# CMIP5 SCIENTIFIC GAPS AND RECOMMENDATIONS FOR CMIP6

R. J. Stouffer, V. Eyring, G. A. Meehl, S. Bony, C. Senior, B. Stevens, and K. E. Taylor

The scientific gaps identified in the fifth phase of the Coupled Model Intercomparison Project (CMIP5) that guided the experiment for its next phase, CMIP6, are identified.

The Coupled Model Intercomparison Project (CMIP) coordinates the comparison of comprehensive climate models and has its roots in earlier model intercomparisons, such as the Atmospheric Model Intercomparison Project (AMIP; Gates 1992; Gates et al. 1999). CMIP has contributed to the evolution and progress of climate science since the mid-1990s, when it was first organized by the World Climate Research Programme (WCRP) Working Group on Coupled Modelling (WGCM). The objective of CMIP is to design coordinated global simulations of the coupled climate system and make available a wide range of model output to advance understanding of past, present, and future climate variability and change. This paper describes factors influencing the experimental design of CMIP6 in the context of the historical development of CMIP as a whole, and as briefly reviewed below.

**AFFILIATIONS:** Stouffer—NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey; Eyring—Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany; Meehl—National Center for Atmospheric Research, Boulder, Colorado; Bony—Laboratoire de Meteorologie Dynamique, IPSL, CNRS, Paris, France; Senior—Met Office Hadley Centre, Exeter, United Kingdom; Stevens—Max Planck Institute for Meteorology, Hamburg, Germany; Taylor—Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California
**CORRESPONDING AUTHOR E-MAIL:** Ronald J. Stouffer, ronald.stouffer@noaa.gov

*The abstract for this article can be found in this issue, following the table of contents.*
DOI:10.1175/BAMS-D-15-00013.1

In final form 3 June 2016


In the first two phases of CMIP (CMIP1 and CMIP2), the experimental design was simple. It encompassed a long control integration where no interannual changes in radiative forcing were allowed and an idealized simulation where atmospheric carbon dioxide ($CO_2$) concentration increased at a rate of 1% per year (doubling near model year 70). Idealized studies like these remain at the heart of CMIP because they allow differences in model responses to be better understood. This understanding is a foundation for confidence in model projections of future climate change. Results from the early CMIP simulations were analyzed through specially created subprojects designed to engage in targeted analyses (Meehl et al. 1997, 2000, 2005). As CMIP evolved, the suite of experiments grew to include a twentieth-century simulation and more detailed and elaborate projections for future changes in climate forcers, in addition to more idealized simulations developed to understand better specific climate processes.

CMIP3 marked a paradigm shift in the climate science community by making model output from state-of-the-art climate change simulations broadly accessible by the scientific community at large (Meehl et al. 2007). This resulted in an explosion of scientific results based on multimodel ensembles of simulations, and it helped establish the concept of robustness as part of the climate modeling lexicon (Held and Soden 2006). The broad availability of model output also enabled its use beyond the traditional climate modeling community. Various communities, including those who study the impact of climate changes from a variety of perspectives, contributed to this expansion of the science.

The intermediate but not widely known CMIP phase (called CMIP4 by the WGCM committee) modestly supplemented the experiments performed in CMIP3 with, for example, single-forcing experiments that held all but one of the twentieth-century forcings fixed. These types of experiments have become the cornerstone of detection and attribution (D&A) studies. An additional rationale for CMIP4 was to align the numbering of the CMIP phases with the IPCC assessments. Not having a major CMIP phase labeled CMIP4 also avoided unnecessary confusion with the existing model intercomparison project (MIP)—the Coupled Climate–Carbon Cycle Model Intercomparison Project (C4MIP), a carbon-focused activity.

The next major phase was CMIP5 (Taylor et al. 2012), which built upon CMIP3 and included more idealized process- and feedback-oriented experiments and output to facilitate understanding of the climate system (Fig. 1). It also was designed to address the physical mechanisms through which the climate system responds to changes in external forcing in the context of internally generated climate variability, including extreme conditions of the more distant past (Braconnot et al. 2012). New for CMIP5 were experiments that enabled investigation of the fast climate responses to perturbed atmospheric $CO_2$ concentrations (Taylor et al. 2012) and the impact of atmospheric chemistry on climate (Lamarque et al. 2013), troposphere–stratosphere interactions (Eyring et al. 2013), and carbon–climate interactions (Arora et al. 2013; Friedlingstein et al. 2014) and feedbacks, as well as idealized model configurations used, for example, in the aquaplanet experiment (Stevens and Bony 2013). For the first time, the suite of experiments was divided into "near term" and "long term" time horizons (Fig. 1). The long-term simulations used the traditional method of starting twentieth-century and twenty-first-century experiments from a preindustrial control run. The near-term experiments included new types of simulations in which climate models were initialized with observations of the climate system at specific times, starting in 1960, allowing for more detailed analysis of predicted changes in the next few decades, which defined the new field of "decadal climate prediction"(Meehl et al. 2009; Smith et al. 2013). Yet another innovation within CMIP5 was the adoption of a new approach to specifying future scenarios by defining representative concentration pathways (RCPs; Moss et al. 2010) developed in collaboration with the integrated assessment modeling community (Kriegler et al. 2012).

In the early years of CMIP (and of its predecessor AMIP), the Program for Climate Model Diagnosis and Intercomparison (PCMDI; http://www-pcmdi.llnl.gov/projects/pcmdi/) largely provided the project management and infrastructure support for CMIP. PCMDI helped champion the establishment of well-defined experiment protocols and data standards, and it hosted several of the early CMIP data archives that facilitated access to and analysis of multimodel output. PCMDI continues to be a key supporter of CMIP, but responsibility for CMIP is now more broadly shared across the community through what is known as the Earth System Grid Federation (Williams et al. 2015).

Over the last two decades, CMIP has expanded and evolved, and the demands it places on modeling groups have also grown. The CMIP Panel, a subgroup of WGCM, traditionally has the responsibility for direct coordination and oversight of CMIP and has defined phase 6 of CMIP (i.e., CMIP6) in consultation with the climate science community (Eyring et al. 2016b). During the last few decades, the climate modeling community itself has also evolved. In the early years, most model analysis was carried out within or in collaboration with individual modeling groups. Today, modeling centers develop models and more routinely release state-of-the-art model output for public scrutiny. Because much of the analysis takes place outside of the modeling centers, the planning phase for CMIP now involves both climate modeling groups and the community of scientists analyzing results. Thus, CMIP plays an important role in helping these groups to exchange insights. Though the CMIP Panel's goal is to define and coordinate the CMIP experiments and to provide model output that can effectively address compelling climate science questions, in the end it is the individual modeling groups—not the CMIP Panel—that consider the list of experiments and requested output and decide which subset of the CMIP experiments to run and what subset of their model output is made public. Likewise, the climate research community at large decides what aspects of the model simulations will be analyzed. These important aspects of CMIP are often misunderstood by people unfamiliar with how CMIP is organized

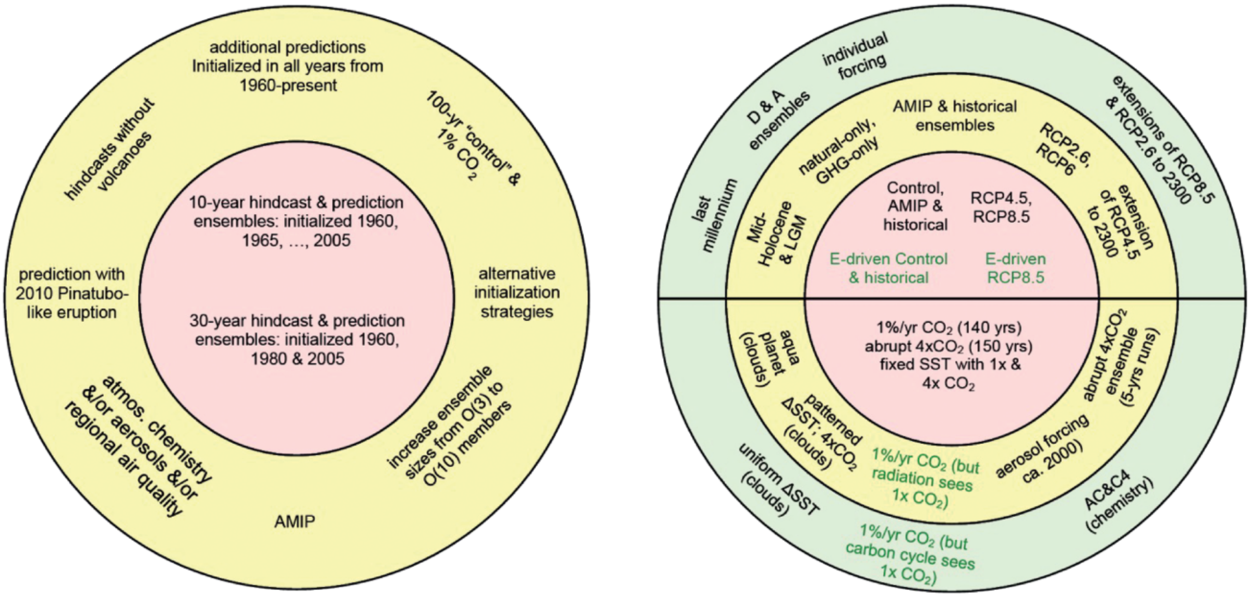


**Fig. 1. Schematic summary of CMIP5 long-term experiments with tier 1 and tier 2 experiments organized around a central core. Green font indicates simulations to be performed only by models with carbon cycle representations. Experiments in the upper hemisphere are suitable either for comparing with observations or for providing projections, whereas those in the lower hemisphere are either idealized or diagnostic in nature and aim to provide better understanding of the climate system and model behavior. Figure taken from Taylor et al. (2012). Hindcasts are model predictions of past climate changes. AC&C4 refers to the atmospheric chemistry and coupled climate–carbon cycle. E-driven is short for emission-driven ESM model runs. LGM is the Last Glacial Maximum (about 21,000 years ago).**

and coordinated. CMIP's central goal is to advance scientific understanding of the Earth system *and* to be a valuable resource for national and international climate assessments, including the Intergovernmental Panel on Climate Change (IPCC).

Planning for CMIP6 started in 2013, again building on the previous phases of CMIP. Experience has shown that the analysis of model output from earlier phases of CMIP is an ongoing process without a specific end date. Planning of a new CMIP phase needs to begin even though the previous phase has not yet completed, so that the lessons learned are fresh and the infrastructure needed to support CMIP remains intact. An important part of the planning process is surveying community needs. Over the past two years, this has taken place in a variety of ways, in part through feedback solicited by an initial proposal for the design of CMIP6 (Meehl et al. 2014) and also as a result of a user survey taken by the CMIP Panel.

The purpose of this paper is to identify science gaps highlighted through this process of community input with a special focus on the user survey. The "Lessons from CMIP5 and a path forward" section summarizes the main findings of the user survey that is used as a basis to provide an outline for a brief discussion of the scientific gaps that are identified and presented in the "Scientific gaps and recommendations for CMIP6" section, and the "Summary and discussion" section closes with a summary and outlook.

**LESSONS FROM CMIP5 AND A PATH FORWARD.** A questionnaire was developed by the CMIP Panel in late spring of 2013 and was sent to the climate modeling groups that participated in CMIP5, the various MIP chairs active in 2013, the cochairs of WCRP and related International Geosphere–Biosphere Programme (IGBP) working groups, the Integrated Assessment Model (IAM) groups, the Earth System Grid Federation (ESGF) data providers, and members of the Climate Services Partnership Coordinating Group, as well as to many IPCC Working Group I authors. Preliminary results of this survey were presented during the summer of 2013 at meetings at the Energy Modeling Forum in Snowmass, Colorado; the Aspen Global Change Institute (AGCI) in Aspen, Colorado; and the WGCM meeting in Victoria, Canada, in the fall of 2013 (see template and presentations under "CMIP5 Survey" at www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6).

An important finding of the survey is that CMIP should focus strongly on specific science questions.

To achieve this goal, CMIP needs to create a scientific framework for a set of coordinated multimodel experiments designed to increase our understanding of the climate system. At the same time CMIP needs to remain relevant to the broader climate change community.

To help emphasize the opportunities to advance scientific understanding, the survey recommended CMIP6 focus on three broad scientific questions, as outlined by Meehl et al. (2014):

1) How does the Earth system respond to changes in forcing?
2) What are the origins and consequences of systematic model biases?
3) How can we assess future climate changes given internal climate variability, predictability, and uncertainties in scenarios?

In the following section, using this framework and the results of the survey, we briefly summarize the main scientific gaps found in CMIP5 that influenced the design of CMIP6 (Eyring et al. 2016b).

## SCIENTIFIC GAPS AND RECOMMENDATIONS FOR CMIP6.

*How does the Earth system respond to changes in forcing?* The question of how the Earth System responds to changes in forcing continues to be raised as the critical issue to advance climate understanding by those surveyed. Often this question is simplified into a single metric, such as the equilibrium climate sensitivity (e.g., Gregory et al. 2004; Andrews et al. 2012), which is defined as the change in global mean surface temperature at equilibrium that is caused by a doubling of the atmospheric $CO_2$ concentration. As noted earlier, CMIP has designed idealized experiments to highlight and understand differences in climate model response to a given common change in radiative forcing. The experiments forced by an abrupt doubling or quadrupling of the $CO_2$ concentration in the atmosphere and the simulation forced by a 1% per year $CO_2$ increase are examples of these types of experiments. Indeed, understanding differences and the spread in climate models' responses is part of nearly all of the CMIP experiments, and this continues to be an important science question addressed by CMIP6.

A related science gap in CMIP5 identified by the survey is how estimates in radiative forcing resulting from human (emissions, land use changes, etc.) or natural (solar changes, volcanoes, etc.) causes can be best quantified. Poor estimates of the forcings can hinder understanding of the causes of different model responses to those forcing changes. In Working Group 1 of the IPCC Fifth Assessment Report (AR5), a concept called the effective radiative forcing (ERF) was extensively used (Boucher et al. 2013; Myhre et al. 2013). It extends the traditional idea of stratospheric adjustment to also allow tropospheric conditions to adjust to the change in atmospheric composition while, ideally, the surface temperature does not change; as such, these adjustments are related to the fixed sea surface temperature (SST) simulations in Hansen et al. (2005). As noted by Myhre et al. (2013) and Chung and Soden (2015), calculations of the radiative forcing (RF) and ERF can differ from each other, and the differences in estimates of radiative forcing among models can be large.

As reviewed by Sherwood et al. (2015), different methods have been proposed for diagnosing the effective radiative forcing. But to understand the differences in the forcings that are eventually diagnosed by these methods, it is important to include more detailed radiative forcing calculations through multiple calls to the radiative transfer code, leaving in or out one or the other constituent through the course of some simulations (Ghan 2013), and through extensive offline comparisons of radiative transfer calculations themselves.

Another issue related to forcings is found in the terrestrial modeling community, where various physical and biogeochemical processes are actively incorporated into the land components of Earth system models (ESMs), but this development still lacks a common framework for intercomparison. Though coordinated time-evolving land-use/land-cover changes were included in the CMIP5 model experiments, the terrestrial components of ESMs used in CMIP5 were quite different from each other in how they modeled the various biological and ecological processes associated with biome type and plant growth and dying and how they handled changes in land use associated with human activities. This led to severe difficulty in evaluating the land components against observations and each other. Ciais et al. (2013, p. 504) concluded, "Broadly, models are still at their early stages in dealing with land use, land use change, and forestry." More quantitatively, Pongratz et al. (2014) showed that up to a factor-of-2 difference in estimates of land carbon emissions can be attributed to the lack of a coordinated terminology. As climate models become more comprehensive, CMIP6 has encouraged the development of model intercomparision projects designed to specifically address these past issues, wherein different components have been added without adequate attention devoted to a strict quantification of what was done.

In summary, a science gap in CMIP5 is the surprisingly poor quantification of radiative forcing in climate models. Without reliable forcing estimates, it is very hard to compare the response of models to changes in the forcings. This gap is particularly notable in the radiative forcing estimates for short-lived atmospheric species, such as aerosols, but it increasingly applies to other components (e.g., the land surface) as ESMs become more and more complex.

*What are the origins and consequences of systematic model biases?* Climate model simulations, when compared to observations, reveal a wide variety of errors on various time and space scales (Flato et al. 2013). Some of these errors have been evident for several decades (prior to the first IPCC assessment in 1990) and together have been called "systematic model biases." Though an ongoing effort has been devoted to reducing or eliminating these biases, such work is often subordinated to the more exciting, easier, and often more rewarding job of expanding the scope of modeling. As a result, systematic biases remain a major climate modeling challenge.

The impact of model errors on the models' climate response has been an important focus of CMIP from the beginning (Meehl et al. 1997). Particularly important and long-standing biases that we hope will be addressed within CMIP6 include the following:

1) The double intertropical convergence zone (ITCZ) is evident to some extent in all models (e.g., Oueslati and Bellon 2015). On Earth, the ITCZ is mainly found in the Northern Hemisphere through most of the annual cycle in the tropical Pacific Ocean. However, in models the ITCZ is much stronger in the Southern Hemisphere and often persists through the seasonal cycle.
2) The double-ITCZ problem may be related to problems simulating the Walker circulation and the associated dry Amazon bias also seen in many models, as well as the representation of tropical variability (Crueger and Stevens 2015; Oueslati and Bellon 2015). The dry Amazon problem is particularly important, as it can lead to large errors in the amount and type of vegetation growing in the Amazon and impact the land carbon fluxes—a cascade of tightly related errors.
3) A third error found in most climate models is the poor simulation of tropical and subtropical low clouds, particularly stratocumulus layers that prevail over the eastern parts of ocean basins. This bias is often related to the sea surface temperatures in the control climates being too warm in the eastern parts of the tropical ocean basins, and it appears to be related to biases in surface humidity and cloudiness that arise from a poor coupling of boundary layer processes to the large-scale climate state (Hourdin et al. 2015). This error can also impact the models' climate sensitivity and transient climate response, since low cloud changes can greatly impact the climate response (Randall et al. 2007; Sherwood et al. 2014). Boucher et al. (2013) noted that this situation had not changed in the newer generation of climate models used in the IPCC AR5.
4) In the ocean, a common model problem is an overly deep tropical thermocline. This problem can adversely impact the model's El Niño–Southern Oscillation (ENSO) simulation in both its magnitude and frequency (Flato et al. 2013; Li and Xie 2012). It can also adversely affect the simulation of biogeochemical tracers and associated marine ecology, and it may be related to the double-ITCZ problem noted above. Further, a too deep thermocline can impact the response of the model to future changes in greenhouse gases (GHGs) and other forcings due to errors in the oceanic heat and carbon uptake and associated heat and carbon storage.
5) A problem common to most climate models is the tendency to predict land surfaces too warm and too dry during summertime (Flato et al. 2013; Klein et al. 2006). This bias appears to result from deficiencies in the representation of land surface schemes and/or cloud schemes (Cheruy et al. 2014; Mueller and Seneviratne 2014).
6) Another common model problem is the position of the Southern Hemisphere atmospheric jet. In many models, this feature is located 5°–10° of latitude equatorward of its observed location (Russell et al. 2006a). The poor position of the atmospheric jet leads to poor simulation of the surface wind stress on the Southern Ocean and to errors in the vertical structure of the water masses found in the Southern Ocean. These simulation limitations then adversely impact the transient oceanic heat (Russell et al. 2006b) and carbon (Frolicher et al. 2015) uptakes of the model, as GHGs are increased in the model atmosphere.

These are just some examples of prominent and systematic model biases that have proven difficult to fix but which CMIP6 should target. The following two strategies should be followed in CMIP6. The first strategy is to develop more innovative experimentation to articulate specific processes, for example, the use of aquaplanets (Williamson et al. 2013) to study factors influencing the position of the ITCZ (e.g., Stevens and Bony 2013), high-resolution approaches to study the influence of resolution on circulation features, and the more systematic evaluation of individual model components (e.g., atmosphere, land surface, or ocean). The second strategy is a more concerted effort to link model output to observations through forward operators (Bodas-Salcedo et al. 2011; Williamson et al. 2015; Yu et al. 1996) and the creation of observation-based datasets whose structure and metadata mirror that of the model-based datasets (e.g., CMIP; Teixeira et al. 2014), as well as more coordinated efforts to apply community-developed evaluation packages. Examples of the latter are the Earth System Model Evaluation Tool (ESMValTool; Eyring et al. 2016a) and the PCMDI metrics package (PMP; Gleckler et al. 2016), which are starting to be used by the community for a more routine assessment of new model versions, and the process-oriented Madden–Julian oscillation diagnostic package (Kim et al. 2014), which has helped many modeling centers focus more attention on their representation of the Madden–Julian oscillation.

*How can we assess future climate changes given climate variability, predictability, and uncertainties in scenarios?* Climate variability. There are several different aspects to this question. One is the need to assess natural variability when detecting or attributing a change in some variable of interest (Santer et al. 1995). A second aspect as noted below (in the "Decadal climate prediction" section) is understanding the physical mechanisms underlying climate variability, which is essential to successful predictions of climate on seasonal to decadal time scales, where both the evolution of the variability and the response to changes in radiative forcing are important.

As a result, clearly identifying the signal (i.e., the response to the forcing) and the noise (variability) is an important part of CMIP. Often an ensemble of integrations generated by a model started from slightly different initial conditions is requested in an attempt to better document the signal and the noise. Since the signal-to-noise ratio varies from variable to variable and on the time and space scales of interest, the number of ensemble members needed to clearly define the signal varies depending on the question being asked. To document the noise when the radiative forcing is not changing, long control integrations are needed. Both of these requests require more computing and archive resources.

Future scenarios. In CMIP5, there was, for the first time, a direct connection between the developers of the future scenarios—for example, integrated assessment modelers—and the modeling centers (Hibbard et al. 2007). In previous phases of CMIP, the uncertainty in each step in the development of the scenario and the uncertainty of the climate model response were combined. In CMIP5, these two uncertainties were separated as part of the implementation process, allowing modelers in each working group to work in parallel (Moss et al. 2010). However, the combination of several mitigation actions in each RCP (a plausible future scenario) made it difficult to assess the costs and benefits of individual mitigation actions in CMIP5. CMIP6 will include a first attempt to address this issue, so that the costs and benefits of any action to limit GHG emissions can be more easily evaluated in any future scenario.

A frequent issue raised in the user survey was that the spread in aerosol scenarios in the four RCPs was too narrow. All four RCPs assume a large reduction in the atmospheric aerosol emissions (and therefore concentrations) by the end of this century (Moss et al. 2010). Even if future changes in aerosol emissions might have relatively little effect on global climate (Stevens 2015), regional effects would still have important implications for air pollution and human health. The new shared socioeconomic pathways (SSPs) that will be used in CMIP6 climate projections will allow this gap to be addressed, since they will cover a much wider range of air pollutant emissions than the RCPs (Riahi et al. 2017).

Another issue is that the spread in model response and climate sensitivity remains large (Collins et al. 2013). Further model improvements to reduce the uncertainty in key climate feedbacks and constraints of the model ensemble with observations (e.g., Cox et al. 2013; Hall and Qu 2006; Sherwood et al. 2014; Wenzel et al. 2014) will remain a challenge for CMIP6 and beyond. There are also still many long-standing issues remaining regarding the use of the CMIP database as an ensemble of opportunity. These issues include the possible use of CMIP data across different phases (Rauser et al. 2015), an improved quantification of uncertainties in the resulting ensemble projections, taking into account both model performance (e.g., Knutti et al. 2010) and model interdependence (e.g., Sanderson et al. 2015), and more generally the statistical interpretation of those ensembles for projections (e.g., Tebaldi and Knutti 2007).

DECADAL CLIMATE PREDICTION. Decadal experiments investigate the predictability and prediction of climate on decadal time scales, starting from observations at a specific time to forecasting the evolution of natural, unforced variability in combination with the response to changes in radiative forcing (Meehl et al. 2009). On time scales of a decade or shorter, the influence of natural variability on the model climate tends to be larger than the response to changes in radiative forcing (Hawkins and Sutton 2009), especially at space scales smaller than hemispheric.

Efforts to estimate what aspects of the climate are predictable 1–10 years into the future and in what regions are areas of current and active research. Many fundamental questions arise in such efforts regarding, for example, how to initialize the climate models, how to determine skill, and how to correct for biases found in the model results (Meehl and Teng 2014; Meehl et al. 2009). Predictability on decadal time scales can arise from the large thermal inertia of the ocean, but attempts to assess the potential information encoded in the state of the ocean are hampered by limitations in the ocean-observing system, hindering model initialization and therefore the evaluation of model skill in the past. The assessment of the decadal climate prediction literature in Kirtman et al. (2013) indicates the most skill at predicting temperature anomalies over the North Atlantic and parts of the Pacific and Indian Oceans. Skill diminishes beyond a few years (Branstator and Teng 2012), over land regions, and for precipitation (Doblas-Reyes et al. 2013; Goddard et al. 2013; Hawkins and Sutton 2011).

In looking toward CMIP6, survey respondents asserted that there should be coordinated experiments to better address our understanding of mechanisms that produce decadal climate variability. However, a lack of historical ocean observations for testing and evaluating models will continue to make it challenging to do so until a more comprehensive and higher-quality observational record is available (Balmaseda et al. 2013).

Another challenge is the surprisingly large computational requirements of decadal prediction. To evaluate model predictions of the future, many integrations of past conditions are needed (hindcasts). In CMIP5, it was recommended that the models be initialized every year starting in 1960 with at least 10 member ensembles for each initial year. This computational demand poses additional challenges and makes coordinated and idealized experiments targeted toward these key issues, without adding to the computational burdens, a paramount concern for CMIP6.

**SUMMARY AND DISCUSSION.** A perusal of literature based on climate results reveals the impact and success of the Coupled Model Intercomparison Project (CMIP) over the last 20 years. As a largely self-organized activity within the international framework of the World Climate Research Programme (WCRP), CMIP has helped improve our understanding of climate variability and change, and it has made state-of-the-art climate model simulations directly available to a broad international community of climate scientists and impacts researchers. Analyses derived from the CMIP multimodel database have been prominently featured in past IPCC assessment reports and various national assessments.

CMIP5 incorporated new paradigms for developing future emission scenarios, introduced experiments to explore carbon (i.e., land and ocean biogeochemical)–climate interactions, and used high-resolution atmosphere-only models to provide more detailed regional climate change information. The future climate change problem was defined by time scale in terms of near-term (out to the mid-twenty-first century) and long-term (to the end of the twenty-first century and beyond). The near-term time frame incorporated experiments from the new field of decadal climate prediction to explore the limits of decadal prediction and predictability. Also, a large range of more idealized experiments was incorporated to help advance understanding of how the Earth system responds to various perturbations and thus helped assess confidence and improve understanding in different aspects of climate predictions. Finally, new diagnostic tools have been used to facilitate the evaluation of simulated physical processes in models (e.g., cloud simulators).

A survey of the various parts of the climate community subsequent to CMIP5 and the Fifth Assessment Report of the IPCC revealed a number of challenges and gaps in the CMIP5 process. For example, infrastructure issues highlighted by the survey are now under the purview of the WGCM Infrastructure Panel (WIP). The science gaps identified by this process have been highlighted here. An important and central finding of the survey and the process of consultation surrounding it is that future CMIP efforts should focus more strongly on specific science questions (Meehl et al. 2014) while continuing to make model output available to a broad scientific community.

Poor quantification and understanding of radiative forcing have been a long-standing problem within CMIP and will be revisited with new approaches in CMIP6, including more coordinated representation of the atmospheric aerosol and land

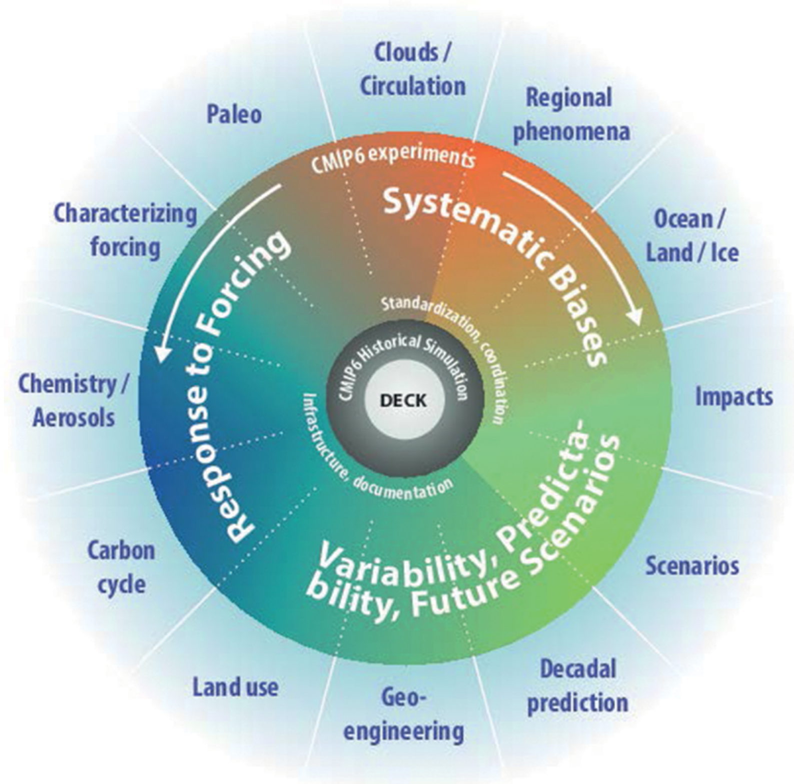


**Fig. 2. Schematic of the CMIP/CMIP6 experiment design. The inner ring and surrounding white text involve standardized functions of all CMIP Diagnostic, Evaluation and Characterization of Klima (DECK) experiments, and the CMIP6 historical simulation. The middle ring shows science topics related specifically to CMIP6 to be addressed by the CMIP6-endorsed MIPs, with MIP topics shown in the outer ring. This framework is superimposed onto the scientific backdrop for CMIP6, which includes the seven WCRP Grand Science Challenges. Figure taken from Eyring et al. (2016b).**

surface processes and much more detailed online and offline intercomparison activities. Likewise, long-standing model biases—such as the structure of circulation systems in the tropics, both in the atmosphere and ocean; the low cloud simulation over tropical oceans; and the position of the Southern Hemispheric jet—should receive increased attention in CMIP6, as well as inclusion of a richer palette of idealized experiments. The problems of comparing models to observations when the signals are small relative to the natural climatic variability, as well as trying to understand and predict the variability on longer than annual time scales (see the "How can we assess future climate changes given climate variability, predictability and uncertainties in scenarios?" section), will also be addressed in CMIP6.

The CMIP5 scientific gaps that were identified in a broad community survey and discussed here were considered in the new CMIP6 experimental design and scientific focus (Eyring et al. 2016b). CMIP6 will consist of individual CMIP6-endorsed MIPs that focus on specific scientific themes that are displayed in Fig. 2. These individual MIPs will build communities around the specific science questions developed within CMIP and tie modeling centers more closely to their specific science interests and the scientific applications of their output. A special issue in *Geoscientific Model Development* (*GMD*) provides detailed information on the new experiment design and organization of CMIP and the suite of CMIP6 experiments in a series of invited contributions. CMIP6 looks forward to building on the long tradition of excellent science in previous CMIP phases, but the ambitious plans of CMIP6 will only be realized through the committed participation and support of a very large community of scientists and their funders.

**ACKNOWLEDGMENTS.** We acknowledge the World Climate Research Programme's (WCRP's) Working Group on Coupled Modelling (WGCM), which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. National funding through the French CNRS, the German Ministry of Education and Research (BMBF), and the Max Planck Society has also made important infrastructural contributions to the success of CMIP. Portions of this study were supported by the Regional and Global Climate Modeling Program (RGCM) of the U.S. Department of Energy's Biological and Environmental Research (BER) program through Cooperative Agreement DE-FC02-97ER62402 to NCAR and under Contract DE-AC52-07NA27344 at LLNL), UK DECC/DEFRA Met Office Hadley Centre Programme (Grant GA01101) and CNRS, UPMC and Labex L-IPSL (Grant ANR-10-LABX-0018).

## REFERENCES


Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.*, **39**, L09712, doi:10.1029/2012GL051607.

Arora, V. K., and Coauthors, 2013: Carbon–concentration and carbon–climate feedbacks in CMIP5 Earth system models. *J. Climate*, **26**, 5289–5314, doi:10.1175/JCLI-D-12-00494.1.

Balmaseda, M. A., K. E. Trenberth, and E. Kallen, 2013: Distinctive climate signals in reanalysis of global ocean heat content. *Geophys. Res. Lett.*, **40**, 1754–1759, doi:10.1002/grl.50382.

Bodas-Salcedo, A., and Coauthors, 2011: COSP: Satellite simulation software for model assessment. *Bull. Amer. Meteor. Soc.*, **92**, 1023–1043, doi:10.1175/2011BAMS2856.1.

Boucher, O., and Coauthors, 2013: Clouds and aerosols. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 571–657.

Braconnot, P., S. P. Harrison, M. Kageyama, P. J. Bartlein, V. Masson-Delmotte, A. Abe-Ouchi, B. Otto-Bliesner, and Y. Zhao, 2012: Evaluation of climate models using palaeoclimatic data. *Nat. Climate Change*, **2**, 417–424, doi:10.1038/nclimate1456.

Branstator, G., and H. Y. Teng, 2012: Potential impact of initialization on decadal predictions as assessed for CMIP5 models. *Geophys. Res. Lett.*, **39**, L12703, doi:10.1029/2012GL051974.

Cheruy, F., J. L. Dufresne, F. Hourdin, and A. Ducharne, 2014: Role of clouds and land-atmosphere coupling in midlatitude continental summer warm biases and climate change amplification in CMIP5 simulations. *Geophys. Res. Lett.*, **41**, 6493–6500, doi:10.1002/2014GL061145.

Chung, E.-S., and B. J. Soden, 2015: An assessment of direct radiative forcing, radiative adjustments, and radiative feedbacks in coupled ocean–atmosphere models. *J. Climate*, **28**, 4152–4170, doi:10.1175/JCLI-D-14-00436.1.

Ciais, P., and Coauthors, 2013: Carbon and other biogeochemical cycles. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 465–570.

Collins, M., and Coauthors, 2013: Long-term climate change: Projections, commitments and irreversibility. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 1029–1136.

Cox, P. M., D. Pearson, B. B. Booth, P. Friedlingstein, C. Huntingford, C. D. Jones, and C. M. Luke, 2013: Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, **494**, 341–344, doi:10.1038/nature11882.

Crueger, T., and B. Stevens, 2015: The effect of atmospheric radiative heating by clouds on the Madden-Julian Oscillation. *J. Adv. Model. Earth Syst.*, **7**, 854–864, doi:10.1002/2015MS000434.

Doblas-Reyes, F. J., and Coauthors, 2013: Initialized near-term regional climate change prediction. *Nat. Commun.*, **4**, 1715, doi:10.1038/ncomms2704.

Eyring, V., and Coauthors, 2013: Long-term ozone changes and associated climate impacts in CMIP5 simulations. *J. Geophys. Res. Atmos.*, **118**, 5029–5060, doi:10.1002/jgrd.50316.

——, and Coauthors, 2016a: ESMValTool (v1.0)—A community diagnostic and performance metrics tool for routine evaluation of Earth system models in CMIP. *Geosci. Model Dev.*, **9**, 1747–1802, doi:10.5194/gmd-9-1747-2016.

——, S. Bony, G. A. Meehl, C. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016b: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, doi:10.5194/gmd-9-1937-2016.

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 741–866.

Friedlingstein, P., M. Meinshausen, V. K. Arora, C. D. Jones, A. Anav, S. K. Liddicoat, and R. Knutti, 2014: Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *J. Climate*, **27**, 511–526, doi:10.1175/JCLI-D-12-00579.1.

Frolicher, T. L., J. L. Sarmiento, D. J. Paynter, J. P. Dunne, J. P. Krasting, and M. Winton, 2015: Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models. *J. Climate*, **28**, 862–886, doi:10.1175/JCLI-D-14-00117.1.

Gates, W. L., 1992: AMIP—The Atmospheric Model Intercomparison Project. *Bull. Amer. Meteor. Soc.*, **73**, 1962–1970, doi:10.1175/1520-0477(1992)073<1962:ATAMIP>2.0.CO;2.

——, and Coauthors, 1999: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull. Amer. Meteor. Soc.*, **80**, 29–55, doi:10.1175/1520-0477(1999)080<0029:AOOTRO>2.0.CO;2.

Ghan, S. J., 2013: Technical note: Estimating aerosol effects on cloud radiative forcing. *Atmos. Chem. Phys.*, **13**, 9971–9974, doi:10.5194/acp-13-9971-2013.

Gleckler, P. J., C. Doutriaux, P. J. Durack, K. E. Taylor, Y. Zhang, D. N. Williams, E. Mason, and J. Servonnat, 2016: A more powerful reality test for climate models.

*Eos, Trans. Amer. Geophys. Union*, **97**, doi:10.1029/2016EO051663.

Goddard, L., and Coauthors, 2013: A verification framework for interannual-to-decadal predictions experiments. *Climate Dyn.*, **40**, 245–272, doi:10.1007/s00382-012-1481-2.

Gregory, J. M., and Coauthors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, L03205, doi:10.1029/2003GL018747.

Hall, A., and X. Qu, 2006: Using the current seasonal cycle to constrain snow albedo feedback in future climate change. *Geophys. Res. Lett.*, **33**, L03502, doi:10.1029/2005GL025127.

Hansen, J., and Coauthors, 2005: Efficacy of climate forcings. *J. Geophys. Res.*, **110**, D18104, doi:10.1029/2005JD005776.

Hawkins, E., and R. Sutton, 2009: Decadal predictability of the Atlantic Ocean in a Coupled GCM: Forecast skill and optimal perturbations using linear inverse modeling. *J. Climate*, **22**, 3960–3978, doi:10.1175/2009JCLI2720.1.

——, and ——, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dyn.*, **37**, 407–418, doi:10.1007/s00382-010-0810-6.

Held, I. M., and B. J. Soden, 2006: Robust responses of the hydrological cycle to global warming. *J. Climate*, **19**, 5686–5699, doi:10.1175/JCLI3990.1.

Hibbard, K. A., G. A. Meehl, P. M. Cox, and P. Friedlingstein, 2007: A strategy for climate change stabilization experiments. *Eos, Trans. Amer. Geophys. Union*, **88**, 217–221, doi:10.1029/2007EO200002.

Hourdin, F., A. Găinuşă-Bogdan, P. Braconnot, J.-L. Dufresne, A.-K. Traore, and C. Rio, 2015: Air moisture control on ocean surface temperature, hidden key to the warm bias enigma. *Geophys Res Lett.*, **42**, 10 885–10 893, doi:10.1002/2015GL066764.

Kim, D., and Coauthors, 2014: Process-oriented MJO simulation diagnostic: Moisture sensitivity of simulated convection. *J. Climate*, **27**, 5379–5395, doi:10.1175/JCLI-D-13-00497.1.

Kirtman, B., and Coauthors, 2013: Near-term climate change: Projections and predictability. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 953–1028.

Klein, S. A., X. Jiang, J. Boyle, S. Malyshev, and S. Xie, 2006: Diagnosis of the summertime warm and dry bias over the U.S. Southern Great Plains in the GFDL climate model using a weather forecasting approach. *Geophys. Res. Lett.*, **33**, L18805, doi:10.1029/2006GL027567.

Knutti, R., G. Abramowitz, M. Collins, V. Eyring, P. J. Gleckler, B. Hewitson, and L. Mearns, 2010: Good practice guidance paper on assessing and combining multi model climate projections. Meeting report of the Intergovernmental Panel on Climate Change Expert Meeting on Assessing and Combining Multi Model Climate Projections, T. Stocker et al., Eds., Rep. 0165-0009, IPCC Working Group I Technical Support Unit, 1–15.

Kriegler, E., B. C. O'Neill, S. Hallegatte, T. Kram, R. J. Lempert, R. H. Moss, and T. Wilbanks, 2012: The need for and use of socio-economic scenarios for climate change analysis: A new approach based on shared socio-economic pathways. *Global Environ. Change*, **22**, 807–822, doi:10.1016/j.gloenvcha.2012.05.005.

Lamarque, J.-F., and Coauthors, 2013: The Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP): Overview and description of models, simulations and climate diagnostics. *Geosci. Model Dev.*, **6**, 179–206, doi:10.5194/gmd-6-179-2013.

Li, G., and S. P. Xie, 2012: Origins of tropical-wide SST biases in CMIP multi-model ensembles. *Geophys. Res. Lett.*, **39**, L22703, doi:10.1029/2012GL053777.

Meehl, G. A., and H. Teng, 2014: CMIP5 multi-model hindcasts for the mid-1970s shift and early 2000s hiatus and predictions for 2016–2035. *Geophys Res Lett.*, **41**, 1711–1716, doi:10.1002/2014GL059256.

——, G. J. Boer, C. Covey, M. Latif, and R. J. Stouffer, 1997: Intercomparison makes for a better climate model. *Eos, Trans. Amer. Geophys. Union*, **78**, 445–451, doi:10.1029/97EO00276.

——, ——, ——, ——, and ——, 2000: The Coupled Model Intercomparison Project (CMIP). *Bull. Amer. Meteor. Soc.*, **81**, 313–318, doi:10.1175/1520-0477(2000)081<0313:TCMIPC>2.3.CO;2.

——, C. Covey, B. McAvaney, M. Latif, and R. J. Stouffer, 2005: Overview of the Coupled Model Intercomparison Project. *Bull. Amer. Meteor. Soc.*, **86**, 89–93, doi:10.1175/BAMS-86-1-89.

——, K. E. Taylor, T. Delworth, R. J. Stouffer, M. Latif, B. McAvaney, and J. F. B. Mitchell, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394, doi:10.1175/BAMS-88-9-1383.

——, and Coauthors, 2009: Decadal prediction: Can it be skillful? *Bull. Amer. Meteor. Soc.*, **90**, 1467–1485, doi:10.1175/2009BAMS2778.1.

——, R. Moss, K. E. Taylor, V. Eyring, R. J. Stouffer, S. Bony, and B. Stevens, 2014: Climate model intercomparisons: Preparing for the next phase. *Eos, Trans. Amer. Geophys. Union*, **95**, 77, doi:10.1002/2014EO090001.

Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature*, **463**, 747–756, doi:10.1038/nature08823.

Mueller, B., and S. I. Seneviratne, 2014: Systematic land climate and evapotranspiration biases in CMIP5 simulations. *Geophys. Res. Lett.*, **41**, 128–134, doi:10.1002/2013GL058055.

Myhre, G., and Coauthors, 2013: Anthropogenic and natural radiative forcing. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 659–740.

Oueslati, B., and G. Bellon, 2015: The double ITCZ bias in CMIP5 models: Interaction between SST, large-scale circulation and precipitation. *Climate Dyn.*, **44**, 585–607, doi:10.1007/s00382-015-2468-6.

Pongratz, J., C. H. Reick, R. A. Houghton, and J. I. House, 2014: Terminology as a key uncertainty in net land use and land cover change carbon flux estimates. *Earth Syst. Dyn.*, **5**, 177–195, doi:10.5194/esd-5-177-2014.

Randall, D. A., and Coauthors, 2007: Cilmate models and their evaluation. *Climate Change 2007: The Physical Science Basis*, S. Solomon et al., Eds, Cambridge University Press, 589–662.

Rauser, F., P. Gleckler, and J. Marotzke, 2015: Rethinking the default construction of multimodel climate ensembles. *Bull. Amer. Meteor. Soc.*, **96**, 911–919, doi:10.1175/BAMS-D-13-00181.1.

Riahi, K., and Coauthors, 2017: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environ. Change*, doi:10.1016/j.gloenvcha.2016.05.009, in press.

Russell, J. L., K. W. Dixon, A. Gnanadesikan, R. J. Stouffer, and J. R. Toggweiler, 2006a: The Southern Hemisphere westerlies in a warming world: Propping open the door to the deep ocean. *J. Climate*, **19**, 6382–6390, doi:10.1175/JCLI3984.1.

——, R. J. Stouffer, and K. W. Dixon, 2006b: Intercomparison of the Southern Ocean circulations in IPCC coupled model control simulations. *J. Climate*, **19**, 4560–4575, doi:10.1175/JCLI3869.1; Corrigendum, **20**, 4287, doi:10.1175/JCLI4326.1.

Sanderson, B. M., R. Knutti, and P. Caldwell, 2015: Addressing interdependency in a multimodel ensemble by interpolation of model properties. *J. Climate*, **28**, 5150–5170, doi:10.1175/JCLI-D-14-00361.1.

Santer, B. D., K. E. Taylor, T. M. L. Wigley, J. E. Penner, P. D. Jones, and U. Cubasch, 1995: Towards the detection and attribution of an anthropogenic effect on climate. *Climate Dyn.*, **12**, 77–100, doi:10.1007/BF00223722.

Sherwood, S. C., S. Bony, and J. L. Dufresne, 2014: Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, **505**, 37–42, doi:10.1038/nature12829.

——, ——, O. Boucher, C. Bretherton, P. M. Forster, J. M. Gregory, and B. Stevens, 2015: Adjustments in the forcing-feedback framework for understanding climate change. *Bull. Amer. Meteor. Soc.*, **96**, 217–228, doi:10.1175/BAMS-D-13-00167.1.

Smith, D. M., and Coauthors, 2013: Real-time multimodel decadal climate predictions. *Climate Dyn.*, **41**, 2875–2888, doi:10.1007/s00382-012-1600-0.

Stevens, B., 2015: Rethinking the lower bound on aerosol radiative forcing. *J. Climate*, **28**, 4794–4819, doi:10.1175/JCLI-D-14-00656.1.

——, and S. Bony, 2013: What are climate models missing? *Science*, **340**, 1053–1054, doi:10.1126/science.1237554.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tebaldi, C., and R. Knutti, 2007: The use of the multimodel ensemble in probabilistic climate projections. *Philos Trans. Roy. Soc. London*, **365A**, 2053–2075, doi:10.1098/rsta.2007.2076.

Teixeira, J., D. Waliser, R. Ferraro, P. Gleckler, T. Lee, and G. Potter, 2014: Satellite observations for CMIP5: The genesis of Obs4MIPs. *Bull. Amer. Meteor. Soc.*, **95**, 1329–1334, doi:10.1175/BAMS-D-12-00204.1.

Wenzel, S., P. M. Cox, V. Eyring, and P. Friedlingstein, 2014: Emergent constraints on climate-carbon cycle feedbacks in the CMIP5 Earth system models. *J. Geophys. Res. Biogeosci.*, **119**, 794–807, doi:10.1002/2013JG002591.

Williams, D. N., and Coauthors, 2015: A global repository for planet-sized experiments and observations. *Bull. Amer. Meteor. Soc.*, **97**, 803–816, doi:10.1175/BAMS-D-15-00132.1.

Williamson, D., and Coauthors, 2013: The Aqua-Planet Experiment (APE): Response to changed meridional SST profile. *J. Meteor. Soc. Japan*, **91A**, 57–89, doi:10.2151/jmsj.2013-A03.

——, A. T. Blaker, C. Hampton, and J. Salter, 2015: Identifying and removing structural biases in climate models with history matching. *Climate Dyn.*, **45**, 1299–1324, doi:10.1007/s00382-014-2378-z.

Yu, W., M. Doutriaux, G. Seze, H. LeTreut, and M. Desbois, 1996: A methodology study of the validation of clouds in GCMs using ISCCP satellite observations. *Climate Dyn.*, **12**, 389–401, doi:10.1007/BF00211685.