostic focuses on the relative role of feedback processes associated  
547 with tropical cyclogenesis by computing the product of MSE anomalies from the mean of a 10°  
548 box surrounding a TC and anomalies of sources and sinks of MSE (such as column radiative  
549 cooling and surface enthalpy fluxes). Figure 8 shows an example of this diagnostic, for the same  
550 GCM simulations and composites as used in Figure 7. The first row shows the squared MSE  
551 anomalies and the bottom two rows show two of the terms in the MSE variance budget – the  
552 radiative and surface flux feedbacks. While the feedbacks have similar structures in all models –  
553 they are generally positive and thus act to amplify MSE anomalies and favor development of the  
554 TC, they tend to be stronger in the models with more intense TCs. This indicates that the  
555 representation of the interaction of spatially varying radiative cooling and surface fluxes with the  
556 developing TC is partially responsible for inter-model spread in TC simulation. Wing et al. (2018)  
557 applied this diagnostic to six different global model simulations.

558

### 559 *Soil moisture control on evapotranspiration*

560 Soil moisture-atmosphere interactions are a key factor modulating surface climate over  
561 land. Soil moisture variations are forced by the atmosphere; in turn, they regulate surface water  
562 and energy fluxes (e.g., evapotranspiration, ET), and thus feedback onto near-surface surface

563 climate (e.g., temperature, humidity, precipitation; e.g., Seneviratne et al. [2010]). One of PODs  
564 focused on the so-called “terrestrial leg” of this coupling, i.e. the dependence of ET on soil  
565 moisture. Models exhibit significant uncertainties in the representation of this relationship (Guo et  
566 al. 2006, Dirmeyer et al. 2006, Berg and Sheffield 2018).

567         The hydrological and radiative controls on ET were assessed with a first-order diagnosis  
568 consisting of the correlations at the interannual time scale between summertime-mean values of  
569 surface (top 10cm) soil moisture (SM) and incoming solar radiation (Rsds), respectively, with ET  
570 (Berg and Sheffield 2018). Regions of positive SM-ET correlations in Figure 9 indicate soil  
571 moisture-limited regions, where soil moisture variability controls ET variability – generally in  
572 drier summer mid-latitude regions. The value of the correlation indicates how strongly SM controls  
573 ET. Conversely, negative values indicate that ET variations drive variations in soil moisture levels:  
574 this occurs in the Tropics and high latitudes, where available soil moisture is sufficient and the  
575 limiting factor for ET becomes atmospheric evaporative demand. This is consistent with the  
576 positive Rsds-ET correlations in the same regions.

577         Figure 9 shows that the CMIP5 multi-model mean qualitatively reproduces the  
578 climatological pattern from the ERA-Interim reanalysis. However, large uncertainties exist  
579 between models in the detailed spatial pattern and amplitude of the soil moisture control on ET.  
580 Overall, model uncertainty in SM–ET coupling tends to be greatest on the outer margins of regions  
581 of positive (from a multimodel mean perspective) coupling, extending into regions of energy-  
582 limited ET. The complementarity across space between hydrological and radiative controls on ET  
583 extends cross models: models that are less soil moisture limited are more radiation limited, and  
584 vice versa (not shown). In regions of greatest model spread, up to half of the inter-model variance  
585 in SM-ET coupling is explained by model differences in model precipitation; the remaining spread

586 may be related to further differences in rainfall characteristics such as intraseasonal distribution,  
587 but differences are also likely to stem from differences in model treatment of land hydrology,  
588 including differences in the simulation of vegetation and the representation of soil water stress. It  
589 is critical that models accurately represent these processes, as differences in SM-ET coupling  
590 affect summertime warming projections over certain land regions of the Tropics and Northern  
591 Hemisphere (Berg and Sheffield 2018).

592

### 593 *Midlatitude Cyclones, Fronts, and Storm Tracks*

594 One element of focus of the task force was extratropical cyclones, which generate  
595 precipitation, winds and clouds in the midlatitudes. GCMs need to capture the dynamics and  
596 thermodynamic properties of both the individual cyclone events and their accumulated behavior.  
597 Eulerian storm track analysis revealed that model sea surface temperature biases impact the  
598 surface storm tracks and precipitation near ocean western boundary currents (Booth et al. 2017;  
599 Small et al. 2018). Targeted analyses of features in cyclones and/or their fronts were carried out  
600 using Lagrangian tracking algorithms and compositing. These metrics facilitated process-  
601 oriented analyses of satellite observations of clouds that lead to: (1) explanations for  
602 relationships between stability and cloud cover (Naud et al., 2016), and (2) pin-pointing the  
603 synoptic locations and conditions where biases in GCM clouds occur (Figure 10). Task force  
604 efforts on cyclone-centered precipitation led to: (1) a satellite-based benchmarks (Naud et al.  
605 2018), and (2) results showing GCMs represent cyclone-total precipitation as well as reanalysis,  
606 but the models have markedly different levels of contributions from their convection  
607 parameterizations (Booth et al. 2018). The Lagrangian metrics require 6-hourly, 3-dimensional  
608 data, some of which is not standard in the CMIP archive.

609

610 *Diurnal cycle as a testbed*

611 Diurnal variations are large in near-surface temperature, pressure, winds, energy fluxes,  
612 precipitation and other fields, especially over land during the warm season. These variations are  
613 linked to many land-surface and atmospheric processes; therefore, they can be used as a testbed  
614 for diagnosing and evaluating weather and climate models (Dai and Trenberth 2004). One POD is  
615 the diurnal cycle in surface temperature and other related fields. Analyses of surface air  
616 temperature in the GFDL AM4 (Zhao et al., 2018a,b) revealed some systematic biases in the daily  
617 minimum (Tmin) and maximum (Tmax) air temperature, despite the relatively small biases in the  
618 daily mean (Tmean) temperature over many land areas (Lu 2018). For example, Tmax in the AM4  
619 showed a cold bias of 1-4°C over most land areas during all seasons except winter northern and  
620 central Asia where a large warm bias exists in Tmin, Tmax and Tmean, while Tmin in the AM4  
621 showed small to slight warm biases when compared with station observations, resulting in greatly  
622 reduced diurnal temperature range (DTR) in the model (Lu 2018).

623 Analyses of surface energy fluxes (Lu 2018) revealed many biases, including higher  
624 surface albedo, higher downward shortwave radiation, and weaker surface winds than the ERA-  
625 Interim reanalysis (Dee et al. 2011). However, large uncertainties in existing surface energy flux  
626 data made it hard to precisely quantify the model biases in these fields. Furthermore, the  
627 inconsistent definitions of 2m air temperature in the model and in observations further complicated  
628 the evaluation because the Tmin, Tmax, and DTR vary with the height of the measurement above  
629 the ground (Figure 11). In the AM4, the reference for the 2m air temperature is close to the  
630 displacement height (rather than the ground, which is the reference for the 2m air temperature from  
631 weather stations), which is about two thirds of the canopy height (a function of vegetation types).

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

640

## 641 **SUMMARY AND PATH FORWARD**

642 This article described in detail an open-source diagnostic package that is portable,  
643 community extensible and usable to aid application of process-oriented diagnostics to the model  
644 development process that was created by the NOAA MAPP MDTF. Moving forward, a renewed  
645 MDTF that will begin its term in Fall of 2018 plans expansion, refinement, and steps to increase  
646 the diagnostic utility of the package. Entrainment of additional diagnostics into the package will  
647 occur. For example, there is a need for standardized basin-scale heat uptake and sea level change  
648 diagnostics. Diagnostics for feedback mechanisms in regional hydroclimate extremes including  
649 cloud feedbacks will be developed, complemented by parameter-perturbation experiments with  
650 models. Diagnostics will be brought into the framework for processes affecting temperature and  
651 precipitation distribution tails, including advanced convective diagnostics and moist-static  
652 energy diagnostics. Collaborations will continue with GFDL and NCAR model development  
653 teams and an expanded number of other centers to refine diagnostics to increase their range and  
654 usability for the model development teams. A particular interest is expanding the diagnostic suite

655 for use with weather forecasting models, which is conceptually attractive given common  
656 physical roots of climate and forecasting models and the shared imperative to reduce biases in  
657 both types of models.

658         The MDTF plans to develop protocols to optimize application of the diagnostic  
659 framework to CMIP6 model simulations. The API already uses standard CF model formats and  
660 variable names used for CMIP6 output, thus the package will read CMIP6 output. Further  
661 development includes developing tools to assist modelers in navigating trade-offs among  
662 multiple observational constraints and expanding functionality to enable ensembles to be  
663 examined. The aim is to emphasize those aspects where the multi-model ensemble provides  
664 information about processes that tend to be ill constrained, and thus should be targeted for close  
665 scrutiny against observations (i.e., there must be added value beyond simply comparing a  
666 development version to existing models). Approaches considered will range from simply placing  
667 the candidate model within a multi-model plot of process-oriented metrics to new means of  
668 assessing parameter perturbation experiments systematically against observations.

669         Given the numerous community efforts related to process-oriented model diagnosis  
670 described above, greater coordination among these efforts would provide efficiencies, optimize  
671 science and technical approaches, and foster the greatest benefit to the climate and modeling  
672 communities. To further this goal, the MDTF plans to be proactive in forging connections to  
673 other efforts, including stronger links to PCMDI to leverage community data standards and  
674 enhance coordination of metrics and diagnostics development across agencies. The project will  
675 provide complementary process diagnosis to PCMDI capabilities that are expected to provide  
676 routine performance evaluation of all CMIP6 Diagnostic, Evaluation and Characterization of  
677 Klima (DECK) and Historical simulations. As MAPP PODs crystalize via experience at GFDL

678 and NCAR, it is expected that some will be entrained into the broader community-based efforts,  
679 including possible collaborations with other modeling groups contributing to CMIP. Hence, the  
680 NOAA MDTF effort will benefit from experience such as PCMDI's working with the broader  
681 modeling community, and in particular its support of the developing standards and protocols.

682

### 683 **ACKNOWLEDGEMENTS**

684 This work was supported by the NOAA Climate Program Office Modeling, Analysis, Predictions  
685 and Projections (MAPP) Program as part of the Model Diagnostics Task Force. The statements,  
686 findings, conclusions, and recommendations do not necessarily reflect the views of NOAA or the  
687 Department of Commerce.

688

689

690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711

## References

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.

Ahn, M.-S., D. Kim, K. Sperber, E. Maloney, D. Waliser, H. Hendon, 2017: MJO Simulation in CMIP5 Climate Models: MJO Skill Metrics and Process-Oriented Diagnosis. *Climate Dyn.*, **49**, 4023. <https://doi.org/10.1007/s00382-017-3558-4>.

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.

Bellenger, H., E. Guilyardi, J. Leloup, M. Lengaigne and J. Vialard, 2014: ENSO representation in climate models: from CMIP3 to CMIP5. *Clim. Dyn.*, **42**,1999-2018

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.

Booth J. F., Y.-K. Kwon, S. Ko, J. Small, R. Madsek, 2017: Spatial Patterns and Intensity of the Surface Storm Tracks in CMIP5 Models. *J. Climate*, **30**, 4965–4981.

Booth J. F., C. M. Naud, 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.

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.

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.

712 Chikira, M., 2014: Eastward-Propagating Intraseasonal Oscillation Represented by Chikira–  
713 Sugiyama Cumulus Parameterization. Part II: Understanding Moisture Variation under  
714 Weak Temperature Gradient Balance. *J. Atmos. Sci.*, **71**, 615-639.

715 Collins, M., R. Knutti, J. Arblaster, J.-L. Dufresne, T. Fichefet, P. Friedlingstein, X. Gao, W.J.  
716 Gutowski, T. Johns, G. Krinner, M. Shongwe, C. Tebaldi, A.J. Weaver, and M. Wehner,  
717 2013: Long-term climate change: Projections, commitments and irreversibility. In  
718 *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to*  
719 *the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. T.F.  
720 Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Doschung, A. Nauels, Y. Xia,  
721 V. Bex, and P.M. Midgley, Eds. Cambridge University Press, pp. 1029-1136

722 Dai, A., and K. E. Trenberth, 2004: The diurnal cycle and its depiction in the Community Climate  
723 System Model. *J. Climate*, **17**, 930-951.

724 Dee, D. P. and Co-authors, 2011: The ERA-Interim reanalysis: configuration and performance of  
725 the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, **137**,  
726 553–597. doi: 10.1002/qj.828.

727 Dirmeyer, P. A., R.D. Koster, and Z. Guo, 2006: Do global models properly represent the feedback  
728 between land and atmosphere? *J. Hydrometeorol.*, **7**, 1177-1198.

729 Eyring, V., and others, 2018: The multiple and growing benefits of confronting climate models  
730 with observations. *Nature Climate Change*, accepted pending revision.

731 Eyring, V., Righi, M., Lauer, A., Evaldsson, M., Wenzel, S., Jones, C., Anav, A., Andrews, O.,  
732 Cionni, I., Davin, E. L., Deser, C., Ehbrecht, C., Friedlingstein, P., Gleckler, P.,  
733 Gottschaldt, K.-D., Hagemann, S., Juckes, M., Kindermann, S., Krasting, J., Kunert, D.,  
734 Levine, R., Loew, A., Mäkelä, J., Martin, G., Mason, E., Phillips, A. S., Read, S., Rio, C.,

735 Roehrig, R., Senftleben, D., Sterl, A., van Ulf, L. H., Walton, J., Wang, S., and  
736 Williams, K. D., 2016a: ESMValTool (v1.0) – a community diagnostic and performance  
737 metrics tool for routine evaluation of Earth system models in CMIP, *Geosci. Model Dev.*,  
738 **9**, 1747-1802.

739 Eyring, V., Gleckler, P. J., Heinze, C., Stouffer, R. J., Taylor, K. E., Balaji, V., Guilyardi, E.,  
740 Joussaume, S., Kindermann, S., Lawrence, B. N., Meehl, G. A., Righi, M., and Williams,  
741 D. N., 2016: Towards improved and more routine Earth system model evaluation in  
742 CMIP, *Earth Syst. Dynam.*, **7**, 813-830.

743 Gettelman, A., V. Eyring, C. Fischer, H. Shiona, I. Cionni, M. Neish, O. Morgenstern, S. W.  
744 Wood, and Z. Li, 2012: A Community Diagnostic Tool for Chemistry Climate Model  
745 Validation, *Geosci. Model. Development* , **5**, 1061-1073.

746 Gleckler, P. J., C. Doutriaux, P. J. Durack, K. E. Taylor, Y. Zhang, D. N. Williams, E. Mason,  
747 and J. Servonnat, 2016: A more powerful reality test for climate models, *Eos*, **97**,  
748 doi:10.1029/2016EO051663.

749 Gonzalez, A. O. and X. Jiang, 2017: Winter Mean Lower-Tropospheric Moisture over the  
750 Maritime Continent as a Climate Model Diagnostic Metric for the Propagation of the  
751 Madden-Julian Oscillation. *Geophys. Res. Lett.*, 10.1002/2016GL072430.

752 Grose, M. R., Brown, J. N., Narsey, S., Brown, J. R., Murphy, B. F., Langlais, C., Gupta, A. S.,  
753 Moise, A. F. and Irving, D. B., 2014: Assessment of the CMIP5 global climate model  
754 simulations of the western tropical Pacific climate system and comparison to CMIP3. *Int.*  
755 *J. Climatol.*. doi: 10.1002/joc.3916.

756 Guo, P. A., and others, 2006: GLACE: The Global Land–Atmosphere Coupling Experiment. Part  
757 II: analysis. *J. Hydrometeorol.*, **7**, 611-625.

758 Hannah, W. M., and E. D. Maloney, 2014: The Moist Static Energy Budget in NCAR CAM5  
759 Hindcasts during DYNAMO. *J. Adv. Modeling Earth Sys.*, **6**,  
760 doi:10.1002/2013MS000272.

761 Henderson, S. A., E. D. Maloney, and S.-W. Son, 2017: Madden-Julian oscillation  
762 teleconnections: The impact of the basic state and MJO representation in general  
763 circulation models. *J. Climate*, **30**, 4567–4587.

764 Holloway, C. E., and J. D. Neelin, 2009: Moisture vertical structure, column water vapor, and  
765 tropical deep convection. *J. Atmos. Sci.*, **66**, 1665-1683,  
766 <https://doi.org/10.1175/2008JAS2806.1>.

767 Hung, M.-P., Lin, J.-L., Wang, W., Kim, D., Shinoda, T. and Weaver, S.J., 2013: MJO and  
768 convectively coupled equatorial waves simulated by CMIP5 climate models. *J. Climate*,  
769 **26**, 6185-6214.

770 Jiang, X., 2017: Key processes for the eastward propagation of the Madden-Julian Oscillation  
771 based on multimodel simulations. *Journal of Geophysical Research: Atmospheres*,  
772 10.1002/2016JD025955.

773 Jiang, X., M. Zhao, E. D. Maloney, and D. E. Waliser, 2016: Convective moisture adjustment  
774 time scale as a key factor in regulating model amplitude of the Madden-Julian  
775 Oscillation. *Geophys. Res. Lett.*, **43**, 10.1002/2016GL070898,10,412-10,419.

776 Jiang, X., Á. F. Adames, M. Zhao, D. Waliser, and E. Maloney, 2018: A Unified Moisture Moist  
777 Framework for Seasonality of MJO Propagation. *J. Climate*, **31**, 4215–4224.

778 Jiang, X., et al., 2015: Vertical structure and physical processes of the Madden-Julian oscillation:  
779 Exploring key model physics in climate simulations. *Journal of Geophysical Research:*  
780 *Atmospheres*, **120**, 10.1002/2014JD022375,4718-4748.

781 Jing, X., K. Suzuki, H. Guo, D. Goto, T. Ogura, T. Koshiro, and J. Mulmenstadt, 2017: A multi-  
782 model study on warm precipitation biases in global models compared to satellite  
783 observations. *J. Geophys. Res. Atmos.*, **122**, doi:10.1002/2017/JD027310.

784 Kim, D., M.-S. Ahn, I.-S. Kang, and A. D. Del Genio, 2015: Role of longwave cloud-radiation  
785 feedback in the simulation of the Madden-Julian oscillation. *J. Climate*, **28**, 6979-6994.

786 Kim, D., A. H. Sobel, E. D. Maloney, D. M. W. Frierson, and I.-S. Kang, 2011: A systematic  
787 relationship between intraseasonal variability and mean state bias. *J. Climate*, **24**, 5506-  
788 5520.

789 Kim, D., P. Xavier, E. Maloney, M. Wheeler, D. Waliser, K. Sperber, H. Hendon, C. Zhang, R.  
790 Neale, Y.-T. Hwang, and H. Liu, 2014: Free Access Process-Oriented MJO Simulation  
791 Diagnostic: Moisture Sensitivity of Simulated Convection. *J. Climate*, **27**, 5379-5395.

792 Kim, D., J.-S. Kug, and A. H. Sobel, 2014: Propagating versus Nonpropagating Madden–Julian  
793 Oscillation Events. *J. Climate*, **27**, 111-125, 10.1175/JCLI-D-13-00084.1.

794 Kim, D., H. Kim, and M.-I. Lee, 2017: Why does the MJO detour the Maritime Continent during  
795 austral summer? *Geophys. Res. Lett.*, 10.1002/2017GL072643.

796 Kim, D., Y. Moon, S.J. Camargo, A.A. Wing, A.H. Sobel, H. Murakami, G.A. Vecchi, M. Zhao,  
797 and E. Page, 2018: Process-Oriented Diagnosis of Tropical Cyclones in High-Resolution  
798 GCMs. *J. Climate*, **31**, 1685–1702, doi: 10.1175/JCLI-D-17-0269.1.

799 Kiranmayi, L. and E. D. Maloney, 2011: Intraseasonal moist static energy budget in reanalysis  
800 data. *J. Geophys. Res.*, **116**, 10.1029/2011JD016031,D21117.

801 Kuo, Y.-H., J. D. Neelin, and C. R. Mechoso, 2017: Tropical convective transition statistics and  
802 causality in the water vapor-precipitation relation. *J. Atmos. Sci.*, **74**, 915-931,  
803 <https://doi.org/10.1175/JAS-D-16-0182.1>.

804 Kuo, Y.-H., K. A. Schiro and J. D. Neelin, 2018: Convective transition statistics over tropical  
805 oceans for climate model diagnostics. Observational baseline. *J. Atmos. Sci.*, **75**, 1553-  
806 1570.

807 Li, G., and S.-P. Xie, 2014: Tropical biases in CMIP5 multi-model ensemble: The excessive  
808 equatorial Pacific cold tongue and double ITCZ problems. *J. Climate*, **27**, 1765-1780.

809 Lu, Y., 2018: Surface air temperature biases and other associated biases in GFDL AM4.0 model.  
810 Master's thesis, Department of Atmospheric and Environmental Sciences, University at  
811 Albany, Albany, NY, pp. 60.

812 Luo, Y. Q., Randerson, J. T., Abramowitz, G., Bacour, C., Blyth, E., Carvalhais, N., Ciais, P.,  
813 Dalmonech, D., Fisher, J. B., Fisher, R., Friedlingstein, P., Hibbard, K., Hoffman, F.,  
814 Huntzinger, D., Jones, C. D., Koven, C., Lawrence, D., Li, D. J., Mahecha, M., Niu, S.  
815 L., Norby, R., Piao, S. L., Qi, X., Peylin, P., Prentice, I. C., Riley, W., Reichstein, M.,  
816 Schwalm, C., Wang, Y. P., Xia, J. Y., Zaehle, S., and Zhou, X. H., 2012: A framework  
817 for benchmarking land models, *Biogeosciences*, **9**, 3857–3874, doi:10.5194/bg-9- 3857-  
818 2012.

819 Maloney, E. D., 2009: The Moist Static Energy Budget of a Composite Tropical Intraseasonal  
820 Oscillation in a Climate Model. *J. Climate*, **22**, 711-729.

821 Mapes, B., and R. Neale, 2011: Parameterizing Convective Organization to Escape the  
822 Entrainment Dilemma, *J. Adv. Model. Earth Syst.*, **3**, M06004, doi:  
823 10.1029/2011MS000042.

824 Marchand, R., G. Mace, T. Ackerman, and G. Stephens, 2008: Hydrometeor detection using  
825 CloudSat – An earth-orbiting 94GHz cloud radar. *J. Atmos. Ocean. Tech.*, **25**, 519-533.

826 Moon, Y., D. Kim, S.J. Camargo, A.A. Wing, A.H. Sobel, H. Murakami, K.A. Reed, E.  
827 Scoccimarro, G.A. Vecchi, M. Wehner, C. Zarzycki, and M. Zhao, 2018: Exploring  
828 dependency of tropical cyclone structures and intensity to horizontal resolution in global  
829 climate model simulations. To be submitted to *J. Climate*.

830 Mundhenk, B. D., E. A. Barnes, E. D. Maloney, and C. F. Baggett, 2018: Skillful subseasonal  
831 prediction of atmospheric river activity based on the Madden-Julian oscillation and the  
832 Quasi-Biennial oscillation. *Nature (npj) Clim. Atmos. Sci.*, **1**, doi:10.1038/s41612-017-  
833 0008-2.

834 Nakajima, T. Y., K. Suzuki, and G. L. Stephens, 2010: Droplet growth in warm water clouds  
835 observed by the A-Train. Part II: A multisensor view. *J. Atmos. Sci.*, **67**, 1897-1907.

836 Naud, C.M., J.F. Booth, and A.D. Del Genio, 2014: Evaluation of ERA-interim and MERRA  
837 cloudiness in the Southern Ocean. *J. Climate*, **27**, 2109-2124, doi:10.1175/JCLI-D-13-  
838 00432.1.

839 Naud, C. M., J. F. Booth, and A. D. Del Genio, 2016: The relationship between boundary layer  
840 stability and cloud cover in the post-cold-frontal region. *J. Climate*, **29**, 8129-8149, DOI:  
841 10.1175/JCLI-D-15-0700.1

842 Naud, C. M., J. F. Booth, L. J. Donner, C. J. Seman, M. Zhao, H. Guo and Y. Ming, 2018:  
843 Extratropical cyclone clouds in the GFDL model: diagnosing biases and the associated  
844 causes. Submitted to *Journal of Climate*.

845 Neelin, J. D., Langenbrunner, B., Meyerson, J.E., Hall, A. & Berg, N., 2013: California winter  
846 precipitation change under global warming in the Coupled Model Intercomparison  
847 Project 5 ensemble. *J. Climate*, **26**, 6238-6256.

848 Neelin, J. D., O. Peters, and K. Hales, 2009: The transition to strong convection. *J. Atmos. Sci.*,  
849 **66**, 2367-2384, <https://doi.org/10.1175/2009JAS2962.1>.

850 Nolan, D.S., Y. Moon, and D.P. Stern, 2007: Tropical Cyclone Intensification from Asymmetric  
851 Convection: Energetics and Efficiency. *J. Atmos. Sci.*, **64**, 3377–3405,  
852 doi:10.1175/JAS3988.1.

853 Peters, O. and J. D. Neelin, 2006: Critical phenomena in atmospheric precipitation. *Nat Phys*, **2**,  
854 393-396.

855 Platnick, S., M. King, S. Ackerman, W. Menzel, B. Baum, J. Riedi, and R. Frey, 2003: The  
856 MODIS cloud products: Algorithms and examples from Terra. *IEEE Trans. Geosci. Rem.*  
857 *Sens.*, **41**, 459-473.

858 Prabhat, Oliver Rübel, Surendra Byna, Kesheng Wu, Fuyu Li, Michael Wehner, Wes Bethel,  
859 2012: TECA: A Parallel Toolkit for Extreme Climate Analysis, *Procedia Computer*  
860 *Science*, **9**, 866-876.

861 Sahany, S., J. D. Neelin, K. Hales, and R. B. Neale, 2012: Temperature-moisture dependence of  
862 the deep convective transition as a constraint on entrainment in climate models. *J. Atmos.*  
863 *Sci.*, **69**, 1340-1358, <https://doi.org/10.1175/JAS-D-11-0164.1>.

864 Sahany, S., J. D. Neelin, K. Hales, and R. B. Neale, 2014: Deep convective transition  
865 characteristics in the Community Climate System Model and changes under global  
866 warming. *J. Climate*, **27**, 9214-9232, <https://doi.org/10.1175/JCLI-D-13-00747.1>.

867 Santos, E., J. Paco, Y. Wei, S. Liu, B. Cook, D. Williams, C. Silva, 2013: UV-CDAT: Analyzing  
868 Climate Datasets from a User's Perspective, *Comp. in Sci. & Eng.*, **15**, 94-103.

869 Schiro, K. A., J. D. Neelin, D. K. Adams, and B. R. Lintner, 2016: Deep convection and column  
870 water vapor over tropical land versus tropical ocean: A comparison between the Amazon

871 and the tropical western Pacific. *J. Atmos. Sci.*, **73**, 4043-4063,  
872 <https://doi.org/10.1175/JAS-D-16-0119.1>.

873 Schiro, K. A., F. Ahmed, S. E. Giangrande, and J. D. Neelin, 2018: GoAmazon2014/5 campaign  
874 points to deep-inflow approach to deep convection across scales. *Proc. Natl. Acad. Sci.*  
875 *USA*, <https://doi.org/10.1073/pnas.1719842115>.

876 Schubert, W.H. and J.J. Hack, 1982: Inertial Stability and Tropical Cyclone Development. *J.*  
877 *Atmos. Sci.*, **39**, 1687–1697, doi:10.1175/1520-0469(1982)039<1687:ISATCD>2.0.CO;2.

878 Seneviratne, S. I., et al., 2010, Investigating soil moisture–climate interactions in a changing  
879 climate: A review, *Earth-Science Reviews*, **99**, 125-161.

880 Sheffield, J. and coauthors, 2013a: North American climate in CMIP5 experiments. Part I:  
881 Evaluation of 20th century continental and regional climatology. *J. Climate*, **26**, 9209-  
882 9245.

883 Sheffield, J. and coauthors, 2013b: North American climate in CMIP5 experiments. Part II:  
884 Evaluation of 20th century intra-seasonal to decadal variability. *J. Climate*, **26**, 9247-9290.

885 Small, J., R. Msadek, Y.-O. Kwon, J. F. Booth, and C. Zarzycki, 2018: Atmosphere surface  
886 storm track response to resolved ocean mesoscale in two sets of global climate model  
887 experiments. *Climate Dynamics*, <https://doi.org/10.1007/s00382-018-4237-9>.

888 Sobel, A. and E. Maloney, 2013: Moisture Modes and the Eastward Propagation of the MJO. *J.*  
889 *Atmos. Sci.*, **70**, Doi 10.1175/Jas-D-12-0189.1,187-192.

890 Sobel, A., S. Wang, and D. Kim, 2014: Moist Static Energy Budget of the MJO during  
891 DYNAMO. *J. Atmos. Sci.*, **71**, 10.1175/JAS-D-14-0052.1,4276-4291.

892 Stephens, G. L., C. Jacob, and G. Tseilioudis, 2015: The GEWEX Process Evaluation Study:  
893 GEWEX-PROES. *GEWEX News*, November 2015.

894 Stevens, B., & Bony, S., 2013: What are climate models missing. *Science*, **340**, 1053-1054.

895 Suzuki, K., T. Nakajima, and G. L. Stephens, 2010: Particle growth and drop collection  
896 efficiency of warm clouds as inferred from joint CloudSat and MODIS observations. *J.*  
897 *Atmos. Sci.*, **67**, 3019-3032.

898 Suzuki, K., J.-C. Golaz, and G. L. Stephens, 2013: Evaluating cloud tuning in a climate model  
899 with satellite observations. *Geophys. Res. Lett.*, **40**, 4464-4468, doi:10.1002/grl.50874.

900 Suzuki, K., G. Stephens, A. Bodas-Salcedo, M. Wang, J.-C. Golaz, T. Yokohata, and T. Koshiro,  
901 2015: Evaluation of the warm rain formation process in global models with satellite  
902 observations. *J. Atmos. Sci.*, **72**, 3996-4014.

903 Tsushima, Y., and Co-authors, 2017: The Cloud Feedback Model Intercomparison Project  
904 (CFMIP) Diagnostic Codes Catalogue – metrics, diagnostics and methodologies to  
905 evaluate, understand and improve the representation of clouds and cloud feedbacks in  
906 climate models. *Geosci. Model Dev.*, **10**, 4285-4305.

907 Wang, C., L. Zhang, S.-K. Lee., L. Wu, and C. R. Mechoso, 2014: A global perspective on  
908 CMIP5 climate model biases. *Nature Climate Change*, **4**, 201-205.

909 Wheeler, M. W., E. D. Maloney, and the MJO Task Force, 2013: Madden-Julian Oscillation  
910 (MJO) Task Force: a joint effort of the climate and weather communities. *CLIVAR*  
911 *Exchanges*. No. 61 (Vol 18 No.1).

912 Williams, D. N., et al., 2016: UV-CDAT v2.4.0, Zenodo, Geneva, Switzerland,  
913 doi:10.5281/zenodo.45136.

914 Wing, A.A. and K.A. Emanuel, 2014: Physical mechanisms controlling self-aggregation of  
915 convection in idealized numerical modeling simulations. *J. Adv. Model. Earth Sys.*, **6**, 59-  
916 74, doi:10.1002/2013MS000269.

917 Wing, A.A., S.J. Camargo, A.H. Sobel, D. Kim, Y. Moon, H. Murakami, K.A. Reed, G.A. Vecchi,  
918 M. Wehner, C. Zarzycki, M. Zhao, 2018: Moist static energy budget analysis of tropical  
919 cyclone formation and intensification in high-resolution climate models. To be submitted  
920 to *J. Climate*.

921 Xu, X., P. B. Rhines, and E. P. Chassignet, 2016, Temperature-salinity structure of the North  
922 Atlantic circulation and associated heat and freshwater transports, *J. Climate*, **29**, 7723-  
923 7742, doi:10.1175/JCLI-D-15-0798.1.

924 Zappa, G., L. C. Shaffrey, and K. I. Hodges, 2013: The Ability of CMIP5 Models to Simulate  
925 North Atlantic Extratropical Cyclones. *J. Climate*, **26**, 5379–5396.

926 Zhao, M., et al. 2018a: The GFDL global atmosphere and land model AM4.0/LM4.0 – Part 1.  
927 Simulation characteristics with prescribed SSTs. *Journal of Advances in Modeling Earth*  
928 *Systems*, **10**. doi:10.1002/2017MS001208

929 Zhao, M., et al. 2018b: The GFDL global atmosphere and land model AM4.0/LM4.0 - Part 2.  
930 Model description, sensitivity studies, and tuning strategies. *Journal of Advances in*  
931 *Modeling Earth Systems*, **10**. doi:10.1002/ 2017MS001209.

932

933

934

### Figure Captions

935 **Figure 1.** Schematic of MDTF Process Oriented Diagnostics (PODs) framework (mdtf.py). (a)

936 Workflow under the API. The model output is that of a candidate version that the development

937 team wants to compare to observations under the various diagnostics. The observations for each

938 POD are supplied in analyzed form within the module. (b) Example webpage (edited for brevity

939 and clarity) for one POD, described in the Convective Transition Statistics subsection, below.

940 **Figure 2.** From Henderson et al. (2017). December-February teleconnection performance

941 averaged across all MJO phases (y-axis) versus (a) MJO skill (MJO E/W ratio) and (b) the RMS

942 error of the 250-hPa December-February zonal wind. The observed E/W precipitation (GPCP)

943 ratio is provided (dashed line, left panel). The MJO skill is derived as the ratio of eastward to

944 westward power of equatorial precipitation in the 30-60 day, zonal wavenumbers 1-3 band (e.g.

945 Ahn et al. 2017). Teleconnection performance was determined as pattern correlation of North

946 Pacific and North America (15°S - 80°N, 130°E - 60°W) MJO composite 250 hPa geopotential

947 height anomalies between CMIP models and ERA-i reanalysis averaged over all MJO phases. In

948 panel (b), the crosses show the model 250hPa zonal wind RMS error over the full Pacific basin,

949 while the closed circles indicate the longitudinal RMS error of the subtropical jet. See the text

950 and Henderson et al. (2017) for more detailed explanations.

951 **Figure 3.** a) Winter (Nov-Apr) mean 650-900hPa specific humidity based on ERA-Interim

952 reanalysis; b) Scatter plot of model skill for eastward propagation of the MJO versus model skill

953 of the mean 650-900hPa moisture pattern over the Maritime Continent (red rectangle in a) based

954 on multi-model simulations from the MJOTF/GASS project. Model MJO propagation skill is

955 defined by the pattern correlation of Homvöller diagrams of model simulated rainfall anomalies

956 associated with the MJO against its observed counterpart following Jiang et al. (2015). Red

957 (blue) dots denote good (poor) MJO models. c) Scatter plot of MJO amplitude and model  
958 convective moisture adjustment time scale in models (black dots) and observations (red dots).  
959 The MJO amplitude in each model is defined by the standard deviation of 20-100 day band-pass  
960 filtered rainfall over the Indian Ocean (75-85°E; 10°S-10°N) during winter. Convective time  
961 scale in a model is defined by the ratio of precipitable water (W) anomaly to precipitation (P)  
962 anomaly associated with the MJO and derived by a regression approach. Before conducting the  
963 regression, both W and P anomalies are subject to 20-100 day filtering and averaged over the  
964 Indian Ocean (75-85°E; 10°S-10°N) region. Adapted from Jiang et al. (2016) and Gonzalez and  
965 Jiang (2017).

966 **Figure 4.** a-b) Relation of the modeled oceanic heat transport with (a) volume transport and (b)  
967 temperature difference (between the upper and lower AMOC limbs) at 26°N; c-d) Relation of the  
968 modeled freshwater transport with (c) volume transport and (d) salinity difference. Colored  
969 symbols and the black circle denote results from 20 CMIP5 historical simulations and their  
970 average, respectively; black dot denotes high-resolution ocean simulation results that represent  
971 well the observed mean heat/freshwater transports and the AMOC structure at 26°N (Xu et al.  
972 2016).

973 **Figure 5:** Scatter plots between CMIP5 models simulated precipitation versus net radiative flux  
974 divergence ( $F_{\text{rad}}$ ): (a) equatorial central Pacific (160°E-160°W; 10°S-5°N) and (b) eastern Pacific  
975 (160°W-80°W; 5°S-5°N). Results are for El Niño winter composites. In each panel, inter-model  
976 correlations and best-fit regressions are also provided.

977 **Figure 6:** The probability density function (color shading in the unit of %/dBZ) of radar reflectivity  
978 (abscissa) normalized as a function of cloud optical depth increasing downward (ordinate), which  
979 is further classified according to different ranges of cloud-top particle sizes for (left to right) 5-

980 10mm, 10-15mm and 15-20mm obtained from (top) A-Train satellite observations, (middle)  
981 GFDL-AM4 and (bottom) MIROC5 (adapted from Jing et al. 2017).

982 **Figure 7:** The first two rows show azimuthally averaged radius-pressure plots of (top) tangential  
983 velocity (shading) and radial velocity (lines); and (middle) relative humidity (shading) and  
984 pressure velocity (lines). The bottom row shows azimuthally averaged rain rate. All panels show  
985 the composites when the simulated tropical cyclones have a similar intensity, in this case between  
986 30-33  $\text{ms}^{-1}$ . Positive/negative values are plotted in red/ blue contour lines. Four different GCM  
987 simulations (GFDL Atmosphere Model 2.5 [AM2.5] and High Resolution Atmospheric Model  
988 [HiRAM] models at  $0.5^\circ$  resolution and National Center for Atmospheric Research (NCAR)  
989 Community Atmosphere Model version 5 - Finite Volume (CAM5-FV) and CAM5-Spectral  
990 Element (CAM-SE) models at  $0.25^\circ$  resolution) are used.

991 **Figure 8:** The first row shows squared anomalies of column-integrated moist static energy (MSE),  
992 in units of  $\text{J}^2\text{m}^{-4}$ , where anomalies are from the mean of a  $10^\circ \times 10^\circ$  box surrounding a TC. The  
993 second and third rows show the radiative and surface flux feedback terms in the MSE variance  
994 budget, plotted as a function of latitude and longitude relative to the TC center. All panels show  
995 the composites when the simulated tropical cyclones have a similar intensity, in this case between  
996 30-33  $\text{ms}^{-1}$ . Four different GCM simulations are shown: GFDL AM2.5 (panels d,j,p) and HiRAM  
997 (panels a,g,m) models at  $0.5^\circ$  degree resolution and NCAR CAM5-FV (panels c,i,o) and CAM5-  
998 SE (panels b,h,n) models at  $0.25^\circ$  resolution.

999 **Figure 9:** (top row) correlation between summertime-mean surface soil moisture and  
1000 evapotranspiration (ET) in ERA-INTERIM (left), in CMIP5 models (middle), and standard  
1001 deviation across CMIP5 models (right). (Bottom row) same as top for the correlation between  
1002 (ET) and incoming surface solar radiation (Rsds). Summertime is defined as June-July-August in

1003 the northern hemisphere and December-January-February in the southern hemisphere.  
1004 Correlations are over 1979-2014 for ERA-INTERIM and 1950-2005 for CMIP5 models, using  
1005 outputs from the historical simulations. Soil moisture and ET outputs were available for 37  
1006 models, and for 43 models for Rsds and ET.

1007 **Figure 10:** Composites of cloud fraction for cyclone-centered map view (top row) and cold-front  
1008 centered transects (bottom row). Observational data: (a) MODIS Aqua and (c) CloudSat-Cloud-  
1009 Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO). Model data: GFDL pre-  
1010 CMIP6 model that uses two-plume convection scheme (i.e. 2PM). Composites are generated  
1011 using cyclones or fronts from 5 years of observations or model data, within 30°-60° latitude band  
1012 in both hemispheres. Plus sign in top row marks the center of the cyclones. The region between  
1013 the black dashed lines in top row is approximate locations for transects used in frontal analysis.  
1014 Black dashed line in bottom row indicates location of the cold front based on 850-hPa potential  
1015 temperature and wind gradient data (see Naud et al. 2014 and Naud et al., 2016 for compositing  
1016 methodology details).

1017 **Figure 11.** The observed mean diurnal cycle of surface air temperatures at 2m (green) and 9m  
1018 (blue) above the ground during July of 2016 and 2017 at 12 New York Mesonet stations,  
1019 compared with the GFDL AM4 model simulated multi-year mean 2m air temperature (red) for  
1020 the grid box containing the station. The daily mean value is shown in the line legend. The 2m  
1021 height in the AM4 is 2m above a reference height, which is close to the displacement height ( $\sim\frac{2}{3}$   
1022 of the canopy height) in the AM4, rather than 2m above the ground as for weather stations.

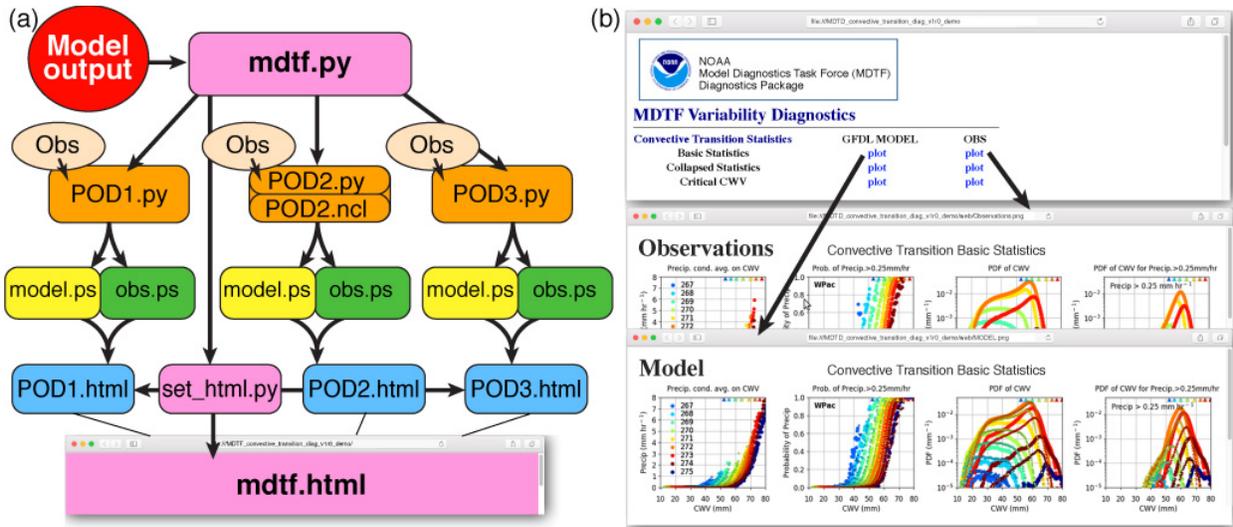
1023

1024

1025

1026

1027



1028

1029

1030

1031 **Figure 1.** Schematic of MDTF Process Oriented Diagnostics (PODs) framework (mdtf.py). (a)

1032 Workflow under the API. The model output is that of a candidate version that the development

1033 team wants to compare to observations under the various diagnostics. The observations for each

1034 POD are supplied in analyzed form within the module. (b) Example webpage (edited for brevity

1035 and clarity) for one POD, described in the Convective Transition Statistics subsection, below.

1036

1037

1038

1039

1040

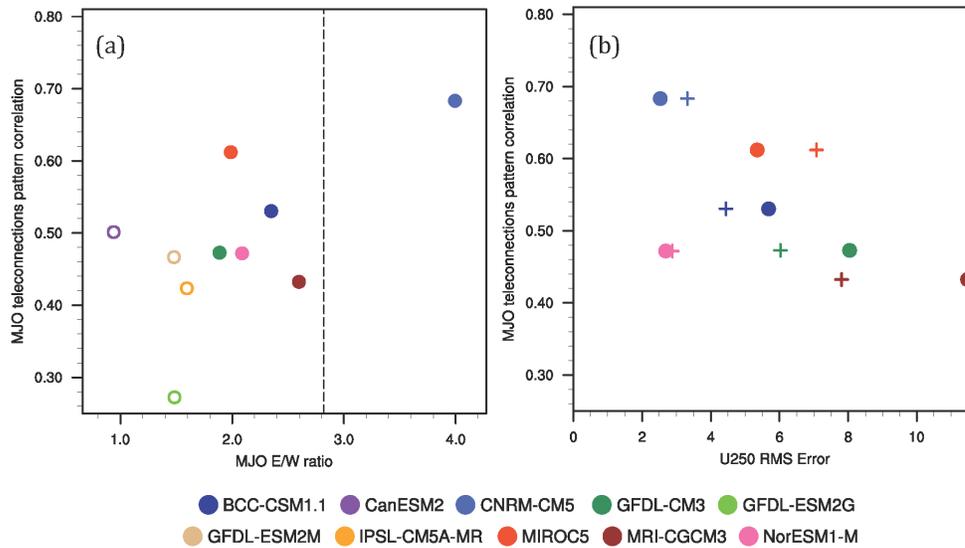
1041

1042

1043

1044

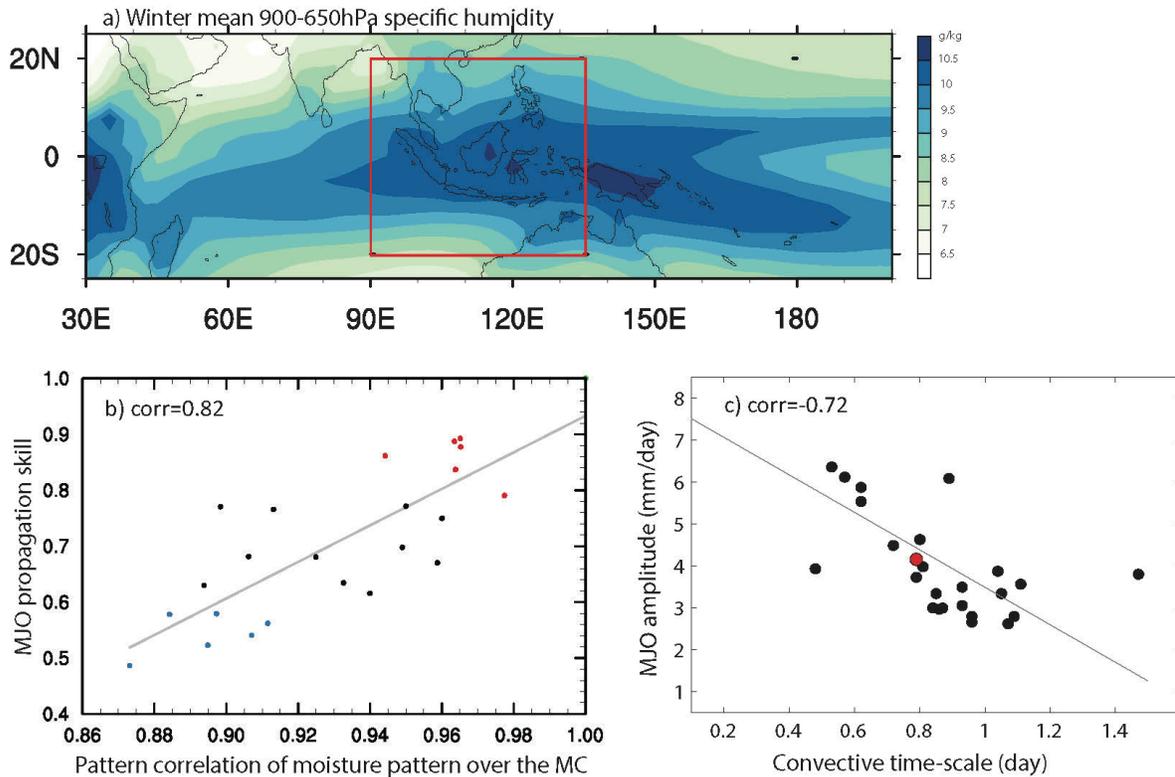
1045



1046  
1047

1048 **Figure 2.** From Henderson et al. (2017). December-February teleconnection performance  
 1049 averaged across all MJO phases (y-axis) versus (a) MJO skill (MJO E/W ratio) and (b) the RMS  
 1050 error of the 250-hPa December-February zonal wind. The observed E/W precipitation (GPCP)  
 1051 ratio is provided (dashed line, left panel). The MJO skill is derived as the ratio of eastward to  
 1052 westward power of equatorial precipitation in the 30-60 day, zonal wavenumbers 1-3 band (e.g.  
 1053 Ahn et al. 2017). Teleconnection performance was determined as pattern correlation of North  
 1054 Pacific and North America (15°S - 80°N, 130°E - 60°W) MJO composite 250 hPa geopotential  
 1055 height anomalies between CMIP models and ERA-i reanalysis averaged over all MJO phases. In  
 1056 panel (b), the crosses show the model 250hPa zonal wind RMS error over the full Pacific basin,  
 1057 while the closed circles indicate the longitudinal RMS error of the subtropical jet. See the text  
 1058 and Henderson et al. (2017) for more detailed explanations.

1059



1060

1061 **Figure 3.** a) Winter (Nov-Apr) mean 650-900hPa specific humidity based on ERA-Interim  
 1062 reanalysis; b) Scatter plot of model skill for eastward propagation of the MJO versus model skill  
 1063 of the mean 650-900hPa moisture pattern over the Maritime Continent (red rectangle in a) based  
 1064 on multi-model simulations from the MJOTF/GASS project. Model MJO propagation skill is  
 1065 defined by the pattern correlation of Homvöller diagrams of model simulated rainfall anomalies  
 1066 associated with the MJO against its observed counterpart following Jiang et al. (2015). Red  
 1067 (blue) dots denote good (poor) MJO models. c) Scatter plot of MJO amplitude and model  
 1068 convective moisture adjustment time scale in models (black dots) and observations (red dots).  
 1069 The MJO amplitude in each model is defined by the standard deviation of 20-100 day band-pass  
 1070 filtered rainfall over the Indian Ocean (75-85°E; 10°S-10°N) during winter. Convective time  
 1071 scale in a model is defined by the ratio of precipitable water (W) anomaly to precipitation (P)  
 1072 anomaly associated with the MJO and derived by a regression approach. Before conducting the

1073 regression, both W and P anomalies are subject to 20-100 day filtering and averaged over the  
1074 Indian Ocean (75-85°E; 10°S-10°N) region. Adapted from Jiang et al. (2016) and Gonzalez and  
1075 Jiang (2017).

1076

1077

1078

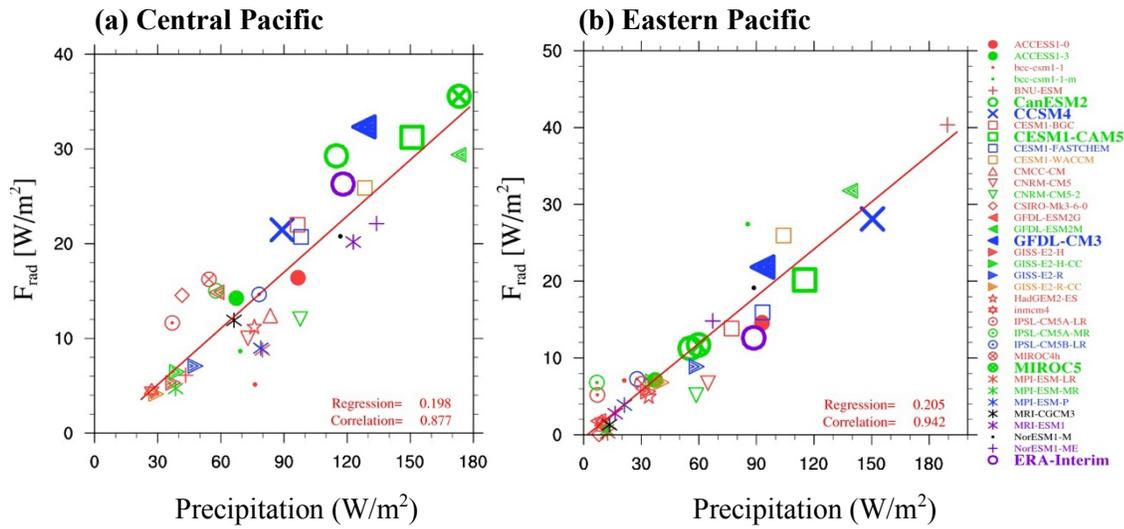
1079

1080

1081



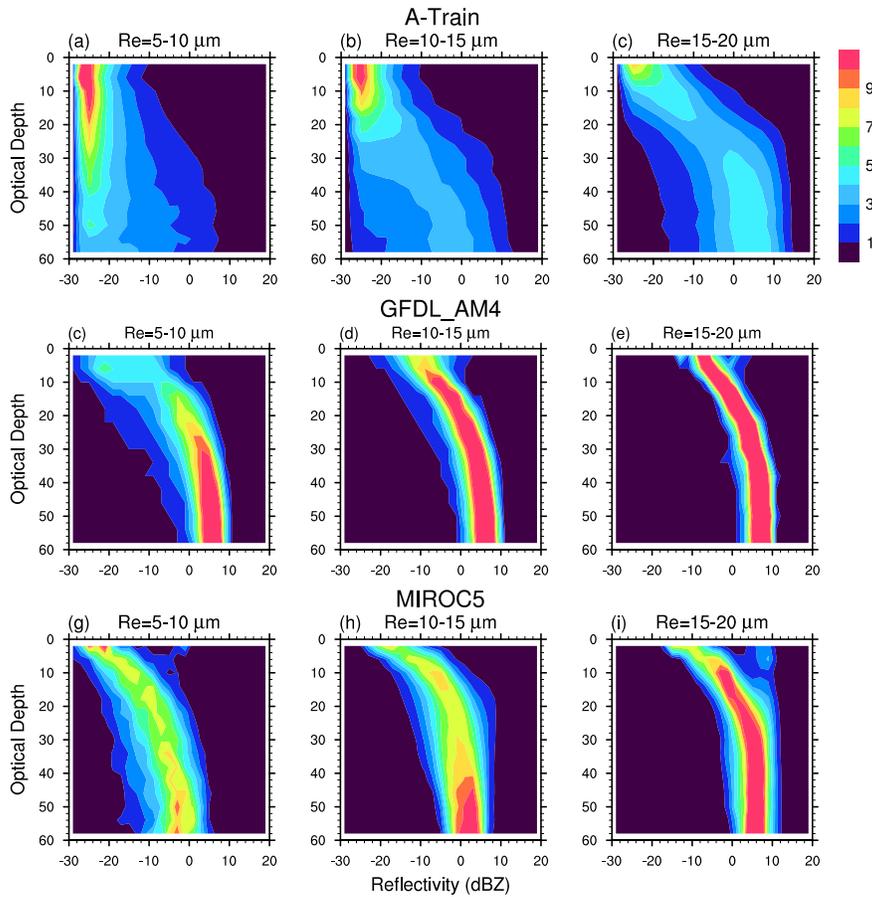
11107  
11108  
11109  
11110  
11111  
11112  
11113



11114  
11115  
11116  
11117  
11118

11119 **Figure 5:** Scatter plots between CMIP5 models simulated precipitation versus net radiative flux  
11120 divergence ( $F_{\text{rad}}$ ): (a) equatorial central Pacific ( $160^{\circ}\text{E}-160^{\circ}\text{W}$ ;  $10^{\circ}\text{S}-5^{\circ}\text{N}$ ) and (b) eastern Pacific  
11121 ( $160^{\circ}\text{W}-80^{\circ}\text{W}$ ;  $5^{\circ}\text{S}-5^{\circ}\text{N}$ ). Results are for El Niño winter composites. In each panel, inter-model  
11122 correlations and best-fit regressions are also provided.

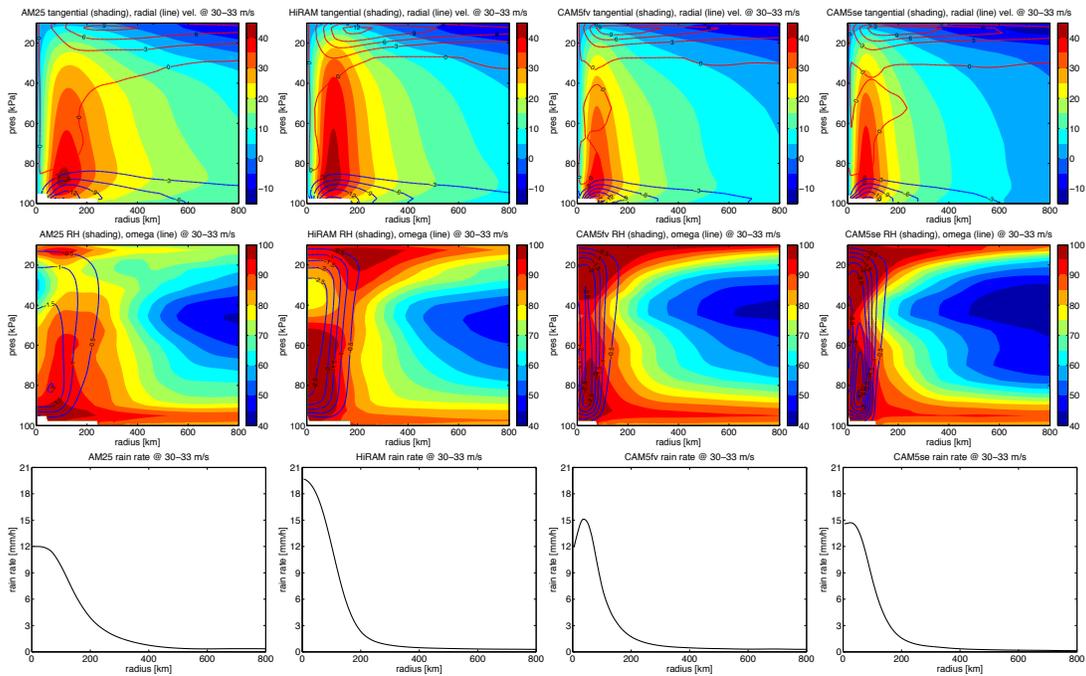
11123  
11124  
11125  
11126  
11127  
11128  
11129  
11130



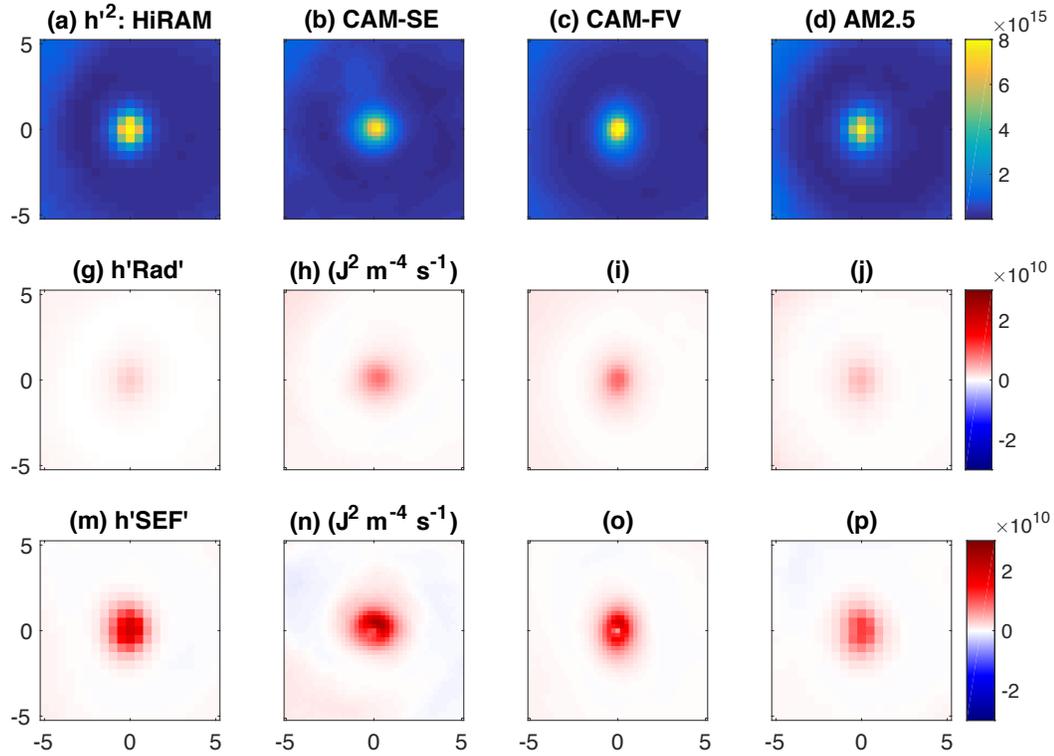
1131  
1132

1133 **Figure 6:** The probability density function (color shading in the unit of %/dBZ) of radar reflectivity  
 1134 (abscissa) normalized as a function of cloud optical depth increasing downward (ordinate), which  
 1135 is further classified according to different ranges of cloud-top particle sizes for (left to right) 5-  
 1136 10mm, 10-15mm and 15-20mm obtained from (top) A-Train satellite observations, (middle)  
 1137 GFDL-AM4 and (bottom) MIROC5 (adapted from Jing et al. 2017).

1138



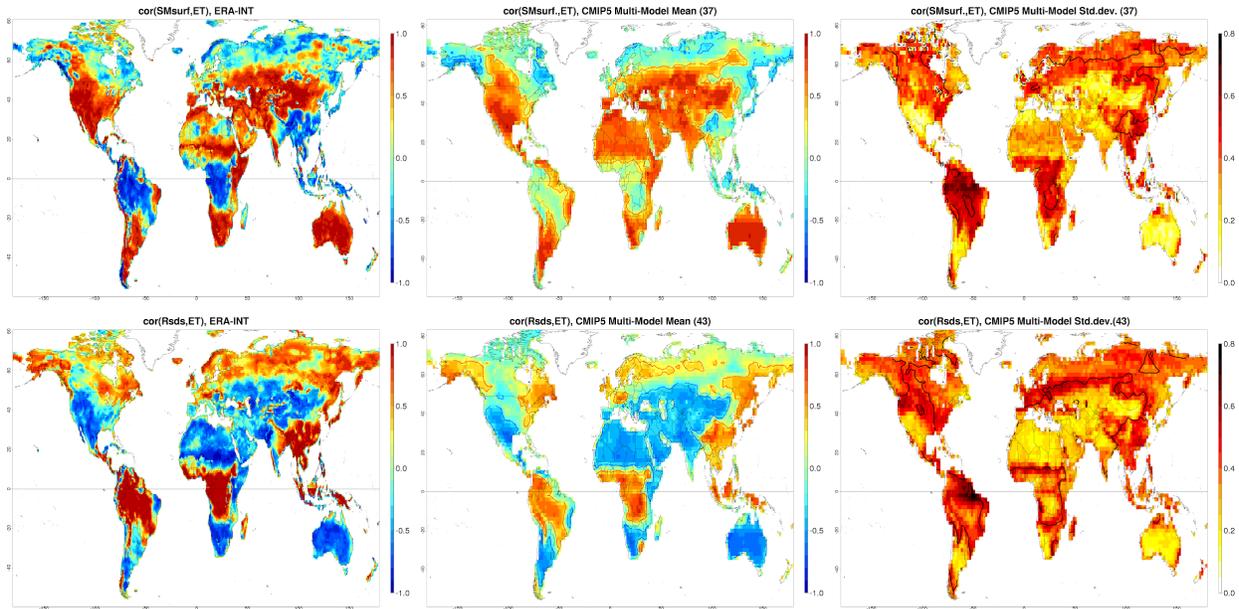
1139  
 1140 **Figure 7:** The first two rows show azimuthally averaged radius-pressure plots of (top) tangential  
 1141 velocity (shading) and radial velocity (lines); and (middle) relative humidity (shading) and  
 1142 pressure velocity (lines). The bottom row shows azimuthally averaged rain rate. All panels show  
 1143 the composites when the simulated tropical cyclones have a similar intensity, in this case between  
 1144  $30\text{-}33\text{ m s}^{-1}$ . Positive/negative values are plotted in red/ blue contour lines. Four different GCM  
 1145 simulations (GFDL Atmosphere Model 2.5 [AM2.5] and HIgh Resolution Atmospheric Model  
 1146 [HiRAM] models at  $0.5^\circ$  resolution and National Center for Atmospheric Research (NCAR)  
 1147 Community Atmosphere Model version 5 - Finite Volume (CAM5-FV) and CAM5-Spectral  
 1148 Element (CAM-SE) models at  $0.25^\circ$  resolution) are used.  
 1149



1150  
1151

1152 **Figure 8:** The first row shows squared anomalies of column-integrated moist static energy (MSE),  
 1153 in units of  $J^2 m^{-4}$ , where anomalies are from the mean of a  $10^\circ \times 10^\circ$  box surrounding a TC. The  
 1154 second and third rows show the radiative and surface flux feedback terms in the MSE variance  
 1155 budget, plotted as a function of latitude and longitude relative to the TC center. All panels show  
 1156 the composites when the simulated tropical cyclones have a similar intensity, in this case between  
 1157  $30\text{-}33\text{ ms}^{-1}$ . Four different GCM simulations are shown: GFDL AM2.5 (panels d,j,p) and HIRAM  
 1158 (panels a,g,m) models at  $0.5^\circ$  degree resolution and NCAR CAM5-FV (panels c,i,o) and CAM5-  
 1159 SE (panels b,h,n) models at  $0.25^\circ$  resolution.

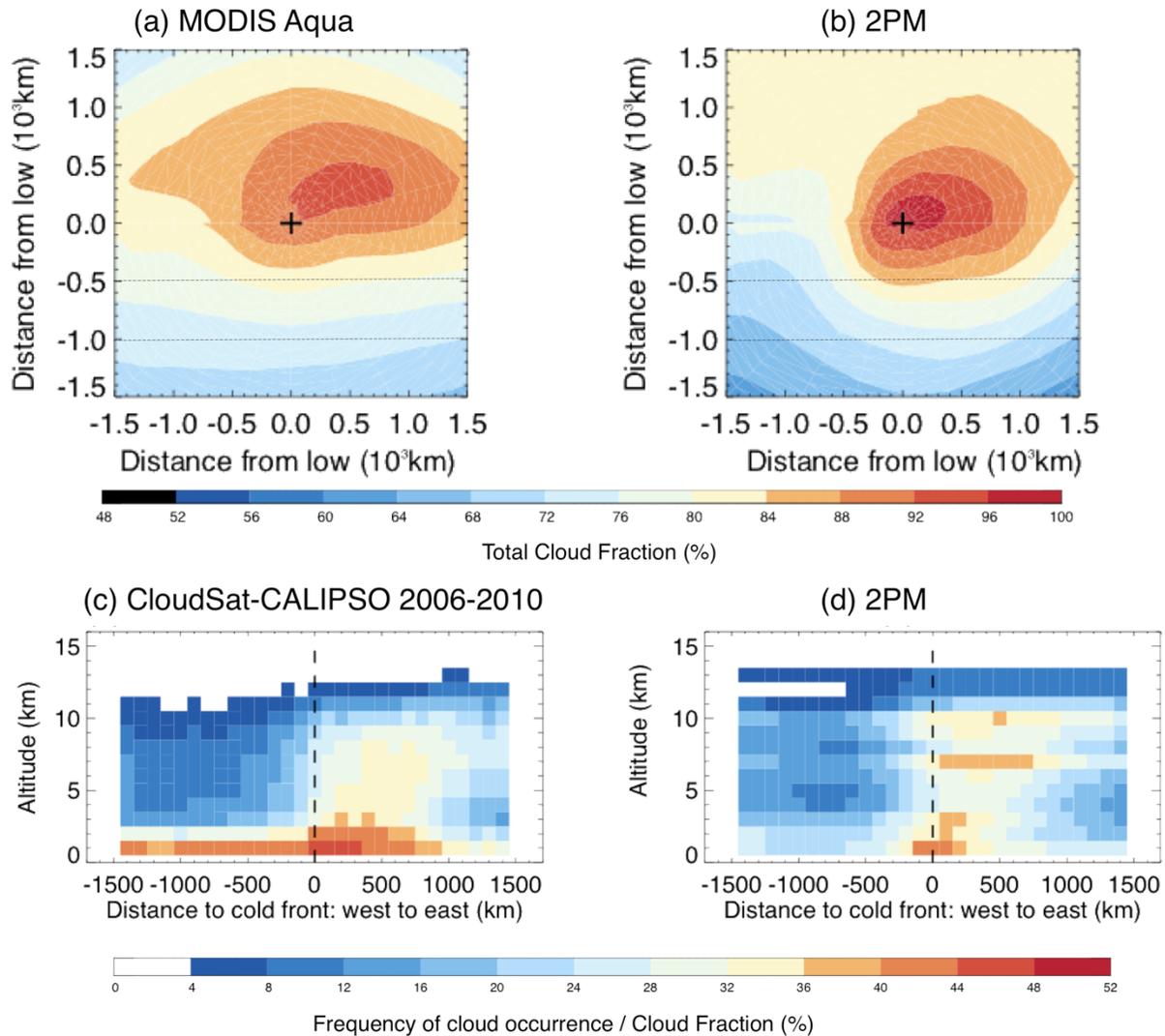
1160  
1161



1162  
1163

1164 **Figure 9:** (top row) correlation between summertime-mean surface soil moisture and  
 1165 evapotranspiration (ET) in ERA-INTERIM (left), in CMIP5 models (middle), and standard  
 1166 deviation across CMIP5 models (right). (Bottom row) same as top for the correlation between  
 1167 (ET) and incoming surface solar radiation (Rsds). Summertime is defined as June-July-August in  
 1168 the northern hemisphere and December-January-February in the southern hemisphere.  
 1169 Correlations are over 1979-2014 for ERA-INTERIM and 1950-2005 for CMIP5 models, using  
 1170 outputs from the historical simulations. Soil moisture and ET outputs were available for 37  
 1171 models, and for 43 models for Rsds and ET.

1172  
1173

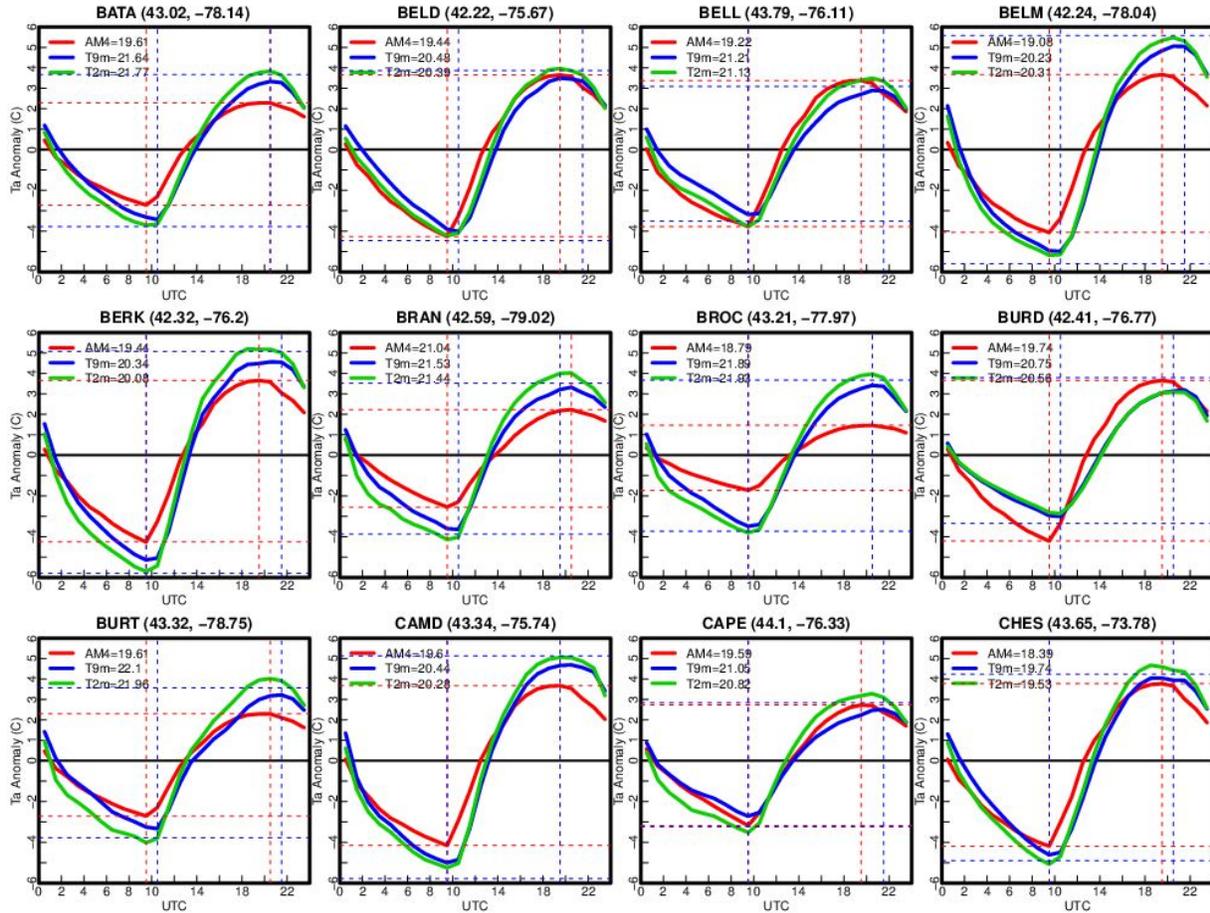


1174  
1175

1176 **Figure 10:** Composites of cloud fraction for cyclone-centered map view (top row) and cold-front  
 1177 centered transects (bottom row). Observational data: (a) MODIS Aqua and (c) CloudSat-Cloud-  
 1178 Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO). Model data: GFDL pre-  
 1179 CMIP6 model that uses two-plume convection scheme (i.e. 2PM). Composites are generated  
 1180 using cyclones or fronts from 5 years of observations or model data, within 30°-60° latitude band  
 1181 in both hemispheres. Plus sign in top row marks the center of the cyclones. The region between  
 1182 the black dashed lines in top row is approximate locations for transects used in frontal analysis.

1183 Black dashed line in bottom row indicates location of the cold front based on 850-hPa potential  
1184 temperature and wind gradient data (see Naud et al. 2014 and Naud et al., 2016 for compositing  
1185 methodology details).

1186



1188

1189

1190 **Figure 11.** The observed mean diurnal cycle of surface air temperatures at 2m (green) and 9m  
 1191 (blue) above the ground during July of 2016 and 2017 at 12 New York Mesonet stations,  
 1192 compared with the GFDL AM4 model simulated multi-year mean 2m air temperature (red) for  
 1193 the grid box containing the station. The daily mean value is shown in the line legend. The 2m  
 1194 height in the AM4 is 2m above a reference height, which is close to the displacement height ( $\sim 2/3$   
 1195 of the canopy height) in the AM4, rather than 2m above the ground as for weather stations.

1196

1197

1198

1199