Topical Editor Decision: Publish subject to minor revisions (Editor review) (01 Apr 2016) by Dr. Julia Hargreaves

Reply to the Topical Editor Report by Julia Hargreaves (Editor)

We thank the editor for the detailed and constructive comments. We have now revised our manuscript in light of these comments. A pointwise reply is given below.

Comments to the Author:

I enjoyed reading the revised paper, which is much improved in terms of the GMD concept of what a model experiment paper should be. The requirement is not that the experiment must be perfect, but that the decisions made should be fully explained in terms of both scientific justification and practical considerations.

You responded to every comment apart from one, which was actually the first made by Gavin Schmidt. Therefore, please provide a response to: "DECK runs: These are a suitable 'entry card' into the process, and the requirements for new DECK entries for variations in physics, resolution, interactivity etc. is necessary (one run that is missing is perhaps a slab ocean equilibrium 2xCO2 run for coherence with previous estimates of the ECS). However, there are some implications of the DECK/Historical approach that need to be addressed. Specifically, because this is a relatively low barrier to entry, more models and model versions will very likely be archived. Thus instead of 60 individual model configurations as were available in CMIP3, there will likely be far more DECK entries over the lifetime of the CMIP6 program. I think this will be a good thing scientifically, but people should be ready for this."

Sorry for missing to provide a response to this comment.

Yes there will be more models and more model versions that will participate in CMIP6 and that will contribute DECK and CMIP6 historical simulations, and there will be substantial documentation efforts. As we already say in Section 3.3, a more routine benchmarking and evaluation of the CMIP6 models is envisaged by using newly developed evaluation tools. These tools will allow to broadly and comprehensively characterize the performance of the wide variety of models and model versions that will contribute to CMIP6. The last paragraph in Section 3.3 has been slightly changed to better clarify this point. Also in response to your comment on documentation below, we have extended the first paragraph in Section 3.3 to include the effort that the diverse models and model versions will be documented, pointing to the WIP contribution to this Special Issue for details.

Historical Simulation:
For simulations that are relevant for model evaluation against observations it is important that the simulations be as realistic as possible. The wider community, for example those using the CMIP ensemble as part of climate impact assessments, surely have the expectation that the historical simulations will be the best current estimates of the climate history, and that the spread in the ensemble can be considered a reasonable estimate of our uncertainty. Otherwise CMIP fails to address its first question, "How does the Earth system respond to forcing?" and instead is considering the less interesting question of, "How do different models of the Earth system respond to certain forcings?" Thus, I find your dismissal of Gavin Schmidt's suggestion that some models

may be more advanced than others and thus able to include new or different forcings to be inadequate. I'm not convinced that you have stepped out of the MIP box sufficiently to consider the wider perception of these runs. Each model should surely be able to submit their best historical run as their prime entry on ESGF. Isn't "conflated uncertainty" exactly what the wider community requires for this particular ensemble?

We have not dismissed Gavin Schmidt's suggestion, rather we had clarified this further already in the revised manuscript that we re-submitted. We disagree that having widely divergent forcings is necessarily helpful or informative. To avoid conflating uncertainty in the response of models to a given forcing, it is strongly preferred for models to be integrated with the same forcing in the entry card historical simulation, and for forcing uncertainty to be sampled in supplementary simulations that are proposed as part of DAMIP and therefore available to the wider community.

The historical simulations that we describe here will therefore indeed be performed with as identical forcings as possible that are centrally provided by the experts. In addition DAMIP includes historical simulations where individual forcings will be varied to cover the respective uncertainty in the forcing dataset, and also includes additional ensemble members. Furthermore modelling groups are free to pursue the response to any new forcings and publish on those results. Together with the CMIP historical simulation as an entry card, the spread in this ensemble can be considered a reasonable estimate of our uncertainty to address the question "How does the Earth system respond to forcing?". We now explicitly point to the DAMIP experiments so that the reader knows this issue on uncertainty in the forcings is covered in CMIP6.

Documentation:
That brings me to the topic of documentation. This is alluded to a number of times in the manuscript, but nowhere (apologies if I have missed it) do I see a description of how the documentation is to be done. This must be addressed. Perhaps it is to be addressed further in one of the other papers in the special issue, but the basic plan needs to be outlined here and that other paper referred to. Speaking personally, as a user of CMIP5, the level of documentation was inadequate, and I often email people to find out details about their runs, a pursuit which is not always successful, as the answer in some cases is already lost in the mists of time. Is there to be a documentation file of a certain format permanently attached to each data file? If this topic is not to be addressed elsewhere, can you include an example documentation file as supplementary information?

We have added a paragraph in Section 3.3 that mentions documentation. For details on ES-DOC we refer to the WIP contribution to this Special Issue.

RFMIP-lite:
All three reviewers are deeply concerned about the radiative forcing and my impression is that they would prefer to see the RFMIP-lite fixed SST experiment included in the DECK. Personally, I think your argument is reasonable (I interpret it as saying that the DECK runs are useful even without RFMIP-lite), but I do not see why these three reviewers would all be unfairly biased the same way, and so I am concerned that your decision to exclude the experiment is actually strongly against the consensus of the community, and that this decision should be carefully considered once more before you finally commit to excluding it.

We agree that RFMIP is extremely important. But the idea that this implies it should be part of the DECK arises from a misconception about the purpose of the DECK. The DECK is not a core set of experiments for CMIP6, rather a thread of continuity that runs through all phases of CMIP. We have modified the text to better emphasize this point, and at the same time also communicate why we think performing the RFMIP-lite experiments is so important, even if they are not part of the DECK.

On more formal grounds, the form of the DECK is based on a community consensus. The CMIP6 experiment design was iterated with the wide consultation of the community through several communication channels and was finally approved at the 18th session of the Working Group on Coupled Modelling (WGCM) in Grainau in October 2014 and once again confirmed at the 19th session of WGCM in Dubrovnik in October 2015 with more than 40 participants attending each session that included representatives of the modelling groups and the CMIP6-Endorsed MIPs. And on practical grounds the DECK consists of a small set of tried and true experiments that have proven their worth in previous phases of CMIP. We hope the simulations proposed in RFMIP prove their value, but until they do it would be premature to argue for their inclusion as part of the standard documentation of a model.

The 1%CO2 run and "effective" Climate Sensitivity:
Given your answers to Gavin Schmidt's questions about these topics, it is apparent that CMIP6 has made the decision to abandon the estimation of equilibrium climate sensitivity. Gavin cites his own paper on GISS, but many have shown that the "Gregory method" based on a short 4xCO2 run is not a very good approximation to equilibrium climate sensitivity (see Feedbacks, climate sensitivity and the limits of linear models, Reto Knutti, Maria A. A. Rugenstein, DOI: 10.1098/rsta.2015.0146 and references therein). This combined with the lack of a stabilisation run after 140 years of the 1% CO2 leaves no means of calculating equilibrium climate sensitivity. This needs to be stated clearly in the paper as a positive decision, as otherwise it may later appear to be an accidental omission. Just a thought, but would replacing the abrupt 4xCO2 with an abrupt 2xCO2 be useful - the 150 years is short, but it would at least be on the way to converging towards the desired value?

We realize that there is some literature which suggests that the non-linear response of the radiative imbalance to temperature and the short duration of the 4xCO2 runs combine to make these simulations inappropriate for estimating ECS. However this literature seems unaware of the work of Geoffroy et al., 2013a,b, Winton et al, Gregory et al., 2015 which largely resolves these issues by showing that fitting the 4xCO2 runs, or even by deconvolving the 1%CO2 runs, it is possible to adequately estimate the ECS of a model even in the presence of a strong non-linear response of N to globally averaged T. Informally in the community we are also aware of a "longrunMIP" with a handful of models which will provide a more stringent test of the two temperature theory that underlies the studies cited above. So we are not neglecting this issue, but we also don't believe that this manuscript is the place to go into depth into these issues, nor is it the place to try and correct misconceptions that have arisen in some other studies.

Abstract:
There is a comment here that surprised me, and I didn't find further explanation in the main body of the manuscript, "The participation in the CMIP6-Endorsed MIPs will be at the discretion of the modelling groups, and will depend on their scientific interests and priorities." I wasn't aware that the MIPs are policed. Why and how would a group be excluded from joining?

This was not what we meant. We have reworded this sentence for clarity.

Section 3.2 last sentence. "..to be evaluated fairly against those simpler models without representation of carbon cycle processes." I am not sure it is reasonable to suggest simplicity is relevant to any of these models, and models that do not have carbon cycle processes may be more complex in other ways. Could just delete the word "simpler"?

Good point. 'simpler' has been removed.

Section 4.2
"A number of MIPs are developments and/or continuation of long standing science themes within CMIP." I do not think this is accurate. At least some of these MIPs were founded pretty much separately from CMIP, and were only included in CMIP for CMIP5. Maybe just delete the "within CMIP". Otherwise it reads as a land grab that fails to give credit to the industry and ingenuity of those who founded the other MIPs.

'within CMIP' removed as suggested.

The Special Issue:
In many places the content of other manuscripts in the special issue is alluded to. GMD has many editors, and their focus is on assessing whether the papers pass the GMD peer review criteria, rather than making sure they include everything that you have promised here. It is therefore important that the authors of this paper make comments in the online review of the other papers in the special issue. Not hard, I think - as you only need check 3 or 4 each, and you'll be done!

Indeed the CMIP Panel will continue to oversee the process and will contribute to the open discussion at GMDD as required.

Summary:
I'd thought the paper had ended rather well at "decade or more". Then I turned the page and found another paragraph which didn't seem to add anything more. Maybe the final paragraph isn't required, or perhaps it could be made more compelling?

The 2nd last paragraph is on the timeline of CMIP6, the last paragraph is on what we expect from CMIP6. We think both are important and kept them but have reformulated the last paragraph.

Data availability:
Did you check whether actual "DOI"s are to be used in the case of ESGF? If you are not using official DOIs, then a different term for the ESGF digital identifiers should be coined.

We have once again checked with the co-chairs of the WIP and yes, the plan is to issue actual DOIs. No changes made to the manuscript.

Finally:::::
"DECK and CMIP6 historical simulation"
I don't understand this. Let's start with DECK - Diagnostic, Evaluation and Characterisation of Klima. What is Klima? I looked it up on Wikipedia and found that

it means “climate” in Czech, Danish, Norwegian (2 dialects), Serbo-Croatian, and Slovene, but that it means “air-conditioning” in Polish and Turkish. You need to provide the translation in order to avoid confusion as it is unfair to expect people, especially those who speak English as a second language, to understand other relatively obscure European languages. There are, incidentally, many stupider acronyms than DECC. So…

In the revised manuscript we now clarify that ‘klima’ is Greek for ‘climate’ since Greek was the original association behind this suggestion. Thanks for noting this omission. We kept the acronym DECK since it reminds a few of us (mostly those familiar with the German language) of a stable boat that is cruising through the various CMIP phases, and also because we find DECC too similar to the abbreviation DEC that is often used for decadal simulation activities.

I think I’ve now understood DECK; but I’m still confused. Why is the historical run outside the DECK? I can’t see where any scientific or practical distinction is made between the DECK and the historical run. They are both equally required as entry cards. Looking at the definition of DECK again - Diagnostic, Evaluation and Characterisation of Climate, it seems to me that the historical runs also belong inside the DECK. Please explain. Are modelling groups actually secretly permitted to exclude the historical run?

We clearly state in the manuscript that the historical simulation is an entry card for CMIP6: *“Under CMIP, credentials of the participating atmospheric-ocean general circulation models (AOGCMs) and ESMs are established by performing the DECK and CMIP historical simulations, so these experiments are required from all models. Together these experiments document the mean climate and response characteristics of models. They should be run for each model configuration used in a CMIP-Endorsed MIP.”* Also in the abstract we say: *‘The DECK and CMIP historical simulations, together with the use of CMIP data standards, will be the entry cards for models participating in CMIP.’* There should therefore be no doubt that a historical simulation has to be performed with each model configuration that contributes to CMIP6. In response to the reviewers’ comments, we had already included additional explanation why the historical simulation is not part of the DECK. We have once again extended this paragraph at the beginning of Section 3 to further clarify this issue. In particular we explicitly mention the criteria for the DECK again: *“The DECK experiments are chosen (1) to provide continuity across past and future phases of CMIP, (2) to evolve as little as possible over time, (3) to be well-established, and incorporate simulations that modelling centres anyway perform as part of their own development cycle, and (4) to be relatively independent of the forcings and scientific objectives of a specific phase of CMIP.”* Since the historical simulation does not meet all of these criteria, we keep it separate from the DECK noting again that it is of course remains an entry card simulations and is part of the CMIP6 characterization experiments. We fully agree with you that the historical simulation has to be an entry card simulation which is already reflected in the CMIP6 design.

Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organisation

Veronika~~.~~ Eyring[1], Sandrine~~.~~ Bony[2], Gerald~~.~~ A. Meehl[3], Catherine A.~~.~~ Senior[4], Bjorn~~.~~ Stevens[5], Ronald ~~.~~J. Stouffer[6], and Karl~~.~~ 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

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) and CMIP historical simulations (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 CMIP historical simulations to address a large range of specific questions and fill the scientific gaps of the previous CMIP phases. The DECK and CMIP historical simulations, together with the use of CMIP data standards, will be the entry cards for models participating in CMIP. Participation in CMIP6-Endorsed MIPs by individual modelling groups will be at their own discretion~~The participation in the CMIP6-Endorsed MIPs will be at the discretion of the modelling groups~~, and will depend on their 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 internal 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 CMIP6 historical simulations, 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; Meehl et al., 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 (e.g., 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 change 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 (Section 2). It also contains a description of a small set of simulations for CMIP which are intended to be common to all participating models (Section 3), details of which are provided in an Appendix. We then present a brief overview of CMIP6 that serves as an introduction to the other contributions to this Special Issue (Section 4), and we close with a summary.

2. CMIP design - a more continuous and distributed organization

In preparing for CMIP6, the CMIP Panel (the authors of this paper), which traditionally has the responsibility for direct coordination and oversight of CMIP, initiated a two year process of community consultation. This consultation involved the modelling centres whose contributions form the substance of CMIP as well as communities that rely on CMIP model output for their work. Special meetings were organized to reflect on the successes of CMIP5 as well as the scientific gaps that remain or have since emerged. The consultation also sought input through a community survey, the scientific results of which are described by Stouffer et al. (2015). Four main issues related to the overall structure of CMIP were identified.

First, we identified a growing appreciation of the scientific potential to use results across different CMIP phases. Such approaches however require an appropriate experimental design to facilitate the identification of an ensemble of models with particular properties drawn from different phases of CMIP (e.g., Rauser et al. (2014)). 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, in which context measures to help establish continuity across MIPs or phases of CMIP would also be welcome.

Second, the scope of CMIP was taxing the resources of modelling centres making it impossible for many to consider contributing to all the proposed experiments. By providing a better basis to help modelling centres decide exactly which subset of experiments to perform, it was thought that it might be possible to minimize fragmented participation in CMIP6. A more federated experimental protocol could also encourage modelling centres to develop intercomparison studies based on their own strategic goals.

Third, some centres expressed the view that the punctuated structure of CMIP had begun to distort the model development process. Defining a protocol that allowed modelling centres to decouple their model development from the CMIP schedule would offer additional flexibility, and perhaps encourage modelling centres to finalize their models and submit some of their results sooner on their own schedule.

Fourth and finally, many groups expressed a 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 for instance articulated by WCRP. By focusing a particular phase of CMIP around specific scientific issues, it was felt that the modelling resources could be more effectively applied to those scientific questions that had matured to a point where coordinated activities were expected to have substantial impact.

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~~)~~ (DECK) experiments (klima is Greek for "climate") and CMIP historical simulations, 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 simulations

The DECK comprises four baseline experiments: (a) a historical Atmospheric Model Intercomparison Project (*amip*) simulation, (b) a pre-industrial control simulation (*piControl* or *esm-piControl*), (c) a simulation forced by an abrupt quadrupling of CO_2 (*abrupt-4xCO2)* and (d) a simulation forced by a 1% yr^{-1} CO_2 increase (*1pctCO2*). CMIP also includes a historical simulation (*historical* or *esm-hist*) that spans the period of extensive instrumental temperature measurements from 1850 to the present. In naming the experiments, we distinguish between simulations with CO_2 concentrations calculated and anthropogenic sources of CO_2 prescribed (*esm-picontrol* and *esm-hist*) and simulations with prescribed CO_2 concentrations (all others). Hereafter, models that can calculate atmospheric CO_2 concentration and account for the fluxes of CO_2 between the atmosphere, the ocean, and biosphere are referred to as Earth System Models (ESMs).

The DECK experiments are chosen (1) to provide continuity across past and future phases of CMIP, (2) to evolve as little as possible over time, (3) to be well-established, and incorporate simulations that modelling centres anyway perform as part of their own development cycle, and (4) to be relatively independent of the forcings and scientific objectives of a specific phase of CMIP. The four DECK experiments and the CMIP historical simulations ~~chosen to be included in the DECK~~ are well suited ~~for evaluating models and~~ for quantifying and understanding important climate change response characteristics. ~~For these reasons, these experiments are already commonly performed by modelling groups as part of their model development cycle.~~ Modelling groups also commonly perform simulations of the historical period, but reconstructions of the external conditions imposed on historical runs (e.g., land-use changes) continue to evolve significantly, influencing the simulated climate. In order to distinguish among the historical simulations performed under different phases of CMIP, the historical simulations are labelled with the phase (e.g., “CMIP5 *historical*” or “CMIP6 *historical*”). A similar argument could be made to exclude the ~~Note that in~~ AMIP ~~runs~~experiments from the DECK. However, the AMIP experiments are simpler, more routine, and~~,~~ the dominating role of sea surface temperatures and the focus on recent decades means that for most purposes AMIP experiments~~runs~~ from different phases of CMIP are more likely to provide the desired continuity~~can be compared near the Earth's surface despite some differences in other imposed conditions~~.

The persistence and consistency of the DECK will make it possible to track changes in performance and response characteristics over future generations of models and CMIP phases. Although the set of DECK ~~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 of the DECK and ~~CMIP~~ the historical simulation~~s~~ were included in the core set of experiments performed under CMIP5 (Taylor et al., 2012), and all but the *abrupt-4xCO2* simulation were included in even earlier CMIP phases.

Under CMIP, credentials of the participating atmospheric-ocean general circulation models (AOGCMs) and ESMs are established by performing the DECK and CMIP historical simulations, so these experiments are required from all models. Together these experiments document the mean climate and response characteristics of 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 an ESM is used in both CO_2 emission-driven mode and CO_2 concentration-driven mode in subsequent CMIP6-Endorsed MIPs, then both emission-driven and concentration-driven control and historical simulations should be done and they will be identical in all forcings except the treatment of CO_2.

The forcing datasets that will drive the DECK and CMIP6 historical simulations are described separately in a series of invited contributions to this Special Issue. These articles also include some discussion of uncertainty in the datasets. The data will be provided by the respective author teams and made publicly available through the ESGF using common metadata and formats.

The historical forcings are based as far as possible on observations and cover the period 1850 to 2014. These include:

- emissions of short-lived species and long-lived greenhouse gases (GHGs),
- GHG concentrations,
- global gridded land-use forcing datasets,

- solar forcing,
- stratospheric aerosol dataset (volcanoes),
- AMIP sea surface temperatures (SSTs) and sea-ice concentrations (SICs),
- for simulations with prescribed aerosols 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 time-varying gridded ozone concentrations and nitrogen deposition.

Some models might require additional forcing datasets (e.g., black carbon on snow or anthropogenic dust). Allowing model groups to use different forcing[1] datasets might better sample uncertainty, but makes it more difficult to assess the uncertainty in the response of models to the best estimate of the forcing, available to a particular CMIP phase. To avoid conflating uncertainty in the response of models to a given forcing, it is strongly preferred for models to be integrated with the same forcing in this entry card historical simulation, and for forcing uncertainty to be sampled in supplementary simulations that are proposed as part of DAMIP. In any case it is important that all forcing datasets are documented and are made available alongside the model output on the ESGF. Likewise to the extent modelling centres simplify forcings, for instance by regridding or smoothing in time or some other dimension, this should also be documented.

For the future scenarios selected by ScenarioMIP, forcings are provided by the integrated assessment model (IAM) community for the period 2015 to 2100 (or to 2300 for the extended simulations). For atmospheric emissions and concentrations as well as for land use, the forcings ~~these~~ are harmonized across IAMs and scenarios using a similar procedure as in ~~to the~~ CMIP5 ~~procedure~~ (van Vuuren et al., 2011). This procedure ~~to~~ ensures consistency with historical forcing datasets and between the different forcing categories. The selection of scenarios and the main characteristics ~~They~~ are described elsewhere in this Special Issue, while the underlying IAM scenarios are described in a Special Issue in Global Environmental Change.

[1] Here we distinguish between an applied input perturbation (e.g. the imposed change in some model constituent, property, or boundary condition), which we refer to somewhat generically as a "forcing", and *radiative* forcing, which can be precisely defined. Even if the forcings are identical, the resulting *radiative* forcing depends on a model's radiation scheme (among other factors) and will differ among models.

An important gap identified in CMIP5, and in previous CMIP phases, was a lack of careful quantification of the radiative forcings from the different specified external forcing factors (e.g., GHGs, sulphate aerosols) in each model (Stouffer et al., 2015). This has impaired attempts to identify reasons for differences in model responses. The "effective radiative forcing" or ERF component of the Radiative Forcing MIP (RFMIP) includes "fixed SST" simulations to diagnose the forcing ('RFMIP-lite'), which are further detailed in the corresponding contribution to this Special Issue. Although not included as part of the DECK, in recognition of this deficiency in past phases of CMIP we strongly encourage all CMIP6 modelling groups to participate in RFMIP-lite. The~~is~~ modest additional effort would enable the radiative forcing to be characterized for both historic and future scenarios across the model ensemble. Knowing this forcing ~~and~~ would lead to a step change in efforts to ~~the~~ understand~~ing of~~ the spread of model responses for CMIP6 and contribute greatly to answering one of CMIP6's science questions.

An overview of the main characteristics of the DECK and CMIP6 historical simulations appears in Table 2. Here we briefly describe these experiments. Detailed specifications for the DECK and CMIP6 historical simulations are provided in Appendix A 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.

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, land, or their interactions, 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 simulations. Large ensembles of AMIP simulations are encouraged as they can help to improve the signal to noise ratio (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* or *esm-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. Historically, the industrial revolution began in the 18th century, and in nature the climate in 1850 was not stable as it was already changing due to prior historical changes in radiative forcings. In CMIP6, however, as in earlier CMIP phases, the control simulation is an attempt to produce a stable quasi-equilibrium climate state under 1850 conditions. When discussing and analysing historical and future radiative forcings, it needs to be recognized that the radiative forcing in 1850 due to anthropogenic greenhouse gas increases alone was already around 0.25 W/m^2 (Cubasch, 2013) although aerosols might have offset that to some extent. In addition, there were other pre-1850 secular changes, for example in land use (Hurtt et al., 2011), and as a result, global net annual emissions of carbon from land use and land-use change already were responsible in 1850 for about 0.6 PgC/yr (Houghton, 2010). Under the assumptions of the control simulation, however, there are no secular changes in forcing, so the concentrations and/or sources of atmospheric constituents (e.g., GHGs and emissions of short-lived species) as well as land use are held fixed, as are Earth's orbital characteristics. Because of the absence of both naturally occurring changes in forcing (e.g., volcanoes, orbital or solar changes) and human-induced changes, the control simulation can be used to study 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. 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 1850 control simulation and are designed to document basic aspects of the climate system response to greenhouse gas forcing. In the first, the CO_2 concentration is immediately and abruptly quadrupled from January 1850 values. This *abrupt-4xCO2* 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 gradually 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 simulations

In addition to the DECK, CMIP ~~challenges~~ requests 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 and its CO_2-emission-driven counterpart, *esm-hist*, branch from the *piControl* and *esm-piControl*, respectively (see details in A1.2). These simulations are 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 simulations provide rich opportunities to assess model ability to simulate climate, including variability and century time-scale trends (e.g., Flato et al. (2013)). These simulations can also be analysed to determine whether climate model forcing and sensitivity are consistent with the observational record, which provides opportunities to better bound the magnitude of aerosol forcing (e.g., Stevens (2015)). In addition they, along with the control run, provide the baseline simulations for performing formal ~~Stevens (2015)When supplemented with additional experiments, the historical simulations can be used in~~ detection and attribution studies (e.g., Stott et al. (2006)) which help uncover the causes of forced climate change~~to help interpret the extent to which observed climate change can be explained by different causes~~.

As in performing control simulations, models that include representation of the carbon cycle should normally perform two different CMIP historical simulations: one with prescribed CO_2 concentration and the other with prescribed CO_2 emissions (accounting explicitly for fossil fuel combustion). In the second CO_2 concentrations are "predicted" by the model. ~~-~~The treatment of other GHGs should be identical in both simulations. Both types of simulation are useful in evaluating how realistically the model represents the response of the carbon cycle anthropogenic CO_2 emissions, 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. In parallel to the development of the CMIP6 experiment design, the ESGF capabilities are being further extended and improved. In CMIP5 with over one thousand different model/experiment combinations, a first attempt was also made to capture structured metadata describing the models and the simulations themselves. Based upon the Common Information Model (CIM, Lawrence et al. (2012)), tools were provided to capture documentation of models and simulations. This effort is now continuing under the banner of the international ES-DOC activity, which is working toward agreements on common Controlled Vocabularies (CVs) to describe models and simulations. Details on these infrastructure related issues can be found in the WIP contribution to this Special Issue.

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 CMIP historical simulations 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 at the ESGF. These efforts utilize observations served by the ESGF contributed from the obs4MIPs (Ferraro et al., 2015; 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., 2016) 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 broadly and comprehensively characterize the performance of the wide variety of models and model versions that will contribute to CMIP6. This evaluation activity~~assess new models, and~~ can, compared with CMIP5, more quickly ~~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. These tools can also be used to compute derived variables or indices alongside the ESGF, and their output could be provided back to the distributed ESGF archive.

4. CMIP6

4.1. Scientific focus of CMIP6

In addition to the DECK and CMIP historical simulations, 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 observational, modelling and analytical 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 following three 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 internal 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 accounting of the impact of short-term forcing agents and land-use on climate, better understanding the mechanisms of decadal climate variability, along with many other issues not 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 on modelling groups, while maximizing the scientific complementarity and synergy among the different MIPs.

4.2. The CMIP6-Endorsed MIPs

Close to 30 suggestions for CMIP6 MIPs have been received so far 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 all are linked to the WCRP Grand Science Challenges (Fig. 3).

Each of the 21 CMIP6-Endorsed MIPs is 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 CMIP6 historical simulations. 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 are discussed and the MIPs are encouraged to help facilitate their use by contributing them to the obs4MIPs/ana4MIPs archives at the ESGF (see Section 3.3). In some MIPs additional forcings beyond those used in the DECK and CMIP6 historical simulations are required, and these are 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), which 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 coordinated activity on high resolution modelling (HighResMIP).

For the first time future scenario experiments, previously coordinated 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 are based on a matrix that uses ~~that are based on~~ the shared socioeconomic pathways (SSPs, O’Neill et al. (2015)) and forcing levels of the ~~–~~Representative Concentration Pathways (RCP) as axes. As a set, they ~~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) are 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 evaluation of the predictions 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 isolating regional land management strategies to study how different surface types respond to climate change and direct anthropogenic modifications, or the geoengineering MIP (GeoMIP), which 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 certain key variables of interest to the VIACS community be delivered in a timely manner to be used by climate services and in 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 to address the specific science questions and evaluation needs of each MIP and to enable a broad characterization of the performance of the CMIP6 models.

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.

The call for MIP applications for CMIP6 is still open and new proposals will be reviewed at the annual WGCM meetings. However, we point out that the additional MIPs suggested after the CMIP6 data request has been finalized will have to work with the already defined model output from the DECK and CMIP6 historical simulations, or work with the modelling group to recover additional variables from their internal archives. We also point out that some experiments proposed by CMIP6-Endorsed MIPs may not be finished until after CMIP6 ends.

5. Summary

CMIP6 continues the pattern of evolution and adaptation characteristic of previous phases of CMIP. To center CMIP at the heart of activities within climate science and encourage links among activities within the World Climate Research Programme (WCRP), CMIP6 has been formulated scientifically around three specific ~~themes~~questions, 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 and the DECK and 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 the response of turbulent cloud processes to radiative forcing, from how the terrestrial biosphere influences the uptake of CO_2 to how much predictability is stored in the ocean, from how to best project near-term to long-term future climate changes while considering interdependences and differences in model performance in the CMIP6 ensemble, 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. By May 2016 the forcings for the DECK and CMIP6 historical simulations will 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~~ questions and the WCRP Grand Science Challenges, ~~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 on the most pressing problems of climate variability and change 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.~~

Data availability

The model output from the DECK and CMIP6 historical simulations described in this paper will be distributed through the Earth System Grid Federation (ESGF) with digital object identifiers (DOIs) assigned. As in CMIP5, the model output will be freely accessible through data portals after registration. In order to document CMIP6's scientific impact and enable ongoing support of CMIP, users are obligated to acknowledge CMIP6, the participating modelling groups, and the ESGF centres (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 are provided by the WGCM Infrastructure Panel (WIP) in their invited contribution to this Special Issue. Along with the data itself, the provenance of the data will be recorded, and DOI~~'~~s will be assigned to collections of output so that they can be appropriately cited. This information will be made readily available so that published research results can be verified and credit can be given to the modelling groups providing the data. The WIP is coordinating and encouraging the development of the infrastructure needed to archive and deliver this information. In order to run the experiments, datasets for natural and anthropogenic forcings are required. These forcing datasets are described in separate invited contributions to this Special Issue. The forcing datasets will be made available through the ESGF with version control and DOIs assigned.

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 CMIP6 historical simulations are 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). For CMIP6, ~~I~~in the special case where the burden of the entry card simulations are prohibitive but the scientific case for including a particular model simulation is compelling (despite only partial completion of the entry card simulations), an exception to this policy can be granted on a model by model basis by the CMIP Panel, which will seek advice from the chairs of the affected CMIP6-Endorsed MIP.

CMIP6 is a cooperative effort across the international climate modelling and climate science communities. The modelling groups have all been involved in the design and implementation of CMIP6, and thus have agreed to a set of best practices proposed for CMIP6. Those best practices include having the modelling groups submit the DECK experiments and the CMIP6 historical simulations to the ESGF, as well as any CMIP6-Endorsed-MIP experiments they choose to run. Additionally, the modelling groups decide what constitutes a new model version. ~~Modelling groups are well aware that their model simulations are under considerable scrutiny. Therefore, we expect that as in the past, modelling groups will in good faith provide their highest quality model version and that it will differ from previous versions by substantive improvements in resolution, physics, or simulation skill.~~ The CMIP Panel will work with the MIP co-chairs and the modelling groups to ensure that these best practices are followed.

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 Section 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.

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 simulations

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 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 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 can be as long as several hundred years and variables that can document the spin-up behaviour should be archived (under the experiment labels *piControl-spinup* or *esm-piControl-spinup*). At the very least the length of the spin-up period should be documented.

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 "controls" for numerical experimentation, the *piControl* and *esm-piControl* are 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 control 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 control, 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 control runs well beyond the minimum recommended number of years.

Because the climate was very likely not in equilibrium with the forcing of 1850 and because different components of the climate system differentially respond to the effects 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 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 due to thermal expansion caused by unrealistic mismatches in conditions in the centennial-scale averaged forcings for the pre- and post-1850 periods. Note that any preindustrial multi-centennial observed trend in global-mean sea level is most likely to be due to slow changes in ice-sheets, which are likely not to be simulated in the CMIP6 model generation.

The first consideration above implies that radiative forcing in the control run 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 transient effects, and practical considerations may rule out conformance with every detail of the control simulation protocol stipulated here. With that understanding, here is a summary of the recommendations for the imposed conditions on the spin-up and control runs, followed by further clarification in subsequent paragraphs:

– 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 (e.g., greenhouse gas concentrations, ozone concentration, surface land conditions) should be representative of Earth around the year 1850.

– Orbital parameters (eccentricity, obliquity, and longitude of the perihelion) should be held fixed at their 1850 values.
– Land use should not change in the control run and should be fixed according to reconstructed agricultural maps from 1850. Due to the diversity of model approaches in ESMs for land carbon, some groups might deviate from this specification, and again this must be clearly documented.
– 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 – 1873 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 the pre-industrial ozone fields from a dataset produced from a pre-industrial control simulation that uses 1850 emissions and a mean solar forcing averaged over solar cycles 8-10, representative of the mean mid-19th century solar forcing.

There are some special considerations that apply to control simulations performed by "emission-driven" ESMs (i.e. runs with atmospheric concentrations of CO_2 calculated prognostically rather than being prescribed). In the *esm-piControl* simulation, emissions of CO_2 from both fossil fuel combustion and land use change are prescribed to be zero. In this run any residual drift in atmospheric CO_2 concentration that arises from an imbalance in the exchanges of CO_2 between the atmosphere and the ocean and land (i.e. by the natural carbon cycle in the absence of anthropogenic CO_2 emissions) will need to be subtracted from perturbation runs to correct for a control state not in equilibrium. It should be emphasized that the *esm-piControl* is an idealized experiment and is not meant to mimic the true 1850 conditions, which would have to include a source of carbon of around 0.6 PgC/yr from the already perturbed state that existed in 1850.

Due to a wide variety of ESMs and the techniques they use to compute land carbon fluxes, it is hard to make statements that apply to all models equally well. A general recommendation, however, is that the land carbon fluxes in the emission and concentration driven control simulations should be stable in time and in approximate balance so that the net carbon flux into the atmosphere is small (less than 0.1 PgC/yr). Further details on ESM experiments with a carbon cycle are provided in the C^4MIP contribution to this Special Issue.

The historical time-average volcanic forcing stipulated above for the control run is likely to approximate the much longer term mean. Crowley's (2000) estimates of volcanic aerosol radiative forcing for the historical period and the last millennium are -0.18 W m^{-2} and -0.22 W m^{-2}, respectively. 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 it will be impossible to eliminate all artefacts in quantities such as historical sea level change. For this reason, and because some models may deviate from these specifications, it is recommended that groups perform an additional simulation of the historical period but with only natural forcing included. With this additional run, which is already called for under DAMIP, the purely anthropogenic effects on sea-level change can be isolated.

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 is accounted for by introducing a time-invariant non-zero volcanic forcing (e.g., the mean volcanic forcing for the *piControl*) into the scenarios. This is further specified in the ScenarioMIP contribution to this Special Issue.

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 the conditions under which this and the other DECK experiments are 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 4xCO_2 simulation was included for the first time as part of CMIP5. Following Gregory et al. (2004), the simulation branches in January of the CO_2-concentration driven *piControl* and abruptly the atmospheric CO_2 concentration is quadrupled and held fixed. 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 effective 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 *abrupt-4xCO2* experiment, the only externally-imposed difference from the *piControl* should be the change in CO_2 concentration. All other conditions should remain 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 *abrupt-4xCO2* simulation should be 150 years, but longer simulations would enable investigations of longer-time scale responses. Also there is value, as in CMIP5, in performing an ensemble of short (~5-year) simulations initiated at different times throughout the year (in addition to the required January run). –Such 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). Different groups will be able to afford ensembles of different sizes, but in any case each realization should be initialized in a different month and the months should be spaced evenly throughout the year.

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). A minimum length of 150 years is requested so that the simulation goes beyond the quadrupling of CO_2 after 140 years. Note that in contrast to previous definitions, the experiment has been simplified so that the 1% CO_2 increase per year is applied throughout the entire simulation rather than keeping it constant after 140 years as in CMIP5. 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 *abrupt-4xCO2* 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 CMIP6 historical simulations (see Section A2), the forcing and response contributions to model differences in simulated climate change cannot be easily isolated.

As in the *abrupt-4xCO2* 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.

Models with a carbon cycle component will be driven by prescribed CO_2 concentrations, but terrestrial and marine surface fluxes and stores 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 simulations

CMIP6 historical simulations of climate change over the period 1850 through 2014 are forced by common datasets that are largely based on observations. They serve as an important benchmark for assessing model performance through evaluation against observations. The historical integration should be initialized from some point in the control integration (with *historical* branching from the *piControl* and the *esm-hist* branching from *esm-piControl*) 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 section A1.2 that the conditions in the control should generally be consistent with the forcing imposed near the beginning of the CMIP historical simulation. This should minimize artificial transient effects in the first portion of the CMIP historical simulation. An exception is that for the CO_2-emission driven experiments, the zero CO_2 emissions from fossil fuel and the land use specifications for 1850 in the *esm-piControl* could cause a discontinuity in land carbon at the branch point.

As described in Section A1.2, the 1850 *esm-piControl* should be developed for an idealized case that is stable in time and balance so that the net carbon flux into the atmosphere is small. Meanwhile, the start of the *esm-hist* in 1850 should be as realistic as possible and attempt to account for the fact the land-surface was not in equilibrium in 1850 due to prior land-use effects (Houghton, 2010; Hurtt et al., 2011). Some modelling groups have developed methods to achieve these twin goals in a computationally efficient manner, for example by performing pre-1850 off-line land model simulations to account for the land carbon cycle disequilibrium before 1850 and to adequately simulate carbon stores at the start of the historical simulation (Sentman et al., 2011). Due to the wide diversity of modelling approaches for land carbon in the ESMs, the actual method applied by each group to account for these effects will differ and needs to be well documented.

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 pre-industrial control 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 Section 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 labelled *historical* (*esm-hist* in the case of the emissions-driven run) and the period from 2015 through near-present will likely be labelled *historical-ext* (*esm-hist-ext*).

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 since 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 which should be documented. 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 of their CMIP historical simulation as requested in DAMIP.

Acknowledgements. We thank the scientific community for their engagement in the definition of CMIP6 and 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, Jonathan Gregory for raising awareness about the treatment of volcanic forcing in the pre-industrial control experiment and its consequence for sea level changes, and Pierre Friedlingstein, George Hurtt, Chris Jones, and David Lawrence for help in defining carbon cycle and land use specifications in the DECK experiments and CMIP6 historical simulations. Norbert Noreiks is thanked for help in drafting the figures. Thanks to our topical editor Julia Hargreaves, to Gavin Schmidt and the other two anonymous reviewers, and to everyone who contributed to the open discussions for constructive comments. GM and KET were ~~was~~ supported by the Regional and Global Climate Modeling Program (RGCM) of the U.S. Department of Energy's Office of Biological & Environmental Research (BER) (through Cooperative Agreement # DE-FC02-97ER62402 for GM), and GM received additional support from the U.S. National Science Foundation. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

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 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, Nature Geosci, 8, 261-268, 2015.

Brasseur, G. and Carlson, D.: Future directions for the World Climate Research Programme, Eos, Transactions American Geophysical Union, 96, 2015.

Cubasch, U., D. Wuebbles, D. Chen, M.C. Facchini, D. Frame, N. Mahowald, and J.-G. Winther: Introduction. 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, Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (Ed.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

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, 2015.

Ferraro, R., Waliser, D. E., Gleckler, P., Taylor, K. E., and Eyring, V.: Evolving obs4MIPs to Support the Sixth Coupled Model Intercomparison Project (CMIP6), B Am Meteorol Soc, doi: 10.1175/BAMS-D-14-00216.1, 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, Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (Ed.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 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 More Powerful Reality Test for Climate Models, Eos Trans. AGU, in press, 2016.

Gregory, J. and Webb, M.: Tropospheric adjustment induces a cloud component in CO2 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, 2010.

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.

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, 2004.

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, Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor and H.L. Miller (Ed.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Houghton, R. A.: How well do we know the flux of CO2 from land-use change?, Tellus B, 62, 337-351, 2010.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., Fisk, J. P., Hibbard, K., Houghton, R. A., Janetos, A., Jones, C. D., Kindermann, G., Kinoshita, T., Goldewijk, K. K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D. P., and Wang, Y. P.: Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, Climatic Change, 109, 117-161, 2011.

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, United Kingdom and New York, NY, USA, 2013.

Kamae, Y. and Watanabe, M.: Tropospheric adjustment to increasing CO2: its timescale and the role of land-sea contrast, Clim Dynam, 41, 3007-3024, 2013.

Lawrence, B. N., Balaji, V., Bentley, P., Callaghan, S., DeLuca, C., Denvil, S., Devine, G., Elkington, M., Ford, R. W., Guilyardi, E., Lautenschlager, M., Morgan, M., Moine, M. P., Murphy, S., Pascoe, C., Ramthun, H., Slavin, P., Steenman-Clark, L., Toussaint, F., Treshansky, A., and Valcke, S.: Describing Earth system simulations with the Metafor CIM, Geosci. Model Dev., 5, 1493-1500, 2012.

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.: The Coupled Model Intercomparison Project (CMIP), B Am Meteorol Soc, 81, 313-318, 2000.

Meehl, G. A., Boer, G. J., Covey, C., Latif, M., and Stouffer, R. J.: Intercomparison makes for a better climate model, Eos, Transactions American Geophysical Union, 78, 445-451, 1997.

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 Trans. AGU, 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. Bréon, 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, Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (Ed.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 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 Environmental Change, doi: 10.1016/j.gloenvcha.2015.01.004, 2015. 2015.

Phillips, A. S., Deser, C., and Fasullo, J.: Evaluating Modes of Variability in Climate Models, Eos Trans. AGU, 95(49), 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.

Sentman, L. T., Shevliakova, E., Stouffer, R. J., and Malyshev, S.: Time Scales of Terrestrial Carbon Response Related to Land-Use Application: Implications for Initializing an Earth System Model, Earth Interactions, 15, 1-16, 2011.

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., Eyring, V., Meehl, G. A., Bony, S., Senior, C., Stevens, B., and Taylor, K. E.: CMIP5 Scientific Gaps and Recommendations for CMIP6, BAMS, submitted, 2015.

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.

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.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: A Summary of the CMIP5 Experiment Design, http://cmip.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf, 2009.

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.

van Vuuren, D. P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G. C., Kram, T., Krey, V., Lamarque, J. F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S. J., and Rose, S. K.: The representative concentration pathways: an overview, Climatic Change, 109, 5-31, 2011.

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, doi: 10.1175/bams-d-15-00132.1, 2015. 150904101253006, 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 simulations.
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 do not 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 CMIP6 historical simulations providing the experiment short names, the CMIP6 labels, brief experiment descriptions, the forcing methods as well as the start and end year and minimum number of years per experiment 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. In the "forcing methods" column, "All" means "volcanic, solar and anthropogenic forcings". All experiments are started on 1 January and end at 31 December of the specified years.

Experiment short name	**CMIP6 label**	**Experiment description**	**Forcing methods**	**Start Year**	**End Year**	**Minimum # Years Per Simulation**	**Major purpose**
DECK Experiments							
AMIP	*amip*	Observed SSTs and SICs prescribed	All; CO_2 concentration prescribed	1979	2014	36	Evaluation, variability
pre-industrial control	*piControl* or *esm-piControl*	Coupled atmosphere/ocean pre-industrial control	CO_2 concentration prescribed or calculated	n/a	n/a	500	Evaluation, unforced variability
abrupt quadrupling of CO_2 concentration	*abrupt-4xCO2*	CO_2 abruptly quadrupled and then held constant	CO_2 concentration prescribed	n/a	n/a	150	Climate sensitivity, feedbacks, fast responses
1% yr^{-1} CO_2 concentration increase	*1pctCO2*	CO_2 prescribed to increase at 1% yr^{-1}	CO_2 concentration prescribed	n/a	n/a	150	Climate sensitivity, feedbacks, idealized benchmark
CMIP6 historical simulation							
past ~1.5 centuries	*historical* or *esm-hist*	Simulation of the recent past	All; CO_2 concentration prescribed or calculated	1850	2014	165	Evaluation

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
GMMIP	Global Monsoons Model	a) Improve understanding of physical processes in global monsoons system; b)	Regional phenomena

	Intercomparison Project	better simulating the mean state, interannual variability and long-term changes of global monsoons.	
HighResMIP	High Resolution Model Intercomparison Project	Assessing the robustness of improvements in the representation of important climate processes with "weather-resolving" global model resolutions (~25km 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 climate and Earth system models contributing to CMIP6. Protocols are provided to perform coordinated ocean/sea-ice/tracer/biogeochemistry 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 4xCO_2 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
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	a) Assessing to what extent responses of the coupled ocean-atmosphere system to strong	Characterizing forcings

	Intercomparison Project	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	
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*, and ScenarioMIP output.	Impacts
DynVarMIP*	Dynamics and Variability Model Intercomparison Project~~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 for CMIP6	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 simulations.

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 – 1873 mean)	Given that the historical simulations start in 1850, the *piControl* should have fixed 1850 atmospheric composition, not true pre-industrial
esm-piControl	As in *piControl*	As in *piControl*	As in *piControl* but with CO_2 concentration calculated, rather than prescribed. CO_2 from both fossil fuel combustion and land use change are prescribed to be zero.
abrupt-4xCO2	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
esm-hist	As in *historical*	As in *historical*	As in *historical* but with CO_2 emissions prescribed and CO_2 concentration calculated (rather than prescribed)

FIGURES

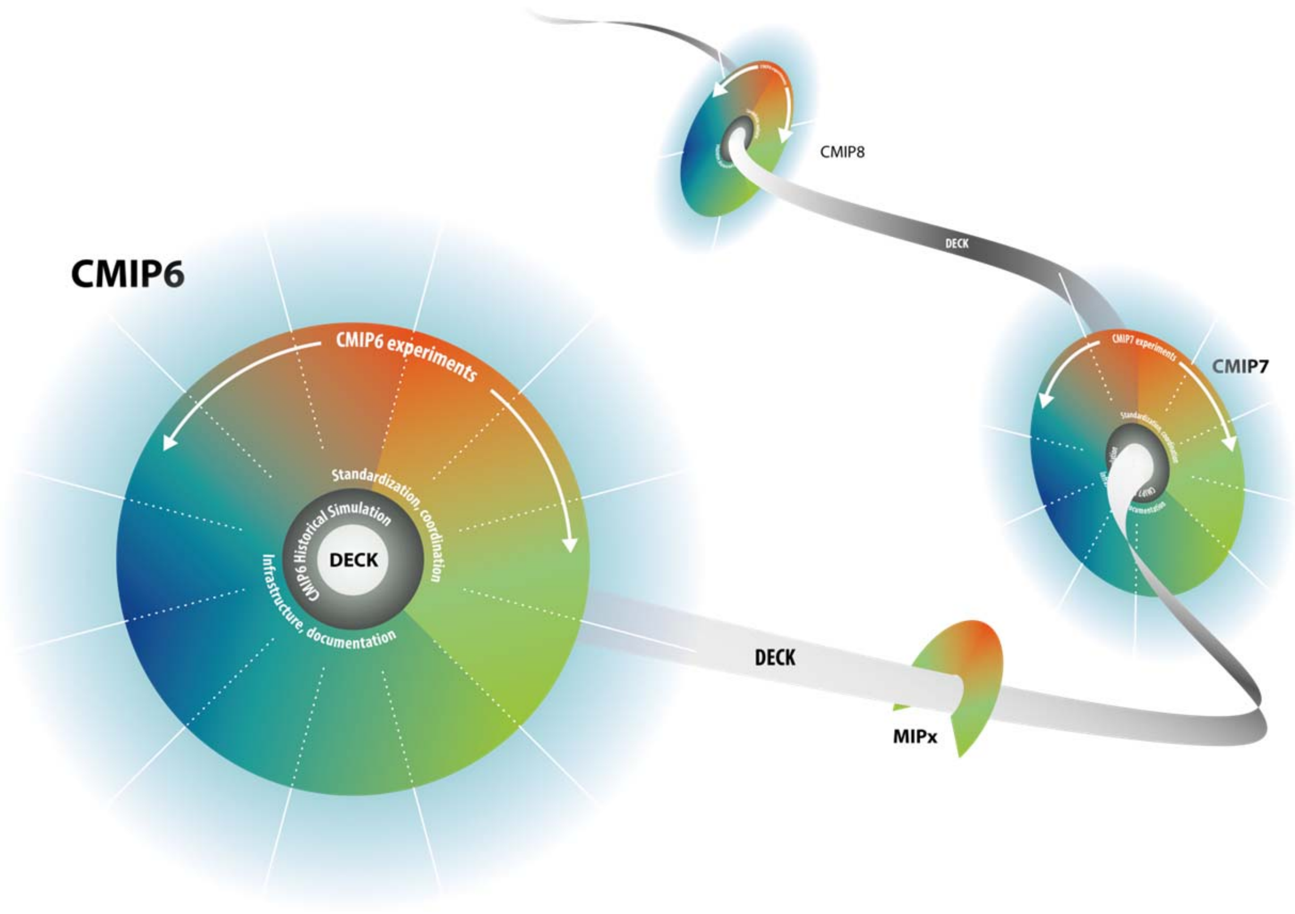

Figure 1. CMIP evolution. CMIP will evolve but the DECK will provide continuity across phases.

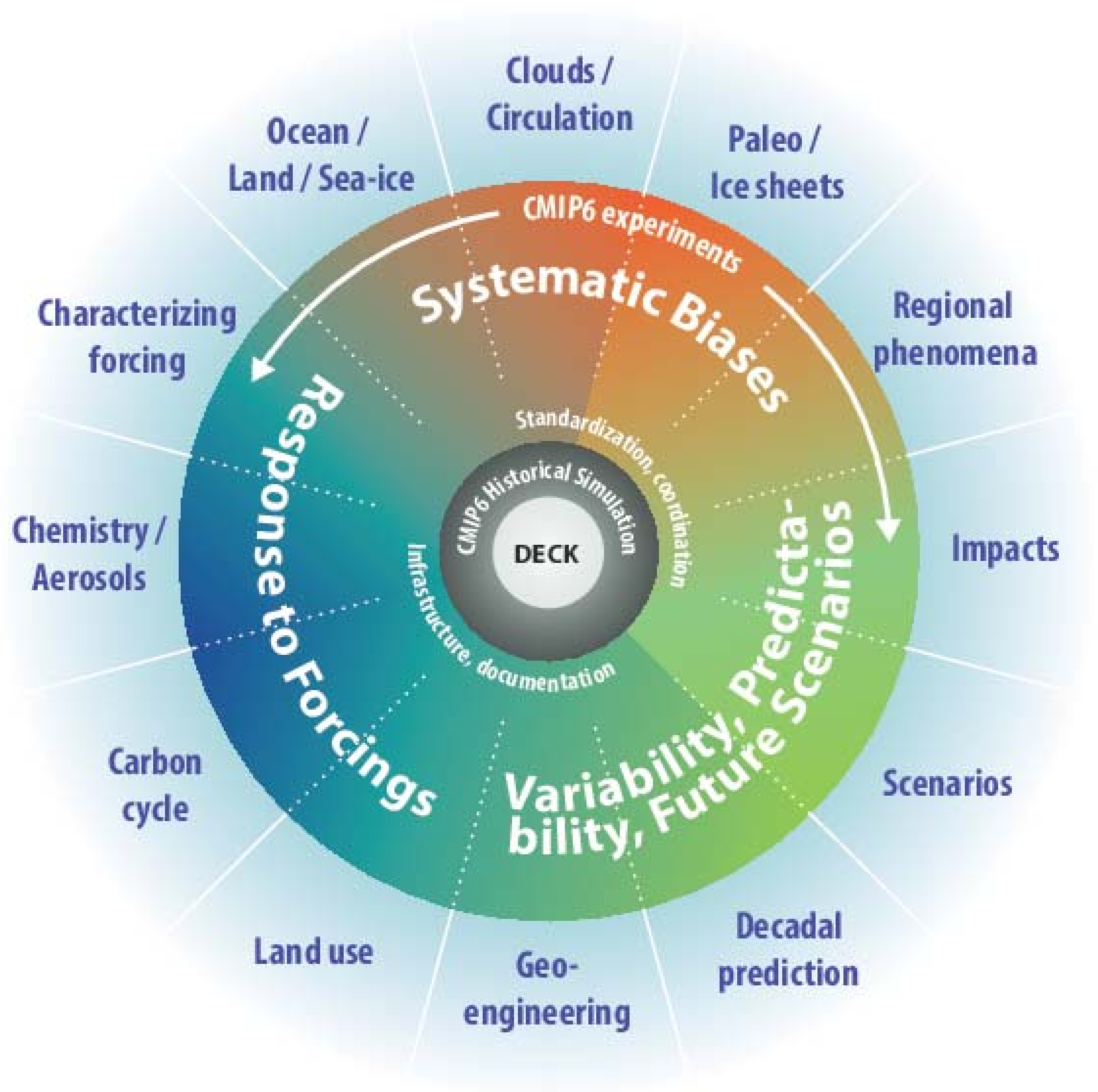

Figure 2. Schematic of the CMIP/CMIP6 experiment design. The inner ring and surrounding white text involve standardized functions of all CMIP DECK experiments and the CMIP6 historical simulation. The middle ring shows science topics related specifically to CMIP6 that are addressed by the CMIP6-Endorsed 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.

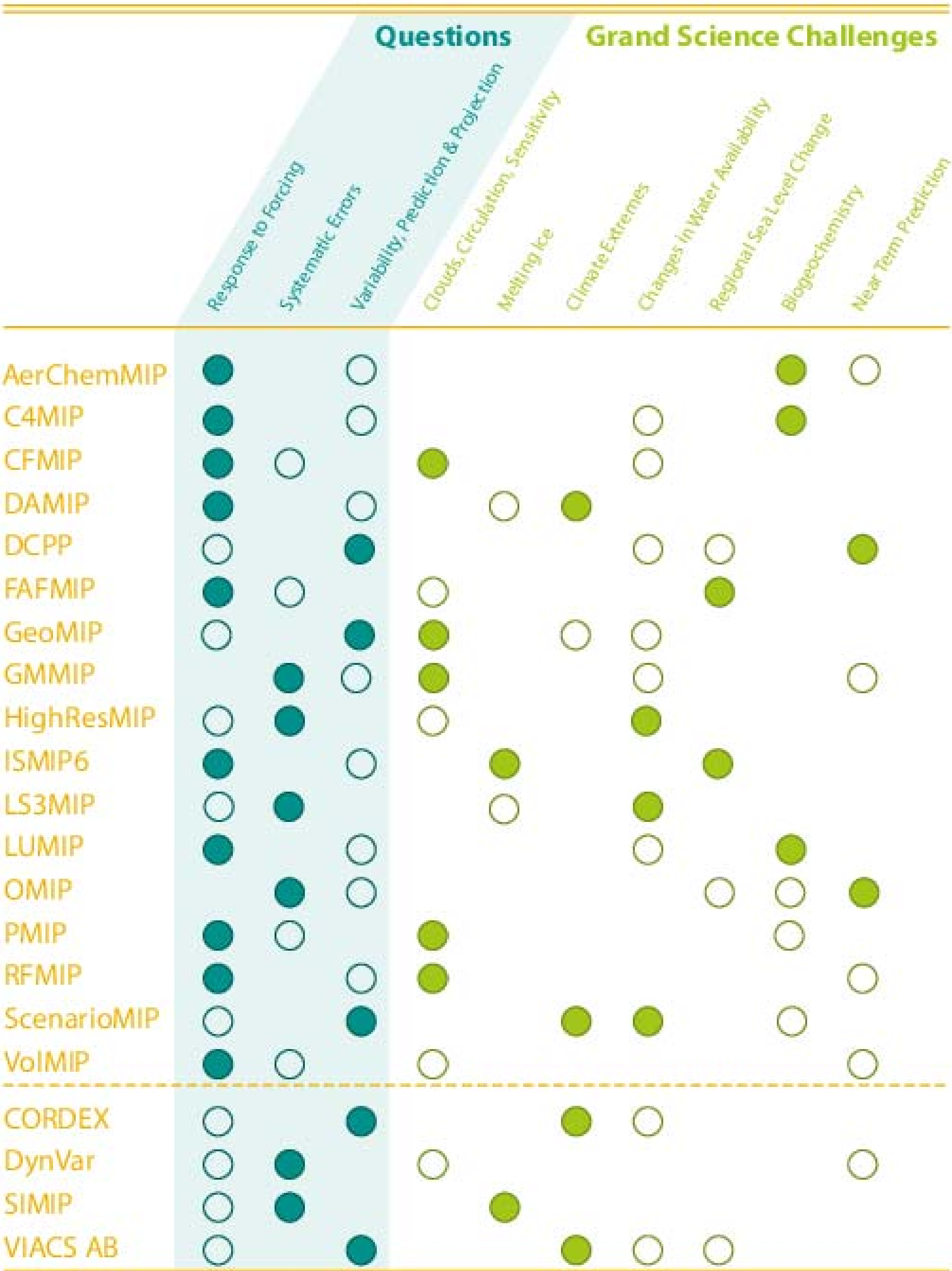

Figure 3. Contributions of CMIP6-Endorsed MIPs to the three CMIP6 science questions and the WCRP Grand Science Challenges. A filled circle indicates highest priority and an open circle, second highest priority. Some of the MIPs additionally contribute with lower priority to other CMIP6 science questions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

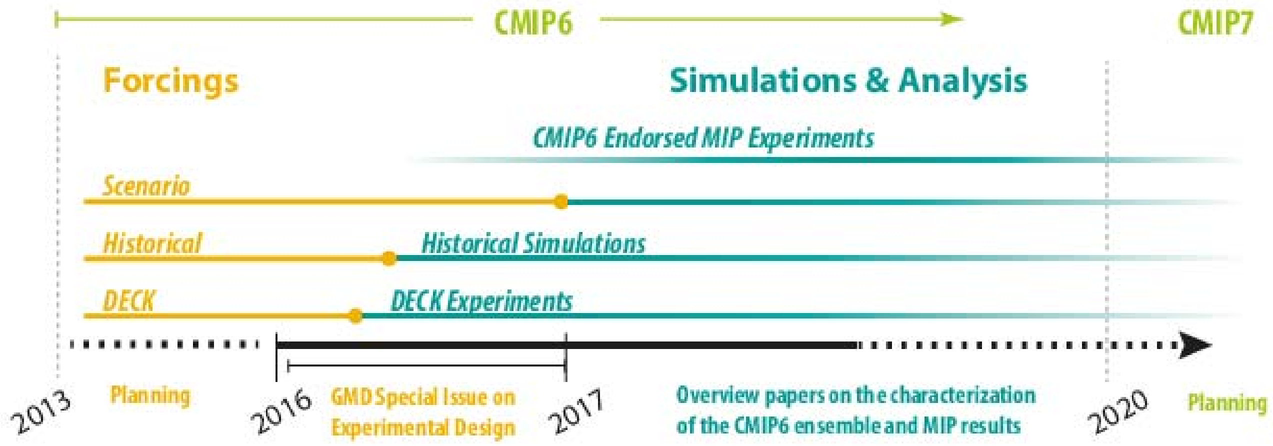

Figure 4. CMIP6 timeline for the preparation of forcings, the realization of experiments and their analysis.