Benchmarking Simulated Precipitation in Earth System Models

WORKSHOP REPORT

U.S. DEPARTMENT OF ENERGY | Office of Science

Published March 2020

DOE/SC-0203

Recommended Citation:

U.S. DOE. 2020. *Benchmarking Simulated Precipitation in Earth System Models Workshop Report,* DOE/SC-0203, U.S. Department of Energy Office of Science, Biological and Environmental Research (BER) Program. Germantown, Maryland, USA.

Benchmarking Simulated Precipitation in Earth System Models

Workshop Report

Convened by
U.S. Department of Energy
Office of Science
Biological and Environmental Research Program

July 1 to 2, 2019
Rockville, Maryland

Organizers
Renu Joseph, Regional and Global Model Analysis

Co-Chairs and Writing Team
Peter Gleckler, Lawrence Livermore National Laboratory
Christian Jakob, Monash University
Ruby Leung, Pacific Northwest National Laboratory
Angeline Pendergrass, National Center for Atmospheric Research

U.S. DEPARTMENT OF ENERGY | Office of Science

Published March 2020
DOE/SC-020X

Acronyms and Abbreviations

AMIP	Atmospheric Model Intercomparison Project
ASoP	Analysing Scales of Precipitation
AR	atmospheric river
ARM	Atmospheric Radiation Measurement
ARTMIP	Atmospheric River Tracking Method Intercomparison Project
BOG	breakout group
CCB	cold conveyor belt
CCI	Commission for Climatology
CCS	cold cloud system
CDD	consecutive dry days
CESM	Community Earth System Model
CLIVAR	Climate and Ocean: Variability, Predictability and Change
CLUBB	Cloud Layers Unified By Binormals
CMEC	Coordinated Model Evaluation Capabilities
CMIP	Coupled Model Intercomparison Project
CORDEX	Coordinated Regional Climate Downscaling Experiment
CWV	column water vapor
DECK	CMIP Diagnostic, Evaluation and Characterization of Klima experiments
DI	dry intrusion
DJF	December–January–February
DOE	U.S. Department of Energy
E3SM	Energy Exascale Earth System Model
ECMWF	European Centre for Medium-Range Weather Forecasts
EESM	Earth and Environmental Systems Modeling
EGU	European Geophysical Union
ENA	Eastern North Atlantic
ENSO	El Niño Southern Oscillation
ERA	ECMWF Re-Analysis
ESM	earth system model
ET	Expert Team
ETC	extra-tropical cyclone
ETCCDI	Expert Team on Climate Change Detection and Indices
FROGS	Frequent Rainfall Observations on Grids
GASS	GEWEX Global Atmosphere and System Studies
GDAP	GEWEX Data Assessment Panel
GEWEX	Global Energy and Water cycle Exchanges
GEV	Generalized Extreme Value
GFDL	Geophysical Fluid Dynamics Laboratory
GISS	Goddard Institute of Space Studies
GPCP	Global Precipitation Climatology Project
HighResMIP	High-Resolution Model Intercomparison Project
IDF	intensity-duration-frequency

ILAMB	International Land Model Benchmarking Project
IMILAST	Intercomparison of Midlatitude Storm Diagnostics
IOC	Intergovernmental Oceanographic Commission
IPWG	International Precipitation Working Group
IVT	integrated vapor transport
JCOMM	Joint Technical Commission for Oceanography and Marine Meteorology
MAE	mean absolute error
MAPP	NOAA Modeling, Analysis, Predictions and Projections
MCS	mesoscale convective systems
MD	monsoon depression
MDTF	Model Diagnostics Task Force
MJO	Madden-Julian Oscillation
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NOAA	National Oceanic and Atmospheric Administration
NSA	North Slope of Alaska
NWP	numerical weather prediction
PCMDI	Program for Climate Model Diagnosis & Intercomparison
PDF	probability distribution function
PF	precipitation feature
PMP	PCMDI Metrics Package
PNA	Pacific North America Pattern
RGMA	Regional and Global Model Analysis
RMCS	robust mesoscale convective system
RMS	root mean square
RMSE	root-mean-square error
Rx1day	monthly maximum 1-day precipitation
Rx3h	monthly maximum three hourly precipitation
Rx5day	monthly maximum consecutive 5-day precipitation
SDII	Simple Daily Intensity Index
SGP	Southern Great Plains
SJ	sting jet
SLP	sea level pressure
SPI	Standardized Precipitation Index
SST	sea surface temperature
TC	tropical cyclone
TRMM	Tropical Rainfall Measuring Mission
UNICON	Unified Convection Scheme
USGCRP	United States Global Change Research Program
WCB	warm conveyor belt
WCRP	World Climate Research Programme
WGNE	Working Group on Numerical Experimentation
WMO	World Meteorological Organization
WS	weather state

Executive Summary

Earth system models (ESMs) bridge observationally based and theoretical understanding of the Earth system. They are among the best tools to study a variety of questions related to variability and changes in the Earth's climate. ESMs realistically simulate observed large-scale precipitation patterns and seasonal cycles that have a multitude of societal and national security implications.

Despite steady improvement in the simulation of precipitation characteristics, persistent errors in several aspects of simulated precipitation preclude higher confidence in using ESMs to understand earth system variability and change and to make decisions.

In July 2019, the Regional and Global Model Analysis (RGMA) Program Area within the Earth and Environmental Systems Modeling (EESM) Program in the Climate and Environmental Sciences Division at the U.S. Department of Energy (DOE) led a two-day Precipitation Metrics Workshop, led by DOE Program Manager Renu Joseph and co-chaired by Peter Gleckler of Lawrence Livermore National Laboratory, Angeline Pendergrass of the National Center for Atmospheric Research (NCAR), and Ruby Leung of Pacific Northwest National Laboratory.

A diverse group of experts participated in the workshop, including model developers, observational experts, scientists with expertise in diagnosing or evaluating simulated precipitation and related processes, and several with experience in objectively summarizing model performance with metrics. Among others, they represented the National Aeronautics and Space Administration (NASA), National Oceanic and Atmospheric Administration (NOAA), DOE national laboratories, and universities.

Key Workshop Expectation

Identify a set of performance metrics that can serve as a baseline to gauge the agreement between observed and simulated precipitation and discuss exploratory metrics for future use.

The impetus for the workshop: Improve modeled precipitation by designing a capability to comprehensively evaluate ESMs—a capability that will help ESM developers better understand their models, providing them with quantitative targets for demonstrating model improvements. Two main thrusts drove the workshop dialogue:

- identify a holistic set of observed rainfall characteristics that could be used to define metrics to gauge the consistency between ESMs and observations
- assess state-of-the-science methods used to evaluate simulated rainfall and identify areas of research for exploratory metrics for improved understanding of model biases and meeting stakeholder needs.

Baseline Metrics

Throughout the workshop, discussions frequently addressed a key expectation of the workshop—identification of a set of observed characteristics to be used for model benchmarking. It was widely recognized that there is no one right way to do this, and multiple viable approaches were discussed. Workshop participants did, however, agree that it was important to establish a starting point and that this effort would improve and expand over time.

During the final plenary session, a set of six large-scale characteristics was agreed upon as an appropriate starting point for developing a set of baseline precipitation metrics. A proposed tiered system includes a wider range of quantitative measures that would provide much more detail than the six scales. These measures are designed to be applied to the common set of simulations requested from all modeling groups participating in the current phase of the Coupled Model Intercomparison Project (CMIP6).

Exploratory Metrics

While the basic function of precipitation metrics is to benchmark model simulations of precipitation for documenting model performance and improvements over time, precipitation metrics are useful for a broad community of researchers and stakeholders with interest in precipitation.

Exploratory metrics go beyond the baseline metrics and often include aspects of model simulations that require higher-temporal-frequency precipitation data or cannot be evaluated based on precipitation data alone.

They can be useful for model developers in guiding model development, for earth system scientists investigating precipitation variability and change, and for researchers and stakeholders interested in specific aspects of precipitation relevant to their applications. Based on the users and their needs, exploratory metrics were grouped into three types according to their functions and characteristics: process-oriented metrics, regime-oriented metrics, and use-inspired metrics.

Charting a Path Forward

Research Community Engagement: Representatives of several related international activities attended and were engaged in the workshop. Both the Global Energy and Water cycle Exchanges (GEWEX) Data Assessment Panel and the International Precipitation Working Group (IPWG) provided crucial expertise with respect to remote and in situ-based measurements.

As progress with precipitation metrics advances with both groups, collaboration with the World Climate Research Programme (WCRP) will help expand the effort to engage with the broader modeling community. Briefings of this activity are expected at upcoming sessions of the Working Group on Numerical Experimentation (WGNE) and GASS (GEWEX Global Atmosphere and System Studies).

Benchmarking: To establish a baseline for climate model precipitation benchmarking, the first step is to apply the initial set of metrics agreed upon to CMIP6. A publication will document the skill, according to these metrics, of precipitation in CMIP6 simulations. Results from earlier generations of climate models will be included in the publication to document progress over the last 20 years.

In five to seven years, this procedure will be revisited to assess the progress made in the intervening period. The goal is to both motivate progress on improving model precipitation and to facilitate it by providing appropriate and holistic observational targets. In the interim, baseline metrics will gradually develop as they are informed by the exploratory efforts.

Exploratory: There are long-term needs to develop and improve exploratory metrics for broad use; there are also strong foundations for metrics applied to climate simulations. A working group on exploratory metrics was established at the workshop to develop coordinated near-term activities to advance the development and use of exploratory metrics.

The first identified activity is a collaborative effort on a manuscript to discuss the need for exploratory metrics, to introduce an initial set of exploratory metrics, and to apply them to simulations produced by CMIP6 models. These model outputs can be useful for different communities of users (model developers, climate scientists, and impacts researchers and stakeholders).

The working group identified the following topics to include from the process-oriented, regime-oriented, and use-inspired metrics for demonstration of an initial set of exploratory metrics:

- Coherence in space and time
- Frontal precipitation
- Top 10 extreme events
- Convection onset
- Orographic enhancement
- Monsoon
- Mesoscale convective systems
- Madden-Julian Oscillation
- Atmospheric rivers.

Near-term activities will be coordinated within the group and with relevant activities supported by other agencies and programs.

Next Steps:

1. Assess the current generation of models to document a baseline.
2. Bring metrics together into a common analysis framework.
3. Make the capability available to modelers and challenge them to improve their models.
4. Work with the WCRP to promote an initiative to stimulate the challenge and—ideally—bring resources to modelers.
5. Revisit with a next generation of models to see how well models have improved.

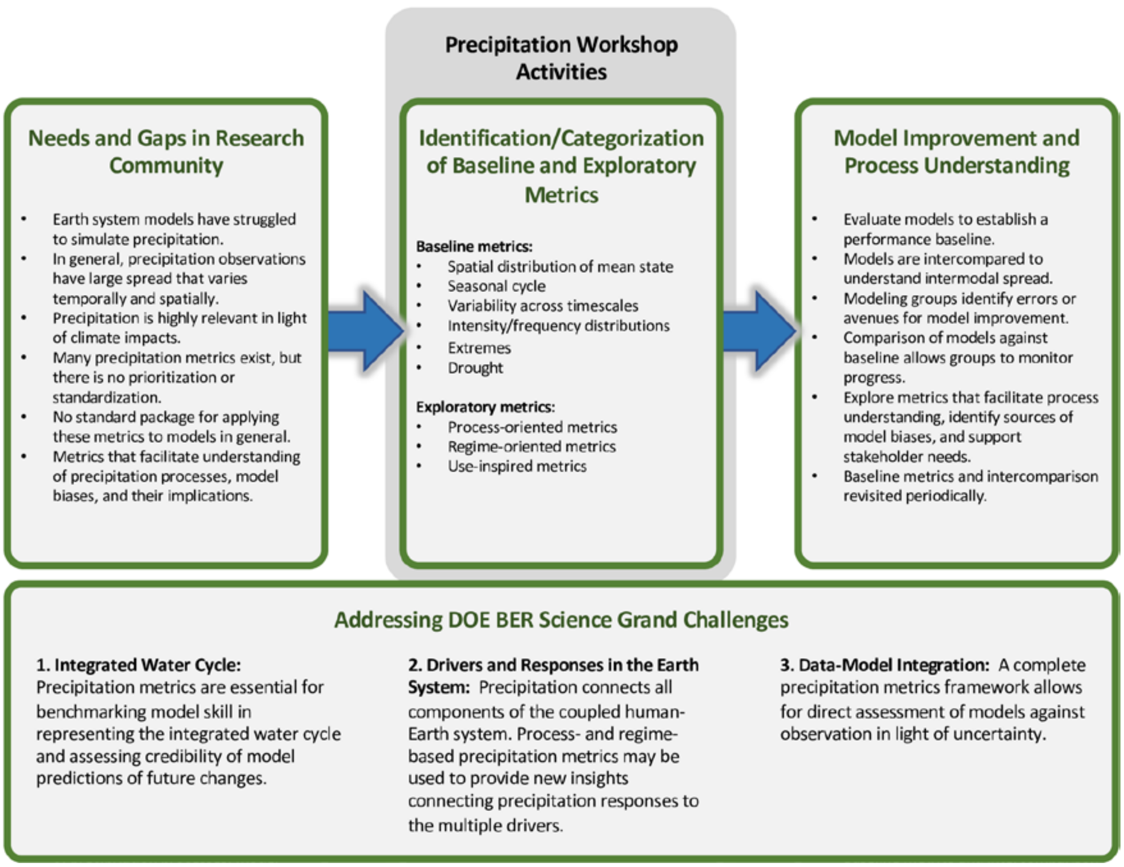

Figure 1. The workshop addressed existing needs and gaps from the research community by defining and prioritizing a set of precipitation metrics and developing a future strategy for model evaluation and intercomparison.

Contents

Introduction

Earth System Models (ESMs) have continuously struggled with simulating precipitation accurately, which has resulted in persistent biases evident across generations of models. A key factor driving this lack of progress on precipitation is the lack of attention that many of its facets receive in the model development and tuning process; namely, insufficient attention is paid to most aspects of precipitation aside from the mean spatial pattern. Instead, the tuning process focuses on the time series of global mean surface temperature, often optimizing temperature at the expense of the fidelity of precipitation. One barrier to addressing these shortcomings is a lack of consensus about which characteristics of precipitation to target, and what the best observations of these characteristics are. The result is that precipitation characteristics, such as intensity, duration, or intermittency, are often incorrect.

Accurately simulating the many processes that contribute to precipitation is not the only challenge—observing precipitation is also a persistent difficulty. Model development groups are less than ideally situated to assess which characteristics of precipitation are observed with the highest confidence. Furthermore, uncertainty in observations of precipitation is usually large—so including uncertainty information along with observations is essential to avoid over-fitting during the model development and tuning process.

The precipitation metrics workshop was prompted by the need to expedite the improvement of precipitation in models. The main thrust of the workshop was to identify benchmarks and metrics to evaluate models, including understanding the limitations in the observations used to characterize reality and track progress in models. To address the gap in information about the quality of simulated precipitation, a prioritized set of precipitation characteristics and a set of observational benchmarks, including uncertainty, has been needed. This workshop thus gathered an expert team on model evaluation and precipitation observations to: identify a set of metrics for precipitation that address salient features, processes, and use cases; prioritize those characteristics that should be targeted for improvement in model development and evaluation; and identify the best available observations of these metrics and their uncertainty. This set of observational benchmarks for precipitation can facilitate a focus on improving precipitation in the model development process.

Faithfully reproducing the many spatial and temporal scales of precipitation (Figure 2) is one of the most important and yet also the most challenging tasks of ESMs. Precipitation is also the signature of atmospheric latent heating, determining circulation features from global to local scales, and is intimately linked to cloud processes and cloud-radiative effects that dominate modeling uncertainties in quantities such as the sensitivity of temperature to radiative forcing. Without progress in modeling precipitation, a multitude of barriers will remain. For example, a lack of fidelity in modeled rainfall will compromise the realism in simulating biogeochemical interactions with the land surface, which in turn will limit our ability to estimate carbon feedbacks. With a carefully selected expert team, we have begun the process of designing and implementing a capability that will enable routine and systematic evaluation of simulated precipitation at all scales in ESMs. This analysis suite will capture a holistic set of precipitation characteristics to quantifiably interrogate models with observations across space and time scales. It will provide clear targets to focus on modeling priorities toward improving key processes directly linked to simulating precipitation deficiencies. Certainly, this effort is an essential component of the process for evaluating and improving models (Figure 3).

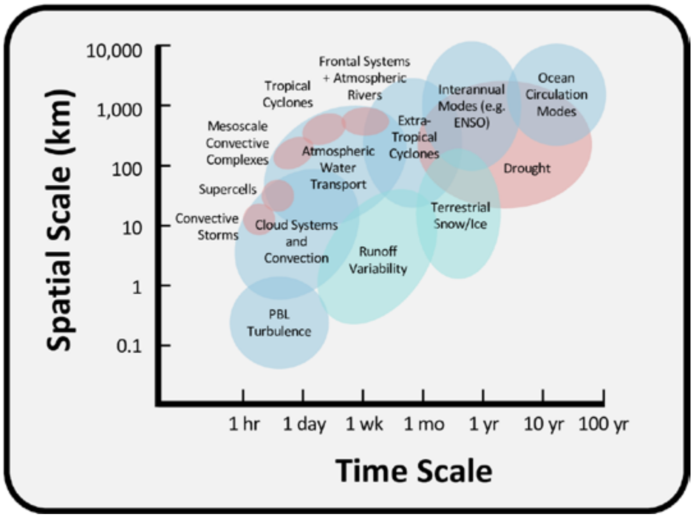

Figure 2. Spatial and temporal characteristics of atmospheric processes and features relevant to precipitation.

Just as precipitation spans many space and time scales, precipitation errors in model outputs can be found on all scales, ranging from large-scale, long-standing rainfall biases in the tropics, to errors in simulating rainfall associated with mid-latitude frontal systems and large-scale tropical circulations such as the Madden Julian Oscillation (MJO), to local errors in the diurnal phase and amplitude of precipitation. It will be essential to probe all relevant phenomena with a unified analysis capability to test the performance of CMIP-class ESMs, with an objective to quantify the added skill of the higher-resolution models on the horizon.

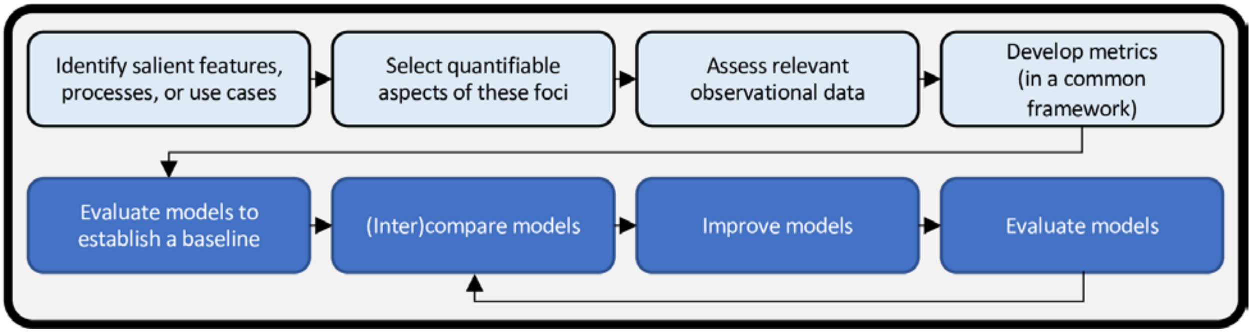

Figure 3. A flowchart depicting a simplified view of the metric and model development process.

Workshop Structure

The two-day precipitation metrics workshop was organized to facilitate extensive participant discussions (Appendix A). Many of the approximately 40 participants gave presentations, and others were given frequent opportunities to contribute to the ideas being developed in breakouts and plenary discussion sessions. Breakouts were held on both days, followed by plenary sessions for breakout reports and continued discussion on advancing the objectives and goals of the workshop.

Objectives and Goals

Despite many years of effort and significant investment, model errors in precipitation have remained large, hindering the use of ESMs in decision making. Progress hinges on mobilizing ideas and resources to address the many problems likely involved in the poor simulation of precipitation, from cumulus parametrization to microphysical processes, and the atmospheric thermodynamic environment and circulation. This workshop has attempted to reinvigorate focus toward improving simulated precipitation, which is urgently needed, by bringing together model developers, observationalists, and theoreticians who focus on different aspects of precipitation, including its spatial distribution and the frequency of occurrence of extreme events.

Overarching Workshop Objectives

To identify precipitation characteristics that will be used to establish a limited set of benchmarks for gauging the consistency between ESMs and observations.

To assess state-of-the-science methods used to evaluate simulated precipitation, and to identify areas of research where well-established metrics are needed but currently lacking.

The immediate goals of the workshop were accomplished by establishing a set of precipitation characteristics for gauging model performance (discussed in Section 3) and identifying research topics for developing exploratory metrics (Section 4). Furthermore, the active discussions and enthusiasm evident during the workshop indicated the group's preparedness for embarking on a long-term initiative.

After the workshop, the short-term goal is to establish a baseline by documenting the performance of the current-generation models with the agreed-upon benchmarks. Related to this goal is the need to create an analysis capability that enables modelers to reapply the performance tests to newer model versions. During this and later phases of the effort, the exploratory group will be tasked with developing increasingly insightful metrics that are sufficiently robust for inclusion in the baseline set. It is envisioned that as these benchmarks mature, the modeling community will recognize them as viable targets for improving their models and that funding agencies will support them in striving to demonstrate improvement. This sequence will facilitate the longer-term goal of establishing performance targets that guide the improvement of models.

Scope

The impetus for this workshop was developed over several years, based on concerns that there has been inadequate progress in improving the quality of precipitation projections using ESMs, particularly during CMIP5 and CMIP6. Parallel to these concerns, it has also been recognized that forecasting precipitation using deterministic and stochastic numerical weather prediction (NWP) models has been lacking, particularly for subseasonal-to-decadal applications. But the research environment for longer-term coupled model projections is unique. In the NWP community, specific metrics for simulation skill have provided a useful target for general improvement, and these simple metrics have motivated and enabled documented progress towards higher-quality weather forecasts over the last several decades. A classic example of a simple metric that has provided a useful benchmark and target for NWP is the skill of simulating 500 hPa geopotential height, quantified by its anomaly correlation. Despite an emphasis on the importance of simulating precipitation in ESMs and a general sense that sufficient accuracy is lacking at present, the community has been more reluctant to take up specific metrics for skill in simulating precipitation. Limited resources have often been used to include additional processes considered important for century-scale simulations, arguably at the expense of progress in precipitation. A comprehensive approach to evaluating simulated precipitation would be useful for many applications (including NWP), but attempting to address all applications at once could limit progress, hence the initial focus here on CMIP-class ESMs. In many ways, a scope of CMIP-class models helps bring focus to the effort. Carefully designed experimental protocols are already in place and used by the modeling community. These include the CMIP6 Diagnostic, Evaluation and Characterization of Klima (DECK) and coupled Historical experiments. The DECK experiments include coupled and uncoupled protocols that are useful for evaluating simulated precipitation, namely the pre-industrial control and prescribed sea surface temperature and sea-ice atmosphere-only simulations defined by the Atmospheric Model Intercomparison Project (AMIP) protocol. Additionally, the standard model output for each of these experiments is well defined, including, e.g., monthly mean, daily, and 3-hourly precipitation. Selected output fields are also available for snow cover.

To create a framework for evaluating simulated precipitation in ESM simulations, one focus of the workshop was to generate a set of metrics that are ready to be applied to CMIP6 simulations to evaluate them against existing observational data sets. To be able to apply the metrics in short order to all model simulations, this initial set is limited in scope to metrics that require only precipitation data, which will be available from CMIP6 DECK and Historical simulations, and is also limited to metrics that are already established in the scientific literature. The set of metrics is intended to evaluate the characteristics of precipitation as holistically as possible within these constraints. The resulting set of metrics identified at the workshop include the spatial distribution of mean-state precipitation, its seasonal cycle, variability on timescales ranging from diurnal to decadal, the intensity and frequency distributions of precipitation, heavy precipitation extremes, and drought.

In parallel with this initial effort to describe benchmark skill in simulating precipitation for the current generation of model simulations with available model output, observational data, and methodologies, a second focus of the workshop was to identify what next steps will be needed beyond this initial benchmarking. The effort towards exploratory metrics considered questions including: what metrics are useful to quantify the processes that generate precipitation? These often require more model output and observational data than just precipitation, and in many cases, fall into gaps in the existing literature. Other exploratory metrics focus on specific weather regimes, which are associated with process-based metrics but occur at a broader scale. Use-inspired metrics specifically focus on evaluating aspects of simulated precipitation that are important for the needs of users of these simulations. Invariably, there is interest in testing the veracity of simulated precipitation beyond the CMIP DECK and Historical experiments. The first test will be to scrutinize simulations performed at higher resolutions as in the CMIP6-endorsed High Resolution Model Intercomparison Project (HighResMIP). One challenge for evaluating simulated precipitation is determining the best way to quantify the skill that is added from increasing model resolution.

Organization of the Workshop

The workshop organization committee was led by DOE Program Manager Renu Joseph and co chaired by Peter Gleckler of Lawrence Livermore National Laboratory, Angeline Pendergrass of the National Center for Atmospheric Research (NCAR), and Ruby Leung of Pacific Northwest National Laboratory. The first morning began with presentations articulating background, motivation, and expectations of the workshop (Appendix A). All attendees were then given an opportunity to provide their reactions to the organizers' presentations as well as the pre-meeting materials. Following these discussions, additional morning presentations emphasized topics relevant to defining the first set of precipitation metrics. These included perspectives from several modelers, who summarized challenges associated with evaluating simulated precipitation against satellite-based observations, and several presentations related to extremes, impacts, and "use-inspired" metrics. After lunch, two breakout groups met to address the same set of discussion topics (in parallel) related to establishing a baseline set of metrics and how to establish them. The first day concluded with summaries from the breakouts and an extended discussion about the next steps.

The second day began with a discussion period to enable participants to share evening discussions and overnight thoughts. A morning session of presentations followed with topics relevant to exploratory metrics. The second set of breakout groups met during the afternoon: one to advance the establishment of an initial set of metrics (Section 3) and a second to identify research topics that could ultimately lead to a more advanced set of precipitation metrics (Section 4).

Community Participation

A diverse group of experts participated in the workshop, including model developers, observational experts, and scientists with expertise in diagnosing or evaluating simulated precipitation and related processes, and several with experience in objectively summarizing model performance with metrics. Recognizing the critical importance of observations in this effort, attendees included scientists involved in the preparation of an assessment of observationally based precipitation products, which is an effort being led by the GEWEX Data Assessment Panel (GDAP). Other examples include the targeted analysis of NOAA's Model Diagnostics Task Force (MDTF; Maloney et al. 2019) emphasizing process-oriented diagnostics, and developers of the software Analyzing Scales of Precipitation (ASoP; Klingaman et al. 2015, Martin et al. 2017) that diagnose structures of coherence at shorter time scales.

Proposed Baseline for Precipitation Metrics

A key expectation of this workshop was to identify a limited set of performance metrics that could serve as a baseline to gauge the agreement between observed and simulated precipitation. The intent is for this to serve as a starting point or set of building blocks upon which newer performance tests would be included as they are vetted over time. More research is needed to establish these newer exploratory metrics as discussed in Section 4.

Benchmarking Strategy

This workshop has been motivated by the lack of progress in improving simulated precipitation. One reason for the limited progress is that insufficient attention has been given to this problem, sometimes resulting from limited resources being focused on modeling areas not directly related to precipitation and its variability. Given that precipitation is highly relevant to society, the intent of this workshop was to re invigorate interest by establishing a collective effort to tackle the problem. The benchmarking discussions during the workshop and subsequent strategic efforts, sequentially, include:

During the workshop

1. Establish a small group of scientists with a diverse set of expertise related to simulated precipitation, including modelers, analysts, and data experts.
2. Identify a holistic set of characteristics where well-established performance tests already exist.
3. Identify research topics that could lead to more advanced performance metrics that can be applied in the future (Section 4).

After the workshop

1. Assess the current generation of models based on these tests to document a baseline.
2. Bring together these metrics into a common analysis framework.
3. Make this capability available to modelers and challenge them improve their models based on the performance benchmarks.
4. Work with the WCRP to promote an initiative to stimulate the challenge and hopefully bring resources to modelers to address it.
5. Revisit with next-generation models to see how well models have improved.

Baseline Metrics

The breakouts and discussions addressed a key expectation of the workshop—identification of a limited set of observed characteristics to be used for model benchmarking. This was challenging as it was widely recognized that there is no one right approach, and multiple viable approaches were discussed. The workshop participants did, however, agree that it was important to establish a starting point and that this would improve and expand over time via experience and community feedback. For example, some metrics require more information than just about precipitation and relate to generation mechanisms or the large-scale environment. Participants decided that these would not be included in the initial set of metrics, but rather classified as exploratory (discussed in Section 4). In the future, these may be reclassified into baseline metrics, after they are more extensively tested for feasibility in the CMIP archive and recognized as serving the needs of users of the metrics. During the final plenary session, a set of six broad categories was agreed upon as an appropriate starting point for developing a set of baseline precipitation metrics. A tiered system was discussed to include a wider range of quantitative measures that would provide more detail than six scalars (Figure 4). For example, global-scale annual-mean characteristics may be appropriate for a

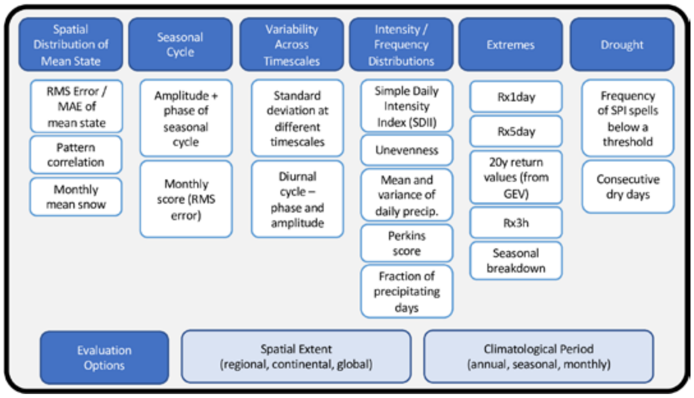

Figure 4. A draft layout of Tiers 1 and 2, resulting from workshop breakout discussions.

top-level tier, but regional and seasonal statistics could be organized into additional tiers. These measures are all designed to be applied to the common set of simulations requested from all modeling groups participating in CMIP6—the Historical and DECK simulations—and focus on only precipitation data at timescales from monthly to 3-hourly, as well as snow. The six characteristics of simulated precipitation are summarized here:

1. How well do models simulate the spatial distribution of average precipitation?

Modelers and analysts have for many years compared global time mean maps of observed and simulated precipitation, and now routinely include difference maps to identify model errors in packages for the model development process as recently highlighted with several different versions of DOE's Energy Exascale Earth System Model (Figure 5; Golaz et al. 2019).

Including standard metrics for the spatial pattern of the mean state enables the metrics set to build on existing workflows as observed and simulated seasonal means are often examined in routine model evaluation. The root-mean-square (RMS) error of the mean state and the pattern correlation are two integrative metrics for the skill of mean precipitation. The mean absolute error (MAE) is a useful complement not heavily weighted by outliers.

While examining the spatial pattern of total precipitation is routine, a lack of trustworthy snowfall measurements has precluded snowfall as being a routine measure. The CloudSat satellite provides a measure of snowfall amount (Figure 6) that can be used to evaluate the spatial pattern of snowfall in model simulations. But there are challenges in making an appropriate comparison between snow characteristics currently observed by satellites and simulated quantities, reducing the overall confidence in evaluation of snow compared to rainfall.

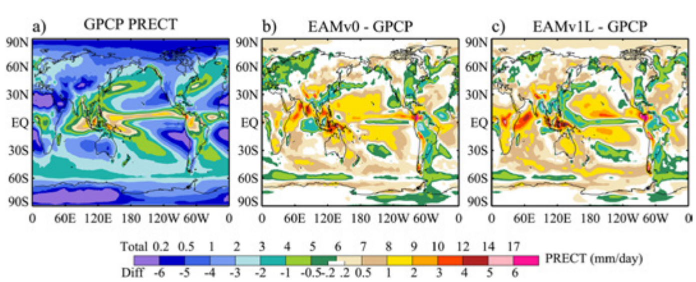

Figure 5. (a) Annual mean Global Precipitation Climatology Project estimated precipitation rates and the biases in several model versions (b) EAMv0 and (c) EAMv1L. The white color in (b) and (c) indicates the regions where differences are less than 0.2 mm/d. Adapted from Xie et al. 2018.

2. How well do models simulate the seasonal cycle of precipitation?

The seasonal means often examined in routine model evaluation reveal the spatial characteristics of a given season, but they do not yield information about the amplitude or phase of the seasonal cycle. A common way to do this is to apply a Fourier transform to the 12-monthly means at each spatial location (grid cell). This provides a map of both the amplitude and phase of the seasonal cycle (Figure 7). Straightforward statistics comparisons between model simulations and observations can then be made analogously to the spatial patterns of the mean state.

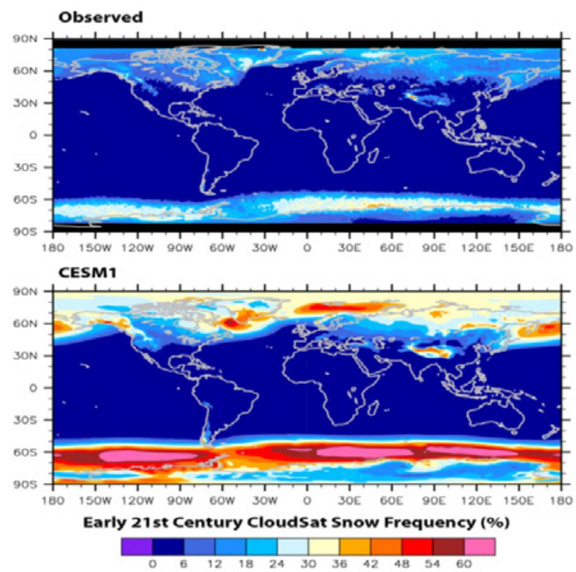

Figure 6. Global map of present-day near-surface snow frequency: observed CloudSat snow (top), Community Earth System Model (CESM)1 CloudSat snow (bottom). CloudSat light snow and CloudSat snow definitions are based on reflectivity and fraction of ice present. (Adapted from Kay et al. 2018)

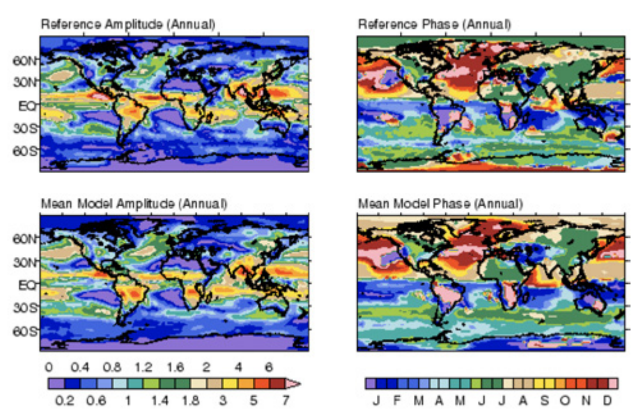

Figure 7. Total precipitation (mm/day). Top row: Observational (CMAP) annual amplitude and phase; Second row: Multi-model ensemble annual amplitude and phase. (Adapted from AchutaRao et al. 2008)

A well-established approach to comparing model and observed patterns is via a Taylor diagram (Taylor 2001), providing a succinct quantitative display of three theoretically related statistics—the pattern correlation, standard deviation, and centered root-mean-square error. Many modeling groups routinely use Taylor diagrams to help summarize performance characteristics. These diagrams are one way the mean and seasonal metrics can be summarized (e.g., Figure 8).

3. How well do models simulate precipitation variability across time scales?

Precipitation varies across a wide range of timescales. At the shortest timescale of model integrations (O[timestep]), there are rich coherence characteristics that are not routinely analyzed but more directly capture modeled precipitation (c.f., Klingaman et al. 2017, Martin et al. 2017). Efforts are advancing to make these analyses more available (e.g., via an open-source package, Analysing Scales of Precipitation – ASoP), but they are not yet routine in model intercomparisons or model evaluation. The expectation is that these performance tests will become more widely adopted and applied in the model development and evaluation process, becoming part of baseline performance tests (see Section 4). With the value of these tests being increasingly recognized, it is expected that the next generation of CMIP will include some output to be saved by model time-step.

The diurnal cycle, a forced component of shorter-timescale variability, is resolved with standard (3-hourly) CMIP output. Models exhibit well-known deficiencies, such as a tendency to produce rainfall too early in the day (Figure 9; Covey et al. 2016, Diaz et al. 2006). Several recent studies have demonstrated that there are also systemic deficiencies in the sub-diurnal or intermittent precipitation in CMIP-class models (Trenberth et al. 2017, Covey et al. 2018), although differences among satellite-derived products are non-trivial. For these and longer timescales, complementary approaches were discussed in breakout groups for evaluating temporal variability: an examination of the temporal standard deviation with different averaging times and band-pass filtered time series. More regime-oriented metrics relevant to the diurnal cycle, such as mesoscale convective systems in the U.S. that exhibit a nocturnal maximum, are discussed in Section 4.

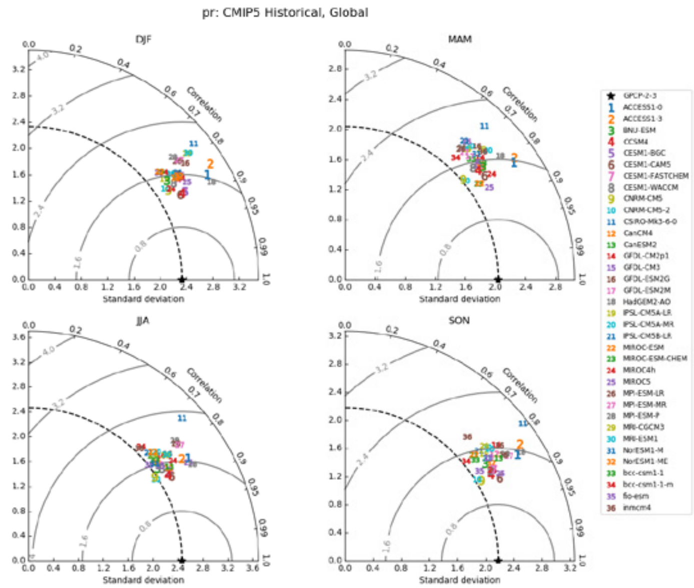

Figure 8. Taylor diagrams quantify the differences between observed and simulated precipitation for the four seasons. Model results are from the CMIP5 and compared to GPCP2.3. (Adapted from the Program for Climate Model Diagnosis & Intercomparison (PCMDI)'s mean-state simulation summaries [https://pcmdi.llnl.gov/research/metrics])

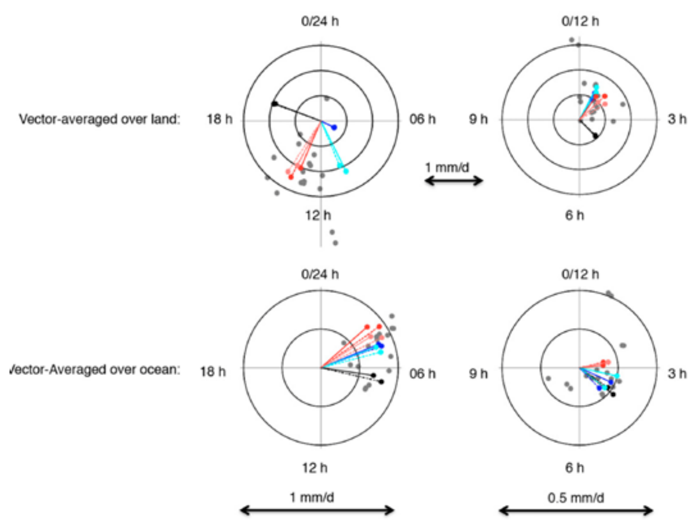

Figure 9. Harmonic dial plots of the amplitude and phase of Fourier components, after vector averaging over land and ocean areas separately, for Tropical Rainfall Measurement Mission (TRMM) 3B42 (black lines and dots), the four highest-resolution CMIP5 models (colored lines), and for the other 17 CMIP5–AMIP models with only July results shown for clarity (gray dots). (Adapted from Covey et al. 2016)

4. How well do models capture observed distributions of intensity and frequency?

Although longstanding systematic biases are evident in the mean-state spatial distributions of precipitation, tougher tests often involve examination of distributions in intensity and frequency (c.f., Pendergrass and Deser 2016). There are various ways to compare observed and simulated distributions. Challenges include qualitative effects of the resolution considered (Figure 10), particularly for reconciling point-like station data with the area-averaged fields from models.

To date, there is no universally adopted approach to quantify the differences between observed and simulated precipitation distributions. As the benchmarking of precipitation is advanced to assess the current generation of CMIP models, it will be necessary to use different approaches to assemble complementary information and determine if any underlying conclusions about changes in model performance depend upon the analysis method chosen. Several examples were discussed at the workshop of varying range in complexity, yielding different information about model agreement with observations.

The Simple Daily Intensity Index (SDII) is a measure of the average intensity of precipitation on days with precipitation. It is part of the World Meteorological Organization (WMO) Commission for Climatology (CCl)/WCRP Climate and Ocean: Variability, Predictability and Change (CLIVAR)/ WMO Intergovernmental Oceanographic Commission (IOC) Joint Technical Commission for Oceanography and Marine Meteorology (JCOMM) Expert Team (ET) on Climate Change Detection and Indices (ETCCDI; Zhang et al. 2011), which is already routinely applied to models and observations—though correct interpretation requires carefully addressing the resolution of the data considered (as mentioned above). Another metric that complements SDII is the fraction of days with precipitation.

The SDII integrates over all types of precipitating events—light, typical, and extreme—obfuscating mechanistic identification and understanding of biases. One metric that focuses on the unevenness of how precipitation falls is the number of wettest days each year, during which half of the total precipitation falls (Figure 11; Pendergrass and Knutti 2018). This measure informs about the relative intensity of heavy precipitation to total precipitation and is not yet part of regular diagnoses.

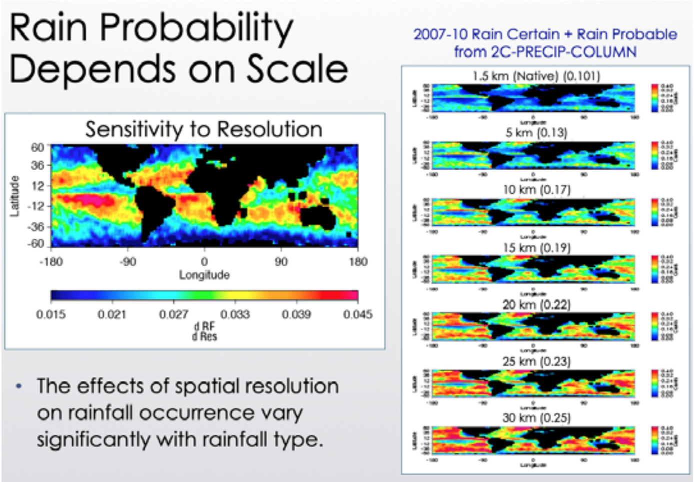

Figure 10. Example thresholds for precipitation occurrence and phase. The frequency of precipitation depends strongly on the scale. On smaller scales, precipitation frequency generally decreases with increasing spatial resolution, as seen clearly with CloudSat observations. (Courtesy T. Lécuyer, University of Wisconsin-Madison)

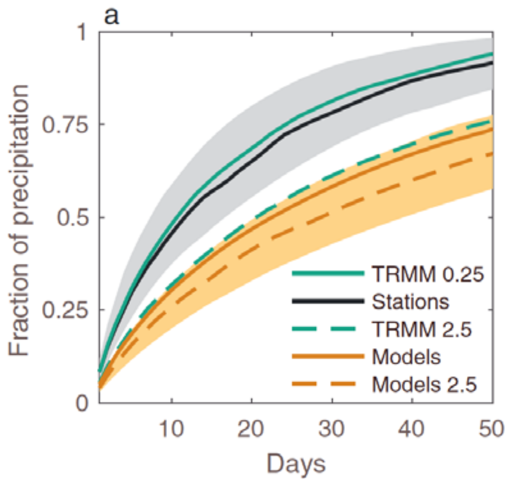

Figure 11. Unevenness of precipitation. Cumulative fraction of total precipitation as a function of the number of wettest days each year. The number of days for half of precipitation corresponds to the value of each line. Present-day observed at stations, according to TRMM 3b42 product at native 0.25° resolution and coarsened to 2.5° and simulated by CMIP5 climate models at native resolution and regridded to 2.5°. Lines show the median across stations. Uncertainty across stations is indicated by the gray shading, which show the 25th and 75th quantiles across stations for station observations. For models, lines show the multi-model median at grid points nearest to stations at native and coarse resolutions. Uncertainty across models is indicated by orange envelopes, which show the range across all models at 2.5° resolution. (Adapted from Pendergrass and Knutti 2018)

Emerging work describes the entire distribution of precipitation with two metrics, which connect modeled and observed precipitation distributions to theoretically based models for precipitation (Martinez-Villalobos and Neelin 2019). A combination of the slope in the bulk of the precipitation distribution (called the "power-law scale") and the rain rate at which the distribution transitions from this regime to a rapidly decreasing one (the "cutoff rate"), combined with a measure of the goodness of fit of this type of distribution, quantifies this integrated measure, which is also connected to a theoretical framework for convective precipitation. As there is room for the theoretical framework to be further developed to facilitate interpretation, more discussion of this type of metric is provided in Section 4.

A final integrated measure to evaluate the skill of the distribution of precipitation is the Perkins score (Perkins et al. 2007). This metric quantifies the difference between two probability distributions, in this case, modeled and observed distributions of precipitation. Applying the Perkins score to different moments of the precipitation distribution (the probability density, the frequency, and the amount/volume distribution) provides information weighted towards different aspects of precipitation.

5. How well do models capture well-observed precipitation extremes?

Precipitation extremes drive climate impacts, so it is crucial for models to simulate them well. Despite their importance, the phenomena associated with precipitation extremes are not always well represented in model simulations, and the fine scales needed to observe highly impactful precipitation events are not always well captured by our observing systems and observational data products. Measures to evaluate simulated precipitation extremes should try to account for these factors.

One set of metrics for extreme precipitation arises from the block-maximum precipitation over various timescales. The maximum daily precipitation accumulation each year, rx1day, describes relatively heavy precipitation and is sufficiently extreme that it behaves differently from mean precipitation (Pendergrass and Knutti 2018). The maximum consecutive 5-day accumulation each year, rx5day, captures protracted synoptic events that can drive flood events. The maximum 3-hourly precipitation each year, rx3h, is a measure of sub-daily extreme precipitation and corresponds more closely to convective timescales than daily precipitation does. These daily indices were first developed by ETCCDI (Zhang et al. 2011) with the sub-daily indices later roughly adapted to be consistent with the ETCDDI daily definitions.

Another set of measures for extreme precipitation are derived from the theory of Generalized Extreme Value (GEV) distributions and rely on statistical modeling of observed and modeled time periods. Long-period return values describe much rarer events than seasonal or annual block maxima. Recent developments in non-stationary methods reduce the statistical uncertainty in their estimation (e.g., Risser et al. 2019), permitting their usage as model evaluation metrics.

Because the phenomena driving precipitation extremes and also the impacts of extremes can be seasonally dependent, examining the seasonal breakdown of extremes will be featured among the tiers of evaluation metrics. Connecting precipitation extremes with their generation mechanisms is considered under exploratory metrics (Section 4).

6. How well do models capture dry periods?

Drought is driven in large part by periods that lack precipitation; meteorological drought is the component of drought that describes this lack of precipitation. The Standardized Precipitation Index (SPI; McKee et al. 1993) quantifies the anomalous precipitation for a given location and can be calculated over different lengths of time, for example, one month to three years, which is normalized to the average precipitation over the time period for the location. Then, the frequency of events falling below a threshold of less-than-normal SPI (such as -1) can be calculated.

Another measure of meteorological drought is the length of dry spells. The ETCCDI indices include the number of Consecutive Dry Days (CDD), which is one measure of dry-spell length.

While many of the basic characteristics highlighted above are at least casually monitored by modelers, a well-organized hierarchy for each of them could be more informative. Partitioning, for example, between the tropics and extra-tropics or over 'wet' and 'dry' land and/or seasons can help better understand the contributions to global-scale statistical comparison. Each measure will be computed over a variety of sub-domains to facilitate this breakdown, forming another component of the second tier of measures.

Observational Data Sets and Their Uncertainty

One challenge for evaluating precipitation in climate models is the uncertainty in observational data sets, and variation among them. For example, even a broad measure like total annual precipitation over a large domain (Figure 12) can vary substantially among data sets. To address observational uncertainty, an evaluation system needs to incorporate multiple observational data sets—an ensemble of in situ and satellite data—to enable quantification of agreement across observational products, since no one data set is uniformly superior across all characteristics of precipitation (Bador et al. 2020, Alexander et al. 2020). Data set choices will be informed by recommendations from the GDAP precipitation assessment. Quantifying the observational uncertainties is one challenge that will continue to be addressed with future research.

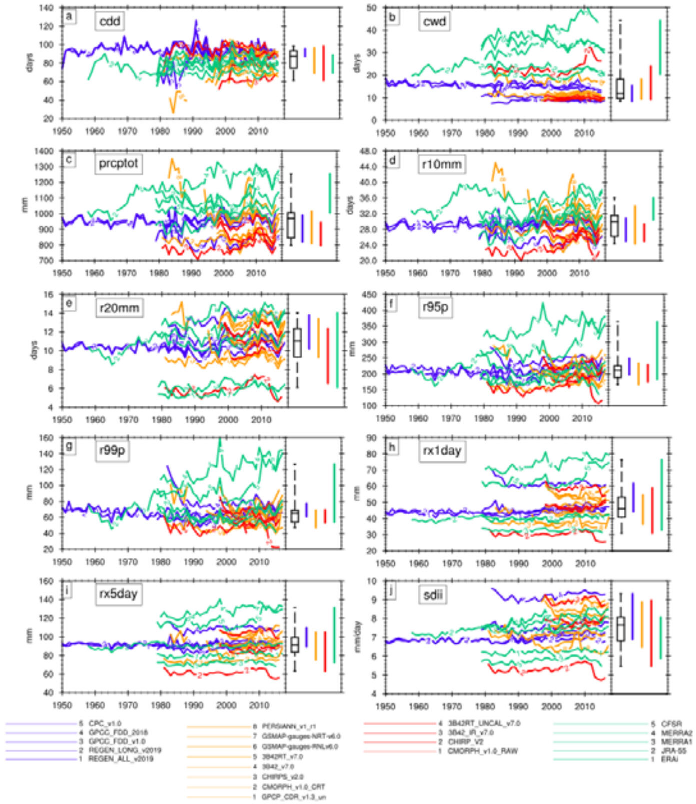

Figure 12. Global land (50S-50N) average timeseries for prcptot (total annual daily precipitation, in mm) in an ensemble of 22 observational products from the Frequent Rainfall Observations on Grids (FROGS) database (Roca et al. 2019). Box and whiskers in each panel provide information on the distribution of prcptot for all products while vertical colored lines indicate the range of values for each type of product: in situ (blue), satellite corrected (orange), satellite uncorrected (red) and reanalyses (green). (Adapted from Alexander et al. 2020)

Exploratory Precipitation Benchmarks

As discussed in Section 3, baseline metrics are performance metrics to be established as a starting point to measure the agreement between observed and model-simulated precipitation. They cover many aspects of precipitation, but because they are intended as building blocks, baseline metrics are initially limited to those that can be calculated using the standard outputs from a common set of simulations from all modeling groups participating in CMIP6. Building on and complementing the baseline metrics, exploratory metrics can serve as benchmarks for increasingly diverse aspects of precipitation to meet the needs of different user communities. Through presentations and breakout group discussions, workshop participants identified the needs for and the types of exploratory metrics that need to be developed to advance the initial set of precipitation metrics in the future. Many of the exploratory metrics require further research and development before they can be implemented in metrics packages for broad community use. A working group on exploratory metrics was established in parallel with the working group on baseline metrics, with a focus on defining coordinated activities towards developing and demonstrating the value of exploratory metrics. These topics are summarized in the subsections below.

The Need for Exploratory Metrics

While the basic function of precipitation metrics is to benchmark model simulations of precipitation for documenting model performance and improvements over time, precipitation metrics are useful for a broad community of researchers and stakeholders with interest in precipitation. Examples of users and their needs for exploratory precipitation metrics are discussed here. Going beyond the baseline metrics (discussed in Section 3), exploratory metrics often require higher-temporal-frequency precipitation output or cannot be evaluated based on precipitation alone.

An important use of precipitation metrics for model developers is to inform or guide model development. Because models are focused around equations that represent individual processes, model developers need process-oriented metrics to diagnose the deficiencies in model parameterizations and gain insights for improving precipitation simulations. For example, they can benefit from metrics that partition the model biases into their component parts to help narrow down aspects of the model parameterizations that may be responsible for biases in each of the precipitation metrics. Relationships that connect precipitation with the thermodynamic and dynamical variables involved in cloud and convection processes may be summarized as metrics to help diagnose model errors in simulating precipitation. Quantifying model biases in the large-scale environment associated with precipitation can also inform model developers whether or how much of the biases in precipitation are a result of model biases in simulating the large-scale environment (e.g., moisture) versus limitations of the model parameterizations in simulating precipitation given a large-scale environment for precipitation that is reasonably captured by the models. There is also a need for metrics that can be applied to short-term initialized forecasts from weather to subseasonal-to-seasonal timescales to differentiate biases associated with fast and slow processes.

Scientists using earth system model simulations to understand the thermodynamic and dynamical contributions to precipitation changes in the future can use more information regarding model skill in variables relevant to precipitation. For example, moisture budget analysis is often used to quantify the thermodynamic and dynamical effects of warming; this type of analysis makes use of information about moisture and vertical motion, which are sensitive to physics parameterizations and model resolutions. Hence, metrics relating precipitation to moisture and vertical motion, or more generally relating precipitation to its generation mechanisms or phenomena such as tropical and extratropical cyclones and different cloud or convection types, are useful for quantifying uncertainty in attributing precipitation changes to thermodynamic and dynamical effects. Quantifying model biases in the specific precipitation regimes of interest is important for guiding the selection of credible models for analysis and characterizing uncertainty in projected changes. Metrics based on emergent relationships that connect model biases in precipitation or closely related variables such as moisture and temperature with the precipitation response to different forcings are also useful for understanding and characterizing uncertainty in projections of precipitation in the future.

Researchers and stakeholders of model precipitation simulations, predictions, and projections often have interests in specific aspects of precipitation relevant to their particular applications. For example, use-inspired precipitation metrics may focus on characteristics related to the space-time variability and coherence as well as intensity, frequency, and duration that impact soil moisture, snowpack, and runoff, with subsequent effects on ecosystems, crops, water resources, and infrastructure. Use-inspired metrics may also include relationships connecting precipitation biases to impacts of precipitation in sectors such as water resources and agricultural production. With a profile of precipitation metric performance, users may be able to identify a subset of more skillful models for their regional applications. Precipitation metrics may potentially provide information that can be incorporated into bias adjustment or scenario creation. The development of use-inspired metrics may initially focus on crop and hydrologic impacts while recognizing that precipitation is relevant to many sectors of our society, so there will be a growing need to expand in this area. Strong and continued engagement with impacts/risk/disaster communities will help identify specific hazard thresholds beyond distribution/percentile metrics.

Types and Examples of Exploratory Metrics

Based on the users and their needs, exploratory metrics can be roughly grouped into a few categories according to their functions and characteristics (Figure 13). Here we summarize the discussion of exploratory metrics under three key categories: **process-oriented metrics, regime-oriented metrics,** and **impacts and use-inspired metrics.** Although different types of metrics may be developed by different communities and, therefore, be more relevant to their needs, many metrics provide complementary information about precipitation so they can serve multiple purposes and communities of users.

a. Process-oriented metrics

To inform model development, a subset of baseline metrics applied to specific sites where many types of measurement data are available can be useful. These differ from the baseline metrics discussed in Section 3 mainly in the spatial and temporal scale of the analysis, as baseline metrics tend to be calculated based on large-to-global domains to provide a broader view of regional differences aggregated to larger scales. However, similar metrics can be applied at a local scale to provide a starting point for more in-depth analysis to reveal the underlying causes of model biases. The subset of baseline metrics may include the diurnal (Figure 14) and seasonal cycles of precipitation and the probability distribution function (PDF) of precipitation rates applied to DOE's Atmospheric Radiation Measurement (ARM) sites where long-term precipitation data and other measurements from in situ and remote-sensing instruments are available. Three heavily instrumented fixed-location atmospheric observatories have been supported by ARM to collect data at the Southern Great Plains (SGP) in the central U.S., North Slope of Alaska (NSA), and Eastern North Atlantic (ENA) capturing different cloud regimes relevant to precipitation. Data from the ARM atmospheric observatories are particularly useful for evaluating present-day climatological CMIP simulations that can be derived from standard coupled and fixed-sea-surface-temperature (SST) model outputs. While field campaigns also provide diverse measurements for model evaluation, the relatively shorter periods they cover require

Exploratory Metrics		
Process-oriented Metrics	Regime-oriented Metrics	Use-inspired Metrics
Diurnal cycle of precipitation	Frontal precipitation	Fraction of wet days
Character of the distribution of precipitation rates	Extratropical + tropical cyclones	Average length of consecutive wet periods
Emergent constraints on precipitation	Mesoscale convective systems	Decorrelation time of precipitation
Orographic precipitation	Monsoonal precipitation	Intensity-duration-frequency curves
Teleconnections	Weather states	Correlated extremes

Figure 13. A categorization of the exploratory metrics considered at the workshop.

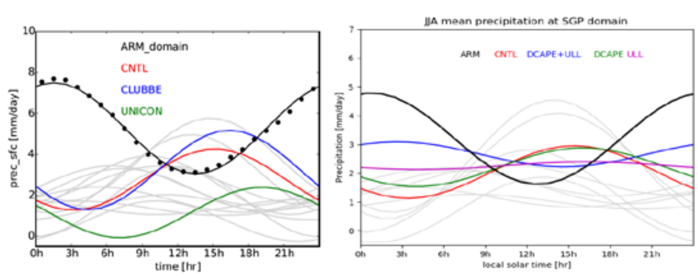

Figure 14. (Left) Mean diurnal cycle of precipitation in June-July-August (JJA) from ARM observations at the Southern Great Plains (black), CMIP5 simulations (grey), and E3SM simulations with the Zhang McFarlane (red), Cloud Layers Unified By Binormals (CLUBB; blue), and Unified Convection Scheme (UNICON; green) convective parameterizations. (Right) Similar to the left panel but including three new E3SM simulations that test two new convection trigger functions (green, purple) and their combination (blue) as described in Xie et al. (2019). (Source: Shaocheng Xie, Lawrence Livermore National Laboratory)

specific numerical experiments such as initialized simulations or simulations constrained by observed large-scale circulation.

On the PDF of precipitation rates, three regimes can be identified from the PDF curve—a non-precipitating regime, and a power-law regime at lower precipitation rates that transitions at a cutoff scale to an exponential regime at higher precipitation rates (Figure 15). Since the exponent of the power-law range and the cutoff scale depend only on the physics of the processes controlling precipitation, and stochastic process models can capture such features to guide physical interpretations (Martinez-Villalobos and Neelin 2019), they are good candidates for process-oriented metrics for precipitation. The exponent and cutoff scale can be determined from the PDF of daily precipitation rates. As the latter generally follows the gamma distribution, the exponent and cutoff are simple functions of the mean and variance of daily precipitation. With satellite or in situ daily precipitation data, the exponent and cutoff scale can be estimated globally or at specific locations and used as metrics for benchmarking an important aspect of precipitation.

Process-oriented metrics may include relationships between precipitation and other variables intimately involved in the processes that generate precipitation. For example, satellite observations have revealed a relationship between column relative humidity and precipitation on daily to monthly time scales (Bretherton et al. 2004, Sobel et al. 2004). On shorter, convective time scales, a similar relationship has also been established between column water vapor (CWV) and precipitation from satellite data (e.g., Peters and Neelin 2006) and in situ data (e.g., Holloway and Neelin 2009, Schiro et al. 2016). The relationship features a sharp increase in precipitation rate, referred to as precipitation pickup, which occurs when the column water vapor exceeds a certain threshold value. As an indication of the interactions between lower tropospheric humidity and the onset of convection, this relationship is sensitive to various aspects of how models represent shallow and deep convection, as well as microphysical processes (e.g., Kuo et al. 2017, Hagos et al. 2018a).

Models have been shown to reproduce the CWV-precipitation relationship to varying degrees, with implications for model skill in simulating a wide range of phenomena such as the Madden-Julian Oscillation (Klingaman et al. 2015, Rushley et al. 2019, Kim et al. 2019), monsoon precipitation (Hagos et al. 2018b), and tropical precipitation variance and extremes (Hagos et al. 2018a). Using satellite and in situ data of CWV and precipitation, this relationship and the PDF of CWV can be determined at locations representing different convection regimes for evaluating model parameterizations of clouds and convection (Figure 16). Information from the relationship can be condensed to derive simple metrics to determine whether precipitation in the models occurs in the right thermodynamic environment.

Besides the CWV-precipitation relationship discussed above, other relationships relating precipitation with temperature, vertical velocity, moist static energy, and

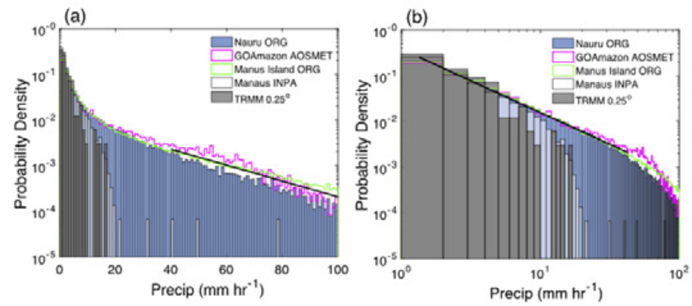

Figure 15. Probability density function of precipitation rate (mm hr^{-1}) based on data from TRMM (solid grey) and ARM (solid blue and open rectangles in magenta, green, and grey) in log-linear plot (left) and log-log plot (right). The fitted linear black line in (a) and (b) shows the exponential range and power law range, respectively. (Source: David Neelin, University of California at Los Angeles)

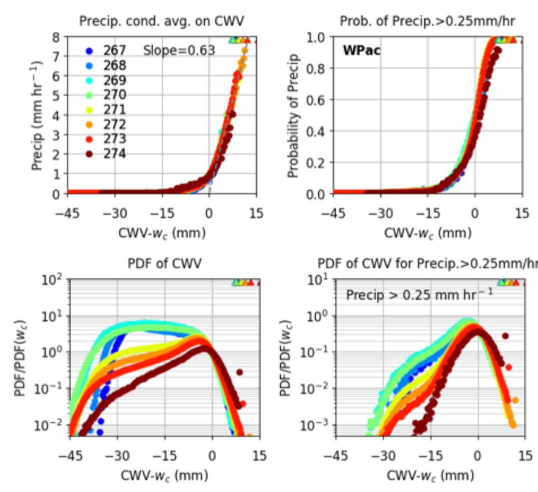

Figure 16. Convective transition collapsed statistics. Conditionally averaged precipitation rate (mm hr^{-1}) (far left) and conditional probability of precipitation (middle left) for various tropospheric-averaged temperatures (colored markers). PDF of all events (middle right) and precipitating events only (i.e., precipitation > 0.25 mm hr^{-1}) (far right) as a function of various tropospheric-averaged temperatures (colored markers). All variables are plotted against CWV-wc where wc is the threshold of CWV above which a rapid increase of precipitation occurs. All plots are based on observations in the tropical (20S 20N) western Pacific. (Source: David Neelin, University of California at Los Angeles)

entrainment and convective triggering can also be explored to derive process-oriented metrics for evaluating precipitation simulations, as well as providing insights on processes that may need improvement in their representation in the models. For example, partitioning the PDF of precipitation rates by regimes of vertical motion (updraft, neutral, downdraft) or conversely, partitioning the PDF of vertical velocity by regimes of precipitation (none, light, moderate, and heavy precipitation) provides useful information for evaluating the relationship between precipitation and vertical motion in the models.

In regions of complex terrain, the amount and phase of orographic precipitation play a dominant role in the regional water cycle, affecting soil moisture, snowpack, and runoff. In models, orographic precipitation is sensitive to model resolution as well as physics parameterizations (e.g., Leung and Qian 2003, Lebassi-Habtezion and Diffenbaugh 2013, Yang et al. 2017). Across a mountain transect, precipitation generally increases with elevation and peaks upwind of the mountain top. Model skill in simulating orographic precipitation can be measured by several metrics, such as the pattern correlation between the observed and simulated precipitation, the orographic enhancement (e.g., measured by the ratio of the maximum and minimum precipitation along topographic transects), and the distance between the upwind precipitation peak and the mountain peak along topographic transects. Orographic blocking, measured by a bulk Froude number that depends on the atmospheric stability and cross-mountain wind speed, has important effects on precipitation in mountainous regions (Leung et al. 1998). The slope of the linear regression between the climatological precipitation and the elevation gradient, reflecting the dominance of blocking versus non-blocking regimes (Hughes et al. 2009), can also be explored as a metric for evaluating orographic precipitation in model simulations.

Precipitation exhibits large variability associated with different large-scale modes of variability. Teleconnection relationships between regional precipitation and modes of variability provide important information about model biases in precipitation variability. As an example, biases in simulating El Niño Southern Oscillation (ENSO) has an important bearing on precipitation biases in North America through the Pacific North America Pattern (PNA). Connections have been established between the MJO and tropical cyclones and atmospheric rivers, so the MJO also has influence on precipitation in different seasons and regions. Such relationships may be used to develop metrics to evaluate regional precipitation, particularly extreme precipitation, simulated by models.

The ASoP package (Klingaman et al. 2017) provides a quantitative approach to evaluation of intensity distributions and coherence in space and time across a wide range of scales and can reveal how models fundamentally produce precipitation. Martin et al. (2017) showed how the package can be used to understand the contribution from model errors on different time and space scales to climatological precipitation biases. Figure 17 shows the contribution to June-September mean rainfall from precipitation intensity distributions calculated at different timescales, from the model timestep (20 minutes) up to 20 days. This illustrates that, for the equatorial Indian Ocean, this model's characteristic wet bias is related to too frequent sub-daily and daily rainfall amounts (not enough dry periods), which leads to an overestimate in 10-day and 20-day totals. Furthermore, too little shift in the distributions to the left with increased averaging period indicates poor intra-seasonal variability. In contrast, for West Africa, considerable intermittency at the timestep level and a poor diurnal cycle leads to 3-hourly rainfall amounts that are smaller than observed. This leads to an underestimate in daily, 10-day, and 20-day totals. Although the lack of shift in the distributions to the left with increased averaging period indicates little day-to-day variability, poor representation of sub-daily and daily rainfall contributes to the climatological dry bias in this model for this region.

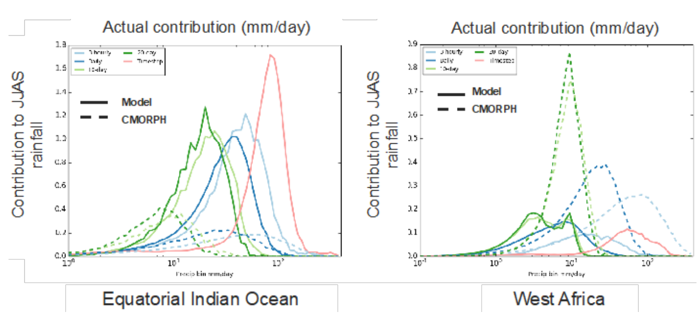

Figure 17. Contribution of June-July-August-September (JJAS) rainfall (mm/day) at different timescales to total rainfall as a function of precipitation amount (mm/day) in (a) equatorial Indian Ocean and (b) West Africa. Timescales are shown for 3-hourly (light blue), daily (dark blue), 10-day (light green), 20-day (dark green), and timestep (red). (Source: Gill Martin, UK Met Office)

b. Regime-oriented metrics

Both moisture and vertical motion are needed to produce precipitation through condensation of water vapor and cloud formation. Dynamical forcing, convective instability, and orographic forcing are three mechanisms for generating upward motion. Hence precipitation, especially extreme precipitation that requires abundant moisture and strong upward motion, exists in distinct meteorological environments that define the precipitation regime. Kunkel et al. (2012) identified several meteorological causes or precipitation regimes that are particularly relevant to extreme precipitation in the U.S. These include fronts, extratropical cyclones (ETCs), monsoon, tropical cyclones (TCs), mesoscale convective systems (MCSs), isolated thunderstorms in convectively unstable air masses, and upslope flow or orographic precipitation. Note that these regimes are not all independent as fronts are usually associated with ETCs, and during spring, MCSs are often embedded in frontal systems. All of the aforementioned precipitation regimes are also relevant to precipitation in different regions around the world.

A set of exploratory metrics can be defined to evaluate how well models simulate precipitation associated with specific precipitation regimes. These regime-oriented metrics provide useful information for model developers and climate scientists to understand model biases, such as those related to model resolution or specific geographical regions. They are also useful for impacts researchers and stakeholders to evaluate and communicate the credibility of model precipitation in terms of meteorological phenomena that are more easily understood.

Automatic methods have been developed to detect and track precipitation or atmospheric circulation features associated with different precipitation regimes. For example, fronts can be identified using the method of Berry et al. (2011) and the thermal front parameter of Hewiston (1998) to identify locations of frontal points based on the maximum gradient of wet-bulb potential temperature in the direction of the moist isentropes (Catto et al. 2012). ETCs (Figure 18) can be tracked using the method of Hodges (1999) based on sea level pressure (SLP) minima and stitching together the SLP minima across time to produce cyclone tracks. TCs can be tracked by identifying the minima in SLP, the existence of a warm core defined by upper-to-mid-level tropospheric temperature, and surface wind speed exceeding a threshold value (e.g., Ullrich and Zarcycki 2017). Similarly, different MCS tracking methods have also been developed to identify contiguous areas of precipitation with certain features such as skewed rain rates reflecting the intense rainfall associated with the convective core of MCSs, and/or high clouds based on outgoing longwave radiation, reflecting the deep convection and the large stratiform area (Feng et al. 2016, 2018; Figure 19).

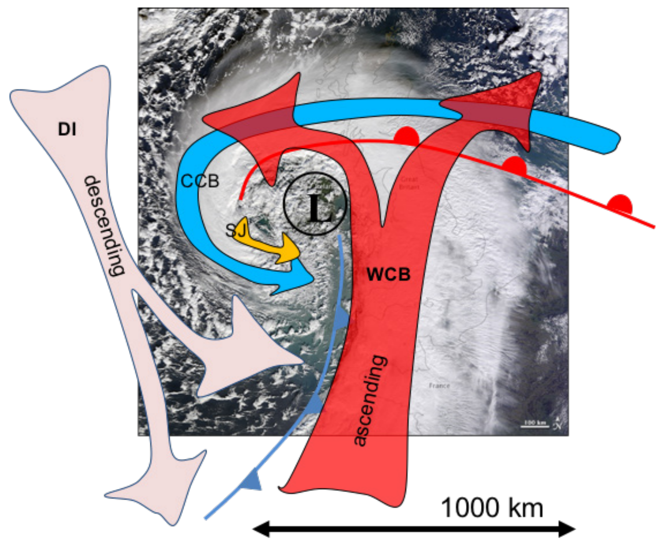

Figure 18. Features of an extratropical cyclone showing the fronts, dry intrusion (DI), warm conveyor belt (WCB), cold conveyor belt (CCB), and sting jets (SJ). (Source: Jennifer Cattoo, University of Exeter)

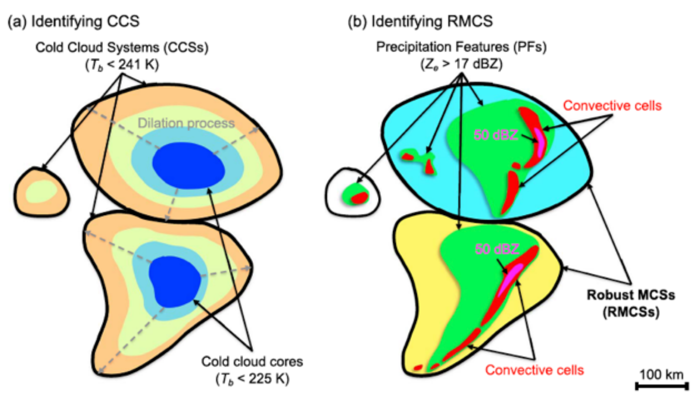

Figure 19. Schematic identification of a robust mesoscale convective system (RMCS). (a) Cold cloud systems (CCS) are identified using satellite Tb data. The thick black contours show the 241 K outline of CCS. The cold cloud cores (Tb < 225 K) shown in blue patches are dilated outward to 241 K to separate CCS with distinct cold cloud cores. (b) Precipitation features (PFs) within the CCS are identified with contiguous area of radar reflectivity >17 dBZ at 2.5 km height. PF major axis length and convective cells with imbedded 50-dBZ echoes anywhere within the PF are used to identify RMCS. The two-colored patches in (b) with thick black outlines denote two RMCSs. (Adapted from Feng et al. 2018)

Atmospheric rivers (ARs) are intense horizontal moisture transport pathways associated with heavy precipitation in many regions worldwide. Many algorithms have been developed to detect and track ARs based on CWV and/or column integrated moisture transport above absolute or percentile thresholds, sometimes with added criteria regarding the geometry (length, width, orientation) of the CWV filaments (Shields et al. 2018; Figure 20).

Large biases in South Asian monsoon precipitation have been noted in two recent generations of CMIP models (Sperber et al. 2015). The precipitation biases can be linked to biases in the thermodynamic states (e.g., moist static energy) of the Asian monsoon (Boos and Hurley 2013). Synoptic-scale vortices related to dynamical instability of the strongly sheared South Asian monsoon basic state (Diaz and Boos 2019) produce much of the extreme precipitation in India (Boos et al. 2015). These vortices contribute around 60% of the seasonal mean precipitation in eastern and central India, but in models this ratio varies between 5% and 60% (Praveen et al. 2015). Hence, quantifying model biases in the low-pressure systems, called Bay of Bengal monsoon depressions (MDs), which generate synoptic-scale vortices and propagate from the Bay of Bengal to produce heavy precipitation in northeastern India (Figure 21), is important for understanding the sources of these model biases. The MDs can be identified and tracked based on the maximum 850 hPa relative vorticity (e.g., Ashfaq et al. 2016, Cohen and Boos 2014, Hurley and Boos 2015, Levine and Martin 2017).

Recently, machine-learning approaches have also been used to detect and track precipitation or circulation features associated with different precipitation regimes. For example, Biard and Kunkel (2019) tested the use of deep-learning neural networks to automate the detection of weather fronts

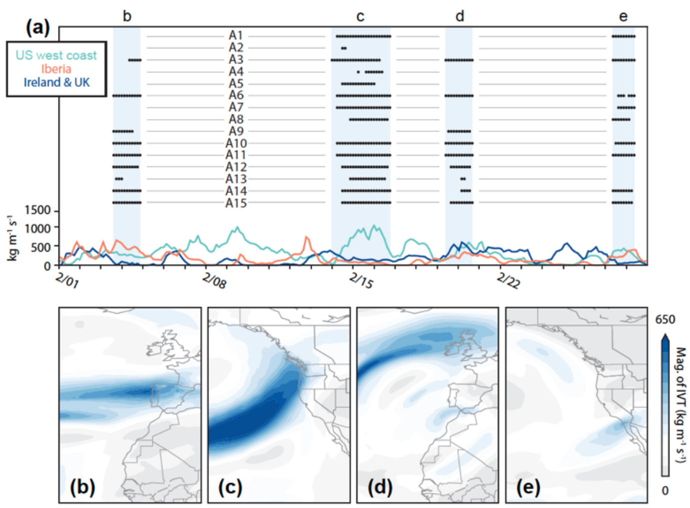

Figure 20. (a) Time series of daily integrated vapor transport (IVT) anomalies for (orange) Iberia, (teal) the U.S. west coast, and (blue) Ireland and the United Kingdom. Four events of varying geometry and intensity are shaded in panel (a), and composites for each event are shown in panels (b)–(e). The black dots above the time series in panel (a) indicate time slices in which each event is detected by an algorithm. (Adapted from Shields et al. 2018)

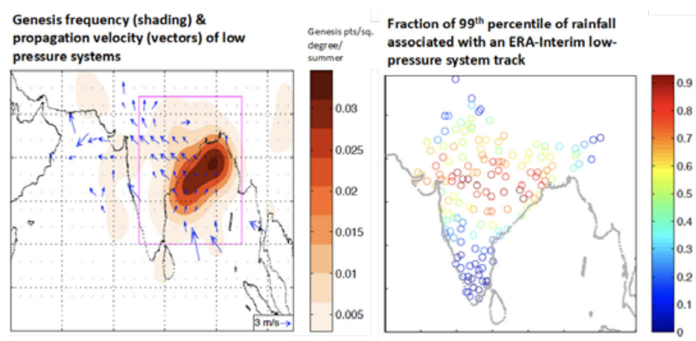

Figure 21. (Left) Shading/colors show the number of genesis points of MDs per square degree (roughly 12 000 km^2) per summer season (June–September), after smoothing with a Gaussian filter. Vectors show the average propagation speed of MDs and are coarsened to a 2◦ × 2◦ grid for clarity; vectors are shown only if the mean zonal or meridional propagation speed at each grid point is statistically significant at the 1% level by a two-tailed t-test. (Adapted from Boos et al. 2015) (Right) Fraction of 99th percentile of rainfall associated with a European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA)-Interim low-pressure system track. (Source: Bill Boos, University of California at Berkeley)

(Figure 22). The U.S. seasonal front climatology developed based on deep learning is comparable to the climatology derived from coded surface bulletins. Neural networks have also been used to segment and track ARs and TCs (Mudigonda et al. 2017).

Once the fronts, ETCs, TCs, MCSs, ARs, or MDs are identified using tracking methods, they can be combined with precipitation data to quantify the mean and extreme precipitation as well as other precipitation characteristics (e.g., diurnal and seasonal variability) associated with each regime regionally or globally. For example, model precipitation biases can be decomposed into biases associated with frontal and non-frontal precipitation, and for frontal precipitation, the biases can further be decomposed into biases associated with frequency versus intensity (Catto et al. 2015). Precipitation biases associated with MCSs can also be further investigated by combining the MCS tracks with large-scale circulation data to identify model biases in simulating the large-scale environment favorable for MCS development (Song et al. 2019, Feng et al. 2019).

Tselioudis et al. (2013) defined global weather states (WSs) using joint histograms of cloud optical thickness and cloud top pressure from satellite data. The 11 WSs they identified exhibit unique distributions of vertical layering of clouds that correspond well to the horizontal structure of cloud properties. Furthermore, the WSs represent a normal progression in dynamic regime from the most convective to the least convective WS. This suggests that the WSs may also be useful for delineating different precipitation regimes. Analysis depicted in Figure 23 shows clear separation of precipitation regimes (no, light, moderate, heavy precipitation) across the WSs, highlighting, for example, the dominance of "no precipitation" under fair weather (WS7) in contrast with deep convection (WS1) that produces significantly heavier precipitation. Representing a cloud-centric approach to delineate precipitation regime, the WS method may offer complementary insights compared to precipitation regimes defined by meteorological causes.

Regime information derived from the various analyses discussed above can be summarized in many forms to be used as precipitation metrics. Examples include monthly distribution of precipitation broken down by regimes, spatial distribution of the top 100 events and their corresponding regimes, and spatial distribution of amplification factor associated with different regimes, where amplification factor may be defined as the ratio of precipitation intensity for a particular regime to the mean intensity of all precipitation events. These regime-based metrics build on baseline metrics discussed in Section 3, providing more information about the distribution of precipitation and also extremes.

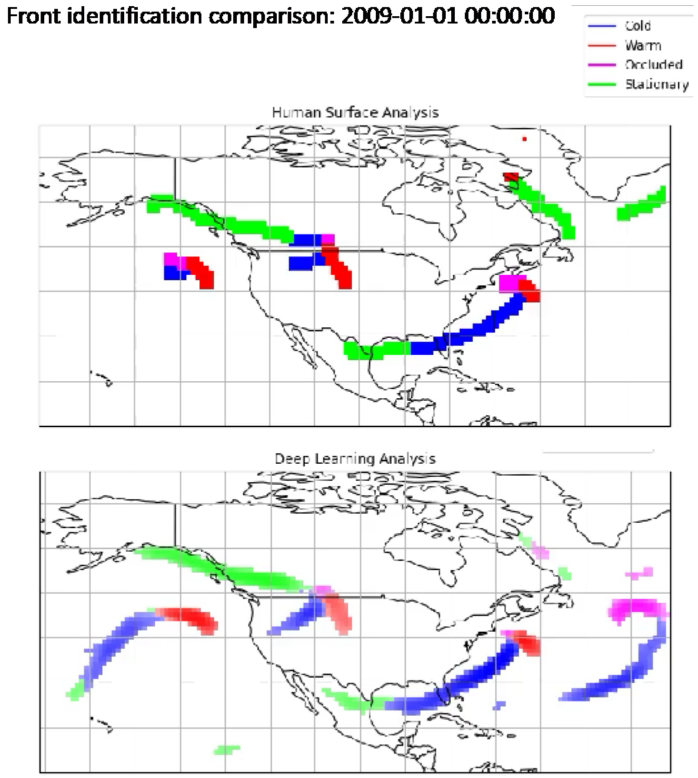

Figure 22. Fronts identified by human surface analysis (left) and deep-learning analysis (right) on 2009 01-01 00Z. Different types of fronts are shown in blue (cold), red (warm), occluded (magenta), and stationary (green). (Source: Ken Kunkel, North Carolina State University)

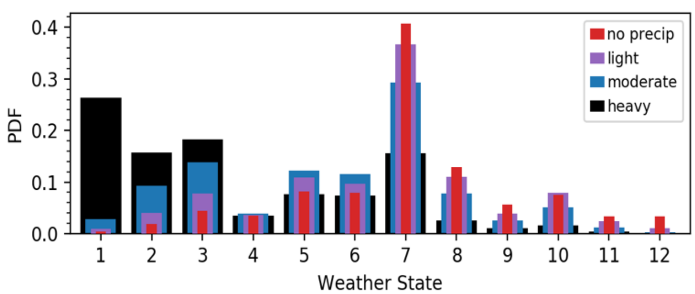

Figure 23. Precipitation distribution in cloud-defined weather states (1–12) as described in Tselioudis et al. (2013). Precipitation regimes are defined as no precipitation (red), and light (purple), moderate (blue), and heavy (black) precipitation. (Source: George Tselioudis, NASA Goddard Institute of Space Studies [GISS])

c. Use-inspired metrics

Precipitation varies across a wide range of spatial and temporal scales, with important impacts on ecosystems, crop productivity, water resources, and other human activities, so evaluating the ability of models to simulate different aspects of precipitation variability is important for stakeholders managing a host of different types of resources. Precipitation variability can be characterized in many ways from the perspectives of both space and time, including coherence and diurnal variability already discussed in Section 3. Going beyond baseline metrics, variability can also be quantified by precipitation sequencing that can be measured by the fraction of wet days and the average length of consecutive wet periods, where wet days can be defined as the lowest precipitation amount recorded by rain gauges, or using different thresholds based on percentile values of interest. The spatial pattern of such statistics can reveal important information about regional climate characteristics and how well they are captured by the models.

Complementary to precipitation sequencing, the contribution of precipitation at different time scales (e.g., sub-daily, daily, pentad) to total precipitation also provides useful information about precipitation variability. As shown in Figure 17, contrasts in the contribution from different time scales can be large across regions such as the Indian Ocean and West Africa. Large contrasts are also found within Europe, and the ability to reproduce the observations varies with models and is sensitive to horizontal resolution (Berthou et al. 2018). Another aspect of precipitation variability can be measured by the space-time coherence of precipitation using metrics such as the de-correlation time of precipitation at different thresholds and timescales and the spatial autocorrelations of precipitation at different distances and lags. These metrics are useful for researchers using model precipitation data for hydrologic and crop modeling, as model biases revealed by the metrics may have important effects on their simulations. They are also useful for model developers to understand more detailed aspects of precipitation biases to set targets for improvements.

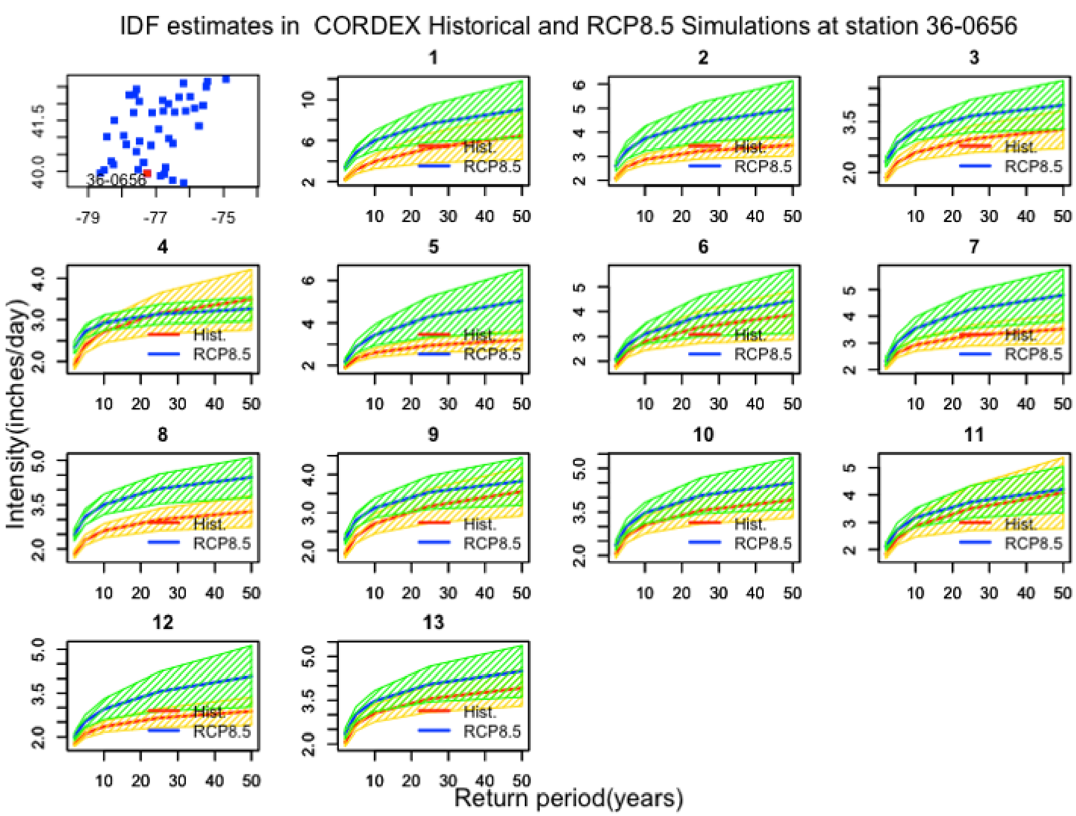

Figure 24. IDF estimates based on observations (upper left) and model simulations for the present climate (red line with yellow shading) and future climate (blue line with green shading) following the RCP8.5 scenario. Simulations are produced by 13 regional climate models that participated in the Coordinated Regional Climate Downscaling Experiment (CORDEX). The IDF curves show the relationships between precipitation intensity (inches/day) and return period (years) for duration of one day at station 36-0656. (Source: Paul Ullrich, University of California at Davis)

Hydraulic structures such as flood drainage systems are designed to accommodate extreme flood events. For this purpose, civil engineers characterize storms by curves that relate the precipitation intensity with its duration and frequency of occurrence, referred to as the intensity-duration-frequency (IDF) curves. As climate change is expected to alter precipitation characteristics, developing IDF curves that consider climate change is an important challenge, requiring assessments of how well model simulated precipitation can reproduce the IDF curves based on observed precipitation. IDF curves can be generated using univariate extreme value analysis to derive the best-fit distribution for exceedance probabilities corresponding to the extreme events with return periods such as 5, 10, 25, 50, and 100 years. IDF curves based on observation and model simulations show a large discrepancy (Figure 24), underscoring the need to improve simulations of precipitation to support infrastructure designs.

Correlated extremes are important concerns for agricultural production and water management. Three types of correlated extremes regarding precipitation should be explored as precipitation metrics. The first type highlights compound extremes such as torrential rain during coastal storm surge that may amplify coastal inundation compared to inundation caused by torrential rain or storm surge alone. Using tropical cyclones as examples, the aforementioned compound extreme requires models to properly simulate both rain and wind simultaneously, in addition to tropical cyclone track and translation speed. The second type of correlated extremes is concurrent extremes, such as simultaneous drought and flood that cause multi-breadbasket failure. Model ability to simulate such concurrent extremes requires realistic representation of system connections and coherence/teleconnections of extreme events. The third type of correlated extreme is sequential extremes, such as floods followed by a dry spell. This type requires models to be capable of capturing the shifting vulnerability owing to conditions prior to event onset.

Research Needs

The examples of process-oriented, regime-oriented, and use-inspired metrics discussed at the workshop and summarized above have been explored and demonstrated in the analysis of observations and model outputs and to diagnose model biases. However, more research is needed to define the metrics before they can be implemented and standardized for broader use. Some research needs identified at the workshop are discussed below.

Synthesizing analysis into succinct metrics. Many examples discussed in Section 4.2 are analyses that can be used to produce visual and numerical comparisons of different aspects of observations and model simulations; however, there is a need to develop metrics that succinctly synthesize key aspects of the analyses. For the power-law and exponential regimes of the PDF of precipitation rates, the exponent of the power law and the cutoff scale of the transition from the power-law regime to the exponential regime are examples of metrics that succinctly encapsulate the physical properties of precipitation regimes. However, other metrics such as those for orographic precipitation are less specific and may be sensitive to the spatial scales of the analysis. For regime-oriented analysis based on meteorological states, various information can be produced regarding the precipitation characteristics and the large-scale environment of the different precipitation regimes (e.g., frontal, ETCs, etc.). For each regime, what information is most important in quantifying model biases and revealing their sources requires more research. For example, what aspects of each precipitation regime most distinctly distinguish it from other regimes? For MCSs, their nocturnal timing and precipitation intensity and area are what distinguish them from non-MCS precipitation, but such distinguishing features may not be as obvious for other regimes such as fronts and ETCs. There are similar research needs to summarize the use-inspired metrics based on information produced by analysis of precipitation variability and IDF curves.

Relating precipitation characteristics with storm characteristics and large-scale environments. In parallel to developing succinct metrics, more research delving into the relationship between precipitation characteristics and storm characteristics could provide important insights on model biases. As an example, for frontal precipitation, relating the precipitation intensity to the strength of the fronts, the presence or absence of frontal features such as warm conveyor belt, dry intrusion, and cyclone depth may reveal how different storm characteristics alter the likelihood of a front producing an extreme precipitation event in observations and model simulations. For MCSs, relating the PDF of rain rates with the properties of the convective core and stratiform regions or storm propagation or lifetime may provide useful information regarding the microphysical processes simulated by the models (Feng et al. 2018). As the large-scale environment provides the meteorological context for the storms and precipitation, understanding and quantifying biases in the large-scale environments can provide important information on precipitation biases under different regimes.

Improving physical interpretation of the metrics. Exploratory metrics are generally more complex than the baseline metrics, as they delve into the processes and meteorological causes of precipitation. Understanding the physical processes behind the metrics and interpreting the benchmarking results are important if the metrics are to inform model developers and other researchers. To this end, more research is needed to understand the physical properties or processes governing the metrics. As an example, stochastic process models based on water vapor and energy equations can be used to understand and interpret the leading aspects of the precipitation PDF. More research on the CWV-precipitation relationship can improve understanding of the processes that control the relationship and the threshold of the CWV for precipitation pickup. As analyses are summarized succinctly using metrics, exploring and documenting the threads connecting the analyses to the metrics is also important to improve the physical interpretation of the metrics.

Developing emergent constraints. Some of the exploratory metrics discussed at the workshop could potentially be used as emergent relationships to constrain future projections of precipitation changes. For example, Hagos et al. (2018b) showed that CMIP5 model spread in projecting South Asian monsoon precipitation changes in the future is related to where each model is situated in the CWV-precipitation curve derived for the equatorial Indian Ocean, which is also reflected in model biases in simulating the present-day monsoon precipitation. Other relationships such as that between precipitation characteristics and storm characteristics for frontal systems may potentially reveal a relationship between model biases in precipitation characteristics and the projected future changes. Given that ESMs are prominently used to understand and project future changes in precipitation, developing emergent relationships that extend the quantification of model biases based on precipitation metrics to constrain future projections is an important research area. The development of emergent relationships can take advantage of the process-oriented and regime-oriented metrics that emphasize process understanding, which aligns with the requirement of emergent constraints that must be physically explainable (Klein and Hall 2015). The numerous examples of exploratory metrics discussed in Section 4.2 offer opportunities for evaluating their use to provide emergent constraints.

Characterizing uncertainty of tracking methods for precipitation regimes. Many detection/tracking methods have been developed and used by researchers to study different precipitation regimes, how well they are simulated or predicted by models, and their variability and change. These methods may differ in many ways, including the variables being tracked, the thresholds used to define the features, methods based on the physics/dynamics of the phenomena versus methods based mainly on geometry or other visual features, etc. There is a need to understand and quantify uncertainty in tracking precipitation regimes and the resulting uncertainty in the metrics derived from the feature tracking. Coordinated efforts such as IMILAST (Intercomparison of Midlatitude Storm Diagnostics; Neu et al. 2013) and ARTMIP (Shields et al. 2018) have investigated algorithm diversity of ETCs and ARs, respectively. However, more research is needed to relate uncertainty in tracking fronts, ETCs, TCs, ARs, etc., to uncertainty in characterizing the precipitation produced by these regimes.

Characterizing uncertainty in observation data. Besides uncertainty in the tracking/detection methods, uncertainty in observation data can also contribute importantly to uncertainty in the metrics used to benchmark precipitation in model simulations. For exploratory metrics, uncertainty in observation data goes beyond precipitation data because many exploratory metrics require information from other variables related to precipitation. As an example, reanalysis data is often used in the analysis of the thermodynamic and dynamical environments of the atmosphere combined with precipitation data from in situ or remote-sensing measurements, as precipitation data from reanalysis is less reliable. However, this may create inconsistency between precipitation and the variables (e.g., moisture) that are being related to precipitation. We need to understand different sources of uncertainty that may be introduced by the observation data, combination of observation data, and their use at different time/space scales.

Engaging with the Broader Research Community

National and International Collaborations

Representatives from several related international activities engaged in this workshop. Both the GEWEX Data Assessment Panel and the International Precipitation Working Group (IPWG) provided crucial expertise with respect to remote and in situ-based measurements. CMIP6 publications from both working groups will serve as a foundation from which a WCRP activity can be formalized. Briefings on the precipitation metrics activity will be included in future sessions of the Working Group on Numerical Experimentation (WGNE) and GASS (GEWEX Global Atmosphere and System Studies). This will likely lead to the formation of a WCRP precipitation metrics panel to establish a longer-term strategy to engage with the broader modeling community. As outreach opportunities arise, the broader community will have an opportunity to engage, for example, through a session proposed for the 2020 meeting of the European Geophysical Union (EGU). Both DOE and NOAA have expressed an interest to sponsor a follow-on workshop to this precipitation metrics workshop as part of the U.S. GEWEX activities under the U.S. Global Change Research Program (USGCRP) to further explore opportunities to quantify precipitation biases and improve modeling and prediction of precipitation across timescales from weather to multidecadal.

ESM Model Evaluation Capabilities

As a first step, the baseline metrics highlighted in Section 3 are being implemented into the PCMDI Metrics Package (PMP; Gleckler et al. 2016), an open-source software package designed for producing objective comparisons between ESMs and observations. A feature of the PMP that is needed for long-term benchmarking is a provenance framework to document versions of all data (simulations and observations), analysis codes, dependencies, and operating conditions. Coordination with other precipitation-related capabilities will help strengthen the precipitation metrics. One example is the DOE-supported ARM Data-Oriented Metrics and Diagnostics Package (Zhang and Xie 2017), which already includes some of the process-oriented metrics highlighted in Section 4. Another is the International Land Model Benchmarking (ILAMB) package for land models that targets uncertainties associated with key biogeochemical processes and feedbacks (ILAMB; Collier et al. 2018). Realistic simulation of precipitation is critical in biogeochemical processes, and a land-based benchmarking component will provide a value bridge with the atmospheric focus. Within NOAA, the Model Diagnostics Task Force (MDTF; Maloney et al. 2019) is developing a model evaluation package that includes process-oriented diagnostics based on precipitation that can provide valuable insight and contributions in the exploratory metrics. The ASoP package (Klingaman et al. 2015) provides a quantitative approach to evaluation coherence in space and time on shorter scales and can reveal how models fundamentally produce precipitation. Progress in systematic model evaluation and benchmarking of precipitation will benefit from synergies across all these capabilities.

One particular effort presently underway that has the potential to standardize and accelerate model evaluation across disparate metric development efforts is the Coordinated Model Evaluation Capabilities project (CMEC; https://cmec.llnl.gov/). This project aims to develop high-level seamless integration of distinct, yet complementary, diagnostics and metrics capabilities. A central expectation of CMEC is that through coordination, these efforts can be made stronger than the sum of their parts. Specifically, CMEC targets accelerated analysis by coordinating multiple capabilities (including PMP, ILAMB, and modules developed by university and laboratory partners) within a single lightweight operational execution and visualization framework. It further aims to enable execution of distinct evaluation capabilities regardless of the structural differences between climate data sets, facilitate version control of observational data sets, and allow users to navigate among evaluation products in a single framework. As such, CMEC provides a pathway towards rapid intercomparison of baseline and exploratory metrics, including those related to precipitation.

Next Steps

Benchmarking

To establish a baseline of skill for earth system model precipitation, the first step is to apply the initial set of metrics agreed upon at the workshop (Section 3) to the current generation of climate model simulations—CMIP6. These metrics will be implemented into the PMP, and they will be applied to the CMIP6 DECK and Historical simulations. A publication will document the skill, according to these metrics, of precipitation in CMIP6 simulations, and compare it against previous generations of ESMs (as archived in CMIP)—CMIP5, CMIP3, and earlier generations as data archival allows. In approximately 5–7 years, this evaluation will be revisited to evaluate progress made in the intervening period. The goal is for this challenge to both motivate progress on improving model precipitation, and also facilitate it by providing appropriate and holistic observational targets. In the interim, we envisage that the baseline metrics will gradually be augmented as informed by the exploratory efforts outlined below.

Exploratory

While there are longer-term research needs to develop and improve the exploratory metrics for broader use, there are strong foundations for some metrics to be applied to climate simulations for demonstration. Parallel to the working group on baseline metrics, a working group on exploratory metrics was established at the workshop with the goal of developing coordinated near-term activities to advance the development and use of exploratory metrics. The first activity identified by the group is a collaborative effort on a manuscript to discuss the need for exploratory metrics, introduce an initial set of exploratory metrics, and apply them to CMIP6 model outputs to demonstrate their usefulness for different communities of users (model developers, climate scientists, and impacts researchers and stakeholders).

The working group identified the following topics to include from the process-oriented, regime oriented, and use-inspired metrics for demonstration of an initial set of exploratory metrics:

- Coherence in space and time
- Frontal precipitation
- Top 10 extreme events
- Convection onset
- Orographic enhancement
- Monsoon
- Mesoscale convective systems
- Madden-Julian Oscillation
- Atmospheric rivers.

As part of CMIP6, HighResMIP (Haarsma et al. 2016) includes simulations at low and high resolution to facilitate analysis of the impacts of model horizontal resolution on climate simulations. Exploratory metrics on the above-selected topics will be applied to a set of low- and high-resolution simulations available from HighResMIP, with the goal to demonstrate their use across multi-models and a range of spatial resolution and insights that can be gained regarding model performance in simulating diverse aspects of precipitation. The near-term activities will be coordinated within the group and with relevant activities supported by other agencies and programs. The working group is open for broader participation by researchers in the community.

References

AchutaRao, KM, C Covey, C Doutriaux, M Fiorino, P Gleckler, T Phillips, K Sperber, and K Taylor. 2004. “An Appraisal of Coupled Climate Model Simulations,” Edited by D. Bader, 2004. UCRL TR 202550.

Alexander, LV, M Bador, R Roca, S Contractor, M Donat, and PL Nguyen. 2020. “Intercomparison of annual precipitation indices over global land areas from in situ, space-based and reanalysis products.” *Environmental Research Letters,* https://doi.org/10.1088/1748-9326/ab79e2

Ashfaq, M, D Rastogi, R Mei, D Touma, and LR Leung. 2016. “Sources of errors in the simulation of south Asian summer monsoon in the CMIP5 GCMs.” *Climate Dynamics* 49(1-2): 193–223, https://doi.org/10.1007/s00382-016-3337-7

Bador, M, LV Alexander, S Contractor, and R Roca. 2020. “Diverse estimates of annual maxima daily precipitation in 22 state-of-the-art quasi-global land observation datasets.” *Environmental Research Letters* 15(3): 035005, https://doi.org/10.1088/1748-9326/ab6a22

Berry, G, C Jakob, and M Reeder 2011. “Recent global trends in atmospheric fronts.” *Geophysical Research Letters* 38(21): L21812, https://doi.org/10.1029/2011GL049481

Berthou, S, EJ Kendon, SC Chan, N Ban, D Leutwyler, C Schär, and G Fosser, 2018. “Pan-European climate at convection-permitting scale: a model intercomparison study.” *Climate Dynamics* https://doi.org/10.1007/s00382-018-4114-6

Biard, JC, and KE Kunkel. 2019. “Automated detection of weather fronts using a deep learning neural network. Advances in Statistical Climatology.” *Meteorology and Oceanography,* in review.

Boos, WR, and JV Hurley. 2013. “Thermodynamic bias in the multimodel mean boreal summer monsoon.” *Journal of Climate* 26(7): 2279–2287, https://doi.org/10.1175/Jcli-D-12-00493.1

Boos, WR, JV Hurley, and VS Murthy. 2015. “Adiabatic westward drift of Indian monsoon depressions.” *Quarterly Journal of the Royal Meteorological Society* 141(689):1035–1048, https://doi.org/10.1002/qj2454

Bretherton, CS, ME Peters, and LE Back. 2004. “Relationships between water vapor path and precipitation over the tropical oceans.” *Journal of Climate* 17: 1517–1528, https://doi.org/10.1175/1520 0442

Catto, JL, C Jakob, G Berry, and N Nicholls. 2012. “Relating global precipitation to atmospheric fronts.” *Geophysical Research Letters* 39(10): L10805, https://doi.org/10.1029/2012GL051736

Catto, JL, C Jakob, and N Nicholls. 2015. “Can the CMIP5 models represent winter frontal precipitation?” *Geophysical Research Letters* 42(20): 8596–8604, https://doi.org/10.1002/2015GL066015

Cohen, NY, and WR Boos. 2014. “Has the number of Indian summer monsoon depressions decreased over the last 30 years?” *Geophysical Research Letters* 41(22): 7846–7853, https://doi.org/10.1002/2014gl061895

Covey, C, PJ Gleckler, C Doutriaux, DN Williams, A Dai, J Fasullo, K Trenberth, and A Berg. 2016. ”Metrics for the diurnal cycle of precipitation: Toward routine benchmarks for climate models.” *Journal of Climate* 29(12): 4461–4471, https://doi.org/10.1175/JCLI-D-15-0664.1

Covey, C, C Doutriaux, PJ Gleckler, KE Taylor, KE Trenberth, and Y Zhang. 2018. “High-frequency intermittency in observed and model-simulated precipitation.” *Geophysical Research Letters* 45(22): 12514–12522, https://doi.org/10.1029/2018GL078926

Collier, N, FM Hoffman, DM Lawrence, G Keppel-Aleks, CD Koven, WJ Riley, M Mu, and JT Randerson. 2018. “The International Land Model Benchmarking System (ILAMB): Design and Theory.” *Journal of Advances in Modeling Earth System* 10(11) 2731–2754, https://doi.org/10.1029/2018MS001354

Dai, A. 2006. “Precipitation characteristics coupled climate models.” *Journal of Climate* 19(18): 4605–4630, https://doi.org/10.1175/JCLI3884.1

Diaz, M, and WR Boos. 2019. “Barotropic growth of monsoon depressions.” *Quarterly Journal of the Royal Meteorological Society* 145(719): 824–844, https://doi.oorg/10.1002/qj.3467

Feng, Z, LR Leung, S Hagos, RA Houze, Jr., CD Burleyson, and K Balaguru. 2016. “More frequent intense and long-lived storms dominate the trend in central US rainfall.” *Nature Communications* 7: 13429, https://doi.org/10.1038/ncomms13429

Feng, Z, LR Leung, RA Houze, Jr., S Hagos, J Hardin, Q Yang. B Han, and J Fan. 2018. “Structure and Evolution of Mesoscale Convective Systems: Sensitivity to Cloud Microphysics in Convection Permitting Simulations over the United States.” *Journal of Advances in Modeling Earth Systems* 10(7): 1470–1494, https://doi.org/10.1029/2018MS001305

Feng, Z, RA Houze, Jr., LR Leung, F Song, J Hardin, J Wang, W Gustafson, Jr., and C Homeyer. 2019. “Spatiotemporal Characteristics and Large-scale Environment of Mesoscale Convective Systems East of the Rocky Mountains.” *Journal of Climate* 32(21): 7303–7328, https://doi.org/10.1175/JCLI-D-19-0137.1

Gleckler, PJ, C Doutriaux, PJ Durack, KE Taylor, Y Zhang, DN Williams, E Mason, and J Servonnat. 2016. "A more powerful reality test for climate models." *Eos* 97, https://doi.org/10.1029/2016EO051663

Golaz, J-C, PM Caldwell, LP Van Roekel, MR Petersen, Q Tang, JD Wolfe, G Abeshu, V Anantharaj, XS Asay-Davis, DC Bader, SA Baldwin, G Bisht, PA Bogenschutz, M Branstetter, MA Brunke, SR Brus, SM Burrows, PJ Cameron-Smith, AS Donahue, M Deakin, RC Easter, KJ Evans, Y Feng, M Flanner, JG Fucar, JG Fyke, BM Griffin, C Hannay, BE Harrop, MJ Hoffman, EC Hunke, RL Jacob, DW Jacobsen, N Jeffery, PW Jones, ND Keen, SA Klein, VE Larson, LR Leung, H-Y Li, W Lin, WH Lipscomb, P-L Ma, S Mahajan, ME Maltrud, A Mametjanov, JL McClean, RB McCoy, RB Neale, SF Price, Y Qian, PJ Rasch, JE Jack Reeves Eyre, WJ Riley, TD Ringler, AF Roberts, EL Roesler, AG Salinger, Z Shaheen, X Shi, B Singh, J Tang, MA Taylor PE Thornton, AK Turner, M Veneziani, H Wan, H Wang, S Wang, DN Williams, PJ Wolfram, PH Worley, S Xie, Y Yang, J-H Yoon, MD Zelinka CS Zender, X Zeng, C Zhang, K Zhang, Y Zhang, X Zheng, T Zhou, and Q Zhu. 2019. "The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution." *Journal of Advances in Modeling Earth Systems* 11(7): 2089–2129, https://doi.org/10.1029/2018MS001603

Haarsma, RJ, M Roberts, PL Vidale, CA Senior, A Bellucci, Q. Bao, P Chang, S Corti, NS Fučkar, V Guemas, J von Hardenberg, W Hazeleger, C Kodama, T Koenigk, LR Leung, J Lu, J-J Luo, J Mao, M Mizielinski, R Mizuta, P Nobre, M Satoh, E Scoccimarro, T Semmler, J Small, and J-S von Storch. 2016. "High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6." *Geoscientific Model Development* 9(11): 4185–4208, https://doi.org/10.5194/gmd-9-4185-2016

Hagos, S, LR Leung, C Zhao, Z Feng, and K Sakaguchi. 2018a. "How do Microphysical Processes Influence Large-scale Precipitation Variability and Extremes?" *Geophysical Research Letters* 45(3): 1661–1667, https://doi.org/10.1002/2017GL076375

Hagos, SM, LR Leung, M Ashfaq, and K Balaguru. 2018b. "South Asian monsoon precipitation in CMIP5: a link between inter-model spread and the representations of tropical convection." *Climate Dynamics* 52(1-2): 1049–1061, https://doi.org/10.1007/s00382-018-4177-4

Hewson, TD. 1998. "Objective fronts." *Meteorological Applications* 5(1): 37–65, https://doi.org/10.1017/S1350482798000553

Holloway, CE, and JD Neelin. 2009. "Moisture vertical structure, column water vapor, and tropical deep convection." *Journal of the Atmospheric Sciences* 66(6): 1665–1683, https://doi.org/10.1175/2008JAS2806.1

Hughes, M, A Hall, and RG Fovell. 2009. "Blocking in Areas of Complex Topography, and its Influence on Rainfall Distribution." *Journal of the Atmospheric Sciences* 66(2): 508–518, https://doi.org/10.1175/2008JAS2689.1

Hurley, JV, and WR Boos. 2015. "A global climatology of monsoon low-pressure systems." *Quarterly Journal of the Royal Meteorological Society* 141(689): 1049–1064, https://doi.org/10.1002/qj.2447

Kay, JE, T L'Ecuyer, A Pendergrass, H Chepfer, R Guzman, and V Yetella. 2018. "Scale-aware and definition-aware evaluation of modeled near-surface precipitation frequency using CloudSat observations." *Journal of Geophysical Research – Atmospheres* 123(8): 4294–4309, https://doi.org/10.1002/2017JD028213

Kim, H, MA Janiga, and K Pegion. 2019. "MJO propagation processes and mean biases in the SubX and S2S reforecasts." *Journal of Geophysical Research – Atmospheres* 124(16): 9314–9331, https://doi.org/10.1029/2019JD031139

Klein, SA, and A Hall. 2017. "Emergent Constraints for Cloud Feedbacks." *Current Climate Change Reports* 1(4): 276–287, https://doi.org/10.1007/s40641-015-0027-1

Klingaman, NP, SJ Woolnough, X Jiang, D Waliser, PK Xavier, J Petch, M Caian, C Hannay, D Kim, H Y Ma, WJ Merryfield, T Miyakawa, M Pritchard, JA Ridout, R Roehrig, E Shindo F Vitart, H Wang, NR Cavanaugh, BE Mapes, A Shelly, and GJ Zhang. 2015. "Vertical structure and physical processes of the Madden-Julian oscillation: Linking hindcast fidelity to simulated diabatic heating and moistening." *Journal of Geophysical Research – Atmospheres* 120(10): 4690–4717, https://doi.org/10.1002/2014JD022374

Klingaman, NP, GM Martin, and A Moise. 2017. "ASoP (v1.0): a set of methods for analyzing scales of precipitation in general circulation models." *Geoscientific Model Development* 10(1): 57–83, https://doi.org/10.5194/gmd-10-57-2017

Kunkel, KE, DR Easterling, DAR Kristovich, B Gleason, L Stoecker, and R Smith. 2012. "Meteorological Causes of the Secular Variations in Observed Extreme Precipitation Events for the Conterminous United States." *Journal of Hydrometeorology* 13(3): 1131–1141, https://doi.org/10.1175/JHM-D-11-0108.1

References (continued)

Kuo, Y-H, JD Neelin, and CR Mechoso. 2017. "Tropical convective transition statistics and causality in the water vapor-precipitation relation." *Journal of the Atmospheric Sciences* 74(3): 915–931, https://doi.org/10.1175/JAS-D-16-0182.1

Lebassi-Habtezion, B, and NS Diffenbaugh. 2013. "Nonhydrostatic nested climate modeling: A case study of the 2010 summer season over the western United States." *Journal of Geophysical Research – Atmospheres* 118(19): 10,944–10,962, https://doi.org/10.1002/jgrd.50773

Leung, LR, and SJ Ghan. 1998. "Parameterizing Subgrid Orographic Precipitation and Surface Cover in Climate Models." *Monthly Weather Review* 126(12): 3271–3291, https://doi.org/10.1175/1520-0493(1998)126<3271:PSOPAS>2.0.CO;2

Leung, LR, and Y Qian. 2003. "The Sensitivity of Precipitation and Snowpack Simulations to Model Resolution via Nesting in Regions of Complex Terrain." *Journal of Hydrometeorology* 4(6): 1025–1043, https://doi.org/10.1175/1525-7541(2003)004<1025:TSOPAS>2.0.CO;2

Levin, RC and GM Martin. 2017. "On the climate model simulation of Indian monsoon low pressure systems and the effect of remote disturbances and systematic biases." *Climate Dynamics* 50: 4721–4743, https://doi.org/10.1007/s00382-017-3900-x

Maloney, ED, A Gettelman, Y Ming, JD Neelin, D Barrie, A Mariotti, C-C Chen, DRB Coleman, Y H Kuo, B Singh, H Annamalai, A Berg, JF Booth, SJ Camargo, A Dai, A Gonzalez, J Hafner, X Jiang, X Jing, D Kim, A Kumar, Y Moon, CM Naud, AH Sobel, K Suzuki, F Wang, J Wang, AA Wing, X Xu, and M Zhao. 2019. "Process-Oriented Evaluation of Climate and Weather Forecasting Models." *Bulletin of the American Meteorological Society* 100(9): 1665–1686, https://doi.org/10.1175/BAMS-D-18-0042.1

Martin, GM, NP Klingaman, and AF Moise. 2017. "Connecting spatial and temporal scales of tropical precipitation in observations and the MetUM-GA6." *Geoscientific Model Development* 10(1): 105–126, https://doi.org/10.5194/gmd-10-105-2017

Martinez-Villalobos, C, and JD Neelin. 2019. "Why do precipitation intensities tend to follow Gamma distributions?" *Journal of the Atmospheric Sciences* 76(11): 3611–3631, https://doi.org/10.1175/JAS-D-18-0343.1

McKee, TB, NJ Doesken, and J Kleist. 1993. "The relationship of drought frequency and duration to time scales." In *Proceedings of the 8th Conference on Applied Climatology* 17(22): 179–183). American Meteorological Society. Boston, Massachusetts.

Neu, U, MG Akperov, N Bellenbaum, R Benestad, R Blender, R Caballero, A Cocozza, HF Dacre, Y Feng, K Fraedrich, J Grieger, S Gulev, J Hanley, T Hewson, M Inatsu, K Keay, SF Kew, I Kindem, GC Leckebusch, MLR Liberato, P Lionello, II Mokhov, JG Pinto, CC Raible, M Reale, I Rudeva, M Schuster, I Simmonds, M Sinclair, M Sprenger, ND Tilinina, IF Trigo, S Ulbrich, U Ulbrich, XL Wang, and H Wernli. 2013. "IMILAST: A community effort to intercompare extratropical cyclone detection and tracking algorithms." *Bulletin of the American Meteorological Society* 94(4): 529–547, https://doi.org/10.1175/bams-d-11-00154.1

Pendergrass, AG, and C Deser. 2017. "Climatological characteristics of typical daily precipitation." *Journal of Climate* 30(15): 5985–6003, https://doi.org/10.1175/JCLI-D-16-0684.1

Pendergrass, AG, and R Knutti. 2018. "The uneven nature of daily precipitation and its change." *Geophysical Research Letters* 45(21): 11980–11988, https://doi.org/10.1029/2018GL080298

Perkins, SE, AJ Pitman, NJ Holbrook, and J McAneney. 2007. "Evaluation of the AR4 Climate Models' Simulated Daily Maximum Precipitation over Australia Using Probability Density Functions." *Journal of Climate* 20(17): 4356–4376, https://doi.org/10.1175/JCLI4253.1

Peters, O, and JD Neelin. 2006. "Critical phenomena in atmospheric precipitation." *Nature Physics* 2: 393–396, https://doi.org/10.1038/nphys314

Praveen, V, S Sandeep, and RS Ajayamohan. 2015. "On the relationship between mean monsoon precipitation and low pressure systems in climate model simulations." *Journal of Climate* 28:5305–5324, https://doi.org/10.1175/JCLI-D-14-00415.1

Risser, M, CJ Paciorek, MF Wehner, TA O'Brien, and WD Collins. 2019. "A probabilistic gridded product for daily precipitation extremes over the United States." *Climate Dynamics* 53(5-6): 2517–2538, https://doi.org/10.1007/s00382-019-04636-0

Roca, R, LV Alexander, G Potter, M Bador, R Jucá, S Contractor, MG Bosilovich, and S Cloché. 2019. "FROGS: a daily 1° × 1° gridded precipitation database of rain gauge, satellite and reanalysis products." *Earth System Science Data* 11(3): 1017–1035, https://doi.org/10.5194/essd-11-1017-2019

Rushley, SS, D Kim, and AF Adames. 2019. "Changes in the MJO under greenhouse gas-induced warming in CMIP5 models." *Journal of Climate* 32(3): 803–821, https://doi.org/10.1175/JCLI-D-18-0437.1

Schiro, KA, JD Neelin, DK Adams, and BR Linter. 2016. "Deep convection and column water vapor over tropical land versus tropical ocean: A comparison between the Amazon and the tropical western Pacific." *Journal of the Atmospheric Sciences* 73(10): 4043–4063, https://doi.org/10.1175/JAS-D-16-0119.1

Shields, CA, JJ Rutz, LR Leung, FM Ralph, M Wehner, B Kawzenuk, J Lora, E McClenny, T Osborne, A Payne, Y Qian, C Sarangi, P Ullrich, A Gershunov, N Goldenson, B Guan, A. Ramos, S Sellars, I Gorodetskaya, K Mahoney, R Pierce, A Subramanian, D Waliser, G Wick, A Wilson, D Lavers, Prabhat, A Collow, F Dominguez, H Krishnan, G Magnusdottir, P Nguyen, T O'Brien, R Silva, and M Tsukernik. 2018. "Atmospheric River Tracking Method Intercomparison Project (ARTMIP): Experimental Design and Project Goals." *Geoscientific Model Development* 11(6): 2455–2474, https://doi.org/10.5194/gmd-11-2455-2018

Sillmann, J, VV Kharin, X Zhang, FW Zwiers and D Bronaugh, 2013. "Climate extremes indices in the CMIP5 multi-model ensemble. Part 1: Model evaluation in the present climate." *Journal of Geophysical Research – Atmospheres* 118(4): 1716–1733, https://doi.org/10.1002/jgrd.50203

Sobel, AH, SE Yuter, CS Bretherton, and GN Kiladis. 2004. "Large-scale meteorology and deep convection during TRMM KWAJEX." *Monthly Weather Review* 132(2): 422–444, https://doi.org/10.1175/1520-0493

Song, F, Z Feng, LR Leung, RA Houze, Jr., J Wang, J Hardin, and C Homeyer. 2019. "Contrasting the Spring and Summer Large-Scale Environments Associated with Mesoscale Convective Systems over the U.S. Great Plains." *Journal of Climate* 32(20): 6749–6767, https://doi.org/10.1175/JCLI-D-18-0839.1

Sperber KR, H Annamalai, I-S Kang, A Kitoh, A Moise, A Turner, B Wang, T Zhou. 2013. "The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century." *Climate Dynamics* 41(9-10): 2711–2744, https://doi.org/10.1007/s00382-012-1607-6

Stephens, G., T L'Ecuyer, R Forbes, A Gettleman, J-C Golaz, A Bodas-Salcedo, K Suzuki, P Gabriel, and J Haynes. 2010. "Dreary state of precipitation in global models." *Journal of Geophysical Research – Atmospheres* 115(D24), https://doi.org/10.1029/2010JD014532

Taylor, KE. 2001. "Summarizing multiple aspects of model performance in a single diagram." *Journal of Geophysical Research* 106(D7): 7183–71, 92, https://doi.org/10.1029/2000JD900719

Trenberth, KE, Y Zhang, and M Gehne. 2017. "Intermittency in precipitation: Duration, frequency, intensity, and amounts using hourly data." *Journal of Hydrometeorology* 18(5): 1393–1412, https://doi.org/10.1175/JHM-D-16-0263.1

Tselioudis, G, W Rossow, Y Zhang, and D Konsta. 2013. "Global weather states and their properties from passive and active satellite cloud retrievals." *Journal of Climate* 26(19): 7734–7746, https://doi.org/10.1175/JCLI-D-13-00024.1

Ullrich, PA, and CM Zarcycki. 2017. "TempestExtremes: a framework for scale-insensitive pointwise feature tracking on unstructured grids." *Geoscientific Model Development* 10(3): 1069–1090, https://doi.org/10.5194/gmd-10-1069-2017

Xie, S, W Lin, PJ Rasch, P-L Ma, R Neale, VE Larson, Y Qian, PA Bogenschutz, P Caldwell, P Cameron-Smith, J-C Golaz, S Mahajan, B Singh, Q Tang, H Wang, J-H Yoon, K Zhang, and Y Zhang. 2018. "Understanding cloud and convective characteristics in version 1 of the E3SM atmosphere model." *Journal of Advances in Modeling Earth Systems* 10(10): 2618–2644. https://doi.org/10.1029/2018MS001350

Xie, S, Y-C Wang, W Lin, H-Y Ma, Q Tang, S Tang, X Zheng, J-C Golaz, GJ Zhang, and M Zhang. 2019. "Improved diurnal cycle of precipitation in E3SM with a revised convective triggering function." *Journal of Advances in Modeling Earth Systems* 11(7): 2290–2310. https://doi.org/10.1029/2019MS001702

Yang, Q, LR Leung, J Lu, Y-L Lin, S Hagos, K Sakaguchi, and Y Gao. 2017. "Exploring the effects of a nonhydrostatic dynamical core in high-resolution aquaplanet simulations." *Journal of Geophysical Research – Atmospheres* 122(6): 3245–3265, https://doi.org/10.1002/2016JD025287

Zhang, C, and S Xie. 2017. ARM Data-Oriented Metrics and Diagnostics Package for Climate Model Evaluation Value-Added Product. United States. https://doi.org/10.2172/1396238

Zhang, X, L Alexander, GC Hegerl, P Jones, AK Tank, TC Peterson, B Trewin, and FW Zwiers. 2011. "Indices for monitoring changes in extremes based on daily temperature and precipitation data." *WIREs Climate Change* 2(6): 851–870, https://doi.org/10.1002/wcc.147

Appendix A – Workshop Agenda

DOE Precipitation Metrics Workshop Agenda | July 1-2, 2019 | Hilton Washington D.C., Rockville

Day 1	
8:00 AM	**Gather**
Welcome and Introductions	
8:20 AM	Welcome from DOE (R. Joseph, G. Geernaert)
Background, Motivation, Starting Point and Expectations	
8:35 AM	Aims of the workshop – Why are we here? (C. Jakob)
8:45 AM	Perspectives on what to include as a baseline (A. Pendergrass)
8:55 AM	A strawman as a starting point (P. Gleckler)
9:20 AM	Perspectives on exploratory metrics (R. Leung)
9:35 AM	Attendee intros, reactions and 1–3-minute perspectives on pre-meeting ideas
10:30 AM	**Break**
Topics relevant to defining Precipitation Benchmarks	
In addition to presenting their expertise as it relates to the workshop, presenters are asked to discuss their views on: 1) the co-chairs' strawman, 2) additional/alternate candidates for an initial set of benchmarks and 3) topics relevant for future research that may lead to a more comprehensive and advanced set of metrics.	
10:45 AM	A modeler's perspective (R. Neale)
11:00 AM	Evaluating Simulated Precipitation (including Snowfall) with Satellite Observations (T. L'Ecuyer)
11:20 AM	Extremes evaluation under observational uncertainty (M. Bador)
11:40 AM	Connecting spatial and temporal scales of precipitation (G. Martin)
11:50 AM	An Impacts-related perspective (A. Ruane)
12:05 PM	Return value extremes and scale mismatch in model evaluation (M. Wehner)
12:25 PM	Use-inspired metrics (P. Ullrich)
12:45 PM	**Lunch**
Break Out Group (BOG) discussion – Identifying initial set of benchmarks	
2:00 PM	Group 1 Chair: Christian Jakob Rapporteur: George Tselioudis Group 2 Chair: Peter Gleckler Rapporteur: Travis O'Brien Discussion topics 1. A set of metrics, comprehensive but also concise and ready to go, and supported by the group, for evaluating precip in models • How should we address observational uncertainty? (building on the GEWEX observational assessment) • How should we address observational uncertainty? (building on the GEWEX observational assessment) • A plan for efficient implementation 2. A plan for turning this into a publication evaluating the CMIP6 models; consider the feasibility of doing this by December 31, 2019 3. A set of (research) priorities for what should go into the next benchmarking round (which won't have the same time pressure as the IPCC/CMIP deadline)

3:30 PM	**Break**
3:45	BOG Report out and plenary discussion
5:30	End of session
Day 2	
9:00 AM	Discussion – overnight thoughts
Topics relevant to exploratory metrics	
9:30 AM	Evaluating simulated precipitation with ARM data (S. Xie)
9:40 AM	Convection onset metrics (D. Neelin)
10:00 AM	Machine learning and frontal systems (K. Kunkel)
10:15 AM	Discussion
10:30 AM	**Break**
10:45 AM	Extratropical and frontal rainfall (J. Catto)
11:00 AM	A regime-oriented perspective on evaluation of precipitation (G. Tselioudis)
11:20 AM	Variability of monsoons (B. Boos)
11:30 AM	Discussion
12:30 PM	**Lunch**
Breakout Group (BOG) discussions – 1) establishing initial set and 2) areas of needed research	
2:00 PM	Group 3: What needs to happen for: 1. Solidifying the strawman – initial set of benchmarks, and other details of the approach 2. Strategy for analysis + publication by IPCC deadline? Chair: Angie Pendergrass Rapporteur: Rich Neale People in this group: Margot Bador, Jiwoo Lee, Tristan L'Ecuyer, Michael Wehner, Christian Jakob, Peter Gleckler, Paul Ullrich Group 4: What should the next generation of metrics look like? - exploratory and process-oriented metrics discussion Chair: Ruby Leung Rapporteur: Shaocheng Xie 1. What aspects (e.g., precipitation types, higher-order moments) of precipitation should the exploratory metrics focus on to complement the standard metrics? 2. Should relationships between precipitation and other quantities be considered in the exploratory metrics? 3. What principles should be used to prioritize the exploratory metrics? 4. How to implement a phased approach for exploratory metrics? People in this group: Bill Boos, Jennifer Catto, Ken Kunkel, Gill Martin, David Neelin, George Tselioudis, Travis O'Brien
3:30 PM	**Break**
4:00 PM	Reports from BOGs Discuss how we will move forward
5:00 PM	**End of workshop**

Appendix B – Participants

Participant	Affiliation
Margot Bador	University of New South Wales (Australia)
Dan Barrie	National Oceanic and Atmospheric Administration
Paul Bayer	U.S. Department of Energy
Bill Boos	University of California, Berkeley
Jennifer Catto	University of Exeter (UK)
Charlotte Demott	Colorado State University
David Easterling	NOAA/National Centers for Environmental Information
Andrew Flatness	U.S. Department of Energy
Gary Geernaert	U.S. Department of Energy
Peter Gleckler	Lawrence Livermore National Laboratory
Wayne Higgins	National Oceanic and Atmospheric Administration
Justin Hnilo	U.S. Department of Energy
Jin Huang	National Oceanic and Atmospheric Administration
Christian Jakob	Monash University (Australia)
Renu Joseph	Department of Energy
Chris Kidd	NASA's Goddard Space Flight Center
Ken Kunkel	North Carolina Institute for Climate Studies
Tristan L'Ecuyer	University of Wisconsin, Madison
Jiwoo Lee	Lawrence Livermore National Laboratory

Participant	Affiliation
Ruby Leung	Pacific Northwest National Laboratory
Gill Martin	Met Office (UK)
Sally McFarlane	U.S. Department of Energy
Jessica Moerman	U.S. Department of Energy
Shaima Nasiri	U.S. Department of Energy
Rich Neale	National Center for Atmospheric Research
David Neelin	University of California, Los Angeles
Travis O'Brien	Lawrence Berkeley National Laboratory
Angie Pendergrass	National Center for Atmospheric Research
Rick Petty	U.S. Department of Energy
Alex Ruane	NASA Goddard Institute for Space Studies
Jennifer Saleem-Arrigo	U.S. Global Change Research Program
Ginny Selz	National Oceanic and Atmospheric Administration
Daniel Stover	U.S. Department of Energy
George Tselioudis	NASA Goddard Institute for Space Studies
Paul Ullrich	University of California, Davis
Bob Vallario	U.S. Department of Energy
Michael Wehner	Lawrence Berkeley National Laboratory
Shaocheng Xie	Lawrence Livermore National Laboratory

For More Information

Regional & Global Model Analysis
climatemodeling.science.energy.gov/rgma
Renu Joseph, renu.joseph@science.doe.gov