Geosci. Model Dev. Discuss., 8, 10539–10583, 2015
www.geosci-model-dev-discuss.net/8/10539/2015/
doi:10.5194/gmdd-8-10539-2015

Geoscientific Model Development Discussions

Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation

V. Eyring[1], S. Bony[2], G. A. Meehl[3], C. Senior[4], B. Stevens[5], R. J. Stouffer[6], and K. E. Taylor[7]

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2]Laboratoire de Météorologie Dynamique, Institut Pierre Simon Laplace (LMD/IPSL), CNRS, Université Pierre et Marie Curie, Paris, France
[3]National Center for Atmospheric Research (NCAR), Boulder, USA
[4]Met Office Hadley Centre, Exeter, UK
[5]Max-Planck-Institute for Meteorology, Hamburg, Germany
[6]Geophysical Fluid Dynamics Laboratory/NOAA, Princeton, NJ, USA
[7]Program for Climate Model Diagnosis and Intercomparison (PCMDI), Lawrence Livermore National Laboratory, Livermore, CA, USA

Received: 3 December 2015 – Accepted: 4 December 2015 – Published: 14 December 2015

Correspondence to: V. Eyring (veronika.eyring@dlr.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

By coordinating the design and distribution of global climate model simulations of the past, current and future climate, the Coupled Model Intercomparison Project (CMIP) has become one of the foundational elements of climate science. However, the need to address an ever-expanding range of scientific questions arising from more and more research communities has made it necessary to revise the organization of CMIP. After a long and wide community consultation, a new and more federated structure has been put in place. It consists of three major elements: (1) a handful of common experiments, the DECK (Diagnostic, Evaluation and Characterization of Klima experiments) and the CMIP Historical Simulation (1850–near-present) that will maintain continuity and help document basic characteristics of models across different phases of CMIP, (2) common standards, coordination, infrastructure and documentation that will facilitate the distribution of model outputs and the characterization of the model ensemble, and (3) an ensemble of CMIP-Endorsed Model Intercomparison Projects (MIPs) that will be specific to a particular phase of CMIP (now CMIP6) and that will build on the DECK and the CMIP Historical Simulation to address a large range of specific questions and fill the scientific gaps of the previous CMIP phases. The DECK and CMIP Historical Simulation, together with the use of CMIP data standards, will be the entry cards for models participating in CMIP. The participation in the CMIP6-Endorsed MIPs will be at the discretion of the modelling groups, and will depend on scientific interests and priorities. With the Grand Science Challenges of the World Climate Research Programme (WCRP) as its scientific backdrop, CMIP6 will address three broad questions: (i) how does the Earth system respond to forcing?, (ii) what are the origins and consequences of systematic model biases?, and (iii) how can we assess future climate changes given climate variability, predictability and uncertainties in scenarios? This CMIP6 overview paper presents the background and rationale for the new structure of CMIP, provides a detailed description of the DECK and the CMIP6 Historical Simulation, and includes a brief introduction to the 21 CMIP6-Endorsed MIPs.

1 Introduction

The Coupled Model Intercomparison Project (CMIP) organized under the auspices of the World Climate Research Programme's (WCRP) Working Group on Coupled Modelling (WGCM) started twenty years ago as a comparison of a handful of early global coupled climate models performing experiments using atmosphere models coupled to a dynamic ocean, a simple land surface, and thermodynamic sea ice (Meehl et al., 1997). It has since evolved over five phases into a major international multi-model research activity (Meehl et al., 2000, 2007; Taylor et al., 2012) that has not only introduced a new era to climate science research, but has also become a central element of national and international assessments of climate change (IPCC, 2013). An important part of CMIP is to make the multi-model output publically available in a standardized format for analysis by the wider climate community and users. The standardization of the model output in a specified format, and the collection, archival, and access of the model output through the Earth System Grid Federation (ESGF) data replication centres have facilitated multi-model analyses.

The objective of CMIP is to better understand past, present and future climate changes arising from natural, unforced variability or in response to changes in radiative forcings in a multi-model context. Its increasing importance and scope is a tremendous success story, but this very success poses challenges for all involved. Coordination of the project has become more complex as CMIP includes more models with more processes all applied to a wider range of questions. To meet this new interest and to address a wide variety of science questions from more and more scientific research communities, reflecting the expanding scope of comprehensive modelling in climate science, has put pressure on CMIP to become larger and more extensive. Consequently, there has been an explosion in the diversity and volume of requested CMIP output from an increasing number of experiments causing challenges for CMIP's technical infrastructure (Williams et al., 2015). Cultural and organizational challenges also arise from the tension between expectations that modelling centres deliver multiple model experiments to CMIP yet at the same time advance basic research in climate science.

In response to these challenges, we have adopted a more federated structure for the sixth phase of CMIP (i.e., CMIP6) and subsequent phases. Whereas past phases of CMIP were usually described through a single overview paper, reflecting a centralized and relatively compact CMIP structure, this GMD Special Issue describes the new design and organization of CMIP, the suite of experiments, and its forcings, in a series of invited contributions. In this paper, we provide the overview and backdrop of the new CMIP structure as well as the main scientific foci that CMIP6 will address. We begin by describing the new organizational form for CMIP and the pressures that it was designed to alleviate (Sect. 2). It also contains a description of a small set of simulations for CMIP which are intended to be common to all participating models (Sect. 3), details of which are provided in an Appendix. We then present a brief overview of CMIP6 serving as an introduction to the other contributions to this Special Issue (Sect. 4) and close with a summary.

2 CMIP design – a more continuous and distributed organization

In preparing for the sixth phase of CMIP, the CMIP Panel (the authors of this paper), which traditionally has the responsibility for direct coordination and oversight of CMIP, initiated an extensive process of community consultation that spanned two years. This consultation not only involved the modelling centres whose contributions form the substance of CMIP, but also the communities that rely on CMIP model output for their work. This consultation included the organization of special meetings to reflect on the success and science gaps that emerged from CMIP5. It also sought input through a community survey, the scientific results of which are described by Stouffer et al. (2015). This process identified four main issues related to the overall structure of CMIP.

First, there was increasing recognition that results from different CMIP phases should not necessarily be used in isolation. For instance, as pointed out by Rauser

et al. (2014), a common set of simulations across different phases of CMIP could be used to construct more rational ensembles than just a single phase ensemble. At the same time it was recognized that an increasing number of Model Intercomparison Projects (MIPs) were being organized independent of CMIP, the data structure and output requirements were often inconsistent, and the relationship between the models used in the various MIPs was often difficult to determine.

Second, the scope of CMIP was taxing the resources of modelling centres making it impossible to consider contributing to all the proposed experiments. At the same time, relatively little guidance was available to help modelling centres decide exactly which subset of experiments to perform. This led to a more fragmented participation in CMIP5, as compared to earlier phases and created the impression that perhaps modelling centre resources were being used sub-optimally. In addition, a monolithic structure to the CMIP design tended to discourage the modelling centres from attempting to design new experiments meant to address specific scientific questions of interest to them. This in turn contributed to the impression that CMIP was a service that the modelling centres provided to the broader community.

Third, the punctuated structure of CMIP has begun to distort the model development process. Whereas in the past modelling centres developed models based on their own scientific goals and released model versions on their own schedule, the visibility and demands of CMIP were beginning to impose a synchronization of model development with different phases of CMIP. Though this might have seemed desirable to those who relied on the different phases of CMIP to measure and pace progress in model development, such a view poorly reflected the reality of model development. The resulting stress this placed on modelling centres often resulted in unnecessary delays in the provision of model output, as it was conditioned on the finalization of new model releases and risked redefining the role of modelling centres in ways that did not best exploit their scientific potential.

Fourth and finally, there was the desire for particular phases of CMIP to be more than just a collection of MIPs, but rather to reflect the strategic goals of the climate science community as developed through the WCRP. By focusing a particular phase of CMIP around a few scientific issues it was felt that the modelling centres could better contribute to selected scientific questions that had matured to a point where coordinated activities could have substantial impact, thereby more rapidly advancing the science and promoting a cohesive strategy across WCRP.

A variety of mechanisms were proposed and intensely debated to address these issues. The outcome of these discussions is embodied in the new CMIP structure, which has three major components. First, the identification of a handful of common experiments, the DECK (Diagnostic, Evaluation and Characterization of Klima experiments) and the CMIP Historical Simulation, which can be used to establish model characteristics and serves as its "entry card" for participating in one of CMIP's phases or in other MIPs organized between CMIP phases, as depicted in Fig. 1. Second, common standards, coordination, infrastructure and documentation that facilitate the distribution of model outputs and the characterization of the model ensemble, and third, the adoption of a more federated structure, building on more autonomous CMIP-Endorsed MIPs.

Realising the idea of a particular phase of CMIP being centred on a collection of more autonomous MIPs required the development of procedures for soliciting and evaluating MIPs, in light of the scientific focus chosen for CMIP6. These procedures were developed and implemented by the CMIP Panel. The responses to the CMIP5 survey helped inform a series of workshops and resulted in a draft experiment design for CMIP6. This initial design for CMIP6 was published in early 2014 (Meehl et al., 2014) and was open for comments from the wider community until mid-September 2014. In parallel to the open review of the design, the CMIP Panel distributed an open call for proposals for MIPs in April 2014. These proposals were broadly reviewed within WCRP with the goal to encourage and enhance synergies among the different MIPs, to avoid overlapping experiments, to fill gaps, and to help ensure that the WCRP Grand Science Challenges would be addressed. Revised MIP proposals were requested and evaluated by the CMIP Panel in summer 2015. The selection of MIPs was based on the CMIP Panel's evaluation of ten endorsement criteria (Table 1). To ensure community engagement,

an important criterion was that enough modelling groups (at least eight) were willing to perform all of the MIP's highest priority (Tier 1) experiments and providing all the requested diagnostics needed to answer at least one of its leading science questions. For each of the selected CMIP6-Endorsed MIPs it turned out that at least ten modelling groups indicated their intent to participate in at least Tier 1 experiments, thus attesting to the wide appeal and level of science interest from the climate modelling community.

3 The DECK and CMIP Historical Simulation

The DECK comprises four baseline experiments: (a) a historical Atmospheric Model Intercomparison Project (AMIP) simulation, (b) a pre-industrial control simulation (*piControl*), (c) a simulation forced by an abrupt quadrupling of CO_2 (*abrupt4×CO2*) and (d) a simulation forced by a $1\,\%\,yr^{-1}$ CO_2 increase (*1pctCO2*). CMIP also includes a Historical Simulation that spans the period of extensive instrumental temperature measurements from 1850 to the present.

Under CMIP, the credentials of the participating atmospheric-ocean general circulation models (AOGCMs) and Earth System Models (ESMs) are established by performing the DECK and the CMIP Historical Simulation, so these experiments are required from all models. They should be run for each model configuration used in a CMIP-Endorsed MIP. A change in "model configuration" includes any change that might affect its simulations other than "noise" expected from different realizations. This would include, for example, a change in model resolution, physical processes, or atmospheric chemistry treatment. If a model is run in emission-driven mode as part of a CMIP6-Endorsed MIP (e.g., C^4MIP) then the *piControl* and the CMIP Historical Simulation should be run in emission-driven mode too, whereas the *amip*, *1pctCO2* and *abrupt4×CO2* simulations would also be required as part of the DECK, but they would be driven by prescribed CO_2 concentrations. Similarly, if a model is used in both emission-driven and concentration-driven mode in subsequent CMIP6-Endorsed MIPs, then both emission-driven and concentration-driven piControl and Historical Simulations should be done.

The forcings that are used in the DECK and in the CMIP6 Historical Simulation are described separately in this Special Issue, and are provided by the respective author teams. These include: (1) historical emissions of short-lived species and long-lived greenhouse gases (GHGs), (2) historical GHG concentrations, (3) global gridded land-use forcing datasets, (4) solar forcing, (5) stratospheric aerosol dataset (volcanoes), and (6) AMIP sea surface temperatures (SSTs) and sea-ice concentrations (SICs). In addition for simulations with prescribed aerosols (7) a new approach to prescribe aerosols in terms of optical properties and fractional change in cloud droplet effective radius to provide a more consistent representation of aerosol forcing, and for models without ozone chemistry (8) time-varying gridded ozone concentrations are provided.

An overview of the main characteristics of the DECK and the CMIP6 Historical Simulation is given in Table 2. Here we briefly describe these experiments. Detailed specifications for the DECK and the CMIP6 Historical Simulation are provided in Appendix A1 and A2, respectively, and are summarized in Table A1.

3.1 The DECK

The AMIP and pre-industrial control simulations of the DECK provide opportunities for evaluating the atmospheric model and the coupled system, and in addition they establish a baseline for performing many of the CMIP6 experiments. Many experiments branch from, and are compared with, the pre-industrial control. Similarly, a number of diagnostic atmospheric experiments use *amip* as a control. The idealized CO_2-forced experiments in the DECK (1 % yr^{-1} CO_2 and abrupt $4xCO_2$ increases), despite their simplicity, can reveal fundamental forcing and feedback response characteristics of models. The experiments of the DECK are already commonly performed by modelling groups in the course of developing a new model, and their protocols are expected to remain essentially unchanged for many years to come. The persistence and consistency of these experiments will make it possible to track over future generations of models

any changes in performance and response characteristics. Although this core set of experiments is not expected to evolve much, additional experiments may become well enough established as benchmarks (routinely run by modelling groups as they develop new model versions) so that in the future they might be migrated into the DECK. The common practice of including the DECK in model development efforts means that models can contribute to CMIP without carrying out additional computationally burdensome experiments. All experiments of the DECK were included in the core set performed under CMIP5 (Taylor et al., 2012), and all but the *abrupt4×CO2* simulation were included in even earlier CMIP phases.

For nearly three decades, AMIP simulations (Gates et al., 1999) have been routinely relied on by modelling centres to help in the evaluation of the atmospheric component of their models. In AMIP simulations, the SSTs and SICs are prescribed based on observations. The idea is to analyse and evaluate the atmospheric and land components of the climate system when they are constrained by the observed ocean conditions. These simulations can help identify which model errors originate in the atmosphere or land components, and they have proven useful in addressing a great variety of questions pertaining to recent climate changes. The AMIP simulations performed as part of the DECK cover at least the period from January 1979 to December 2014. The end date will continue to evolve as the SSTs and SICs are updated with new observations. Besides prescription of ocean conditions in these simulations, realistic forcings are imposed that should be identical to those applied in the CMIP Historical Simulation. Large ensembles of AMIP simulations are encouraged as they can help separate the signal of forced responses (Li et al., 2015).

The remaining three experiments in the DECK are premised on the coupling of the atmospheric and oceanic circulation. The pre-industrial control simulation (*piControl*) is performed under conditions chosen to be representative of the period prior to the onset of large-scale industrialization with 1850 being the reference year. There are no secular changes in forcing, so the concentrations and/or sources of atmospheric constituents (e.g., GHGs and other forcings) are held fixed, as are Earth's orbital characteristics.

External human influences on the land surface are likewise excluded. Because of the absence of both naturally occurring changes in forcing (e.g., volcanoes, orbital or solar changes) and human-induced changes, the *piControl* simulation gives insight into the unforced internal variability of the climate system.

An initial climate "spin-up" portion of a control simulation, during which the climate begins to come into balance with the forcing, is usually performed and discarded. The length of this "spin-up" period is model and resource dependent. At the end of the "spin-up" period, the *piControl* starts. The *piControl* serves as a baseline for experiments that branch from it. To account for the effects of any residual drift, it is required that the *piControl* simulation extends as far beyond the branching point as any experiment to which it will be compared. Only then can residual climate drift in an experiment be removed, so that it is not misinterpreted as part of the model's forced response. The recommended minimum length for the *piControl* is 500 years.

The two DECK "climate change" experiments branch from some point in the *piControl* and are designed to document basic aspects of the climate system response to GHG forcing. In the first, the CO_2 concentration is immediately and abruptly quadrupled. This *abrupt4×CO2* simulation has proven to be useful for characterizing the radiative forcing that arises from an increase in atmospheric CO_2 as well as changes that arise indirectly due to the warming. It can also be used to estimate a model's equilibrium climate sensitivity (ECS, Gregory et al., 2004). In the second, the CO_2 concentration is increased gradual at a rate of 1 % per year. This experiment has been performed in all phases of CMIP since CMIP2, and serves as a consistent and useful benchmark for analysing model transient climate response (TCR). The TCR takes into account the rate of ocean heat uptake which governs the pace of all time-evolving climate change (e.g., Murphy and Mitchell, 1995). In addition to the TCR, the 1 % CO_2 integration with ESMs that include explicit representation of the carbon cycle allows the calculation of the transient climate response to cumulative carbon emissions (TCRE) defined as the transient global average surface temperature change per unit of accumulated CO_2 emissions (IPCC, 2013). Despite their simplicity, these experiments

provide a surprising amount of insight into the behaviour of models subject to more complex forcing (e.g., Bony et al., 2013; Geoffroy et al., 2013).

3.2 CMIP Historical Simulation

In addition to the DECK, CMIP challenges models to simulate the historical period, defined to begin in 1850 and extend to the near present (i.e., 2014 in CMIP6). The CMIP Historical Simulation branches from the *piControl* and is forced, based on observations, by evolving, externally-imposed forcings such as solar variability, volcanic aerosols, and changes in atmospheric composition (GHGs, and aerosols) caused by human activities. The CMIP Historical Simulation provides rich opportunities to assess model ability to simulate climate, including variability and century time-scale trends (e.g., Flato et al., 2013), and it has also proven essential in reducing uncertainty in radiative forcing associated with short lived species such as the atmospheric aerosol (e.g., Stevens, 2015). When supplemented with additional experiments, the Historical Simulation can be used in detection and attribution studies (e.g., Stott et al., 2006) to help interpret the extent to which observed climate change can be explained by different causes.

As in the *piControl* simulation, models that include representation of the carbon cycle should normally perform two different CMIP Historical Simulations: a prescribed CO_2 concentration and a prescribed emissions simulation (accounting explicitly for fossil fuel combustion), in which concentrations are then "predicted" by the model. Both types of simulation are useful in evaluating how realistically the model represents the carbon cycle, but the prescribed concentration simulation enables these more complex models to be evaluated fairly against those simpler models without representation of carbon cycle processes.

3.3 Common standards, infrastructure and documentation

A key to the success of CMIP and one of the motivations for incorporating a wide variety of coordinated modelling activities under a single framework in a specific phase of CMIP (now CMIP6) is the desire to reduce duplication of effort, minimize operational and computational burdens, and establish common practices in producing and analysing large amounts of model output. To enable automated processing of output from dozens of different models, CMIP has led the way in encouraging adoption of data standards (governing structure and metadata) that facilitate development of software infrastructure in support of coordinated modelling activities. The ESGF has capitalized on this standardization to provide access to CMIP model output hosted by institutions around the world. As the complexity of CMIP has increased and as the potential use of model output expands beyond the research community, the evolution of the climate modelling infrastructure requires enhanced coordination. To help in this regard, the WGCM Infrastructure Panel (WIP) was set up (see details in the corresponding contribution to this Special Issue), and is now providing guidance on requirements and establishing specifications for model output, model and simulation documentation, and archival and delivery systems for CMIP6 data.

A more routine benchmarking and evaluation of the models is envisaged to be a central part of CMIP6. As noted above, one purpose of the DECK and the CMIP Historical Simulation is to provide a basis for documenting model simulation characteristics. Towards that end an infrastructure is being developed to allow analysis packages to be routinely executed whenever new model experiments are contributed to the CMIP archive. These efforts utilize observations served by the ESGF contributed from the obs4MIPs (Teixeira et al., 2014) and ana4MIPs projects. Examples of available tools that target routine evaluation in CMIP include the PCMDI metrics software (Gleckler et al., 2015) and the Earth System Model Evaluation Tool (ESMValTool, Eyring et al., 2015), which brings together established diagnostics such as those used in the evaluation chapter of IPCC AR5 (Flato et al., 2013). The ESMValTool also integrates

other packages, such as the NCAR Climate Variability Diagnostics Package (Phillips et al., 2014), or diagnostics such as the cloud regime metric (Williams and Webb, 2009) developed by the Cloud Feedback MIP (CFMIP) community. These tools can be used to assess new models, and can help inform users of model output, as well as the modelling centres, as to the strengths and weaknesses of the simulations, including the extent to which long-standing model errors remain evident in newer models. Building such a community-based capability is not meant to replace how CMIP research is currently performed but rather to complement it.

4 CMIP6

4.1 Scientific focus of CMIP6

In addition to the DECK and the CMIP Historical Simulation, a number of additional experiments will colour a specific phase of CMIP, now CMIP6. These experiments are likely to change from one CMIP phase to the next. To maximize the relevance and impact of CMIP6, it was decided to use the Grand Science Challenges (GCs) of the WCRP as the scientific backdrop of the CMIP6 experimental design. By promoting research on critical science questions for which specific gaps in knowledge have hindered progress so far, but for which new opportunities and more focused efforts raise the possibility of significant progress on the timescale of 5–10 years, these GCs constitute a main component of the WCRP strategy to accelerate progress in climate science (Brasseur and Carlson, 2015). Five such GCs have been identified, and two additional ones are under consideration. They relate to advancing (1) understanding of the role of clouds in the general atmospheric circulation and climate sensitivity (Bony et al., 2015), (2) assessing the response of the cryosphere to a warming climate and its global consequences, (3) understanding the factors that control water availability over land (Trenberth and Asrar, 2014), (4) assessing climate extremes, what controls them, how they have changed in the past and how they might change in the future

(Alexander et al., 2015), (5) understanding and predicting regional sea-level change and its coastal impacts, (6) improving near-term climate predictions, and (7) determining how biogeochemical cycles and feedbacks control greenhouse gas concentrations and climate change.

These GCs will be using the full spectrum of observation, modelling and analysis expertise across the WCRP, and in terms of modelling most GCs will address their specific science questions through a hierarchy of numerical models of different complexities. Global coupled models obviously constitute an essential element of this hierarchy, and CMIP6 experiments will play a prominent role across all GCs by helping to answer the three following CMIP6 science questions: How does the Earth system respond to forcing? What are the origins and consequences of systematic model biases? How can we assess future climate change given climate variability, climate predictability, and uncertainties in scenarios?

These three questions will be at the centre of CMIP6. They will be addressed through a range of CMIP6-Endorsed MIPs that are organized by the respective communities and overseen by the CMIP Panel (Fig. 2). Through these different MIPs and their connection to the GCs, the goal is to fill some of the main scientific gaps of previous CMIP phases. This includes in particular facilitating the identification and interpretation of model systematic errors, improving the estimate of radiative forcings in past and future climate change simulations, facilitating the identification of robust climate responses to aerosol forcing during the historical period, better taking into account the impact of short-term forcing agents and land-use on climate, better understanding the mechanisms of decadal climate variability, and many other issues that could not be addressed satisfactorily in CMIP5 (Stouffer et al., 2015). In endorsing a number of these MIPs the CMIP panel acted to minimize overlaps among the MIPs and to reduce the burden of modelling groups, while maximizing the scientific complementarity and synergy among the different MIPs.

4.2 The CMIP6-Endorsed MIPs

Close to thirty suggestions for CMIP6 MIPs were received of which 21 MIPs were eventually endorsed and invited to participate (Table 3). Of those not selected some were asked to work with other proposed MIPs with overlapping science goals and objectives. Of the 21 CMIP6-Endorsed MIPs, four are diagnostic in nature, which means that they define and analyse additional output, but do not require additional experiments. In the remaining 17 MIPs, a total of around 190 experiments have been proposed resulting in 40 000 model simulation years with around half of these in Tier 1. The CMIP-Endorsed MIPs show broad coverage and distribution across the three CMIP6 science questions and links to the WCRP Grand Science Challenges (Fig. 3).

Each of the 21 CMIP6-Endorsed MIPs will be described in a separate invited contribution to this Special Issue. These contributions will detail the goal of the MIP and the major scientific gaps the MIP is addressing, and will specify what is new compared to CMIP5 and previous CMIP phases. The contributions will include a description of the experimental design and scientific justification of each of the experiments for Tier 1 (and possibly beyond), and will link the experiments and analysis to the DECK and the CMIP6 Historical Simulation. They will additionally include an analysis plan to fully justify the resources used to produce the various requested variables, and if the analysis plan is to compare model results to observations, the contribution will highlight possible model diagnostics and performance metrics specifying whether the comparison entails any particular requirement for the simulations or outputs (e.g. the use of observational simulators). In addition, possible observations and reanalysis products for model evaluation will be discussed and the MIPs are encouraged to help contributing them to the obs4MIPs/ana4MIPs archives at the ESGF (see Sect. 3.3). In some MIPs additional forcings beyond those used in the DECK and CMIP6 Historical Simulation will be required, and these will be described in the respective contribution as well.

A number of MIPs are developments and/or continuation of long standing science themes within CMIP. These include MIPs specifically addressing science questions related to cloud feedbacks and the understanding of spatial patterns of circulation and precipitation (CFMIP), carbon cycle feedbacks and the understanding of changes in carbon fluxes and stores (C^4MIP), detection and attribution (DAMIP) that newly includes 21st-century GHG-only simulations allowing the projected responses to GHGs and other forcings to be separated and scaled to derive observationally-constrained projections, and paleoclimate (PMIP) that assesses the credibility of the model response to forcing outside the range of recent variability. These MIPs reflect the importance of key forcing and feedback processes in understanding past, present and future climate change and have developed new experiments and science plans focused on emerging new directions that will be at the centre of the WCRP Grand Science Challenges. A few new MIPs have arisen directly from gaps in understanding in CMIP5 (Stouffer et al., 2015), for example poor quantification of radiative forcing (RFMIP), better understanding of ocean heat uptake and sea-level rise (FAFMIP) and understanding of model response to volcanic forcing (VolMIP).

Since CMIP5, other MIPs have emerged as the modelling community has developed more complex ESMs with interactive components beyond the carbon cycle. These include the consistent quantification of forcings and feedbacks from aerosols and atmospheric chemistry (AerChemMIP), and, for the first time in CMIP, modelling of sea-level rise from land-ice sheets (ISMIP6).

Some MIPs specifically target systematic biases focusing on improved understanding of the sea-ice state and its atmospheric and oceanic forcing (SIMIP), the physical and biogeochemical aspects of the ocean (OMIP), land, snow and soil moisture processes (LS3MIP), and improved understanding of circulation and variability with a focus on stratosphere–troposphere coupling (DynVar). With the increased emphasis in the climate science community on the need to represent and understand changes in regional circulation, systematic biases are also addressed on a more regional scale by the Global Monsoon MIP (GMMIP) and a first co-ordinated activity on high resolution modelling (HighResMIP).

For the first time future scenario experiments, previously co-ordinated centrally as part of the CMIP5 "core" experiments, will be run as a MIP ensuring clear definition and well-coordinated science questions. ScenarioMIP will run a new set of future long-term (century time scale) integrations engaging input from both the climate science and Integrated assessment modelling communities. The new scenarios that are based on the shared socioeconomic pathways (SSPs, O'Neill et al., 2015) – Representative Concentration Pathways (RCP) matrix span the same range as the CMIP5 RCPs (Moss et al., 2010), but fill critical gaps for intermediate forcing levels and questions, for example, on short-lived species and land-use. The near-term experiments (10–30 years) will be coordinated by the decadal climate prediction project (DCPP) with improvements expected for example from the initialization of additional components beyond the ocean and from a more detailed process understanding and verification of the models to better identify sources and limits of predictability.

Other MIPs include specific future mitigation options, e.g. the land use MIP (LUMIP) that is for the first time in CMIP looking at regional land management strategies to study how different surface types respond to climate change and direct anthropogenic modifications, or the geoengineering MIP (GeoMIP) that examines climate impacts of newly proposed radiation modification geoengineering strategies.

The diagnostic MIP CORDEX will oversee the downscaling of CMIP6 models for regional climate projections. Another historic development in our field that provides, for the first time in CMIP, an avenue for a more formal communication between the climate modelling and user community is the endorsement of the vulnerability, impacts and adaptation and climate services advisory board (VIACS AB). This diagnostic MIP requests key outputs from CMIP6 models to deliver to the VIACS communities in rapid time for application to climate services and impact studies.

All MIPs define output streams in the centrally coordinated CMIP6 data request for each of their own experiments as well as the DECK and CMIP6 Historical Simulations (see the CMIP6 data request contribution to this Special Issue for details). This will ensure that the required variables are stored at the frequency and resolution required so that the specific science questions and evaluation needs of each MIP can be addressed, and a broad characterization of the performance of the CMIP6 models and ensemble can be made.

We note that only the Tier 1 MIP experiments are overseen by the CMIP Panel, but additional experiments are proposed by the MIPs in Tier 2 and 3. We encourage the modelling groups to participate in the full suite of experiments beyond Tier 1 to address in more depth the scientific questions posed.

5 Summary

CMIP6 continues the pattern of evolution and adaptation characteristic of previous phases of CMIP. To address the importance of broadly centring CMIP at the heart of activities within climate science, yet link activities within the World Climate Research Programme (WCRP) more specifically, CMIP6 has been formulated scientifically around three specific themes, amidst the backdrop of the WCRP's seven Grand Science Challenges. To meet the increasingly broad scientific demands of the climate-science community, yet be responsive to the individual priorities and resource limitations of the modelling centres, CMIP has adopted a new, more federated organizational structure.

CMIP has now evolved from a centralized activity involving a large number of experiments to a federated activity, encompassing many individually designed MIPs. CMIP6 comprises 21 individual CMIP6-Endorsed MIPs, the DECK and the CMIP6 Historical Simulations. Four of the 21 CMIP6-Endorsed MIPs are diagnostic in nature, meaning that they require additional output from models, but not additional simulations. The total amount of output from CMIP6 is estimated to be between 20 and 40 Petabytes, depending on model resolution and the number of modelling centres ultimately participating in CMIP6. Questions addressed in the MIPs are wide ranging, from the climate of distant past to turbulent cloud processes influence by the response to radiative forcing, from how the terrestrial biosphere influences the uptake of carbon-dioxide to how

much predictability is encoded in the ocean, and from what regulates the distribution of tropospheric ozone, to the influence of land-use changes on water availability.

The last two years have been dedicated to conceiving and then planning what we now call CMIP6. Starting in 2016, the first modelling centres are expected to begin performing the DECK and uploading output on the ESGF. Around April 2016 the forcings for the historical simulations should be ready, and by the end of 2016 the diverse forcings for different scenarios of future human activity will become available. Past experience suggests that most centres will complete their CMIP simulations within a few years while the analysis of CMIP6 results will likely go on for a decade or more (Fig. 4).

Through an intensified effort to align CMIP with specific scientific themes and activities we expect CMIP6 to continue CMIP's tradition of major scientific advances. CMIP6 simulations and scientific achievements are expected to support the IPCC Sixth Assessment Report (AR6) as well as other national and international climate assessments or special reports. Ultimately scientific progress will be the best measure of the success of CMIP6. Measures of success will include improved understanding of how the climate system works through the quantification of forcings and feedbacks, improved understanding and interpretation of systematic model biases and corresponding identification of ways to alleviate them for model improvements, and robust climate projections and uncertainty estimates for adaptation and mitigation policies.

Appendix A: Experiment Specifications

A1 Specifications for the DECK

Here we provide information needed to perform the DECK, including specification of forcing and boundary conditions, initialization procedures, and minimum length of runs. This information is largely consistent with but not identical to the specifications for these experiments in CMIP5 (Taylor et al., 2009).

The DECK and the CMIP6 Historical Simulation is requested from all models participating in CMIP. The expectation is that this requirement will be met for each model configuration used in the subsequent CMIP6-Endorsed MIPs (an entry card). In the special case where the burden of the entry card simulations are prohibitive but the scientific case for including a particular model is strong (despite only partial completion of the entry card simulations), an exception to this policy can be granted on a model by model basis based on a specific recommendation to the CMIP Panel made by the chairs of the affected CMIP6-Endorsed MIP.

A1.1 AMIP simulation

As in the first simulations performed under the Atmospheric Model Intercomparison Project (AMIP, Gates et al., 1999), SSTs and SICs in AMIP experiments are prescribed consistent with observations (see details on this forcing dataset in the corresponding contribution to this Special Issue). Land models should be configured as close as possible to that used in the CMIP6 Historical Simulation including transient land use and land cover. Other external forcings including volcanic aerosols, solar variability, GHG concentrations, and anthropogenic aerosols should also be prescribed consistent with those used in the CMIP6 Historical Simulation (see Appendix A2 below). Even though in AMIP simulations models with an active carbon cycle will not be fully interactive, surface carbon fluxes should be archived over land. This will enable evaluation of the carbon cycle component of the model when climate conditions are more similar to the observed than in coupled atmosphere–ocean simulations.

AMIP integrations can be initialized from prior model integrations or from observations or in other reasonable ways. Depending on the treatment of snow cover, soil water content, the carbon cycle, and vegetation, these runs may require a spin-up period of several years. One might establish quasi-equilibrium conditions consistent with the model by, for example, running with ocean conditions starting earlier in the 1970's or cycling repeatedly through year 1979 before simulating the official period. Results from

the spin-up period (i.e., prior to 1979) should be discarded, but the spin-up technique should be documented.

For CMIP6, AMIP simulations should cover at least the period from January 1979 through December 2014, but modelling groups are encouraged to extend their runs to the end of the observed period. Output may also be contributed from years preceding 1979 with the understanding that surface ocean conditions were less complete and in some cases less reliable then.

The climate found in AMIP simulations is largely determined by the externally-imposed forcing, especially the ocean conditions. Nevertheless, unforced variability ("noise") within the atmosphere introduces some non-deterministic variations that hamper unambiguous interpretation of apparent relationships between, for example, the year-to-year anomalies in SSTs and their consequences over land. To assess the role of unforced atmospheric variability in any particular result, modelling groups are encouraged to generate an ensemble of AMIP simulations. For most studies a three-member ensemble, where only the initial conditions are varied, would be the minimum required, with larger size ensembles clearly of value in making more precise determination of statistical significance.

A1.2 Multi-century pre-industrial control simulation

Like laboratory experiments, numerical experiments are designed to reveal cause and effect relationships. A standard way of doing this is to perform both a "control" experiment and a second experiment where some externally-imposed experiment condition has been altered. For many CMIP experiments, including the rest of the experiments discussed in this Appendix, the "control" is a simulation with atmospheric composition and other conditions prescribed and held constant consistent with best estimates of the forcing estimated from the historical period.

Ideally the pre-industrial control (*piControl*) experiment for CMIP would represent a near-equilibrium state of the climate system under the imposed conditions. In reality, simulations of hundreds to many thousands of years would be required for the ocean's depths to equilibrate and for biogeochemical reservoirs to fully adjust. Available computational resources generally preclude integrations long enough to approach equilibrium, so in practice shorter runs must suffice. Usually, a *piControl* simulation is initialized from the control run of a different model or from near present-day observations, and then run until at least the surface climate conditions stabilize using 1850 forcings (see Stouffer et al., 2004 for further discussion). This spin-up period, which is discarded, can be as long as several hundred years. Note that the length of the spin-up period should be documented. The *piControl* used in CMIP begins at this point and generally continues for at least a few hundred years.

Although equilibrium is generally not achieved, the changes occurring after the spin-up period are usually found to evolve at a fairly constant rate that presumably decreases slowly as equilibrium is approached. After a few centuries, these "drifts" of the system mainly affect the carbon cycle and ocean below the main thermocline, but they are also manifest at the surface in a slow change in sea level. The climate drift must be removed in order to interpret experiments that use the pre-industrial simulation as a control. The usual procedure is to assume that the drift is insensitive to CMIP experiment conditions and to simply subtract the control run from the perturbed run to determine the climate change that would occur in the absence of drift.

Besides serving as a "control" for numerical experimentation, the *piControl* is used to study the naturally occurring, unforced variability of the climate system. The only source of climate variability in a control arises from processes internal to the model, whereas in the more complicated real world, variations are also caused by external forcing factors such as solar variability and changes in atmospheric composition caused, for example, by human activities or volcanic eruptions. Consequently, the physical processes responsible for unforced variability can more easily be isolated and studied using the control run of models, rather than by analysing observations.

A DECK *piControl* simulation is required to be long enough to extend to the end of any perturbation runs initiated from it so that climate drift can be assessed and possibly removed from those runs. If, for example, a historical simulation (beginning in

1850) were initiated from the beginning of the control simulation and then were followed by a future scenario run extending to year 2300, a control run of at least 450 years would be required. As discussed above, control runs are also used to assess model-simulated unforced climate variability. The longer the *piControl*, the more precisely can variability be quantified for any given time scale. A control simulation of many hundreds of years would be needed to assess variability on centennial time-scales. For CMIP6 it is recommended that the control run should be at least 500 years long (following the spin-up period), but of course the simulation must be long enough to reach to the end of the experiments it spawns. It should be noted that those analysing CMIP6 simulations might also require simulations longer than 500 years to accurately assess unforced variability on long time-scales, so modelling groups are encouraged to extend their *piControl* runs well beyond the minimum recommended number of years.

Because the climate was very likely not in equilibrium with its forcing in the year 1850, and different components of the climate system differentially measure the effect of the forcing prior to that time, there is some ambiguity in deciding on what forcing to apply for the control. For CMIP6 we recommend a specification of this forcing that attempts to balance partially conflicting objectives to

- Minimize artificial climate responses to discontinuities in radiative forcing at the time a historical simulation is initiated.
- Minimize artefacts in sea level change caused by unrealistic mismatches in conditions in the centennial-scale averaged forcings for the pre- and post-1850 periods.

The first consideration above implies that radiative forcing in the *piControl* should be close to that imposed at the beginning of the CMIP Historical Simulation (i.e., 1850). The second implies that a background volcanic aerosol and time-averaged solar forcing should be prescribed in the control run, since to neglect it would cause an apparent drift in sea-level associated with the suppression of heat uptake due to the net effect of, for instance, volcanism after 1850, and this has implications for sea level changes (Gregory, 2010; Gregory et al., 2013). We recognize that it will be impossible to entirely avoid artefacts and artificial transients, and practical considerations may rule out conformance with every aspect of the *piControl* protocol stipulated here. With that understanding, here are the recommendations for the imposed conditions on the *piControl*:

- Conditions must be time-invariant except for those associated with the mean climate (notably the seasonal and diurnal cycles of insolation).
- Unless indicated otherwise (e.g., the background volcanic forcing), experiment conditions should be representative of Earth ca. 1850.
- Orbital parameters (eccentricity, obliquity, and longitude of the perihelion) should be held fixed at their 1850 values.
- The solar constant should be fixed at its mean value (no 11 year solar cycle) over the first two solar cycles of the historical simulation (i.e., the 1850–1871 mean).
- A background volcanic aerosol should be specified that results in radiative forcing matching, as closely as possible, that experienced, on average, during the historical simulation (i.e., 1850–2014 mean).
- Models without interactive ozone chemistry should specify ozone as in the mean of the first decade of the CMIP Historical Simulation.

Because the mean volcanic forcing between 1850 and 2014 is small, the discontinuity associated with transitioning from a mean forcing to a time-varying volcanic forcing is also expected to be small. Even though this is the design objective, it is likely that all artefacts in quantities such as historical sea level change will not be entirely prevented. For this reason, and because some models may deviate from these specifications, it is recommended for all groups to perform an additional simulation of the historical period but with only natural forcing included. This natural-only historical simulation is called for under DAMIP. Modelling groups are urged to perform this experiment even if they elect not to participate in DAMIP as doing so will most effectively separate the role of natural vs. anthropogenic drivers of climate change and variability since 1850.

The forcing specified in the *piControl* also has implications for simulations of the future, when solar variability and volcanic activity will continue to exist, but at unknown levels. These issues need to be borne in mind when designing and evaluating future scenarios, as a failure to include volcanic forcing in the future will cause future warming and sea-level rise to be over-estimated relative to a *piControl* experiment in which a non-zero volcanic forcing is specified. This could be addressed by re-introducing the mean volcanic forcing for the *piControl* into the scenarios.

These issues, and the potential of different modelling centres adopting different approaches to account for their particular constraints, highlight the paramount importance of adequately documenting how this and other DECK experiments were performed.

A1.3 Abruptly quadrupling CO_2 simulation

Until CMIP5, there were no experiments designed to quantify the extent to which forcing differences might explain differences in climate response. It was also difficult to diagnose and quantify the feedback responses, which are mediated by global surface temperature change (Sherwood et al., 2015). In order to examine these fundamental characteristics of models – CO_2 forcing and climate feedback – an abrupt 4xCO2 simulation was included for the first time as part of CMIP5. Following Gregory et al. (2004), the simulation branches from the *piControl* and the atmospheric CO_2 concentration is abruptly quadrupled and then held constant. As the system subsequently evolves toward a new equilibrium, the imbalance in the net flux at the top of the atmosphere can be plotted against global temperature change. As Gregory et al. (2004) showed, it is then possible to diagnose both the effective radiative forcing due to a quadrupling of CO_2 and also equilibrium climate sensitivity (ECS). Moreover, by examining how individual flux components evolve with surface temperature change, one can learn about the relative strengths of different feedbacks, notably quantifying the importance of various feedbacks associated with clouds.

In the *abrupt4×CO2* experiment, the only externally-imposed difference from the *piControl* should be the change in CO_2 concentration. All other conditions should remain

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

as they were in the *piControl*, including any background volcanic aerosols. By changing only a single factor, we can unambiguously attribute all climatic consequences to the increase in CO_2 concentration.

The minimum length of the *abrupt4×CO2* simulation should be 150 years, but longer simulations with CO_2 held constant after quadrupling are of value for investigating longer-time scale responses. Also there is value in performing an ensemble of short (~ 5 year) simulations initiated at different times throughout the year, as called for as part of a Tier 1 set of experiments in CMIP5. An ensemble would reduce the statistical uncertainty with which the effective CO_2 radiative forcing could be quantified and would allow more detailed and accurate diagnosis of the fast responses of the system under an abrupt change in forcing (Bony et al., 2013; Gregory and Webb, 2008; Kamae and Watanabe, 2013; Sherwood et al., 2015).

A1.4 1 % CO_2 increase simulation

The second idealized climate change experiment was introduced in the early days of CMIP (Meehl et al., 2000). It is designed for studying model responses under simplified but somewhat more realistic forcing than an abrupt increase in CO_2. In this experiment, the simulation is branched from the *piControl*, and CO_2 concentration is gradually increased at a rate of 1 % yr^{-1} (i.e., exponentially). Since the radiative forcing is approximately proportional to the logarithm of the CO_2 increase, the radiative forcing linearly increases over time. Drawing on the estimates of effective radiative forcing (for definitions see Myhre et al., 2013) obtained in the *abrupt4×CO2* simulations, analysts can scale results from each model in the 1 % CO_2 increase simulations to focus on the response differences in models, largely independent of their forcing differences. In contrast, in the CMIP6 Historical Simulation (see Sect. A2), the forcing and response contributions to model differences in simulated climate change cannot be easily isolated.

As in the *abrupt4×CO2* experiment, the only externally-imposed difference from the *piControl* should be the change in CO_2 concentration. The omission of changes in

aerosol concentrations is the key to making these simulations easier to interpret. The 1 % yr^{-1} CO_2 increase simulation should be run for a minimum of 150 years (ten years after the time of quadrupling).

Models with a carbon cycle component will be driven by prescribed CO_2 concentrations, but terrestrial and marine surface fluxes of carbon will become a key diagnostic from which one can infer emission rates that are consistent with a 1 % yr^{-1} increase in model CO_2 concentration. This DECK baseline carbon cycle experiment is built upon in C^4MIP to diagnose the strength of model carbon climate feedback and to quantify contributions to disruption of the carbon cycle by climate and by direct effects of increased CO_2 concentration.

A2 The CMIP6 Historical Simulation

The CMIP6 Historical Simulation is meant to reproduce observed climate and climate change starting in the year 1850 and extending to the present (i.e., 2014 in CMIP6). It serves as an important benchmark for assessing model performance. The historical integration should be initialized from some point in the *piControl* integration and be forced by time-varying, externally-imposed conditions that are based on observations. Both naturally-forced changes (e.g., due to solar variability and volcanic aerosols) and changes due to human activities (e.g., CO_2 concentration, aerosols, and land-use) will lead to climate variations and evolution. In addition there is unforced variability which can obscure the forced changes and lead to expected differences between the simulated and observed climate variations (Deser et al., 2012).

The externally-imposed forcing datasets that should be used in CMIP6 cover the period 1850 through the end of 2014 are described in detail in various other contributions to this Special Issue. Recall from Sect. A1.2 that the conditions in the *piControl* should generally be consistent with the forcing imposed near the beginning of the CMIP Historical Simulation. This should minimize artificial transients in the first portion of the CMIP Historical Simulation.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

As discussed earlier, there will be a mismatch in the specification of volcanic aerosols between control and historical simulations that especially affect estimates of ocean heat uptake and sea level rise in the historical period. This can be minimized by prescribing a background volcanic aerosol in the *piControl* that has the same cooling effect as the volcanoes included in the CMIP6 Historical Simulation. Any residual mismatch will need to be corrected, which requires a special supplementary simulation (see Sect. A1.2) that should be submitted along with the CMIP6 Historical Simulation.

For model evaluation and for detection and attribution studies (the focus of DAMIP) there would be considerable value in extending the CMIP6 Historical Simulations beyond the nominal 2014 ending date. To include the more recent observations in model evaluation, modelling groups are encouraged to document and apply forcing data sets representing the post-2014 period. For short extensions (up to a few years) it may be acceptable to simply apply forcing from one of the future scenarios defined by ScenarioMIP. To distinguish between the portion of the historical period when all models will use the same forcing data sets (i.e., 1850–2014) from the extended period where different data sets might be used, the experiment for 1850 through 2014 will be referred to as *historical* and the period from 2015 through near-present will be referred to as *historical-extension*.

Even if the CMIP6 Historical Simulations are extended beyond 2014, all future scenario simulations (called for by ScenarioMIP and other MIPs) should be initiated from the end of year 2014 of the CMIP6 Historical Simulation. The “future” in CMIP6 begins in 2015.

Due to interactions within and between the components of the Earth system, there is a wide range of variability on various time and space scales (Hegerl et al., 2007). The time scales vary from shorter than a day to longer than several centuries. The magnitude of the variability can be quite large relative to any given signal of interest depending on the time and space scales involved and on the variable of interest. To more clearly identify forced signals emerging from natural variability, multiple model integrations (comprising an “ensemble”) can be made where only the initial conditions

are perturbed in some way. A common way to do this is to simply branch each simulation from a different point in the control run. Longer intervals between branch points will ensure independence of ensemble members on longer time-scales. By averaging many different ensemble members together, the signal of interest becomes clear because the natural variations tend to average out if the ensemble size and averaging period are long enough. If the variability in the models is realistic, then the spread of the ensemble members around the ensemble average is caused by unforced (i.e., “internal”) variability. To minimize the number of years included in the entry card simulations, only one ensemble member is requested here. However, we strongly encourage model groups to submit at least three ensemble members for the CMIP Historical Simulation as requested in DAMIP.

Data availability

The model output from the DECK and the CMIP6 Historical Simulations described in this paper will be distributed through the Earth System Grid Federation (ESGF). As in CMIP5, the model output will be freely accessible through data portals after registration. In order to document CMIP6’s impact and enable ongoing support of CMIP, users are obligated to acknowledge CMIP6 and the participating modelling groups (see details on the CMIP Panel website at http://www.wcrp-climate.org/index.php/wgcm-cmip/about-cmip). Further information about the infrastructure supporting CMIP6, the metadata describing the model output, and the terms governing its use will be provided by the WGCM Infrastructure Panel (WIP) in their invited contribution to this Special Issue. In order to run the experiments, datasets for natural and anthropogenic forcings are required. These forcing datasets will be described in separate invited contributions to this Special Issue. The forcing datasets will be made available through the ESGF with version control and digital object identifiers (DOI’s) assigned, and can additionally be provided as a supplement to the corresponding documentation paper. Links to all forcings datasets will be made available via the CMIP Panel website.

Acknowledgements. We thank the scientific community for their engagement in the definition of CMIP6, in particular for the broad participation in the CMIP5 survey in 2013. We thank the co-chairs and steering committee members of the CMIP6-Endorsed MIPs for their continuous engagement in defining CMIP6, and the modelling groups and wider community for reviewing the CMIP6 design and organisation. We thank the WGCM Infrastructure Panel (WIP) for overseeing the CMIP6 infrastructure, Martin Juckes for taking the lead in preparing the CMIP6 data request, and Jonathan Gregory for raising awareness about the treatment of volcanic forcing in the preindustrial control experiment and its consequence for sea level changes. Norbert Noreiks is thanked for help in drafting the figures.

The article processing charges for this open-access publication were covered by a Research Centre of the Helmholtz Association.

References

Alexander, L. V., Kumar, A., Naveau, P., Seneviratne, S. I., Sivakumar, M. V. K., Zhang, X., and Zwiers, F. W.: The World Climate Research Program Grand Challenge on Extremes – WCRP-ICTP Summer School on Attribution and Prediction of Extreme Events, 21 July to 1 August 2014, Trieste, Italy, 2015.

Bony, S., Bellon, G., Klocke, D., Sherwood, S., Fermepin, S., and Denvil, S.: Robust direct effect of carbon dioxide on tropical circulation and regional precipitation, Nat. Geosci., 6, 447–451, 2013.

Bony, S., Stevens, B., Frierson, D. M. W., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma, A. P., Sobel, A. H., Watanabe, M., and Webb, M. J.: Clouds, circulation and climate sensitivity, Nat. Geosci., 8, 261–268, 2015.

Brasseur, G. and Carlson, D.: Future directions for the World Climate Research Programme, Eos T. Am. Geophys. Un., 96, doi:10.1029/2015EO033577, 2015.

Deser, C., Knutti, R., Solomon, S., and Phillips, A. S.: Communication of the role of natural variability in future North American climate, Nat. Clim. Change, 2, 775–779, 2012.

Eyring, V., Righi, M., Evaldsson, M., Lauer, A., Wenzel, S., Jones, C., Anav, A., Andrews, O., Cionni, I., Davin, E. L., Deser, C., Ehbrecht, C., Friedlingstein, P., Gleckler, P., Gottschaldt, K.-D., Hagemann, S., Juckes, M., Kindermann, S., Krasting, J., Kunert, D., Levine, R., Loew, A.,

Mäkelä, J., Martin, G., Mason, E., Phillips, A., Read, S., Rio, C., Roehrig, R., Senftleben, D., Sterl, A., van Ulft, L. H., Walton, J., Wang, S., and Williams, K. D.: ESMValTool (v1.0) – a community diagnostic and performance metrics tool for routine evaluation of Earth System Models in CMIP, Geosci. Model Dev. Discuss., 8, 7541–7661, doi:10.5194/gmdd-8-7541-2015, 2015.

Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., Cox, P., Driouech, F., Emori, S., Eyring, V., Forest, C., Gleckler, P., Guilyardi, E., Jakob, C., Kattsov, V., Reason, C., and Rummukainen, M.: Evaluation of Climate Models, in: Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P. M., Cambridge University Press, Cambridge, UK and New York, NY, USA, 741–866, 2013.

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., Fiorino, M., Gleckler, P. J., Hnilo, J. J., Marlais, S. M., Phillips, T. J., Potter, G. L., Santer, B. D., Sperber, K. R., Taylor, K. E., and Williams, D. N.: An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I), B. Am. Meteorol. Soc., 80, 29–55, 1999.

Geoffroy, O., Saint-Martin, D., Olivie, D. J. L., Voldoire, A., Bellon, G., and Tyteca, S.: Transient climate response in a two-layer energy-balance model. Part I: analytical solution and parameter calibration using CMIP5 AOGCM experiments, J. Climate, 26, 1841–1857, 2013.

Gleckler, P. J., Doutriaux, C., Durack P. J., Taylor K. E. , Zhang, Y., Williams, D. N., Mason, E., and Servonnat, J.: A package to make performance summaries from climate model intercomparisons more accessible, Eos Trans. AGU, submitted, 2015.

Gregory, J. and Webb, M.: Tropospheric adjustment induces a cloud component in CO_2 forcing, J. Climate, 21, 58–71, 2008.

Gregory, J. M.: Long-term effect of volcanic forcing on ocean heat content, Geophys. Res. Lett., 37, L22701, doi:10.1029/2010gl045507, 2010.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D.: A new method for diagnosing radiative forcing and climate sensitivity, Geophys. Res. Lett., 31, L15312, doi:10.1029/2004gl020258, 2004.

Gregory, J. M., Bi, D., Collier, M. A., Dix, M. R., Hirst, A. C., Hu, A., Huber, M., Knutti, R., Marsland, S. J., Meinshausen, M., Rashid, H. A., Rotstayn, L. D., Schurer, A., and Church, J. A.: Climate models without preindustrial volcanic forcing underestimate historical ocean thermal expansion, Geophys. Res. Lett., 40, 1600–1604, 2013.

Hegerl, G. C., Zwiers, F. W., Braconnot, P., Gillett, N. P., Luo, Y., Marengo Orsini, J. A., Nicholls, N., Penner, J. E., and Stott, P. A.: Understanding and Attributing Climate Change, in: Climate Change 2007: The Physical Science Basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M. and Miller, H. L., Cambridge University Press, Cambridge, UK and New York, NY, USA, 663–745, 2007.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, UK and New York, NY, USA, 2013.

Kamae, Y. and Watanabe, M.: Tropospheric adjustment to increasing CO_2: its timescale and the role of land–sea contrast, Clim. Dynam., 41, 3007–3024, 2013.

Li, C., Stevens, B., and Marotzke, J.: Eurasian winter cooling in the warming hiatus of 1998–2012, Geophys. Res. Lett., 42, 8131–8139, 2015.

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J.: Intercomparison makes for a better climate model, Eos T. Am. Geophys. Un., 78, 445–451, 1997.

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J.: The Coupled Model Intercomparison Project (CMIP), B. Am. Meteorol. Soc., 81, 313–318, 2000.

Meehl, G. A., Covey, C., Taylor, K. E., Delworth, T., Stouffer, R. J., Latif, M., McAvaney, B., and Mitchell, J. F. B.: The WCRP CMIP3 multimodel dataset: a new era in climate change research, B. Am. Meteorol. Soc., 88, 1383–1394, 2007.

Meehl, G. A., Moss, R., Taylor, K. E., Eyring, V., Stouffer, R. J., Bony, S., and Stevens, B.: Climate Model Intercomparisons: preparing for the Next Phase, Eos T. Am. Geophys. Un., 59, 77, 2014.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., van Vuuren, D. P., Carter, T. R., Emori, S., Kainuma, M., Kram, T., Meehl, G. A., Mitchell, J. F. B., Nakicenovic, N., Riahi, K., Smith, S. J., Stouffer, R. J., Thomson, A. M., Weyant, J. P., and Wilbanks, T. J.: The next generation of scenarios for climate change research and assessment, Nature, 463, 747–756, 2010.

Murphy, J. M. and Mitchell, J. F. B.: Transient–response of the Hadley-center coupled ocean–atmosphere model to increasing carbon-dioxide 2. Spatial and Temporal Structure of Response, J. Climate, 8, 57–80, 1995.

Myhre, G., D. Shindell, F.-M. Breìon, W. Collins, J. Fuglestvedt, J. Huang, D. Koch, J.-F. Lamarque, D. Lee, B. Mendoza, T. Nakajima, A. Robock, G. Stephens, Takemura, T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V. and Midgley, P. M., Cambridge University Press, Cambridge, UK and New York, NY, USA, 659–740, 2013.

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., and Solecki, W.: The roads ahead: narratives for shared socioeconomic pathways describing world futures in the 21st century, Global Environ. Chang., in press, doi:10.1016/j.gloenvcha.2015.01.004, 2015.

Phillips, A. S., Deser, C., and Fasullo, J.: Evaluating modes of variability in climate models, Eos T. Am. Geophys. Un., 95, 453–455, 2014.

Rauser, F., Gleckler, P., and Marotzke, J.: Rethinking the default construction of multimodel climate ensembles, B. Am. Meteorol. Soc., 96, 911–919, 2014.

Sherwood, S. C., Bony, S., Boucher, O., Bretherton, C., Forster, P. M., Gregory, J. M., and Stevens, B.: Adjustments in the forcing–feedback framework for understanding climate change, B. Am. Meteorol. Soc., 96, 217–228, 2015.

Stevens, B.: Rethinking the lower bound on aerosol radiative forcing, J. Climate, 28, 4794–4819, 2015.

Stott, P. A., Mitchell, J. F. B., Allen, M. R., Delworth, T. L., Gregory, J. M., Meehl, G. A., and Santer, B. D.: Observational constraints on past attributable warming and predictions of future global warming, J. Climate, 19, 3055–3069, 2006.

Stouffer, R. J., Weaver, A. J., and Eby, M.: A method for obtaining pre-twentieth century initial conditions for use in climate change studies, Clim. Dynam., 23, 327–339, 2004.

Stouffer, R. J., Eyring, V., Meehl, G. A., Bony, S., Senior, C., Stevens, B., and Taylor, K. E.: CMIP5 scientific gaps and recommendations for CMIP6, B. Am. Meteorol. Soc., submitted, 2015.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: A summary of the CMIP5 experiment design, available at: http://cmip.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf (last access: 14 December 2015), 2009.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, B. Am. Meteorol. Soc., 93, 485–498, 2012.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Teixeira, J., Waliser, D., Ferraro, R., Gleckler, P., Lee, T., and Potter, G.: Satellite observations for CMIP5: the genesis of Obs4MIPs, B. Am. Meteorol. Soc., 95, 1329–1334, 2014.
Trenberth, K. and Asrar, G.: Challenges and opportunities in water cycle research: WCRP contributions, Surv. Geophys., 35, 515–532, 2014.
Williams, D. N., Balaji, V., Cinquini, L., Denvil, S., Duffy, D., Evans, B., Ferraro, R., Hansen, R., Lautenschlager, M., and Trenham, C.: A Global Repository for Planet-Sized Experiments and Observations, B. Am. Meteorol. Soc., e-View, doi:10.1175/BAMS-D-15-00132.1, 2015.
Williams, K. and Webb, M.: A quantitative performance assessment of cloud regimes in climate models, Clim. Dynam., 33, 141–157, 2009.

Table 1. Main criteria for MIP endorsement as agreed with representatives from the modelling groups and MIPs at the WGCM 18th Session in Grainau, Germany in October 2014.

Nr	MIP Endorsement Criterion
1	The MIP and its experiments address at least one of the key science questions of CMIP6.
2	The MIP demonstrates connectivity to the DECK experiments and the CMIP6 Historical Simulation.
3	The MIP adopts the CMIP modelling infrastructure standards and conventions.
4	All experiments are tiered, well-defined, and useful in a multi-model context and don't overlap with other CMIP6 experiments.
5	Unless a Tier 1 experiment differs only slightly from another well-established experiment, it must already have been performed by more than one modelling group.
6	A sufficient number of modelling centres (~ 8) are committed to performing all of the MIP's Tier 1 experiments and providing all the requested diagnostics needed to answer at least one of its science questions.
7	The MIP presents an analysis plan describing how it will use all proposed experiments, any relevant observations, and specially requested model output to evaluate the models and address its science questions.
8	The MIP has completed the MIP template questionnaire.
9	The MIP contributes a paper on its experimental design to the GMD CMIP6 Special Issue.
10	The MIP considers reporting on the results by co-authoring a paper with the modelling groups.

Table 2. Overview of DECK and the CMIP6 Historical Simulation providing the experiment, the CMIP6 label, a brief experiment description, the forcing methods as well as the start and end year and minimum number of years per experiments and its major purpose. The DECK and CMIP6 Historical Simulation are used to characterize the CMIP model ensemble. Given resource limitations, these entry card simulations for CMIP include only one ensemble member per experiment. However, we strongly encourage model groups to submit at least three ensemble members for the CMIP Historical Simulation as requested in DAMIP. Large ensembles of AMIP simulations are also encouraged. "All" in the "forcing methods" column means volcanic, solar and anthropogenic forcings.

Experiment	CMIP6 label	Experiment description	Forcing methods	Start Year	End Year	Minimum # years per simulation	Major purpose
DECK Experiments							
Historical AMIP	*amip*	Observed SSTs and SICs prescribed	All; CO_2 concentration-driven	1979	2014	36	Evaluation, variability
Pre-industrial control	*piControl*	Coupled atmosphere/ocean pre-industrial control	CO_2 emission- or concentration-driven	1850	n/a	500	Evaluation, unforced variability
Quadruple CO_2 abruptly, then hold fixed	*abrupt4×CO2*	CO_2 abruptly quadrupled and then held constant	CO_2 concentration-driven	n/a	n/a	150	Climate sensitivity, feedbacks, fast responses
1 % yr^{-1} CO_2 increase	*1pctCO2*	CO_2 prescribed to increase at 1 % yr^{-1}	CO_2 concentration-driven	n/a	n/a	150	Climate sensitivity, feedbacks, idealized benchmark
CMIP6 Historical Simulation							
Past ~ 1.5 centuries	*historical*	Simulation of the recent past	All; CO_2 emission- or concentration-driven	1850	2014	165	Evaluation

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

Table 3. List of CMIP6-Endorsed MIPs along with the long name of the MIP, the primary goal(s) and the main CMIP6 science theme as displayed in Fig. 2. Each of these MIPs is described in more detail in a separate contribution to this Special Issue. MIPs marked with * are Diagnostic-MIPs.

Short name of MIP	Long name of MIP	Primary Goal(s) in CMIP6	Main CMIP6 Science Theme
AerChemMIP	Aerosols and Chemistry Model Intercomparison Project	(a) Diagnosing forcings and feedbacks of tropospheric aerosols, tropospheric ozone precursors and the chemically reactive WMGHGs; (b) documenting and understanding past and future changes in the chemical composition of the atmosphere; (c) estimating the global to regional climate response from these changes.	Chemistry/Aerosols
C^4MIP	Coupled Climate Carbon Cycle Model Intercomparison Project	Understanding and quantifying future century-scale changes in the global carbon cycle and its feedbacks on the climate system, making the link between CO_2 emissions and climate change.	Carbon cycle
CFMIP	Cloud Feedback Model Intercomparison Project	Improved assessments of cloud feedbacks via (a) improved understanding of cloud- climate feedback mechanisms and (b) better evaluation of clouds and cloud feedbacks in climate models. Also improved understanding of circulation, regional-scale precipitation and non-linear changes.	Clouds/Circulation
DAMIP	Detection and Attribution Model Intercomparison Project	(a) Estimating the contribution of external forcings to observed global and regional climate changes; (b) observationally constraining future climate change projections by scaling future GHG and other anthropogenic responses using regression coefficients derived for the historical period.	Characterizing forcings
DCPP	Decadal Climate Prediction Project	Predicting and understanding forced climate change and internal variability up to 10 years into the future through a coordinated set of hindcast experiments, targeted experiments to understand the physical processes, and the ongoing production of skilful decadal predictions.	Decadal prediction
FAFMIP	Flux–Anomaly–Forced Model Intercomparison Project	Explaining the model spread in climate projections of ocean climate change forced by CO_2 increase, especially regarding the geographical patterns and magnitude of sea-level change, ocean heat uptake and thermal expansion.	Ocean/Land/Ice
GeoMIP	Geoengineering Model Intercomparison Project	Assessing the climate system response (including on extreme events) to proposed radiation modification geoengineering schemes by evaluating their efficacies, benefits, and side effects.	Geoengineering

Table 3. Continued.

Short name of MIP	Long name of MIP	Primary Goal(s) in CMIP6	Main CMIP6 Science Theme
GMMIP	Global Monsoons Model Intercomparison Project	(a) Improve understanding of physical processes in global monsoons system; (b) better simulating the mean state, interannual variability and long-term changes of global monsoons.	Regional phenomena
HighResMIP	High Resolution Model Intercomparison Project	Assessing the robustness of improvements in the representation of important climate processes with "weather-resolving" global model resolutions (~ 25 km or finer), within a simplified framework using the physical climate system only with constrained aerosol forcing.	Regional phenomena
ISMIP6	Ice Sheet Model Intercomparison Project for CMIP6	Improving confidence in projections of the sea level rise associated with mass loss from the ice sheets of Greenland and Antarctica.	Ocean/Land/Ice
LS3MIP	Land Surface, Snow and Soil Moisture	Providing a comprehensive assessment of land surface, snow, and soil moisture-climate feedbacks, and diagnosing systematic biases in the land modules of current ESMs using constrained land-module only experiments.	Ocean/Land/Ice
LUMIP	Land-Use Model Intercomparison Project	Quantifying the effects of land use on climate and biogeochemical cycling (past-future), and assessing the potential for alternative land management strategies to mitigate climate change.	Land use
OMIP	Ocean Model Intercomparison Project	Provide a framework for evaluating, understanding, and improving ocean, sea-ice, and biogeochemical (including inert tracers) components of AOGCMs and ESMs. Protocols are provided to perform coordinated ocean/sea-ice/tracer/biogeochemisty simulations forced with common atmospheric datasets.	Ocean/Land/Ice
PMIP	Paleoclimate Modelling Intercomparison Project	(a) Analysing the response to forcings and major feedbacks for past climates outside the range of recent variability; (b) assessing the credibility of climate models used for future climate projections.	Paleo
RFMIP	Radiative Forcing Model Intercomparison Project	(a) Characterizing the global and regional effective radiative forcing for each model for historical and 4xCO2 simulations; (b) assessing the absolute accuracy of clear-sky radiative transfer parameterizations; (c) identifying the robust impacts of aerosol radiative forcing during the historical period.	Characterizing forcings

Table 3. Continued.

Short name of MIP	Long name of MIP	Primary Goal(s) in CMIP6	Main CMIP6 Science Theme
ScenarioMIP	Scenario Model Intercomparison Project	(a) Facilitating integrated research on the impact of plausible future scenarios over physical and human systems, and on mitigation and adaptation options; (b) addressing targeted studies on the effects of particular forcings in collaboration with other MIPs; (c) help quantifying projection uncertainties based on multi-model ensembles and emergent constraints.	Scenarios
VolMIP	Volcanic Forcings Model Intercomparison Project	(a) Assessing to what extent responses of the coupled ocean–atmosphere system to strong volcanic forcing are robustly simulated across state-of-the-art coupled climate models; (b) Identifying the causes that limit robust simulated behaviour, especially differences in their treatment of physical processes	Characterizing forcings
CORDEX*	Coordinated Regional Climate Downscaling Experiment	Advancing and coordinating the science and application of regional climate downscaling (RCD) through statistical and dynamical downscaling of CMIP DECK, CMIP6 Historical Simulation and ScenarioMIP output.	Impacts
DynVar*	Dynamics and Variability of the Stratosphere–Troposphere System	Defining and analysing diagnostics that enable a mechanistic approach to confront model biases and understand the underlying causes behind circulation changes with a particular emphasis on the two-way coupling between the troposphere and the stratosphere.	Clouds/Circulation
SIMIP*	Sea-Ice Model Intercomparison Project	Understanding the role of sea-ice and its response to climate change by defining and analysing a comprehensive set of variables and process-oriented diagnostics that describe the sea-ice state and its atmospheric and ocean forcing.	Ocean/Land/Ice
VIACS AB*	Vulnerability, Impacts, Adaptation and Climate Services Advisory Board	Facilitating a two-way dialogue between the CMIP6 modelling community and VIACS experts, who apply CMIP6 results for their numerous research and climate services, towards an informed construction of model scenarios and simulations and the design of online diagnostics, metrics, and visualization of relevance to society.	Impacts

Table A1. Specifications in the DECK and CMIP6 Historical Simulation.

Experiment	Volcanic Stratospheric Aerosol	Solar Variability	Anthropogenic forcings
amip	Time-dependent observations	Time-dependent observations	Time-dependent observations
piControl	Background volcanic aerosol that results in radiative forcing matching, as closely as possible, that experienced, on average, during the historical simulation (i.e., 1850–2014 mean)	Fixed at its mean value (no 11 year solar cycle) over the first two solar cycles of the historical simulation (i.e., the 1850–1871 mean)	Given that the *historical* starts in 1850, the *piControl* should have fixed 1850 atmospheric composition, not true pre-industrial
abrupt4×CO2	As in *piControl*	As in *piControl*	As in *piControl* except CO_2 that is four times *piControl*
1pctCO2	As in *piControl*	As in *piControl*	As in *piControl* except CO_2 that is increasing at 1 % yr^{-1}
historical	Time-dependent observations	Time-dependent observations	Time-dependent observations

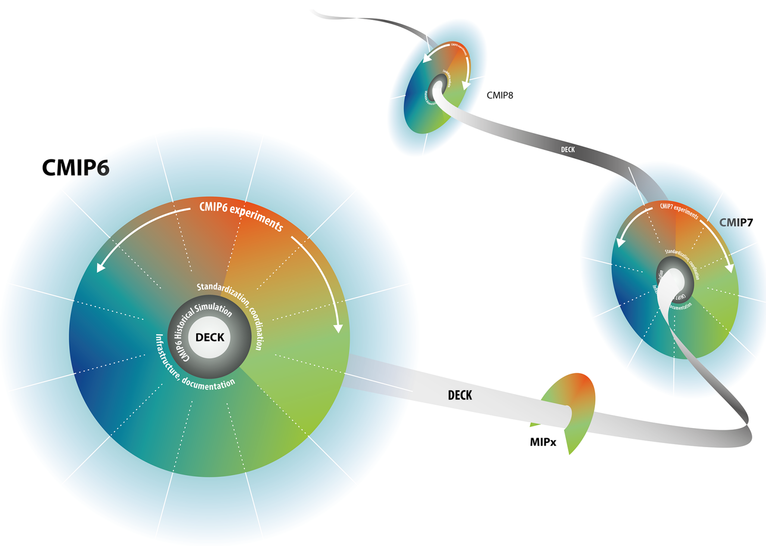

Figure 1. CMIP continuity across different phases.

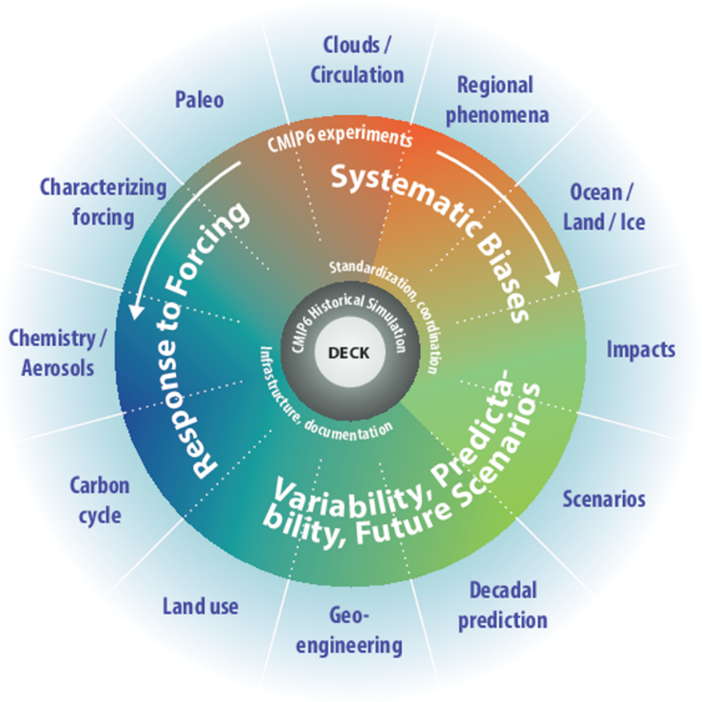

Figure 2. Schematic of the CMIP/CMIP6 experiment design. The inner ring and surrounding white text involve standardized functions of all CMIP DECK experiments and the CMIP6 Historical Simulation. The middle ring shows science topics related specifically to CMIP6 to be addressed by the MIPs, with MIP topics shown in the outer ring. This framework is superimposed on the scientific backdrop for CMIP6 which are the seven WCRP Grand Science Challenges.

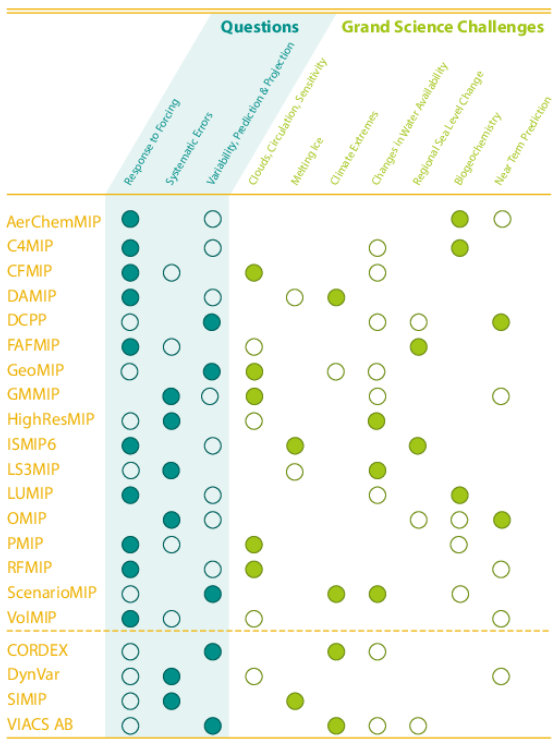

Figure 3. Contributions of CMIP6-Endorsed MIPs to the three CMIP6 science questions and the WCRP Grand Science Challenges. A filled circle indicates highest and an open second highest priority. Some of the MIPs additionally contribute with lower priority to other CMIP6 science questions or WCRP Grand Science Challenges.

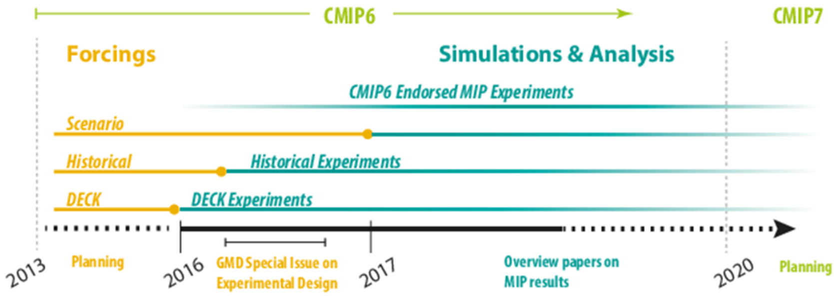

Figure 4. CMIP6 Timeline for the preparation of forcings, the realization of simulations and their analysis.