Geosci. Model Dev. Discuss., 8, C4169–C4181, 2016
www.geosci-model-dev-discuss.net/8/C4169/2016/

Geoscientific Model Development Discussions

Interactive comment on "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organisation" *by* V. Eyring et al.

V. Eyring et al.

veronika.eyring@dlr.de

Received and published: 11 March 2016

Reply to Gavin Schmidt (Referee #1)

We thank the reviewer for the helpful comments. We have now revised our manuscript in light of these and the other comments we have received. A pointwise reply is given below.

This paper describes the organization and choices being made in preparation for Phase 6 of the Coupled Model Intercomparison Project. CMIP has been

an enormously successful set of projects, and yet has always failed to quite match the full expectations of the community. Thus at each stage, improvements and enhancements have been proposed, some of which have been implemented and some which have not. This iteration is no different, and the increasing scope of the proposals have necessitated a radical overhaul in the organization which is basically reported here.

The ambition of the project is commendable, but it is to be expected that implementation will inevitably fall short. Some of these issues are very predictable and I mention below a few of the ones I can foresee. The biggest problem is of course the reliance on ESGF for data delivery, of which more below.

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.

AMIP: With the large changes in the Arctic over the AMIP period, particularly in ice thickness, modellers may need to start offering sea ice thickness as well as concentration as an input field. Has this been discussed/considered?

We agree that some models now might make use of sea ice thickness. To our knowledge, however, there is not a single standard sea ice thickness monthly mean data set

agreed upon by the international community as a standard. There is a “synthetic” sea ice thickness data set that has been used in the past, but it is not based on any direct observations (see http://www-pcmdi.llnl.gov/projects/amip/RESOURCES/synice.php). Until a “reference” dataset has been developed, we do not expect to supply sea ice thickness for AMIP simulations.

piControl: (p10563) specification of land use components (crops/pasture/irrigation) also have to be set to 1850 conditions. Background volcanic is best set to the mean the 1850-1900 period rather than the open-ended full period - since there is in fact a long term trend related to volcanic forcing (i.e. PAGES2K and associated papers).

Land use: The sentence ’Unless indicated otherwise (e.g., the background volcanic forcing), experiment conditions should be representative of Earth ca. 1850.’ is implying that this should be done for all forcings. However, we now explicitly specify the land use component in a separate bullet: land use is set to 1850 conditions but land use changes are not included in *piControl*.

Background volcanic: The average volcanic RF over the CMIP historical period is quite close to the average over the last millennium in the forcing dataset of Crowley (2000); they are -0.22 Wm^{-2} and -0.18 Wm^{-2} respectively (Gregory et al., 2013). Of course there is uncertainty about this but the similarity means that it is reasonable to use the historical period average in *piControl* on the assumption that it is typical. Given the uncertainty in the observations this assumption seems okay and it has the advantage of being well-defined, implying that the model will by design not have a volcanic spin-up drift in ocean heat content during the historical period. This can also be tested by running historicalNat (assuming that solar forcing has a smaller long-term effect) or historicalVol (even better, but not part of anyone’s Tier 1). We therefore stick to the historical mean period, but have added additional explanation to the text.

Historical Simulation: The CMIP protocol should not be limiting forcings, or specifying what forcings groups have to use. Providing input to help groups without their own capacity to generate ozone datasets etc. is of course helpful, but since the historical runs are the most requested and the most likely to be compared to observations, groups must be free to choose to use their best estimates of all changes that they think important. For instance, orbital forcing is small, but to maintain coherence with past1000k runs, should be included. Irrigation, black carbon on snow, anthropogenic dust, direct heating etc. might all be possible forcings next time around for certain groups and this should not be precluded from the design. Similarly, facilities must be made to allow for variations in forcing datasets as a function of real uncertainty, for instance in aerosol composition and distribution through time. The authors should explicitly acknowledge this here, and in the upcoming specific paper related to that experiment.

The forcing datasets for the DECK and the CMIP6 historical simulations will indeed be made available through the ESGF for common use in the simulations. One reason to do this is obviously that it is convenient for the participating modelling groups that datasets are provided by the corresponding experts since it is a lot of work to produce forcings datasets. We don’t believe it is helpful to purposefully conflate uncertainty in the specification of forcing with uncertainty in the response to a given forcing, and rather encourage groups to sample the latter through supplementary simulations.

It is true that in the future other forcings might be required. But it could also be the other way round, some of the forcings that are provided as boundary condition to the runs might disappear since they are simulated interactively by the models. In any case, any deviations from the specified forcings as well as any additional forcings used

in the models should be documented, as we state in the manuscript.

The last point on uncertainty will be addressed by some of the MIPs (e.g. DAMIP and DCPP). We have also asked the groups who produce the forcings to discuss uncertainty in the forcing datasets in their contributions to the GMD Special Issue.

The discussion on page 10547 has been changed and extended in response to this comment.

In the section A1.2, the authors call for a single 'HistoricalMisc'/DAMIP run to be done as well as the historical simulation ('Nat forcing only'). I don't disagree that this is useful, but it elevates the Tier 1 of DAMIP above all other MIPs, and I'm not sure that is sensible. (Additionally, why is this in the description of piControl and not in section A2?). If any MIP should be so elevated, it should be RFMIP (see below).

We discuss this here because it directly relates to the choice of volcanic forcing in the control run. We have reworded the paragraph and now the reference to DAMIP is just informational and the explanation of the value to understanding the *piControl* and historical runs is the focus.

In Section 3 we have now expanded the discussion of the problem in CMIP5 that forcing was not well quantified and we now specifically encourage modeling groups to do the most important of the RFMIP experiments.

p10567 line 22. Is it not possible to move this to 2016?

We have extensively discussed this with all groups involved, in particular those that produce the forcing datasets mostly in form of voluntary and unfunded work. It

is already a success that they can provide forcings extending until the end of 2014, made available early in 2016. Forcings for the CMIP6 historical simulations are as much as possible based on observations. In the CMIP6 timeline, DECK and the CMIP6 historical simulations can be run in 2016 (see Figure 4). It will take substantial time to update the forcings to end of 2015 which would cause a significant delay of CMIP6 from the start. It would also delay the harmonization with the future scenarios that takes several months, so this would just mean that the entire CMIP6 process is substantially delayed. Therefore no, this cannot be moved. However, several groups are working towards updating their forcings on a more regular basis, and modelling groups might get updates during the next coming years to extend the historical simulation. But the "start of the future in CMIP6" will remain to be 2015 and the period from 2015 through near-present will be referred to as "historical-extension" as said in the manuscript.

p10568. Is there a recommendation for the interval to use between successive ICs? i.e 20 years? 30 years? The term 'longer' on p10567 is not well-defined.

The statement as stands indicates that the larger the interval between ICs the more independent the resulting simulations will be. It is difficult to say much beyond this. There are many considerations, both practical and scientific, that might affect the interval. If, for example, a model has a strong 50 yr oscillation in the AMOC then using a 50 yr interval could lead to problems. It is therefore hard to give a general recommendation but we request that groups document what they did and why and have added this recommendation to the text.

MIPs: I strongly support the panel's decision to move towards a federation of the MIP organization since it draws in a far wider community of interested parties than just the modelling groups or the CMIP panel. But I am concerned about RFMIP being run as a separate project. One of the key missing analyses

in CMIP5 was a coherent test of the forced response across the ensemble. This was hampered because while the specified input files or concentrations over time of atmospheric constituents were available, exactly what the forcing related to those changes was not. The forcing in any specific model depends on the radiative transfer code, the background climatology of water vapour and clouds, and on many model-specific indirect effects and the specific forcing definition. To my knowledge, only GISS have made available full radiative forcing diagnostics for their CMIP5 runs (both iRF and ERF) (Miller et al, 2014; Marvel et al, 2015) and given the importance of this for judging responses, this should be greatly extended in CMIP6. Thus of all the MIPs, RFMIP should be very tightly coordinated with the historical simulations, and indeed, the RF for every Historical run should be archived as soon as possible afterwards.

We had extensive discussions with the modelling groups and MIP co-chairs on the design of CMIP6, and in particular on the aspect which simulations to include in the DECK. In the interest of keeping the DECK small and restricted to those simulations the modelling groups do anyway as part of their model development cycles, the four DECK simulations were selected, and it was decided to have all other experiments except the CMIP6 historical simulations in one of the CMIP6-Endorsed MIPs. We agree that the quantification of forcings and feedbacks in RFMIP (and also AerChemMIP) is essential, which is reflected in the CMIP6 design since it is one of the three main CMIP6 science questions. In our revised version, we encourage the participation in RFMIP-lite to fill gaps identified in CMIP5.

Abrupt4xCO2: p10564 line 23. "effective" ECS, since it is demonstrated at least in some cases that the Gregory method is biased low relative the true ECS (i.e. Schmidt et al, 2014).

'effective' has been added.

1%CO2: previous CMIPs called for stabilized versions (ie 1%CO2 until 2xCO2 or 4xCO2 and then constant thereafter). Has there been a specific decision to not do this? If so it should be stated.

Most important for analysis of the 1pctCO2 experiment is that the 1% CO_2 increase is continued until at least CO_2 quadrupling after 140 years. We are not aware that the stabilization period after 4xCO2 has been analysed and have decided to simplify the experiment to avoid mistakes in the setup of the simulation. The text has been extended to clarify this change.

Data requirements: Of the 3PB in CMIP5, has the panel assessed the downloading and utilization of specific diagnostics? My sense is that while some diagnostics were very heavily used - surface fields, the historical simulations etc., there were many diagnostics that were requested that never got used, not even by the people who requested them in the first place. This might be because the package as a whole was not coherent (for instance the full energy budget) or ultimately, the diagnostic was too obscure or too difficult to compare across models. While it's hard to say that these lesser-used diagnostics will never be useful, the modeling groups would benefit from this ranking as they work to prepare the diagnostic packages for CMIP6.

The way the CMIP6 Data Request is created differs from how this was done in CMIP5. All CMIP6-Endorsed MIPs were asked to specify the variables they need for their own simulations, from other MIP simulations, and from the DECK and CMIP6 historical simulations. The MIPs also commit to analyse the data they request which will avoid that large amount of data that is requested will never be looked at.

ESGF: Much of the success of CMIP6 will be tied to the usability and ac-

cessibility of the ESGF. This paper takes it for granted that this will be available. Given the intermittent access over the last 6 months, the clunky interface, the notoriously difficult scripting options for systematic downloading, and general unhappiness in the wider community, does the panel want to address a backup option? i.e. a federated set of no-frills ftp sites - one per modeling group perhaps? Ideally, we should be discussing setting up intelligent data analysis sites that sit on top of the datasets to reduce the need for downloads, but I appreciate this goes beyond the scope of this paper.

We agree that the shut-down of CMIP5 ESGF for an extended period was unacceptable. As you note, this has little to do with the experiment design for CMIP6, which is the focus of this paper. The ESGF will be described and discussed further in the WIP contribution to this Special Issue.

Additionally, a vital improvement to CMIP and an accelerator for scientific discovery would be providing an archive for derived datasets, and perhaps even code for producing that derived data. Examples would be indices such as global mean temperatures, NAO indices, NINO3.4, Max Atl. Overturning, forward modeled brightness temperatures (for MSU + SSU satellite observations), ocean heat content anomalies, etc. I have long pushed for this to be part of ESGF, but this has not happened for a variety of reasons. The CMIP panel however and the authors here should be at the forefront of making this work, and this paper would be a good place to describe their initiatives and aims in this direction.

Indeed many users would benefit from enhanced ESGF capabilities to perform server side calculations for derived variables such as global mean temperatures and precipitation, zonal or annual means. This was mentioned also in the CMIP5 survey, has been encouraged during WGCM meetings, and is something that the ESGF teams will try to establish. Details on what actually is planned to achieve for CMIP6 will be

given in the WIP contribution to this Special Issue. In addition, evaluation tools that will be installed at some ESGF nodes 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. We have added one sentence at the end of Section 3.3. that addresses this.

DOIs: To document the impacts of CMIP6, we should be ensuring that a) every simulation has a doi for the package of diagnostics at the time of deposit, and b) every paper should have a data table listing the doi's used. This will allow forward referencing for every group and simulation, allowing for much improved accountability and feedback. This did not work at all in CMIP5 (because the unscalable bottleneck of individual filelevel 'quality control' was (IMHO) a disaster) and we should be ensuring that this does not happen again. This has to be built in to the design explicitly. The only mention of DOI's in the section on p10568 for the forcing datasets and not the simulations which I find very odd. This has to be made explicit right from the get-go and it has to explicit that this will be a 'on-release' system (as opposed to a 'post QC' system in CMIP5). (Note, if the authors for whatever reason get hung up on the nature of a 'doi' for the simulation package, please replace this acronym with an identifier of their choice that is digital and refers to an object).

Good point, thanks. Again this will be further detailed in the WIP contribution, but we have added a paragraph making a few points about this to the 'Data Availability' section.

Minor edits:

p10541 line 20: will depend on THEIR scientific interests

Changed as suggested.

p10541 line 25: INTERNAL climate variability

Changed as suggested, here and in abstract.

p10542 line 9: central element -> central INPUT

Changed as suggested.

p10544 "In addition, a monolithic structure to the CMIP design tended to discourage the modelling centres from attempting to design new experiments meant to address specific scientific questions of interest to them." - this might be better phrased as a reflection of some peoples opinions, rather than an absolute truth. From our point of view, we did not feel inhibited from expanding the scope of CMIP5 experiments (via HistoricalMisc, different 'physics-versions', forcings etc.) and exploring our scientific interests. "This in turn contributed to the impression that CMIP was a service that the modelling centres provided to the broader community." - there are many reasons why the interaction is not two-way and there are a number of issues that could be proposed to deal with that. In my opinion, it has very little to do with the monolithic structure, and far more to do with the inability to track where output is used, a lack of archiving possibilities for derived data and code, and a traditional publication schedule that is so long that it makes many analyses obsolete before they are even available.

This section has been considerably revised in light of the reviewers comments.

p10544 "Third, the punctuated structure of CMIP has begun to distort the

model development process. Whereas in the past modelling centres developed models based on their own scientific goals and released model versions on their own schedule, the visibility and demands of CMIP were beginning to impose a synchronization of model development with different phases of CMIP." - this is strangely phrased. It is clear that there is a synchronisation (i.e. it hasn't just begun). Indeed, it has been this way since CMIP3. I don't see why this is considered a problem though. Indeed, without external deadlines, I fear models would almost never be released. Frankly this just seems like some people in the community are whining and it detracts from the paper.

See previous comment.

p10548 line 23. "the signal FROM THE forced responses (Li et al., 2015)." – Note here that 'forced responses' in AMIP includes forcing by SST/SIC in addition to the external forcing. The authors should be clear the term is being used differently here than elsewhere in the paper.

We have removed the explicit mentioning of forcings in this sentence and replaced it with 'to improve the signal to noise ratio'.

p10548 line 27. A word perhaps about what 'pre-industrial' means. It is not the same in this context as zero anthropogenic influence. GHGs, LU etc. are all already modified in 1850. There is ongoing discussion about defining it to be in the late 18th Century as well - but presumably CMIP is not going to move the start date for the historical runs back to 1750 to account for this.

This has been clarified in the revised version.

p10549 line 1: "External human influences on the land surface are likewise

excluded. " This cannot be true. You would have a shock to the runs if you had zero LU difference in the piControl and then suddenly jumped to 1850 conditions in the historical transient. Presumably, the authors simply mean that further transient changes to LU are not made in the piControl runs.

Thanks for spotting this. LU is set to 1850 conditions as specified now in the appendix. Land use changes are however excluded in *piControl*. The sentence has been deleted to avoid misinterpretation.

p10550 line 5. We should already be aiming to have 2015 forcing included, and for this to be updated on an annual basis.

See response to your comment on p10567 line 22 above.

p10556 line 12. use 'evaluation of the predictions' instead of 'verification of the models".

Changed as suggested.

Interactive comment on Geosci. Model Dev. Discuss., 8, 10539, 2015.