Response to Interactive comment on by H. Shiogama on “The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6” by Mark J. Webb et al.

Reviewer comments below are shown in bold and our responses are in italics.

Dear Hideo,

This paper provides a clear description of the design of CFMIP3/CMIP6. The proposed experiments and outputs are interesting and will be important contributions to CMIP6.

I have only a few minor comments.

Thank you for your careful consideration of our manuscript and for these helpful comments.

I assume that all the CFMIP experiments are CO2 concentration driven. Should ESMs turn off dynamic vegetation and chemistry schemes?

The CFMIP experiments are indeed driven by CO2 concentration rather than CO2 emissions. Many of the CFMIP experiments are based on the DECK experiments (e.g. amip, piControl, abrupt-4xCO2). Experiments such as amip-p4K and abrupt-2xCO2 should be configured consistently with the DECK experiments that they are based on.

We have added additional text at line 176 as follows:

“Most of the CFMIP-3 experiments are based on CO₂ concentration forced amip, piControl and abrupt-4xCO₂ CMIP DECK (Diagnostic, Evaluation and Characterization of Klima) experiments (Eyring et al., 2016). Unless otherwise specified below, the CFMIP-3 experiments should be configured consistently with the DECK experiments on which they are based, using consistent model formulation, and forcings and boundary conditions as specified by Eyring et al., 2016.”

Line 213 “Sea ice and SSTs under sea ice remain the same as in the amip DECK experiment.”: How should we set SSTs in grids with 50% concentration of sea ice?

We have modified the text at L268 as follows. L269, L356 also amended similarly.

“Sea ice and SSTs in grid boxes containing sea ice remain the same as in the amip DECK experiment.”

Line 263 “As such we hope that these experiments will provide useful synergies with Palaeoclimate Model Intercomparision Project (PMIP)” : If there are any experiments that are directly related to the CFMIP experiments, please specify.
We have modified this text as follows (L361):

“As such we hope that these experiments will provide useful synergies with the Palaeoclimate Model Intercomparison Project (PMIP) CMIP6 experiments (e.g. in interpreting differing cloud feedbacks between future CO₂ forced experiments and those representing the Last Glacial Maximum, as highlighted by Yoshimori et al., 2009).”

Line 302 “cloud-radiative effects are switched off in the longwave part of the radiation code”: Is the shortwave part retained?

We have modified this text as follows: (L409)

“cloud-radiative effects are switched off in the longwave part of the radiation code while retaining those in the shortwave (Fermepin and Bony, 2014).”

2.4 Abrupt +/-4% solar forced runs: Not only TSI but also spectral solar irradiance (SSI) are provided for CMIP6 (http://solarisheppa.geomar.de/cmip6). I assume that many ESMs use the SSI data for their DECK experiments. How to add +/-4% solar forcing on SSI?

We have added a line to section 2.4 which states: (L439)

“ When changing the solar constant, the shape of the spectral solar irradiance distribution should remain consistent with that in the piControl experiment.”

Line 411 piSST: Do we use the monthly mean values of each year of piControl?
Monthly mean climatology would lead to better S/N.

We have added the following to Section 2.7 (L550)

“These are forced with monthly- and annually-varying monthly mean SSTs and sea ice, which reproduce regional precipitation patterns more accurately than is possible using climatological SST forcing (Skinner et al., 2012).”

Line 550 “allowing a detailed evaluation clouds”: allowing a detailed evaluation of clouds?

We have corrected that (L711)
Response to Interactive comment on by F. Brient on “The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6” by Mark J. Webb et al.

Reviewer comments below are shown in bold and our responses are in italics.

Dear Florent,

This paper summarizes the objectives of CFMIP and the contribution of CFMIP-3 to CMIP6. CFMIP helps to explain the spread of cloud feedbacks, adjustments and processes across climate models. This updated contribution goes a step forward and suggests additional experiments to allow the community to tackle in more detail the physical reasons underlying dynamical and regional biases seen in climate models. By proposing experiments that test especially the atmospheric components of climate models, CFMIP provides a relevant framework to understand and improve cloud parameterizations and processes which remain the principal sources of surface and atmospheric model biases.

First, the authors summarised well how former CFMIP/CMIP5 experiments helped to improve our scientific understanding of climate feedbacks. It thus provides a relevant background supporting the additional experiments that they advise the modelling groups to perform. I particularly appreciated (1) the will to promote the analysis of experiments when cloud radiative effects are switched off, (2) the pertinent time slice experiments aiming to understand regional climate responses and (3) the encouragement of a more extensive distribution and use of physical tendencies which are a signature of the atmospheric components of climate models.

Thank you for your careful consideration of our manuscript and for these helpful comments.

Below, I have listed a number of minor points which might be addressed to clarify the text (if the authors find them useful)

- Some acronyms are not defined : AOGCM (l.83), GCM (l.92), RFMIP (l.378), TOA (l.637), PMC (l.777)

We have defined these in the revised manuscript (L92, L101, L494, L802, L985.)

L. 196 : I have trouble understanding the meaning of “known answer”.

We have modified this sentence to read (L255):

“Aqua-planet simulations (and other idealized) experiments are particularly effective at highlighting model differences, for instance in the placement of the tropical rain bands, or in the representation of cloud changes with warming, as it is not possible to tune them to observations in the same way as is for more realistic configurations (e.g., Stevens and Bony, 2013).”
L. 217: The amip-future4K experiments used the CMIP3 pattern of SST increase. Is this pattern consistent with the one derived from CMIP5 models?

*We haven’t looked into this, because we consider consistency with the CMIP5 protocol to be more important than using SSTs from CMIP5 rather than CMIP3.*

*We have added the following to Appendix C (L1043)*

“We have retained the SST forcing based on the CMIP3 coupled models because we consider it more important to be able to compare CMIP5 and CMIP6 models forced with the same SST pattern than to use a pattern which is consistent with, say, the CMIP5 coupled response.”

L. 222-225 and L. 419-422: I’m a little bit confused about all 4xCO2 experiments. The amip4xCO2 experiment involves the CO2 effect on the atmospheric component and land warming without the vegetation feedback. It is thus “equivalent” to the piSST-4xCO2-rad experiment listed in section 2.7 (but not to piSST-4xCO2). I guess abrupt4xCO2 takes into account the vegetation feedback. So, the amip4xCO2 experiment should be named amip4xCO2-rad, doesn’t it?

*We agree that this would be a more consistent naming of this experiment. However, we think that the experiment descriptions are clear. Unfortunately however we understand that CMIP6 experiment names have now been finalised and propagated to the ESG and so it is not now possible to change them.*

L.257-264: You could also add the reference “Block and Mauristen (13) JAMES - Forcing and Feedback in the MPI-ESM-LR coupled model under abruptly quadrupled CO2”, which highlights the utility of diverse amip-pXk and abrupt2xCO2 experiments.

L. 288-299:

*We have added a citation to this paper in section 2.5. (L479)*

(1) It is thus right that LW effects are the most important contributor to cloud atmospheric radiative effects, and SW effects play a minor role (e.g. Takahashi 09). Nevertheless, local SW cloud effects exist (Pendergrass and Hartmann, 2014). It might thus be interesting to point this fact out in the text and leave the discussion about SW effects sufficiently open.

*We have added the following to section 2.3 (L413)*

“We note that the presence of clouds does affect the shortwave radiative heating of the atmosphere, although this is a much smaller effect than its longwave equivalent (e.g. Pendergrass and Hartmann, 2014).”

(2) Since only LW radiative effects are removed, does it mean that models still have a SW cloud feedback but no LW cloud feedback?
Yes. We have clarified this by adding the following to section 2.3 (L413):

“In this configuration, the models will have a shortwave cloud feedback but no longwave cloud feedback.”

(3) “and the radiation code only”. Does this mean that, for instance, a boundary-layer parameterization based on LW cloud-top radiative cooling continues to see LW effects?

We have added the following comment to section 2.3 (L418):

“Care should also be taken to remove the effects of cloud on any longwave cooling used in other model schemes (e.g. turbulent mixing) if these are calculated independently of the radiation scheme.

L.326-328: Contrary to CO2 effects, the radiative forcing of solar insolation depends on latitude. Is this dependency taken into account when the authors state that a 4% change results in a “radiative forcing of a similar magnitude to that due to CO2 quadrupling”?

Yes this has been taken into account. We have modified the text as follows to make it clear that this gives a similar magnitude in global mean forcing (L448)

“...resulting in a global mean radiative forcing of a similar magnitude to that due to CO2 quadrupling.”

L. 482: Single Column Model already defined line 91-92.

Duplication removed.

L. 600-601: Is it normal that “cfDay-2d” is named by CMIP5 and not CFMIP? Why is there no CMIP5 or CFMIP prefix for “cfDay-3d”?

The different prefixes represent detail in the formal data request which is not required here. In the manuscript We have deleted the prefixes to avoid confusion, and will add the following sentence (L782-802)

“(Please note that in the full data request these variable groups are in many cases split into a number of sub-tables. As noted above, the formal data request provides the definitive specification of the model outputs.)”

Fig.1 : The DECK is written in the caption but not highlighted in the graph.

We have updated the figure and caption to be consistent in this regard (See Figure 1 and L1052)
Fig.1: I consider lwoff experiments as part of the “Clouds” analysis. You may consider making the arrow longer.

*We have done this. (See Figure 1)*
Response to Interactive comment on by Anonymous Referee #1 on “The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6” by Mark J. Webb et al.

Reviewer comments below are shown in bold and our responses are in italics.

Dear Referee,

In this paper, the authors state the goals and motivation of CFMIP, review the major accomplishments of previous CFMIPs, and describe the proposed experiments and diagnostics for CFMIP3. The coordinated experiments proposed for CFMIP3 will target a number of outstanding questions for which previous model intercomparisons were not equipped to address, in addition to sustaining a number of highly useful experiments from earlier MIPs that will help to characterize and understand the response of the CMIP6-generation of models to external forcing (in addition to help quantify the forcing itself). Advanced diagnostics (e.g., satellite simulators and high frequency tendency terms) will aid in dissecting model results, and the authors have proposed that they be used more broadly (e.g., COSP turned on for longer durations and in more experiments). The emphasis on (mostly) atmosphere-only simulations in CFMIP3 should hopefully make it appealing for modeling centers to take part in several of the experiments despite the high volume of requested diagnostics.

The scientific questions to be addressed by CFMIP3 are well articulated and the various proposed experiments seem well designed to address these questions, and will advance the community’s knowledge. The presentation of the paper is not particularly concise, nor are the figures particularly insightful, but the writing is clear and overall the presentation seems appropriate for a paper proposing a model intercomparison project. Thus, in my opinion the manuscript represents a substantial contribution to modeling science within the scope of Geoscientific Model Development, and I recommend publication following consideration of some minor comments detailed below.

Thank you for your careful consideration of our manuscript and for these helpful comments.

Specific Comments:

*piSST and a4SST: It is not clear to me whether (a) monthly- and annually-varying SSTs from the relevant 30 years in the piControl run, or (b) a monthly-resolved climatology of SSTs over the relevant 30 years in the piControl run are prescribed in piSST. Same question for a4SST.

We have modified the text to read ‘monthly- and –annually varying SSTs....’ in the descriptions for piSST, a4SST, a4SSTice and a4SSTice-4xCO2. For consistency we also refer to the AMIP SSTs and sea ice in the amip-a4SST-4xCO2 description as monthly- and –annually varying. (L566, L571, L576, L583, L585, L587, L596).

*a4ip-piForcing: I’m curious whether there was any interest in performing a similar experiment, but with present-day (rather than preindustrial) forcing held fixed. An example application that occurs to me is that a model with large aerosol-cloud interactions would
presumably have brighter clouds with smaller droplets downwind of aerosol sources if
the forcing were fixed at present-day, and its temperature-mediated changes in clouds
might therefore be different than that occurring in an atmosphere with fewer aerosols.
Having these two experiments would allow one to explore this effect (and others related
to other forcing agents).

This is an interesting idea. However, to be recommended for CMIP by CFMIP, we generally require
new experiments to have been piloted and ideally written up with at least one GCM previously. If
such an experiment can be demonstrated to provide new insights which are relevant to the objectives
of CFMIP then we will certainly consider it in the future.

*Given its implications for understanding apparent state- or time-dependent changes in
effective climate sensitivity, I was a little surprised to see no experiments designed to
explore causes of nonlinearity in the Gregory plot, perhaps using warming experiments
in which the SST pattern is fixed in time (with various patterns), similar to those con-
ducted in Andrews et al, J. Climate (2015). Is there a reason for not proposing these,
or are these effects already captured in other proposed experiments?

We do consider the causes of non-linearity in abrupt4xCO2 experiments to be an important area to
be investigated. The experiments in Andrews at al (2015) were based on actual SSTs from individual
models. Pilot studies are ongoing to devise future experiments for CFMIP relevant to this question
based on SST pattern responses more representative CMIP5 ensemble mean. We plan to organise a
pilot intercomparison based on this, although this might initially be arranged informally within CFMIP
rather than as part of CFMIP/CMIP6.

*Line 126: should be “
...
meetings AND international
...
”
We have corrected this (L157)

*Line 426: “a4SST-4xCO2-all” should be “a4SSTice-4xCO2-all”. There may be other
instances of this; please verify that they are also changed.

This was incorrectly named—the correct name is in fact a4SSTice-4xCO2. Now corrected (L587,589).

*Line 512: What is the reason for dispensing with the cloud tendency terms in CFMIP3?

We have added the following at L680:

“We have dispensed with the cloud water tendency terms because these have been less widely used
than the temperature and humidity tendencies.”
*Lines 597-608: it is not clear to me why some of these have a CMIP5 prefix, a CFMIP prefix, or no prefix at all (cfDay-3d). Why would a CMIP5 prefix be appropriate at all?

The different prefixes represent detail in the formal data request which is not required here. In the manuscript we have deleted the prefixes to avoid confusion, and will add the following sentence (L782-802)

“(Please note that in the full data request these variable groups are in many cases split into a number of sub-tables. As noted above, the formal data request provides the definitive specification of the model outputs.)”

*Line 611: should be “
... for 140 years OF the piControl
... ”

We have corrected that (L796).

*Appendix A: I don’t understand what is meant by “Lead coordinator”. Is this the person who has “first dibs” on writing papers based on these experiments? Are interested investigators expected to contact this person to avoid duplicating work that others are doing with output from these experiments?

We have added the following to Appendix A (L933)

“We plan to scientifically analyze, evaluate and exploit the proposed experiments and diagnostic outputs, and have identified lead coordinators within CFMIP for different aspects of this activity. The lead coordinators are responsible for encouraging analysis of the relevant experiments as broadly as possible across the scientific community. While they may lead some analysis themselves, they do not have any first claim on analysing or publishing the results. All interested investigators are encouraged to exploit the data from these experiments. While investigators may wish to liaise with the lead coordinators to avoid duplicating work that others are doing, this is not a requirement.”

*Figure 1: I think “CMIP6” should be deleted before “historical”. If it is supposed to be there, I don’t understand why it is only there.

This is the correct naming. Please see Eyring et al. for the justification.

It is also unclear to me why the “Clouds” arrow only extends as far as abrupt-0p5xCO2. I think both the clouds arrow and the circulation and precipitation arrows should include all experiments, but in that case, what is the point of showing them?
In response to a comment from F. Briant, we have extended the cloud arrow to encompass the lwoff experiments. The timeslice experiments in the bottom group are designed to look at circulation and precipitation responses rather than cloud feedbacks. (See updated Figure 1)

*Table 1: should be “This IS a single
...
*

We have corrected that (See Table 1).

*Table 3: Several of the observational datasets end many years ago despite the fact that these satellites are still in orbit. Are there plans to extend these records, especially since the AMIP runs end in 2015?

We have added the following end of Section 3.2 (L814):

“These datasets are periodically updated to include more recent data from the relevant satellites, many of which are still operational. Please refer to the CFMIP-OBS website for updates.”
Response to Interactive comment by Anonymous Referee #2 on "The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6" by Mark J. Webb et al.

Reviewer comments below are shown in bold and our responses are in italics.

Dear Referee,

This manuscript outlines the CFMIP-3 experimental strategies, the associated model output, and the motivation and anticipated results of these experiments. Overall the manuscript is clearly written and accurately summarizes the plans for CFMIP-3, and in many ways represents more of a review of past CFMIP achievements, which in itself is a useful contribution. I recommend acceptance with only minor revisions as outlined in my suggestions below.

Thank you for your careful consideration of our manuscript and for these helpful comments.

The authors are very generous in their citations of other work, which is commendable, but it detracts from the readability of the manuscript. I recommend the authors consider focusing on a few select highlights of the previous CFMIPs that illustrate the main contributions, rather than attempting an exhaustive summary of everything that's been learned from CFMIP experiments. In the current form, it's difficult to identify what the key contributions of CFMIP have been.

We appreciate that the many citations do make the manuscript difficult to read in places. We are glad that the review of the main CFMIP achievements is appreciated, and agree that this could be achieved with fewer citations. However, we also consider it important to communicate the full breadth of studies arising from CFMIP, as this will we think help to inform the decisions made by modelling groups on which CFMIP experiments to perform and which model outputs to provide. Following guidance in the subsequent interactive comment from the Editor (Julia Hargreaves) we have reduced the number of citations in the introduction, in particular where there is duplication with Section 2. Throughout, where several citations are made together, we have broken them into smaller groups as suggested to give the reader a better idea of what distinguishes them. (See for example L230-239, L248, L673).

Section 2.1 reads more as a review of all previous studies that used CFMIP data, rather than a description of the CFMIP-3. I would recommend moving much of this to the previous section which reviews past CFMIPs and identify any changes/deletions from the past CFMIPs before then proceeding to describe the new additions to the CFMIP-3 set of experiments.

We appreciate that there is some duplication between the text in the introduction and in Section 2.1, in particular in the case of citations. We have addressed this by modifying the text in the introduction, as described above. We considered the referee's suggestion to move the bulk of this to the introduction, and to then to describe these Tier I experiments in terms of changes/deletions compared to those in previous CFMIPs. However, as pointed out in the subsequent comment by the Editor, it is important that, as a MIP documentation paper, we document the experiments in such a way as to allow a third party could set up each run from the information provided. We think that recapping on the CFMIP-2 experiment protocol in the introduction and then introducing aspects of the CFMIP-3 protocol as changes relative to this would make it harder for modelling groups to use this paper as the definitive specification for the CFMIP-3 experiments, and so prefer to leave the structure as it is presently.
It would also be useful to define what a "DECK" is.

We have modified the text at L175 to read:

"Most of the CFMIP-3 experiments are based on CO2 concentration forced amip, piControl and abrupt-4xCO2 CMIP DECK (Diagnostic, Evaluation and Characterization of Klima) experiments (Eyring et al., 2016)."

There is rightfully considerable attention within CFMIP devoted to isolating and quantifying the fast adjustments. However the fast adjustments arise from both atmospheric radiative heating changes and land warming.

We agree. We checked the manuscript, and all references to tropospheric adjustments do also refer to land warming.

It would be useful to isolate these contributions (beyond the use of aqua planets, whose utility in quantifying CGCM feedbacks is a little over sold here IMO). Has there been any efforts to develop experiments for this? If not, this issue might warrant some discussion in reference to the experiments designed to quantify adjustments.

We agree that experiments designed to separate the effects of land warming and atmospheric heating in realistic experiments would be useful. However we are not aware of any published studies which demonstrate a way to do this. To be recommended for CMIP by CFMIP, we generally require new experiments to have been piloted and ideally written up with at least one GCM previously. If such an experiment can be demonstrated to provide new insights which are relevant to the objectives of CFMIP then we will certainly consider it in the future.
Response to Interactive comment on by A. Voigt on “The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6” by Mark J. Webb et al.

Reviewer comments below are shown in bold and our responses are in italics.

Dear Aiko,

The authors provide a concise and well-written presentation of the CFMIP experiments proposed for CMIP6, which will continue the successful CFMIP activities over the last 15 years. I enjoyed reading the paper, in particular the historical context given in the introduction, and find that it nicely presents the scientific motivation and chosen simulation strategy at a level amenable to both CFMIP experts and climate scientists with other backgrounds. I recommend publication in GMD after my following minor comments have been addressed.

Thank you for your careful consideration of our manuscript and for these helpful comments.

Line 217, amip-future4K simulations: Why is the CMIP3 SST pattern used and not an updated pattern from CMIP5 AOGCM runs?

We have added the following to Appendix C (L1068):

“We have retained the SST forcing based on the CMIP3 coupled models because we consider it more important to be able to compare CMIP5 and CMIP6 models forced with the same SST pattern than to use a pattern which is consistent with, say, the CMIP5 coupled response.”

Line 227: I am very glad to hear that the CMIP-3 aquaplanet simulations will be extended to 10 years. This will be beneficial for studies of extratropical dynamics, for which internal variability is larger than in the tropics.

Thank you.

Line 238, amip-m4k simulations: I am wondering to what extent some models might have problems with SSTs below freezing? Maybe this might require code changes in some models in case they employ a fixed lower threshold for the SST used in the calculation of surface fluxes? Such a problem would, of course, not occur for the p4K simulations?

We have added the following at L356:

“In models which employ a fixed lower threshold near freezing for the SST used in the calculation of the surface fluxes, this should ideally also be reduced by 4K.”

Lines 279: The authors might consider to also refer to Voigt and Shaw (2015, Nature Geoscience) here for the extratropical circulation. The study showed that cloud-
radiative feedbacks contribute substantially to the poleward jet shifts under 4K warming in aquaplanet simulations.

We have added that reference at L379.

Line 266, lwoff experiments: Just an idea, but I though it’s worthwhile bringing it up here: While the surface cloud effect is stronger in the shortwave than the longwave domain, the longwave can still substantial. I am wondering whether an experiment with clear-sky heating in the atmosphere and all-sky heating at the surface would be even better to isolate the effect of atmospheric cloud-radiative heating. I suspect it’s too late to change the experimental protocol, and maybe there is a reason why lwoff is still better. If so, it might be worthwhile to briefly discuss this.

This is an interesting idea thank you. However, to be recommended for CMIP by CFMIP, we generally require new experiments to have been piloted and ideally written up with at least one GCM previously. The lwoff are experiments currently proposed are very similar to those piloted by Fermepin and Bony, 2014, and technically easier to implement than what is proposed. If such an experiment can be demonstrated to provide new insights which are relevant to the objectives of CFMIP then we will certainly consider it in the future.

We have added the following to the manuscript at L417:

“An alternative method (proposed by A. Voigt) was also considered, in which clear-sky heating rates would be applied in the atmosphere while retaining the all-sky fluxes at the surface. Although this approach would potentially isolate the effects of cloud heating in the atmosphere more cleanly than the lwoff experiments proposed here, it is yet to be demonstrated in a pilot study, and is considered more technically difficult to implement than the lwoff experiments, which are very similar to those piloted by Fermepin and Bony, 2014.”

Line 342: Non-linearity was also shown in the CMIP5 ensemble by Meraner et al. (2013, GRL, doi:10.1002/2013GL058118). Meraner et al. showed non-linear climate sensitivity across the multi-model CMIP5 ensemble, whereas the other cited work used single models if I am not mistaken. So maybe worthwhile including here?

We have added that reference at L455:

Line 368: Maybe specify the reason why the CFMIP2/CMIP5 runs did not allow such an estimate. I.e., I assume that one would use SST-driven simulations for this and that the usual amip period is too short to reliably calculate feedbacks?

We have modified line to read (L485):

“The previous CFMIP-2/CMIP5 design was unable to diagnose the time-variation of feedbacks of explicit relevance to the historical period, because this requires the removal of the time varying forcing.”
Sect. 2.7: The time slice experiments ask an interesting question but given that 8 experiments are demanded, I was wondering how they ought to be combined to answer the questions in mind. Maybe the authors can give an example?

We have added two examples in the text of how these experiments are combined at L576:

“The time slice experiments can be combined in various ways to isolate the climate response to each individual aspect of forcing and warming. For example the response to SST pattern change is given by taking the difference between a4SST and piSST-pxK, and the plant physiological response is found by taking the difference between piSST-4xCO2 and piSST-4xCO2-rad.”

Line 487: I would be curious to know about the reasons to no longer ask for cfSites output in the aquaplanet ensemble and amip-future4K. Is it the lack of observational data to compare to, or a choice to avoid asking for too much data?

We have amended the manuscript at L640 to read:

“We have dispensed with the cfSites outputs in the aquaplanet and amip-future4K experiments because these have been less widely used compared to those from the other experiments.”

Line 685: and --> an

We have amended that.

Figure 1: Why does the vertical cloud bar on the right side not include the lwoff simulations?

This point was also raised by F. Brient. We have extended the arrow to include the lwoff experiment (See Figure 1).

For some of the proposed simulations the link to clouds, which are the prime motivation for CFMIP, is not very evident and maybe could be made clearer? I am thinking of the simulations in Sect. 2.7 (time slice experiments) and Sect. 2.5 (nonLinMIP).

The primary objective of CFMIP is to inform future assessments of cloud feedbacks through improved understanding of cloud-climate feedback mechanisms and better evaluation of cloud processes and cloud feedbacks in climate models. However, the CFMIP approach is also increasingly being used to understand other aspects of climate change, and so a second objective has now been introduced, to improve understanding of circulation, regional-scale precipitation, and non-linear changes. For this reason, not all experiments need to be relevant to clouds. We have modified the text in the abstract, introduction and conclusions to state this explicitly (L26,L127,L849).
Response to Interactive comment by J. C. Hargreaves (Editor) on “The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6” by Mark J. Webb et al.

Editor comments below are shown in bold and our responses are in italics.

Dear Julia,

This seems to me to be a pretty good MIP manuscript. You do have the advantage that the protocols for the experiments are relatively simple to describe, but I still think it very well organised.

Thank you.

I haven’t checked all the protocols - just a couple that I am particularly interested in - but the information seemed complete for those. However, please do check through your revised manuscript to make sure that a third party could set up each run from the information provided.

We have done this, and have made a few changes to clarify some issues:

We have updated the amip-future4K experiment definition at L274 to include the sentence:

“Care should be taken to ensure that SSTs are increased in any inland bodies of water and near coastal edges, for example by linearly interpolating the provided warming pattern dataset to fill in missing data before re-gridding to the target resolution. “

We have inserted the word ‘open’ at L562 into the sentence:

“The magnitude of the uniform increase is taken from each model’s global, climatological annual mean open SST change between abrupt-4xCO2 and piControl (using the mean of years 111-140 of abrupt-4xCO2, and the parallel 30-year section of piControl).”

We have also re-written the text between L496-500 to make it clearer:

“Time-varying feedbacks in the amip experiment could alternatively be diagnosed by subtracting a time-varying radiative forcing diagnosed from RFMIP experiments. However, the amip-piForcing approach has the benefit of diagnosing the time-varying feedbacks over the full 1870-present period rather than the last 36 years, and does so with reference to a single experiment, which reduces noise compared to that which would be present with a double difference of the amip experiment and two RFMIP experiments.”

We have updated the information on which versions of COSP are available to reflect recent developments (L778):


“COSP is available via the CFMIP website (https://www.earthsystemcog.org/projects/cfmi). Version 1.4 is a stable code that was made available well in advance of CMIP6 at the request of the modelling groups. Small updates are required to enable some new diagnostics requested by CFMIP3/CMIP6, most notably joint histograms of particle size and optical thickness from the MODIS simulator; with these updates the code is known as version 1.4.1. Modeling centers are encouraged to update to COSP 1.4.1 to provide these new diagnostics but may provide results from COSP 1.4. Developed over the last few years, COSP 2 substantially revises the infrastructure for integrating satellite simulators in climate models. COSP 2 makes many fewer inherent assumptions about the model representation of clouds than do previous versions but contains an optional interface allowing it to be used as a drop-in replacement for COSP 1.4 or COSP 1.4.1. At the time of this writing COSP 2 is undergoing final testing in two climate models. Availability of the final version will be announced on the CFMIP website and modelling groups are free to adopt it for use in CFMIP at that time.”

We have also updated the following text in Appendix B (L1018)

“The ozone distribution is the same as used in APE and CFMIP2/CMIP5, and is derived from the climatology used in AMIP II (Gates et al., 1999), and is constant in time and symmetric zonally and about the equator. This ozone distribution is provided as a netCDF file which is archived on the Earth System Grid and available via the DOI http://dx.doi.org/10.5065/D61834Q6. Ozone values are provided up to 0.28hPa (about 60km altitude in mid-latitudes). For models with tops above this level, a high top ozone dataset is also provided, which is available via the DOI http://doi.org/10.5065/D64X5653. The ozone climatologies provided use pressure as a vertical coordinate. Most models use a sigma or hybrid vertical coordinate in pressure or altitude, which will mean that the pressure on a given model level varies in time, near the surface at the very least. Although the ozone climatology can in theory be interpolated to the pressure of each model level as it varies in time within the model, for simplicity we recommend interpolating the ozone dataset onto the model vertical grid before the experiment is performed, and then specifying ozone values which are constant in time on each model level. This vertical interpolation will require a zonally symmetric climatology of pressure on model levels which is as consistent as possible with that expected in the aqua-control experiment. This could for example be produced by initially running a test version of the aqua-control experiment with an ozone climatology taken from a more realistic model configuration such as the AMIP DECK experiment.”

The remaining peculiarity is the reference to boundary conditions that will become available through other papers in this special issue - are there now references that can be provided for these papers?

We have added the following at L176:

“Most of the CFMIP-3 experiments are based on CO2 concentration forced amip, piControl and abrupt-4xCO2 CMIP DECK (Diagnostic, Evaluation and Characterization of Klima) experiments (Eyring et al., 2016). Unless otherwise specified below, the CFMIP-3 experiments should be configured consistently with the DECK experiments on which they are based, using consistent model formulation, and forcings and boundary conditions as specified by Eyring et al., 2016.”

The thing I spotted in the reviewers’ comments that I am unsure about is the suggestion
to abbreviate the citations to increase readability. Here’s an example,
"Temperature and humidity tendency terms in particular have been shown to be useful for understanding the roles of different parts of the model physics in cloud feedbacks and adjustments (Kamae and Watanabe 2012; Williams et al., 2013; Webb and Lock 2013; Demoto et al., 2013; Sherwood et al., 2014; Ogura et al., 2014; Brient et al., 2015)"

I generally don’t like the idea of reducing the citations, but that is an awful lot of references all apparently showing the same thing! As a reader I’d want to know what the difference is between these papers, and which one I should look up in order to learn about the thing I am specifically interested in. The obvious solution would be to add a little more description, so that the reader has more knowledge about the content of the references. Doing so will make the manuscript longer, which could get out of hand, but maybe there is a middle way which produces a more readable and more useful manuscript.

We have addressed this in the manuscript as follows. We have reduced the number of citations in the introduction, in particular where there is duplication with Section 2. Throughout, where several citations are made together, we have broken them into smaller groups as suggested to give the reader a better idea of what distinguishes them. Please see L55-75, L105-113, L233-240, L249-250, L656-657.

For example, in the case highlighted above, we have updated the manuscript at L656 to read:

“Temperature and humidity tendency terms in particular have been shown to be useful for understanding the roles of different parts of the model physics in cloud feedbacks (e.g. Webb and Lock 2013; Demoto et al., 2013; Sherwood et al., 2014; Brient et al., 2015) and cloud adjustments (e.g. Kamae and Watanabe 2012; Ogura et al., 2014) as well as in understanding clouds and circulation in the present climate (e.g. Williams et al., 2013; Oueslati and Bellon, 2013; Xavier et al., 2015).”

We hope that these changes strike the required balance effectively.
The Cloud Feedback Model Intercomparison Project (CFMIP) contribution to CMIP6.

Mark J. Webb1, Timothy Andrews1, Alejandro Bodas-Salcedo1, Sandrine Bony3, Christopher S. Bretherton1, Robin Chadwick1, Hélène Chepfer2, Hervé Douville3, Peter Good4, Jennifer E. Kay1, Stephan A. Klein5, Roger Marchand3, Brian Medeiros6, A. Pier Siebesma4, Christopher B. Skinner4, Bjorn Stevens16, George Tselioudis17, Yoko Tsushima1, Masahiro Watanabe12.

1Met Office Hadley Centre, Exeter, United Kingdom.
2LMD/IPSL, CNRS, Université Pierre et Marie Curie, Paris, France.
3University of Washington, Seattle, USA.
4Centre National de Recherches Météorologiques, Toulouse, France.
5University of Colorado at Boulder, Boulder, USA.
6Lawrence Livermore National Laboratory, Livermore, USA.
7National Center for Atmospheric Research, Boulder, USA.
8Royal Netherlands Meteorological Institute, De Bilt, The Netherlands.
9University of Michigan, Ann Arbor, USA.
10Max Planck Institute for Meteorology, Hamburg, Germany.
11NASA Goddard Institute for Space Studies, New York, USA.
12Atmosphere and Ocean Research Institute, Tokyo, Japan.

Correspondence to: Mark Webb (mark.webb@metoffice.gov.uk) Revised for Geoscientific Model Development (GMD) 15th October, 2016.

Abstract

The primary objective of CFMIP is to inform future assessments of cloud feedbacks through improved understanding of cloud-climate feedback mechanisms and better evaluation of cloud processes and cloud feedbacks in climate models. However, the CFMIP approach is also increasingly being used to understand other aspects of climate change, and so a second objective has now been introduced, to improve understanding of circulation, regional-scale precipitation, and non-linear changes. CFMIP is supporting ongoing model inter-comparison activities by coordinating a hierarchy of targeted experiments for CMIP6, along with a set of cloud related output diagnostics. CFMIP contributes primarily to addressing the CMIP6 questions “How does the Earth System respond to forcing?” and “What are the origins and consequences of systematic model biases?” and supports the activities of the WCRP Grand Challenge on Clouds, Circulation and Climate Sensitivity. A compact set of Tier 1 experiments is proposed for CMIP6 to address the question: “1) What are the physical mechanisms underlying the range of cloud feedbacks and cloud adjustments predicted by climate models, and which models have the most credible cloud feedbacks?” Additional Tier 2 experiments are proposed to address the following questions: 2) Are cloud feedbacks consistent for climate cooling and warming, and if not, why? 3) How do cloud-radiative effects impact the structure, the strength and the variability of the general atmospheric circulation in present and future climates? 4) How do responses in the climate system due to changes in solar forcing differ from changes due to CO₂ and is the response sensitive to the sign of the forcing? 5) To what extent is regional climate change per CO₂ doubling state-dependent (nonlinear), and why? 6) Are climate feedbacks during the 20th century different to those acting on long term climate change and climate sensitivity? 7) How do regional climate responses (e.g. in precipitation) and their uncertainties in coupled models arise from the combination of different aspects of CO₂ forcing and sea surface warming?

CFMIP also proposes a number of additional model outputs in the CMIP DECK, CMIP6 Historical and CMIP6 CFMIP experiments, including COSP simulator outputs and process diagnostics to address the following questions: 1) How well do clouds and other relevant variables simulated by models agree with observations? 2) What physical processes and mechanisms are important for a credible simulation of clouds, cloud feedbacks and cloud adjustments in climate models? 3) Which models have the most credible representations of processes relevant to the simulation of clouds? 4) How do clouds and their changes interact with other elements of the climate system?

1 Introduction

Inter-model differences in cloud feedbacks continue to be the largest source of uncertainty in predictions of equilibrium climate sensitivity (Boucher et al., 2013). Although the ranges of cloud feedbacks and climate sensitivity from comprehensive climate models have not reduced in recent years, considerable progress has been made in understanding (a) which types of clouds contribute most to this spread (e.g. Bony and Dufresne 2005; Webb et al., 2006; Zelinka et al., 2013), (b) the role of cloud adjustments in climate sensitivity (e.g. Gregory and Webb, 2008; Andrews and Forster, 2008; Kamae and Watanabe, 2012; Zelinka et al., 2013), (c) the processes and mechanisms which are (and are not) implicated in cloud
feedbacks, both in fine resolution models (e.g. Rieck et al., 2012; Bretherton et al., 2015) and in comprehensive climate models (e.g. Brient and Bony 2012; Sherwood et al., 2014; Zhao, 2015; Webb et al., 2015b), (d) the inconstancy of cloud feedbacks and effective climate sensitivity (e.g. Senior and Mitchell, 2000; Williams et al., 2008; Andrews et al., 2012; Geoffroy et al., 2013; Armour et al., 2013; Andrews and Gregory, 2014) and (e) the extent to which models with stronger or weaker cloud feedbacks or climate sensitivities agree with observations (e.g. Panullo and Tredearh, 2012; Su et al., 2014; Qu et al., 2014; Sherwood et al., 2014; Myers and Norris, 2016). Additionally, our ability to evaluate model clouds using satellite data has benefited from the increasing use of satellite simulators. This approach, first introduced by Yu et al. (1996) for use with data from the International Satellite Cloud Climatology Project (ISCCP) attempts to reproduce what a satellite would observe given the model state. Such approaches enable more quantitative comparisons to the satellite record (e.g. Yu et al., 1996; Klein and Jakob, 1999; Webb et al., 2001; Bodas-Salcedo et al., 2011; Cesana and Chepfer, 2013). Much of our improved understanding in these areas would have been impossible without the continuing investment of the scientific community in successive phases of the Coupled Model Intercomparison Project (CMIP), and its co-evolution in more recent years with the Cloud Feedback Model Intercomparison Project (CFMIP).

CFMIP started in 2003 and its first phase (CFMIP-1) organised an intercomparison based on perpetual July SST forced Cess style +2K experiments and 2xCO2 equilibrium mixed-layer model experiments including ISCCP simulator in parallel with CMIP3 (McAvaney and Le Treut, 2003). CFMIP-1 had a substantial impact on the evaluation of clouds in models and in the identification of low level cloud feedbacks as the primary cause of inter-model spread in cloud feedback, which featured prominently in the fourth and fifth IPCC assessments (Randall et al., 2007; Boucher et al., 2013).

The subsequent objective of CFMIP-2 was to inform improved assessments of climate change cloud feedbacks by providing better tools to support evaluation of clouds simulated by climate models and understanding of cloud-climate feedback processes. CFMIP-2 organized further experiments as part of CMIP5 (Bony et al., 2011; Taylor et al., 2012), introducing seasonally varying SST perturbation experiments for the first time, as well as fixed SST CO2 forcing experiments to examine cloud adjustments. CFMIP-2 also introduced idealized ‘aquaplanet’ experiments into the CMIP family of experiments. These experiments were motivated by extensive research in the framework of the aqua-planet experiment (Neale and Hoskins, 2000, Blackburn and Hoskins, 2013) and the particular finding, based on a small subset of models, that the global mean cloud feedback of more realistic model configurations could be reproduced, and more easily investigated, using the much simpler aqua-planet configuration (Medeiros et al., 2008). CFMIP-2 proposed the inclusion of the abrupt CO2 quadrupling AOGCM (atmosphere–ocean general circulation model) experiment in the core experiment set of CMIP5, based on the approach of Gregory et al. (2004), which consequently formed the basis for equilibrium climate sensitivity estimates from AOGCMs (Andrews et al., 2012). Additionally CFMIP-2 introduced satellite simulators to CMIP via the CFMIP Observation Simulator Package (COSP, Bodas-Salcedo et al., 2011); not only the ISCCP simulator, but additional simulators to facilitate the quantitative evaluation clouds using a new generation of active radars and lidars in space. CFMIP-2 also introduced CMIP5 process diagnostics such as temperature and humidity budget tendency terms and high frequency ‘cfSites’ outputs at 120 locations around the globe. In an effort less directly connected to CMIP, CFMIP organized a joint project with the GEWEX Global Atmosphere Observing System (GEOSS) (the CMIP-GEOSS project). Intercomparison of LES and SCM models to develop cloud feedback intercomparison cases to assess the physical credibility of cloud feedbacks in climate models by comparing Single Column Model (SCM) versions of General Circulation Models (GCMs) with high resolution Large Eddy Simulations (LES) models. CFMIP-2 also developed the CFMIP-OBS data portal and the CFMIP diagnostic code catalogue. For more details, and for a full list of CFMIP related publications, please refer to the CFMIP website (http://www.earthsystemcog.org/projects/cfmip).

Studies arising from CFMIP-2 include numerous single and multi-model evaluation studies which use COSP to make quantitative and fair comparisons with a range of satellite products (e.g. Kay et al., 2012; Franklin et al., 2013; Klein et al., 2013, Lin et al., 2014, Chepfer et al., 2014). COSP has also enabled studies attributing cloud feedbacks and cloud adjustments to different cloud types (e.g. Zelinka et al., 2013; Zelinka et al., 2014; Tsushima et al., 2015). CFMIP-2 additionally enabled the finding that idealized ‘aquaplanet’ experiments without land, seasonal cycles or Walker circulations are able to reproduce the essential differences between models’ global cloud feedbacks and cloud adjustments in a substantial ensemble of models (Ringer et al., 2014; Medeiros et al., 2015). Process outputs from CFMIP have also been used to develop and test physical mechanisms proposed to explain and constrain inter-model spread in cloud feedbacks in the CMIP5 models (e.g. Sherwood et al., 2014; Brient et al., 2015; Webb et al., 2015a; Nuijens et al., 2015a,b; Dal Gesso et al., 2015). CGILS has demonstrated a consensus in the responses of LES models to climate forcings and identified shortcomings in the physical representations of cloud feedbacks in climate models (e.g. Blossey et al., 2013; Zhang et al., 2013; Dal Gesso et al., 2015). The CFMIP experiments have additionally formed the basis for coordinated experiments to explore the impact of cloud radiative effects on the circulation (Stevens et al., 2012; Fermin and Bony 2014; Crueger and Stevens 2015; Li et al., 2015; Harrop and Hartmann 2016), the impact of parametrized convection on cloud feedback (Webb et al., 2015b) and the mechanisms of negative shortwave cloud feedback in mid to high latitudes (Cimpf et al., 2015). Additionally the CFMIP experiments have, due to their idealized nature, proven useful in a number of studies not directly related to clouds, but instead analyzing the responses of regional precipitation and circulation patterns to CO2 forcing and climate change (e.g. Rong et al., 2013; Chadwick et al., 2014; He and Soden 2015; Oueslati et al., 2016). Studies using CFMIP-2 outputs from CMIP5 remain ongoing and further results are expected to feed into future assessments of the representation of clouds and cloud feedbacks in climate models.

The primary objective of CFMIP is to inform future assessments of cloud feedbacks through improved understanding of cloud-climate feedback mechanisms and better evaluation of cloud processes and cloud feedbacks in climate models. However, the CFMIP approach is also increasingly being used to understand other aspects of climate change, and so a second objective has been introduced, to improve understanding of circulation, regional-scale precipitation, and non-linear changes.
This involves bringing climate modelling, observational and process modelling communities closer together and providing better tools and community support for evaluation of clouds and cloud feedbacks simulated by climate models and for understanding of the mechanisms underlying them. This is achieved by:

- Coordinating model inter-comparison activities which include experimental design as well as specification of model output diagnostics to support quantitative evaluation of modelled clouds with observations (e.g. COSP) and in-situ measurements (e.g. cSites) as well as process-based investigation of cloud maintenance and feedback mechanisms (e.g. cSites, temperature and humidity tendency terms)
- Developing and improving support infrastructure including COSP, CFMIP-OBS and the CFMIP diagnostic codes catalogue.
- Fostering collaboration with the observational and cloud process modelling communities via annual CFMIP meetings and international funded projects.

This paper describes and documents the CFMIP contribution to the current phase on the Coupled Model Intercomparison Project (CMIP6, Eyring et al., 2016). It is anticipated that CFMIP-3 will eventually be broader than what is described here, for instance including studies with process models, but for the purposes of this document CFMIP-3 should be considered to be synonymous with the CFMIP contribution to CMIP6. CFMIP-3 touches, to differing degrees, on each of the three questions around which CMIP6 is organized, with its focus on cloud feedback. CFMIP-3 is central to CMIP6’s attempt to answer the question: ‘How does the Earth system respond to forcing?’ But as illustrated in the remainder of this document, CFMIP-3 also offers the opportunity to contribute to the other two guiding questions of CMIP6. Through its strong model evaluation component it stands to help answer the question: ‘What are the origins and consequences of systematic model biases?’ CFMIP-3 will also help answer the question: ‘How can we assess future climate changes given climate variability, climate predictability, and uncertainties in scenarios?’ For example the amip-piForcing experiment proposed below will support studies relating cloud variability and feedbacks on observable timescales to long term cloud feedbacks (Andrews, 2014; Gregory and Andrews, 2016). The CFMIP-3 experiments proposed for CMIP6 are outlined below in Section 2. Section 3 describes the diagnostics outputs proposed by CFMIP for the CFMIP-3 experiments and other experiments within CMIP. We provide a summary of the CFMIP-3 contribution to CMIP6 in Section 5.

2 CFMIP-3 Experiments

The CFMIP-3 experiments are summarised in Figure 1 and Tables 1 and 2, and are described in detail below. Most of the CFMIP-3 experiments are based on CO2 concentration forced amip, piControl and amip4xCO2 CMIP DECK-Diagnostic, Evaluation and Characterization of Klima) experiments (Eyring et al., 2016). Unless otherwise specified below, the CFMIP-3 experiments should be configured consistently with the DECK experiments on which they are based, using consistent model formulation, and forcings and boundary conditions as specified by Eyring et al., 2016. Following the CMIP6 design protocol, groups of experiments are motivated by science questions and are separated into Tiers 1 and 2 (Eyring et al., 2016). It is a requirement for participation in CMIP6/CMIP-3. Tier 1 experiments are performed and published through the ESGF, to support the CFMIP-3 Tier 1 science question. Tier 2 experiments are optional, and are associated with additional science questions. Any subset of Tier 2 experiments may be performed. All model output archived by CFMIP/CMIP6 is expected to be made available under the same terms as CMIP output. Most modelling groups currently release their CMIP data for unrestricted use. Our analysis of plans for the CFMIP-3 experiments are summarised in Appendix A.

2.1 CFMIP-3 Tier 1 Experiments

Lead Questionator: Mark Webb

Science Question: What are the physical mechanisms underlying the range of cloud feedbacks and cloud adjustments predicted by climate models, and which of the cloud responses are the most credible?

Equilibrium climate sensitivity (ECS) can be estimated using an idealized AOGCM experiment such as the abrupt-4xCO2 experiment in the CMIP5 DECK, at the same time statistically separating the global mean contributions from climate feedbacks and adjusted radiative forcing due to CO2 (Gregory et al. 2004, Andrews et al. 2012). However understanding the physical processes underlying cloud feedbacks and adjustments requires diagnosis in SST forced experiments with atmosphere-only general circulation models (AGCMs), which can resolve cloud feedbacks and adjustments independently from each other and with minimal statistical noise at regional scales, while faithfully reproducing the inter-model differences in global values from the fully coupled models (Ringer et al., 2014). The ability of these AGCM experiments to reproduce the inter-model differences in global cloud feedbacks and adjustments from coupled models indicates that they do not strongly depend on different ocean model formulations or SST biases). The CFMIP-2 amip4xCO2 experiments in CMIP5, which quadrupled CO2 while leaving SSTs at present day values (Bony et al., 2011), allowed the land/tropospheric adjustment process and the cloud adjustment to CO2 to be examined in this way for the first time in the multi-model context (Kamae and Watanabe, 2012; Ringer et al., 2014; Kamae et al. 2015) in conjunction with the CMIP5 sstClim/sstClim4xCO2
experiments which were based on climatological preindustrial SSTs (Andrews et al., 2012; Zelinka et al., 2013; Vial et al., 2013). These experiments have additionally formed the basis for more in-depth studies with individual models (e.g. Wyatt et al., 2012; Kamae and Watanabe, 2013; Bretherton et al., 2014; Ogura et al., 2014). The CFMIP-2/CMIP5 amip-K and amipFuture SST perturbed atmosphere-only experiments (Bony et al., 2011) have been used to examine cloud feedbacks in greater detail (e.g. Brient and Bony, 2012; Bretherton et al., 2014; Lacagina et al., 2014; Bellomo and Clement, 2015; Webb et al., 2015b), often in conjunction with simulator outputs (e.g. Gordon and Klein, 2014; Chepfer et al., 2014; Tsushima et al., 2015; Ceppi et al., 2016) and CFMIP process diagnostics (e.g. Webb and Lock, 2013; Sherwood et al., 2014; Brient et al., 2015; Webb et al., 2015a; Dal Gesso et al., 2015). Similarly, these experiments have been used to investigate regional responses of various quantities to direct radiative forcing due to increasing CO₂ concentrations and/or increases in SST, including precipitation (e.g. Ma and Xie, 2013; Huang et al., 2013; Widlansky et al., 2013; Kent et al., 2015; Long et al., 2016), circulation (e.g. He et al., 2014; Zhou et al., 2014; Kamae et al., 2014; Bellomo and Clement, 2013; Shaw and Voigt, 2015) and stability (e.g. Qu et al., 2015).

A more idealized set of fixed SST experiments proposed by CFMIP-2 for CMIP5 (aquaControl, aqua4xCO2, and aquaK) based on zonally symmetric, fixed season ‘aquaplanet’ configurations without land have been shown to reproduce the inter-model differences in global mean cloud adjustments and feedbacks from realistic experiments surprisingly effectively (Medeiros et al., 2008; Ringer et al., 2014; Medeiros et al., 2015) as well as many aspects of the zonal mean circulation response (Medeiros et al., 2015). This indicates that those features of the climate system excluded from these experiments (i.e. the ocean, land, seasonal cycle, monsoon and Walker circulations) are not central to understanding inter-model differences in global mean feedbacks and adjustments, and demonstrates the value of aquaplanet experiments for investigating the origin of such differences, as well as differences in zonally averaged precipitation and circulation and their responses to climate change (e.g. Stevens et al., 2012; Bony et al., 2013; Oueslati and Bellon, 2013; Feng and Bony, 2014; Voigt and Shaw 2015). The aquaplanet experiments have the benefit not only of being less computationally expensive than alternative experiments (requiring only 5-10 years to get a robust signal); they are also much more straightforward to analyse, as their behaviour can mostly be characterized by examining zonal means, avoiding the analysis overhead of compositing which is generally required in realistic model configurations to isolate the various cloud regimes. Aqua-planet simulations (and other idealized) experiments are particularly effective at highlighting model differences, for instance in the placement of the tropical rain bands, or in the representation of cloud changes with warming, as it is not possible to tune them to observations in the same way as for more realistic configurations (e.g., Stevens and Bony, 2013).

The CMIP5/CFMIP-2 experiments and diagnostic outputs have thus enabled considerable progress on a number of questions. However, participation by a larger fraction of modelling groups is desired in CMIP6 to enable a more comprehensive assessment of the uncertainties across the full multi-model ensemble. Our proposal is therefore to retain the CFMIP-2/CMIP5 experiments (known in CMIP5 as amip-K, amip4xCO2, amipFuture, aquaControl, aqua4xCO2 and aquaK) in Tier 1 for CFMIP/CMIP6. These are summarised in Table 1 (the names have been changed slightly compared to the CMIP5 equivalents to fit in with a wider naming convention of CMIP6). The set up for each of these experiments is described below. (For output requirements from these and other experiments please refer to Section 3).

amip: This is a single ensemble member of the CMIP DECK amip experiment which contains additional outputs which are required both for model evaluation using COSP, and for interpretation of feedbacks and adjustments in conjunction with the amip-p4K, amip-4xCO2, amipFuture and amip-mK experiments.

amip-p4K (formerly amip-K): The same as the amip DECK experiment, except that SSTs are subject to a uniform warming of 4K. This warming should be applied to the ice free ocean surface only. Sea ice and SSTs in grid boxes containing sea ice remain the same as in the amip DECK experiment.

amipFuture-kE (formerly amipFuture): The same as the amip DECK experiment, except that a composite SST warming pattern derived from the CMIP3 coupled models is added to the AMIP SSTs (see Appendix C for details). As with the amip-p4K experiment, the warming pattern should only be applied to the ice free ocean surface, and sea ice and SSTs in grid boxes containing sea ice should remain the same as in the amip DECK experiment. The warming pattern should be scaled to ensure that the global mean SST increase averaged over the ice free oceans is 4K. Care should be taken to ensure that SSTs are increased in any inland bodies of water and near coastal edges, for example by linearly interpolating the provided warming pattern dataset to fill in missing data before re-gridding to the target resolution.

amip-4xCO2 (formerly amip4xCO2): The same as the amip experiment within the DECK, except that the CO₂ concentration seen by the radiation scheme is quadrupled. The CO₂ seen by the vegetation should be the same as in the amip DECK experiment. This experiment gives an indication of the adjusted radiative forcing due to CO₂ quadrupling, including stratospheric, land surface, tropospheric and cloud adjustments.

The configuration of the aqua-control, aqua-p4K and aqua-4xCO2 experiments are unchanged compared to their equivalents in CFMIP-2/CMIP5, except that the simulation length has been extended to 10 years to improve the signal to noise ratio. Further details of their experimental set up are included in Appendix B.

We also propose to use the Tier 1 experiments as the foundation for further experiments planned in the context of the Grand Challenge on Clouds, Circulation and Climate Sensitivity (Bony et al., 2015). These will include for example sensitivity experiments to assess the impacts of different physical processes on cloud feedbacks and regional circulation/precipitation responses and also to test specifically proposed cloud feedback mechanisms (e.g. Webb et al., 2015b; Ceppi et al., 2015). Additional experiments further idealizing the aquaplanet framework to a non-rotating rotationally symmetric case are also under development (e.g. Popke et al., 2013). These will be proposed as additional Tier 2 experiments at a future time, or coordinated by CFMIP outside of CMIP6.
2.2 amip minus 4K Experiment (Tier 2)

Lead Coordinators: Mark Webb and Bjorn Stevens

Science Question: Are cloud feedbacks consistent for climate cooling and warming, and if not, why?

There is some evidence to suggest that cloud feedbacks might operate differently in response to cooling rather than warming. For example, Yoshimori et al., 2009 found a positive shortwave cloud feedback in a CO2 doubling experiment with a particular GCM, but noted a tendency for it to become weaker or even negative in cooling experiments designed to replicate the climate of the last glacial maximum. They suggested that this might be related to different displacements of mixed-phase clouds in the two scenarios. For small enough changes where linearity is a good approximation, one would expect the cloud response to cooling and warming to be the same, differing only in sign, resulting in an identical cloud feedback expressed per degree of global temperature change. But for larger perturbations this symmetry of response may no longer hold. A warming or cooling of the atmosphere of equal magnitude while maintaining relative humidity will for example generate different changes in absolute humidity, and its horizontal and vertical gradients, which have been linked to cloud feedbacks (Brient and Bony, 2013; Sherwood et al., 2014), the atmospheric lapse rate and circulation which influences clouds and depends in part on the absolute humidity (Held and Soden, 2006; Ku et al., 2015) and additionally on extratropical cloud optical depth feedbacks which may be related to adiabatic cloud liquid water contents (Gordon and Klein, 2014) or phase changes that depend upon whether a given volume crosses the 0 degree isotherm in the climate change (Ceppi et al. 2015).

The configuration of the amip-p4K experiment will be the same as the amip-p0K experiment, except that the sea surface temperatures are uniformly reduced by 4K rather than increased. This cooling should be applied to sea ice free grid boxes only. Sea ice and SSTs in grid boxes containing sea ice should remain the same as in the amip DECK experiment. In models which employ a fixed lower threshold near freezing for the SST used in the calculation of the surface fluxes, this should ideally also be reduced by 4K. This experiment will contain CFMIP COSP and process outputs so as to support the investigation of inconsistent responses of clouds to a cooling vs. a warming climate in a controlled way through comparison with the amip-p4K experiment. This experiment also complements the abrupt 0.5XCO2 and the -4% solar experiments in that one can identify asymmetries in the warming/cooling response with and without interactions with the ocean. As such we hope that these experiments will provide useful synergies with the Palaeoclimate Model Intercomparision Project (PMIP) CMIP6 experiments (Kageyama et al., 2016), for example in interpreting differing cloud feedbacks between future CO2 forced experiments and those representing the Last Glacial Maximum, as highlighted by Yoshimori et al., 2009.

2.3 Atmosphere-only experiments without longwave cloud radiative effects. (Tier 2)

Lead Coordinators: Sandrine Bony and Bjorn Stevens

Science question: How do cloud-radiative effects impact the structure, the strength and the variability of the general atmospheric circulation in present and future climates?

It is increasingly recognized that clouds, and atmospheric cloud-radiative effects in particular, play a critical role in the general circulation of the atmosphere and its response to global warming or other perturbations: they have been found to modulate the structure, the position and shifts of the ITCZ (e.g. Slingo and Slingo 1988; Randall et al., 1989; Sherwood et al 1994; Bergman and Hendon 2000; Hwang and Frierson, 2013; Fermepin and Bony 2014; Voigt et al., 2014; Loeb et al., 2015), the organisation of convection in tropical waves, Madden-Julian Oscillations and other forms of convective aggregation (e.g. Lee et al., 2001; Lin and Mapes, 2004; Bony and Emanuel, 2005; Zorin and Jevic et al., 2006; Crueger and Stevens, 2015; Muller and Bony, 2015), the extra-tropical-circulation and the position of eddy-driven jets (e.g.,Ceppi et al., 2012; Ceppi et al., 2014; Grise and Polvani 2014; Li et al., 2015; Voigt and Shaw, 2015), and modes of interannual to decadal climate variability (e.g. Bellomo et al., 2015; Radel et al., 2016; Yuan et al., 2016). A better assessment of this role would greatly help to interpret model biases (how much do biases in cloud-radiative properties contribute to biases in the structure of the ITCZ, in the position and strength of the storm tracks, in the lack of intra-seasonal variability, etc) and to inter-model differences in simulations of the current climate and in climate change projections (especially changes in regional precipitation and extreme events). More generally, a better understanding of how clouds couple to the circulation is expected to improve our ability to answer the four science questions raised by the WCRP Grand Challenge on Clouds, Circulation and Climate Sensitivity (Bony et al., 2015).

These questions provided the scientific motivation for the Clouds On/Off Klima Intercomparison Experiment (COOKIE) project proposed by the European consortium EUCLIPSE and CFMIP (Stevens et al., 2012). The COOKIE experiments, which have been run by four to eight climate models (depending on the experiment), switched off the cloud-radiative effects (clouds seen by the radiation code -and the radiation code only- were artificially made transparent) in an atmospheric model forced by prescribed SSTs. By doing so, the atmospheric circulation could feel the lack of cloud-radiative heating within the atmosphere, but the land surface could also feel the lack of cloud shading, which led to changes in land surface temperatures and land-sea contrasts. The change in circulation between On and Off experiments resulted from both effects, obscuring to some degree the mechanisms through which the atmospheric cloud-radiative effects interact with the circulation for given surface boundary conditions. As the longwave cloud-radiative effects are felt mostly within the troposphere (representing most of the net atmospheric cloud-radiative heating) while the shortwave effects are felt mostly at the surface (e.g. L’Ecuyer and McGarragh 2010; Haynes et al., 2013), we could better isolate the role of tropospheric cloud-radiative effects on the...
Together with solar forcing experiments (Tier 2) we propose in Tier 1 a set of simple experiments similar to the amip, amip-p4K, aqua-control and aqua-p4K experiments within Tier 1, but in which cloud-radiative effects are switched off in the longwave part of the radiation code while retaining those in the shortwave (Fermepin and Bony, 2014). Care should also be taken to remove the effects of cloud on any longwave cooling used in other model schemes (e.g. turbulent mixing) if these are calculated independently of the radiation scheme. These experiments will be referred to as amip-lwoff, amip-p4K-lwoff, aqua-control-lwoff and aqua-p4K-lwoff. The analysis of idealised (aqua) experiments will allow us to assess the robustness of the impacts found in more realistic (AMIP) configurations. It will also facilitate the interpretation of the results using simple dynamical models or theories, in collaboration with large-scale dynamists (e.g. DynVar). The comparison of the inter-model spread of simulations between the standard and ‘lwoff’ experiments for present-day and warmer climates will help to identify which aspects of the inter-model spread depend on the representation of cloud-radiative effects, and which aspects do not, thus better highlighting other sources of spread. An alternative method (proposed by Aiko Voigt) was also considered, in which clear-sky heating rates would be applied in the atmosphere while retaining the all-sky fluxes at the surface. Although this approach would potentially isolate the effects of cloud heating in the atmosphere more clearly than the lwoff experiments proposed here, it is yet to be demonstrated in a pilot study, and is considered more technically difficult to implement than the lwoff experiments, which are very similar to those piloted by Fermepin and Bony, 2014.

2.4 Abrupt +/-4% Solar Forced AOGCM experiments (Tier 2)

Science Question: How do responses in the climate system due to changes in solar forcing differ from changes due to CO2, and is the response sensitive to the sign of the solar forcing?

While rapid adjustments in clouds and precipitation can easily be separated from conventional feedbacks in SST forced experiments, such a separation in coupled models is complicated by various issues, including the response of the ocean on decadal timescales. A number of studies have examined cloud feedbacks in coupled models subject to a solar forcing, which is generally associated with much smaller global cloud and precipitation adjustment, due to a smaller atmospheric absorption for a given top of atmosphere forcing (e.g. Lambert and Faull, 2007; Andrews et al., 2010), but the regional cloud and precipitation changes have yet to be rigorously investigated across models. Solar forcing also differs from greenhouse forcing through its different fingerprint on the vertical structure of warming (Santer et al., 2013) and small changes in the radiative heating near the tropopause may project measurably on tropospheric climate (e.g., Butler et al., 2010), for instance by influencing the baroclinicity in the upper troposphere and thus the storm-tracks (Bony et al., 2015).

A +/-4% solar experiment abrupt-solldr is proposed which is analogous to the abrupt-4xCO2 experiment but rather than changing CO2, it would abruptly increase the solar constant by four percent and keep it fixed for 150 years, resulting in a global mean radiative forcing of a similar magnitude to that due to CO2 quadrupling. When changing the solar constant, the shape of the spectral solar irradiance distribution should remain consistent with that in the piControl experiment. This experiment complements the DECK abrupt-4xCO2 experiment, tests the forcing feedback framework for analyzing climate change, and would support our understanding of regional responses of the coupled system with and without CO2 adjustments. The complementary -4% abrupt solar forcing experiment (abrupt-solldr) would allow the examination of feedback asymmetry under climate cooling, and would also help with the interpretation of model responses to geo-engineering scenarios and volcanic forcing, and of past climate signals.

2.5 NonLinMIP abrupt 2xCO2 and abrupt 0.5xCO2 Experiments (Tier 2)

Lead Coordinator: Peter Good

Science Question: To what extent is regional-scale climate change per CO2 doubling state-dependent (nonlinear); what are the associated mechanisms; and how does this affect our understanding of climate model uncertainty?

Recent studies with individual, or a small number of climate models, have found substantial nonlinearities in regional-scale precipitation change (Good et al., 2012; Chadwick and Good, 2013), associated with robust physical mechanisms (Chadwick and Good, 2013). Significant nonlinearity has also been found in global and regional-scale warming (e.g. Colman and McAvaney, 2009; Jonko et al., 2013; Good et al., 2015; Meraner et al., 2013) and ocean heat uptake (Bouttes et al., 2015).

To address this science question we propose two new experiments for Tier 2, abrupt 2xCO2 and abrupt 0.5xCO2. These are the same as the DECK abrupt-4xCO2 experiment except that CO2 concentrations are doubled and halved respectively relative to the preindustrial control. These experiments are based on a proven analysis approach, including traceability of these experiments to transient-forcing simulations (Good et al., 2016) to explore global and regional-scale nonlinear responses, highlighting different behaviour under business-as-usual scenarios, mitigation scenarios and palaeoclimate simulations.

Additionally comparisons of the abrupt 2xCO2 and abrupt 4xCO2 experiments will help to establish the extent to which the
latter accurately estimates the equilibrium climate sensitivity to CO2 doubling (e.g. Gregory et al., 2004, Block and
Mauritsen, 2013). Additional experiments (Good et al., 2016) may be proposed for Tier 2 in the future, or coordinated via
CFMIP outside of CMIP6. These include 100-year extensions to abrupt-4xCO2 and abrupt-2xCO2; a 1% ramp-down from
the end of the 1pctCO2 experiment; an abrupt step-down to 1xCO2 from year 100 of the abrupt-4xCO2. These would be used to
explore longer-timescale responses, quantify nonlinear mechanisms more precisely and understand the reversibility of
climate change.

2.6 Feedbacks in AMIP experiments (Tier 2)

Lead Coordinator: Timothy Andrews

Science question: Are climate feedbacks during the 20th century different to those acting on long term climate change?

Recent studies have shown significant time variation in climate feedbacks in response to CO2 quadrupling (e.g. Andrews et
al., 2012; Geoffroy et al., 2013; Armour et al., 2013; Andrews et al., 2015). This raises the possibility that feedbacks during
the 20th century may be different to those acting on long term change, and hence has the potential to alleviate the apparent
discrepancy between estimates of climate sensitivity from comprehensive climate models and from simple climate models
fitted to observed warming trends (Collins et al., 2013). For example Gregory and Andrews, 2016 found that two models
forced with observed monthly 20th century SST and sea-ice variations simulated effective climate sensitivities of about 2K,
whereas these same models forced with patterns of long term SST change simulated effective climate sensitivities of over 3K
and 4K.

The previous CFMIP-2/CMIP5 design was unable to diagnose the time-variation of feedbacks of explicit relevance to the
historical period, because this requires the removal of the time varying forcing. To address this we propose an additional
experiment called ‘amip-piForcing’ (amip pre-industrial forcing) following the design of Andrews 2014 and Gregory and
Andrews, 2016. This experiment is the same as the amip DECK experiment (i.e. using observed monthly updating SSTs and
sea-ice), but run for the period 1870-present and with constant pre-industrial forcings (i.e. all anthropogenic and natural
forcing boundary conditions identical to the piControl experiment). Since the forcing constituents do not change in this
experiment it readily allows a simple diagnosis of the simulated atmospheric feedbacks to observed SST and sea ice changes,
which can then be compared to feedbacks representative of long term change and climate sensitivity (e.g. from abrupt-4xCO2
or amip-p4K). The experiment has the additional benefit, by differing with the standard amip run that includes time-
varying forcing agents, of providing detailed information on the transient effective radiative forcing and adjustments in
models during the AMIP period (Andrews, 2014). This can then be compared to the forcings diagnosed in the Radiative
Forcing Model Intercomparison Project (RFMIP, Pincus et al., 2016, who use a pre-industrial climate baseline) to test for any
dependence of forcing and adjustments on the climate state. Time-varying feedbacks in the amip experiment could
alternatively be diagnosed by subtracting a time-varying radiative forcing diagnosed from RFMIP experiments. However, the
amip-pForcing approach has the benefit of diagnosing the time-varying feedbacks over the full 1870-present period rather
than the last 36 years, and does so with reference to a single experiment, which reduces noise compared to that which would
be present with a double difference of the amip experiment and two RFMIP experiments. Also, the inclusion of CFMIP
process diagnostics in the amip-pForcing experiment will enable a deeper understanding of the factors underlying forcing
and feedback differences in the present and future climate.

2.7 Time slice experiments for understanding regional climate responses to CO2 (Tier 2)

Lead Coordinators: Robin Chadwick, Hervé Douville and Christopher Skinner

Science questions:

- How do regional climate responses (e.g. of precipitation) in a coupled model arise from the combination of
  responses to different aspects of CO2 forcing and sea surface warming (uniform SST warming, patterned SST
  warming, sea-ice change, direct CO2 effect, plant physiological effect)?
- Which aspects of forcing/warming are most important for causing inter-model uncertainty in regional climate
  projections?
- Can inter-model differences in regional projections be related to underlying structural or resolution differences
  between models through improved process understanding, and could this help us to constrain the range of regional
  projections?
- What impact do coupled model SST biases have on regional climate projections?

The CFMIP-2/CMIP5 set of idealised amip experiments (e.g. amip4K, amipFuture) have allowed the contribution of different
aspects of SST warming and increased CO2 concentrations to the projections of fully coupled GCMs to be examined (e.g.
Bony et al., 2013; Chadwick et al., 2014; He and Soden, 2015). However the amip experiments were not designed to replicate
coupled GCM responses on a regional scale, and large discrepancies exist between the two in many regions, particularly
when individual models are examined instead of the ensemble mean (Chadwick, 2016). This is largely due to the choice of
The experiments are:

1) piSSST – An AGCM experiment with monthly- and annually-varying SSTs, sea-ice, atmospheric constituents and any other necessary conditions (e.g. vegetation if required) taken from a section of each model's own piControl run, using the 30 years of piControl that are parallel to years 111-140 of its abrupt-4xCO2 run. Note that dynamic vegetation (if included in the model) should not be turned on in any of the piSSST set of experiments;

2) piSSST-paK – same as piSSST, but with a global spatially and temporally uniform SST anomaly applied on top of the monthly- and annually-varying piSSST SSTs. The magnitude of the uniform increase is taken from each model's global, climatological mean open SST change between abrupt-4xCO2 and piControl (using the mean of years 111-140 of abrupt-4xCO2, and the parallel 30-year section of piControl). Sea-ice is unchanged from piSSST values;

3) piSSST-4xCO2-rad – same as piSSST but CO2 as seen by the radiation scheme is quadrupled;

4) piSSST-4xCO2 – same as piSSST but with CO2 quadrupled, and this increase is seen by both the radiation scheme and the plant physiological effect. If a model does not include the plant physiological response to CO2, then piSSST-4xCO2 can be omitted from the set of piSSST experiments for that model;

5) a4SSST – same as piSSST, but with monthly- and annually-varying SSTs taken from years 111-140 of each model's own abrupt-4xCO2 experiment instead of from piControl (sea ice is unchanged from piSSST);

6) a4SSSTice – same as piSSST, but with monthly- and annually-varying SSTs and sea-ice taken from years 111-140 of each model's own abrupt-4xCO2 experiment instead of from piControl; CO2 is also quadrupled, and is seen by both the radiation scheme and the plant physiological effect (if included in the model). a4SSSTice-4xCO2 is used to establish whether a time slice experiment can adequately recreate the coupled abrupt-4xCO2 response in each model, and then forms the basis for a decomposition using the other experiments. The time slice experiments can be combined in various ways to isolate the climate response to each individual aspect of forcing and warming. For example the response to SST pattern change is given by taking the difference between a4SSST and piSSST-paK, and the plant physiological response is found by taking the difference between piSSST-4xCO2 and piSSST-4xCO2-rad.

8) We also propose an additional amip based experiment, amip-a4SSST-4xCO2: the same as amip, but a patterned SST anomaly is applied on top of the monthly- and annually-varying amip SSTs. This anomaly is a monthly climatology, taken from each model's own abrupt-4xCO2 run minus piControl (using the mean of years 111-140 of abrupt-4xCO2, and the parallel 30-year section of piControl). CO2 is quadrupled, and the increase in CO2 is seen by both the radiation scheme and vegetation. Comparison of amip-a4SSST-4xCO2 and a4SSSTice-4xCO2 should help to illuminate the impact of SST biases on regional climate responses in each model, and how this contributes to inter-model uncertainty.

3 CFMIP Recommended Diagnostic Outputs for CMIP experiments

The CFMIP-3 specific diagnostic request is designed to address the following questions: 1) How well do clouds and other relevant variables simulated by models agree with observations? 2) What physical processes and mechanisms are important for a credible simulation of clouds, cloud feedbacks and cloud adjustments in climate models? 4) Which models have the most credible representations of processes relevant to the simulation of clouds? 5) How do clouds and their changes interact with other elements of the climate system?

The set of diagnostic outputs recommended for CFMIP-3 is based on that from CFMIP-2, with some modifications. The request outlined below is in three parts. The first part describes an updated set of CFMIP process diagnostics (based on those in CFMIP-2 which are documented at http://cmip-pcmdi.llnl.gov/cmip5/output_req.html) in terms of the various groups of variables and the experiments in which they are requested. This set was drawn up by the CFMIP committee and ratified by the modelling groups following a presentation at the 2014 CFMIP meeting. The second part describes recommendations for COSM outputs in the CFMIP-3, CMIP DECK and CMIP6 Historical experiments. The third part describes additional diagnostics requested for evaluation of mean diurnal cycle of tropical clouds and radiation. The summaries below give an overview of the diagnostic request; however the definitive and detailed specification is documented in the CMIP6 data request, available at https://www.earthsystemco2.org/projects/wp/CMIP6DataRequest (Juckes et al., in preparation.)
In the case where CFMIP-3 specific outputs are requested in DECK and CMIP6-Historical experiments, and modelling groups run more than one ensemble member of an experiment, we request that each set of CFMIP-3 specific outputs are submitted for one ensemble member only. Having different CFMIP variables in different ensemble members is acceptable, but submitting them all in the same ensemble member is preferable. We request that the modelling groups provide information on which CFMIP diagnostic sets are submitted in which ensemble members so that this information can be made available to those who may be analyzing the output. Our analysis plans for the CFMIP diagnostic outputs in the CMIP DECK, CMIP6 Historical and CFMIP-3 experiments, including details of the CFMIP Diagnostics Code Catalogue are summarised in Appendix A.

3.1 Process outputs

In CFMIP-2, instantaneous high frequency ‘cfSites’ outputs were requested for 120 locations in the amip, amip4K, amipFuture and amip4xCO2 experiments, and for 73 locations along the Greenwich meridian in the aquaplanet experiments, to support understanding and evaluation of clouds and their interactions with convection and other processes. The 120 locations include the locations of instrumented sites (ARM and CloudNet stations, Dome C, etc), the transect associated with the GCSS Pacific Cross-section Intercomparison (GPC), past field campaigns (DYNAMO-IL, NARVAL, HOPE, VOCALS, ASTEX and AMMA transects, TOGA-COARE, RICO, etc) and a number of climate regimes that contribute substantially to the inter-model spread of cloud feedbacks in climate change (Webb et al., 2015a). These outputs have so far been used to evaluate the models with in-situ measurements (e.g. Nuijens et al., 2015a, Nuijens et al., 2015b, Ngeggers et al., 2015), to investigate the diurnal cycle of cloud feedbacks (Webb et al., 2015a) and to compare cloud feedbacks in climate models with available or in the planning stage, for example from the Barbados Cloud Observatory (Stevens et al., 2015) the ARM and cloud adjustments. An increasing wealth of observational data with which to evaluate the models using these outputs is also being used to understand regional warming patterns such as polar amplification in coupled models (e.g. Yoshimori et al., 2015), to evaluate the models with in-situ measurements (e.g. Nuijens et al., 2015a, Nuijens et al., 2015b, Neggers et al., 2015), to investigate the diurnal cycle of cloud feedbacks (Webb et al., 2015a) and to compare cloud feedbacks in climate models with available or in the planning stage, for example from the Barbados Cloud Observatory (Stevens et al., 2015) the ARM and cloud adjustments. An increasing wealth of observational data with which to evaluate the models using these outputs is also being used to understand regional warming patterns such as polar amplification in coupled models (e.g. Yoshimori et al., 2015), to evaluate the models with in-situ measurements (e.g. Nuijens et al., 2015a, Nuijens et al., 2015b, Neggers et al., 2015). We have added St. Helena to the list of locations in light of ongoing field work, increasing the total number of locations to 121 for CFMIP-3. A text file containing the list of locations is available in the Supplementary Information and on the CFMIP website; these are also presented graphically in Figure 2.

For CFMIP-3 ‘cfSites’ outputs are now requested for one ensemble member of the amip DECK experiment, and the amip4K and amip4xCO2 experiments. Outputs should be provided for the full duration of each experiment. The sampling interval should be the integer multiple of the model time step that is nearest to 30 minutes and divides into 60 minutes with no remainder: e.g. 30 minutes for a 30, 15 or 10 minute time step or 20 minutes for a 20 minute time step. Outputs should be instantaneous (i.e. not time means) and from nearest grid box (i.e. not spatial interpolation). We have dispensed with the ‘cfSites’ outputs in the aquaplanet and amip-future4K experiments because these have been less widely used compared to those from the other experiments.

The ‘cfSites’ outputs from CFMIP-3 provide instantaneous outputs of a range of quantities (including temperature and humidity tendency terms) in experiments which can be used to evaluate the present day relationships of clouds to convection controlling factors using in situ measurements, and at the same time explore how these relationships affect cloud feedbacks and cloud adjustments. An increasing wealth of observational data with which to evaluate the models using these outputs is available or in the planning stage, for example from the Barbados Cloud Observatory (Stevens et al., 2015) the ARM Program (e.g. Wood et al., 2015; Marchand et al., 2015) or within the German national project on high-definition clouds and precipitation for climate-prediction, HD(CP)², inclusive of its observational prototype experiment (HOPE), and which has collected observations over Germany following conventions adopted for CMIP (Andrea Lammert, personal communication).

CFMIP-2 also requested cloud, temperature and humidity tendency terms from convection, radiation, dynamics etc. in the amip, amip4K, amipFuture and amip4xCO2, aquaControl, aqua4xCO2 and aqua4K experiments, as global monthly mean outputs and high frequency outputs at fixed locations (Bony et al., 2011). Upward and downward radiative fluxes on model levels were also requested in these experiments, and for instantaneous CO₂ quadrupling in the amip experiment only.

Temperature and humidity tendency terms in particular have been shown to be useful for understanding the roles of different parts of the model physics in cloud feedbacks (e.g. Webb and Lock 2013; Demott et al., 2013; Sherwood et al., 2014; Brown and Lock 2013). We have dispensed with the cloud tendency terms because these have been less widely used than the temperature and humidity tendencies.

In CFMIP-3 we have improved the definitions of the temperature and humidity tendency terms, and added some additional terms such as clear-sky radiative heating rates to more precisely quantify the contributions of different processes to the temperature and humidity budget changes underlying cloud feedbacks and adjustments. We have dispensed with the cloud water tendency terms because these have been less widely used than the temperature and humidity tendency terms.

A shortcoming of the CMIP5 protocol was that we were unable to interpret the physical feedback mechanisms in coupled model experiments due to a lack of process diagnostics. For this reason in CMIP6 we are requesting these budget terms in the DECK, abrupt-4xCO2 experiment and the pre-industrial control as well as one ensemble member of the amip DECK experiment, and all of the CFMIP-3 experiments listed in Sections 2.1-2.6.

Clustering approaches (e.g., Jakob and Tselioudis, 2003) are now commonly used for assessing the contributions of different cloud regimes (e.g. stratocumulus, trade cumulus, frontal clouds, etc) to present day biases in cloud simulations and to inter-model differences in cloud feedbacks (e.g. Williams and Webb 2009, Tsushima et al., 2013, Tsushima et al., 2015). We have also added some additional daily 2D fields to the standard package of CFMIP daily outputs to allow further investigation of feedbacks between clouds and aerosols associated with the changing hydrological cycle (aerosol loadings and cloud top effective radii/number concentrations) and a clearer diagnosis of the roles of convective and stratiform clouds (convective vs. stratiform ice and condensed water paths and cloud top effective radii/number concentrations).
3.2 COSP outputs

This section motivates and summarizes the COSP outputs requested from the DECK, and CMIP6 historical and CFMIP-3 experiments as well as a corresponding set of observations.

There is no unique definition of clouds or cloud types, neither in models nor in observations. Therefore, to compare models with observations, and even to compare models with each other, it is necessary to use a consistent definition of clouds between the model and the satellite product in question (i.e., “definition-aware”). Further complicating matters - climate model grid boxes (typically 1 degree) are much larger than the scales over which many satellite observations are made (typically <10 km). As a result, one must downscale the climate model cloud properties to the observation scale (i.e., be “scale-aware”). The CFMIP Observation Simulator Package (COSP) enables definition-aware and scale-aware comparisons between models and multiple sets of observations by producing cloud diagnostics from model simulations that are quantitatively comparable to a variety of satellite products from ISCCP, CloudSat, CALIPSO, MODIS, MISR and Parasol (Bodas-Salcedo et al., 2011). COSP enables a more quantitative comparison of model outputs with satellite cloud products, which often sub-sample low level clouds in the presence of high level clouds due to the effects of cloud overlap and attenuation (e.g. Yu et al., 1996). COSP also provides histograms of various cloud properties as a function of height or pressure which are directly comparable with satellite products and cannot be calculated correctly from time mean model outputs. The multiple simulators within COSP allow a multi-faceted evaluation of clouds in models whereby the strengths and weaknesses of different satellite products may be considered together.

COSP is increasingly being used not only for model intercomparison activities but as part of the model development and evaluation process by modelling groups (e.g. Marchand et al., 2009; Zhang et al., 2010; Kay et al., 2012; Franklin et al., 2013; Lacagnina and Sefeter, 2014; Num et al., 2014; Williams et al., 2015; Konsta et al., 2015). Many of the standard monthly and daily COSP outputs have been shown to be valuable in the CMIP5 experiments, not only for cloud evaluation, allowing a detailed evaluation of clouds and precipitation, and their interaction with radiation (e.g. Nam et al., 2012; Cesana and Chepfer, 2012; Kay et al. 2012; Klein et al., 2013; Tsushima et al., 2013; Gordon and Klein, 2014; Lin et al., 2014; Bodas-Salcedo et al., 2014; Bellomo and Clement, 2015), but also in quantifying the contributions of different cloud types to cloud feedbacks and forcing adjustments in climate change experiments (e.g. Zelinka et al., 2013; Zelinka et al., 2014; Chepfer et al., 2014; Tsushima et al., 2015). For a full list of studies that use COSP diagnostics for model evaluation and feedback analysis please refer to the ‘CFMIP publications’ section of the CFMIP website.

Here we will give only a brief overview of the COSP request; readers interested in the complete details of the data request are referred to the Earth System CoG website (https://www.earthsystemcog.org/projects/wip/CMIP6DataRequest). The COSP data request for the CMIP DECK and CMIP6 has been designed to span model evaluation across different space and time scales. Monthly-mean diagnostics allow for the evaluation and intercomparison of large-scale distributions of cloud properties and their interaction with radiation. High-frequency model outputs (daily, 3-hourly) are aimed at a process-oriented evaluation (e.g. Bodas-Salcedo et al., 2012) and offer the opportunity of exploiting the synergy between multiple instruments (e.g. Konsta et al., 2015). Recent observational developments have improved our capability to retrieve cloud radiative properties. In particular, new methodologies for cloud phase identification are available for CALIPSO and MODIS, and COSP has been enhanced to provide diagnostics that are compatible with these new observational datasets (Cesana and Chepfer, 2013). These new diagnostics will help elucidate some open questions regarding the role of cloud phase in model biases (Ceppi et al., 2016; Bodas-Salcedo et al., in press).

Within CFMIP-3 COSP output is requested from six simulators as follows:

- ISCCP: pseudo-retrievals of cloud top pressure (CTP) and cloud optical thickness (tau) (Klein and Jacob 1999; Webb et al., 2001).
- CloudSat: a forward model for radar reflectivity as a function of height (Haynes et al., 2007).
- CALIPSO (Chepfer et al., 2008; Cesana and Chepfer, 2013): forward model for lidar scattering ratio as function of height, and cloud phase retrieval.
- MODIS: pseudo-retrievals of CTP, effective particle size and tau as function of phase (Pincus et al., 2012).
- MISR: pseudo-retrievals of cloud top height (CTH) and tau (Marchand and Ackerman, 2010).
- PARASOL: simple forward model of mono-directional reflectance (Konsta et al., 2015).

The main difference to CFMIP-2 is that output is requested from a greater number of simulators and longer periods of simulated time. MISR provides more accurate retrievals of cloud-top-height for low-level and mid-level clouds, and more reliable discrimination of mid-level clouds from other clouds, while MODIS provides better retrievals of high-level clouds. ISCCP and MISR histograms can be combined to separate optically-thin high-level clouds into multi-layer and single-layer categories (Marchand et al. 2010). Aerosol schemes are becoming more complex, with more elaborate representations of cloud-aerosol interactions. This makes the evaluation of the phase partitioning an important aspect of model evaluation, and height-resolved partitioning estimates from the CALIPSO simulator are included in the COSP request. Cloud phase and particle size estimates from the MODIS simulator were not available in CFMIP-2 but may prove a useful complement to investigate cloud-aerosol interactions by virtue of greater geographic sampling and longer time records. Many of the COSP diagnostics are now requested for the entire lengths of the DECK, CMIP6 Historical and CFMIP-3 experiments to support the quantification and interpretation of cloud feedbacks and cloud adjustments in a broader context. The new inclusion in this COSP request of a long time series of three-dimensional cloud fractions will facilitate the comparison of cloud trends with the observational record (Chepfer et al., 2014). More details of all the changes with respect to CFMIP-2 can be found in the
The primary goal of CFMIP is to inform improved assessments of cloud feedbacks on climate change. This involves bringing climate modelling, observational and process modelling communities closer together and providing better tools and community support for understanding and evaluation of clouds and cloud feedbacks simulated by climate models. CFMIP
supports ongoing coordinated model inter-comparison activities by recommending experiments and model output diagnostics for CMIP, designed to support the understanding and evaluation of climate processes and cloud feedbacks in models. The CFMIP approach is also increasingly being used to understand other aspects of climate change, and so a second objective has now been introduced, to improve understanding of circulation, regional-scale precipitation, and non-linear changes. CFMIP proposes a number of experiments and model outputs for CMIP6, building on and extending those which were part of CMIP5. A compact set of Tier 1 experiments are proposed address the question: “1) What are the physical mechanisms underlying the range of cloud feedbacks and cloud adjustments predicted by climate models, and which models have the most credible cloud feedbacks?” The Tier 1 experiments (amip-p4K, amip-4xCO2, amip-future4K, aqua-control, aqua-4xCO2 and aqua-p4K) retain the idealized experimental hierarchy of the CFMIP-2/CMIP5 experiments while building on the DECK AMIP experiment. A number of Tier 2 experiments are proposed to address additional science questions. An amip minus 4K experiment is proposed to address the question “2) Are cloud feedbacks consistent for climate cooling and warming, and if not, why?” Atmosphere-only experiments with clouds made transparent to longwave radiation address the question “3) How do cloud-radiative effects impact the structure, the strength and the variability of the general atmospheric circulation in present and future climates?” Abrupt +/-4% Solar Forced AOGCM experiments are proposed for the question “4) How do responses in the climate system due to changes in solar forcing differ from changes due to CO2, and is the response sensitive to the sign of the solar forcing?” abrupt 2xCO2 and abrupt 0.5xCO2 experiments are proposed to address the question “5) To what extent is regional-scale climate change per CO2 doubling state-dependent (nonlinear), and why?” Other experiments and questions proposed include: AMIP with preindustrial forcing “6) Are climate feedbacks during the 20th century different to those acting on long-term climate change and climate sensitivity?”. Time slice experiments forced with SSTs from preindustrial and abrupt-4xCO2 simulations “7) How do regional climate responses (e.g. precipitation) in a coupled model arise from the combination of responses to different aspects of CO2 forcing and warming (uniform SST warming, pattern SST warming, direct CO2 effect, plant physiological effect, sea-ice change)?” The CFMIP experiments in CMIP6 will continue to include outputs from the CFMIP Observational Simulator Package (COSP) to support robust scale-aware and definition-aware evaluation of modelled clouds with observations and to relate cloud feedbacks to observed quantities. COSP outputs are also proposed for inclusion in the DECK and CMIP6 Historical experiments. Process diagnostics including ‘cSites’ high frequency outputs at selected locations and temperature and humidity budget terms from radiation, convection, dynamics, etc. are also retained from CMIP5. These will help to address the following questions: 1) How well do clouds and other relevant variables simulated by models agree with observations? 2) What physical processes and mechanisms are important for a credible simulation of clouds, cloud feedbacks and cloud adjustments in climate models? 4) Which models have the most credible representations of processes relevant to the simulation of clouds? 5) How do clouds and their changes interact with other elements of the climate system? By continuing the CFMIP experiments and diagnostic outputs within CMIP6 we hope to apply the well established aspects of the CFMIP approach to a larger number of climate models. Additionally we have proposed new experiments to investigate a broader range of questions relating to the Grand Challenge on Clouds, Circulation and Climate Sensitivity. We hope that the modelling community will participate fully in CFMIP via CMIP6 so as to maximize the relevance of our findings to future assessments of climate change.

Code and Data Availability

COSP is published under an open source license via GitHub (please see the CFMIP website for details). The model output from the DECK, CMIP6 historical and CFMIP-3 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 DOIs 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.

Acknowledgements: We are grateful to Florent Brient, Hideo Shiogama, Aiko Voigt, Mark Ringer and two anonymous referees for helpful comments on the manuscript. We thank the modelling groups and wider CFMIP community for reviewing and supporting the CFMIP contribution to CMIP6, the CMIP Panel for their coordination of CMIP6, the WGCM Infrastructure Panel (WIP) overseeing the CMIP6 infrastructure, and Martin Jackes for taking the lead in preparing the CMIP6 data request. We are also grateful to Robert Pincus and Yuying Zhang for their contributions to COSP and to CFMIP- OBS, to Dustin Swales for his development work for COSP-2, and to Gregory Cesana and Matheus Reverdy for their contributions to CFMIP-OBS. We are grateful to Brian Soden for producing the CMIP3 composite pattern dataset used for the CMIP5 amipFuture and CMIP6 amip-future4K experiments, and to PMIP representatives Pascale Braconnot, Masa
We plan to scientifically analyze, evaluate and exploit the proposed experiments and diagnostic outputs, and have Challenge spanning the duration of CMIP6. An overview analysis of regional responses and model uncertainty in the piSST set of experiments will be carried out by Andrews. Analysis of abrupt-4xCO2 experiment. Analysis will primarily involve comparing the \( \text{op}5\times\text{CO}_2 \) Andrews (submitted). We propose to use this as a starting point for a multi-model analysis. Lead coordinator: Timothy cloud feedbacks in climate models. The wide range of analysis activities described above in the context of CFMIP-2 will be continued in CFMIP-3 using the CMIP DECK and CMIP6 experiments, allowing the techniques developed in CFMIP-2 to be applied to an expanded number of models, including the new generation of models currently under development. These activities will include evaluation of clouds using additional simulators, investigation of cloud processes and cloud feedback/adjustment mechanisms using process outputs (cfSites, tendency terms, etc). The inclusion of COSP and budget tendency terms in additional DECK experiments (e.g. abrupt-4xCO2) will enable the CFMIP approach to be applied to a wider range of experimental configurations. Lead coordinator: Mark Webb.

Analysis of the +/-4% solar forcing runs will include an evaluation of both rapid adjustments and longer-term responses on global and regional top-of-atmosphere radiative fluxes, cloud types (using ISCCP and other COSP simulators) and precipitation characteristics, as well as comparison of these responses with responses in DECK abrupt-4xCO2 experiments. GeoMIP and SolarMIP have expressed a strong interest in these CFMIP experiments and joint analysis of these CFMIP experiments with GeoMIP and SolarMIP experiments is anticipated, specifically with the goal of determining to what degree results from abrupt solar forcing only experiments and abrupt CO2 only experiments can be used to predict what happens when both forcing are applied simultaneously, as done in the GeoMIP experiments. Lead coordinators: Chris Bretherton, Roger Marchand and Bjorn Stevens.

Analysis of nonlinear climate processes is discussed in detail by Good et al., 2016. This includes a method for validating traceability of abrupt CO2 experiments to transient simulations, which is also recommended as a standard test of the DECK abrupt-4xCO2 experiment. Analysis will primarily involve comparing the abrupt-4xCO2, abrupt-2xCO2 and abrupt-0p5xCO2 experiments over the same timescale. Lead coordinator: Peter Good.

Analysis of amip-piForcing has already been performed in detail for two models in Andrews, 2014 and Gregory and Andrews (submitted). We propose to use this as a starting point for a multi-model analysis. Lead coordinator: Timothy Andrews.

An overview analysis of regional responses and model uncertainty in the piSST set of experiments will be carried out by the coordinators, in collaboration with members of contributing modelling groups. We anticipate that further detailed analysis on the processes at work in different regions will be carried out by a variety of research groups with interest and expertise in a particular region: for example a set of similar experiments has previously been used to examine the climate response of the West African monsoon in CCSM3 (Skinner et al., 2012). The piSST set of experiments have already been successfully run using the Met Office, NCAR and CNRM CMIP5 models. Lead Coordinators: Robin Chadwick, Hervé Douville and Christopher Skinner.

The analysis of the COOKIE experiments will be reviewed by the coordinators in collaboration with members of the contributing modelling groups. The role of longwave atmospheric cloud-radiative effects in large-scale circulations, regional precipitation patterns and the organisation of tropical convection will be investigated in the current climate and in climate change, with the aim of highlighting both robust effects and sources of uncertainties in the model responses. Lead coordinators: Sandrine Bony and Bjorn Stevens.

When analyzed together with the amip-pi4K experiment, the amip-n4K experiment allows the CFMIP process diagnostics to be used to understand for asymmetries in the climate response to warming and cooling which have been noted in PMIP experiments. These might arise from cloud phase responses in middle- and high-latitude clouds or from the adiabatic cloud liquid water path response feedback which is important over land regions and which would be expected to be weaker with cooling because of the non-linearity in the Clausius-Clapeyron relation. Lead coordinators: Mark Webb and Bjorn Stevens.

The COSP data request for the amip DECK experiment will allow a comprehensive multi-model evaluation of clouds and radiation, following on from CMIP5 studies (e.g. Klein et al., 2013; Bodas-Salcedo et al., 2014). The COSP data request for the other experiments (e.g. amip-p4K, abrupt-4xCO2, etc.) permits evaluation of cloud feedbacks and adjustments by cloud
type (Zelinka et al., 2013, Tsushima et al., 2015) or cloud trends (Chepfer et al., 2014). New COSP diagnostics have been used in single-model analyses: cloud phase diagnostics (Cesana and Chepfer, 2013); MISR simulator outputs to evaluate cloud fraction and multilayer clouds (Marchand and Ackerman, 2010); CALIPSO vertical distribution of cloud fraction for the study of cloud trends (Chepfer et al., 2014). These studies will be used as starting points for multi-model analyses. The COSP Project Management Committee co-chairs will coordinate and encourage the exploitation of these resources. Lead coordinators: Alejandro Bodas-Salcedo and Steve Klein.

Analysis of output from CFMIP and CMIP6 experiments will also be facilitated by sharing of diagnostic codes via the CFMIP Diagnostics Code Catalogue (accessible via the CFMIP website http://www.earthsystemcog.org/projects/cfmip/). This is a catalogue of programs written by various members of the CFMIP community, implementing a number of diagnostic approaches from published studies. These include daily cloud clustering evaluation metrics based on ISCCP and ISCCP simulator outputs (Williams and Webb, 2009, Tsushima et al., 2013), error metrics for total cloud amount, longwave and shortwave cloud properties (Klein et al., 2013), process oriented evaluation of clouds using A-train instantaneous observations (Konsta et al., 2012), quality control and low-cloud diagnostics (Nam et al., 2012; Nam and Quaas, 2012), sensitivity of low cloud cover to estimated inversion strength and SST (Qu et al., 2012), and cloud radiative kernels (Zelinka et al., 2012). Any codes which implement diagnostics which are relevant to analysing clouds, circulation and climate sensitivity are also discussed in the peer reviewed studies and are eligible for inclusion in the catalogue, and we welcome additional contributions to further support community analysis of CMIP6 outputs.

### APPENDIX B: Aquaplanet Experimental Design

Aquaplanets are Earth-like planets with completely water-covered surfaces. They are often used as idealized configurations of atmospheric GCMs, and in this context the usual convention is that landmasses and topography are removed. Although many flavours of aquaplanet configurations exist, another convention is to retain as much of the atmospheric model's formulation as possible. That is, the numerical grid, dynamical core, and parameterized physics are all used just as in realistic climate simulations.

The Tier 1 aquaplanet experiments follow the same experimental design as CMIP5/CFMIP-2 (Medeiros et al., 2015). Those, in turn, were closely related to previous aquaplanet descriptions. In particular, the control configuration closely follows the AquaPlanet Experiment protocol (Blackburn and Hoskins, 2013) using a prescribed SST pattern described by Neale and Hoskins (2000). Two additional runs parallel the CFMIP-2 amip4K and amip4xCO2 experiments: a uniform 4K warming and a quadrupling of atmospheric CO2.

Here we provide the detailed experimental protocol for the three aquaplanet simulations that are part of Tier 1. We again that these follow the APE protocol and CMIP5/CFMIP-2, and therefore largely mirror previous descriptions in Blackburn and Hoskins (2013), Williamson et al. (2012), and Medeiros et al. (2015).

Orbital parameters are set to perpetual equinox conditions. This is usually achieved by setting eccentricity and obliquity to zero to define a circular orbit and insolation independent of calendar. The diurnal cycle is retained. Insolation is based on a non-varying solar constant of 1365 W m2.

The SST is non-varying and zonally uniform. The longitudinal variation is specified using the “Qobs” SST pattern from Neale and Hoskins (2000), given by:

\[
T(\phi) = \begin{cases} 
\frac{1}{2} \left( 1 - \sin^2 \phi \right) - \sin^2 \phi \delta T + T_{\text{min}} \text{ if } |\phi| < \frac{\pi}{2} \\
0 \text{, otherwise}
\end{cases}
\]  

where \( \delta T \) is the SST variation, \( \phi \) is latitude, \( \frac{\pi}{2} \) is the average latitude of the equatorial region, \( \phi_{\text{max}} \) is the latitude of the SST gradient, \( T_{\text{max}} = 27^\circ \text{C} \), and \( T_{\text{min}} = 0^\circ \text{C} \).

Because results are sensitive to the specification of the SSTs, groups that use a prognostic equation for the surface skin temperature are asked to set this skin temperature to the specified SST. No sea ice is prescribed, so the surface temperature is spatially uniform at 0°C poleward of 60° for the control simulation.

Radiatively active trace gases are well-mixed with mixing ratios following the AMIP II recommendations: CO2: 348 ppmv; CH4: 1650 ppbv; N2O: 306 ppbv; Halocarbon yield of approximately 0.24 W m-2 radiative forcing. The ozone distribution is the same as used in APE and CFMIP2/CMIP5, and is derived from the climatology used in AMIP II (Gates et al., 1999), and is constant in time and symmetric zonally about the equator. This ozone distribution is provided as a netCDF file which is archived on the Earth System Grid and available via the DOI http://dx.doi.org/10.5065/D6153406. Ozone values are provided up to 0.28 hPa (about 60 km altitude in mid-latitudes). For models with tops above this level a high top ozone dataset is also provided, which is available via the DOI http://dx.doi.org/10.5065/D64X5653. The ozone climatologies provided use pressure as a vertical coordinate. Most models use a sigma or hybrid vertical coordinate in pressure or altitude, which will mean that the pressure on a given model level varies in time, near the surface at the very least. Although the ozone climatology can be interpolated to the pressure of each model level as it varies in time within the model, for simplicity we recommend interpolating the ozone dataset onto the model vertical grid before the experiment is performed, and then specifying ozone values which are constant in time on each model level. This vertical interpolation will require a zonal symmetric climatology of ozone on model levels which is as consistent as possible with that expected in the aqua-control experiment. This could for example be produced by initially running a test version of the aqua-control experiment with an ozone climatology taken from a more realistic model configuration such as the AMIP DECK experiment.
Aerosols are removed to the extent possible to remove aerosol-radiation interaction (aka direct effects) and aerosol-cloud interaction (aka indirect effects). No external surface emissions are to be prescribed. Models requiring aerosol for cloud condensation should use a constant oceanic climatology that is symmetric about the equator and zonally. Alternatively, models with the capability should set the cloud droplet and crystal numbers to \(100 \times 10^6 \text{ m}^{-3}\) and \(0.1 \times 10^6 \text{ m}^{-3}\), respectively (as in Medeiros et al., 2016). As in APE, it is recommended that the atmospheric dry mass be adjusted to yield a global mean of 101080 Pa. It is also recommended to adopt the APE recommended values for geophysical constants, as listed in Table 2 of Williamson et al. (2012).

The aqua-4K experiment follows the above protocol, but with SST derived by adding 4K to Eq. B1. The aqua-4xCO2 experiment replaces the CO2 mixing ratio with 1392 ppmv. The SST is unchanged from the control simulation (Eq. B1).

Model runs should be 10 years. We recommend discarding the initial spin up period of a few months.

APPENDIX C: SST Pattern for CFMIP amip-future4K/amipFuture experiments

The amip-future4K (formerly amipFuture) experiment is the same as the amip DECK experiment, except that the SSTs are subject to a composite SST warming pattern derived from the CMIP3 coupled models. The patterned SST forcing dataset is available in a netcdf file called cfmp2_4k_patterned_sst_forcing.v1.0.nc which is available in the supplementary information for this paper, and via the CFMIP website. This is a normalised multi-model ensemble mean of the ocean surface temperature response pattern (the change in ocean surface temperature (TOS) between years 0-20 and 140-160, the time of CO2 quadrupling in the 1% runs) from thirteen CMIP3 AOGCMs (cica, ccmr, gfdlcm20, gfdlcm21, gisser, inmcm3, ipsl, miroc-medres, miub, mpi, mri, ncar-ccsm3, and ncar-pcm1.) Before computing the multi-model ensemble mean, each model's TOS response was divided by its global mean and multiplied by 4. This guarantees that the pattern information from all models is weighted equally and the global mean SST forcing is the same as in the uniform +4K experiment. We have retained the SST forcing based on the CMIP3 coupled models because we consider it more important to be able to compare CMIP5 and CMIP6 models forced with the same SST pattern than to use a pattern which is consistent with, say, the CMIP3 coupled response.
Figure 1. Summary of CFMIP-3 experiments and DECK + CMIP6 Historical experiments.
Figure 2. CFMIP-3 cfSites locations. The contours give an indication of inter-model spread in cloud feedback from the CFMIP-2 amip/amip4K experiments (please refer to Webb et al., 2015a for details).
### Table 1. Summary of CFMIP Tier 1 experiments.

<table>
<thead>
<tr>
<th>Experiment Name</th>
<th>Experiment Description / Design</th>
<th>Configuration</th>
<th>Start Year</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>amip</td>
<td>This is a single ensemble member of the AMIP DECK experiment which contains additional outputs which are required for model evaluation using COSP, and as control values for model outputs in the amip-p4K, amip-4xCO2, amip-future4K and amip-m4K experiments.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>36</td>
</tr>
<tr>
<td>amip-p4K</td>
<td>As CMIP5/CFMIP-2 amip-p4K experiment. AMIP experiment where SSTs are subject to a uniform warming of 4K.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>36</td>
</tr>
<tr>
<td>amip-4xCO2</td>
<td>As CMIP5/CFMIP-2 amip-4xCO2 experiment. AMIP experiment where SSTs are held at control values and the CO\textsubscript{2} seen by the radiation scheme is quadrupled.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>36</td>
</tr>
<tr>
<td>amip-future4K</td>
<td>As CMIP5/CFMIP-2 amipFuture experiment. AMIP experiment where SSTs are subject to a composite SST warming pattern derived from coupled models, scaled to an ice-free ocean mean of 4K.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>36</td>
</tr>
<tr>
<td>aqua-control</td>
<td>Extended version of CMIP5/CFMIP-2 aquaControl experiment. Aquaplanet (no land) experiment with no seasonal cycle forced with specified zonally symmetric SSTs.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>10</td>
</tr>
<tr>
<td>aqua-p4K</td>
<td>Extended version of CMIP5/CFMIP-2 aqua-p4K experiment. Aquaplanet experiment where SSTs are subject to a uniform warming of 4K.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>10</td>
</tr>
<tr>
<td>aqua-4xCO2</td>
<td>Extended version of CMIP5/CFMIP-2 aqua-4xCO2 experiment. Aquaplanet experiment where SSTs are held at control values and the CO\textsubscript{2} seen by the radiation scheme is quadrupled.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>10</td>
</tr>
</tbody>
</table>
Table 2. Summary of CFMIP Tier 2 experiments.

<table>
<thead>
<tr>
<th>Experiment Name</th>
<th>Experiment Description / Design</th>
<th>Configuration</th>
<th>Start Year</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>amip-m4K</td>
<td>As amip experiment but SSTs are subject to a uniform cooling of 4K.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>36</td>
</tr>
<tr>
<td>amip-lwoff</td>
<td>As amip experiment, but with cloud-radiative effects switched off in the LW radiation code.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>36</td>
</tr>
<tr>
<td>amip-p4K-lwoff</td>
<td>As amip-p4K experiment, but with cloud-radiative effects switched off in the LW radiation code.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>10</td>
</tr>
<tr>
<td>aqua-control-lwoff</td>
<td>As aqua-control experiment, but with cloud-radiative effects switched off in the LW radiation code.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>10</td>
</tr>
<tr>
<td>abrupt-solp4p</td>
<td>Conceptually similar to abrupt 4xCO2 DECK experiment, except that the solar constant rather than CO2 is abruptly increased by 4%.</td>
<td>Coupled AOGCM</td>
<td>1850</td>
<td>150</td>
</tr>
<tr>
<td>abrupt-solm4p</td>
<td>Same as abrupt-solp4p, except solar constant is reduced by 4% rather than increased.</td>
<td>Coupled AOGCM</td>
<td>1850</td>
<td>150</td>
</tr>
<tr>
<td>abrupt-2xCO2</td>
<td>Identical to the DECK abrupt4xCO2, but at 2xCO2.</td>
<td>Coupled AOGCM</td>
<td>1850</td>
<td>150</td>
</tr>
<tr>
<td>abrupt-0p5xCO2</td>
<td>Identical to the DECK abrupt4xCO2, but at 0.5xCO2.</td>
<td>Coupled AOGCM</td>
<td>1850</td>
<td>150</td>
</tr>
<tr>
<td>amip-piForcing</td>
<td>Identical to AMIP DECK experiment but from 1870-present with constant pre-industrial forcing levels (anthro &amp; natural).</td>
<td>Atmos-only</td>
<td>1870</td>
<td>145</td>
</tr>
<tr>
<td>piSST</td>
<td>An AGCM experiment with monthly-varying SSTs, sea-ice, atmospheric constituents and any other necessary boundary conditions (e.g. vegetation if required) taken from each model's own piControl run (using the 30 years of piControl that are parallel to years 111-140 of its abrupt4xCO2 run). Dynamic vegetation should be turned off in all the piSST set of experiments.</td>
<td>Atmos-only</td>
<td>Year 111</td>
<td>30 of abrupt-4xCO2</td>
</tr>
<tr>
<td>piSST-p4K</td>
<td>Same as piSST, but with a spatially and temporally uniform SST anomaly applied on top of the monthly-varying piSST SSTs. The magnitude of the uniform increase is taken from each model's global, climatological annual mean open SST change between abrupt4xCO2 minus piControl (using the mean of years 111-140 of abrupt4xCO2, and the parallel 30-year section of piControl).</td>
<td>Atmos-only</td>
<td>Year 111</td>
<td>30 of abrupt-4xCO2</td>
</tr>
<tr>
<td>piSST-4xCO2-rad</td>
<td>Same as piSST but CO2 as seen by the radiation scheme is quadrupled.</td>
<td>Atmos-only</td>
<td>Year 111</td>
<td>30 of abrupt-4xCO2</td>
</tr>
<tr>
<td>piSST-4xCO2</td>
<td>Same as piSST but CO2 is quadrupled. The increase in CO2 is seen by both the radiation scheme and vegetation.</td>
<td>Atmos-only</td>
<td>Year 111</td>
<td>30 of abrupt-4xCO2</td>
</tr>
<tr>
<td>a4SST</td>
<td>As piSST, but with monthly-varying SSTs taken from years 111-140 of each model's own abrupt4xCO2 experiment instead of from piControl. Sea-ice is unchanged from piSST.</td>
<td>Atmos-only</td>
<td>Year 111</td>
<td>30 of abrupt-4xCO2</td>
</tr>
<tr>
<td>a4SSTice</td>
<td>As piSST, but with monthly-varying SSTs and sea-ice taken from years 111-140 of each model's own abrupt4xCO2 experiment instead of from piControl.</td>
<td>Atmos-only</td>
<td>Year 111</td>
<td>30 of abrupt-4xCO2</td>
</tr>
<tr>
<td>a4SSTice-4xCO2</td>
<td>As a4SSTice, but CO2 is quadrupled, and the increase in CO2 is seen by both the radiation scheme and vegetation.</td>
<td>Atmos-only</td>
<td>Year 111</td>
<td>30 of abrupt-4xCO2</td>
</tr>
<tr>
<td>amip-a4SST-4xCO2</td>
<td>Same as amip, but a patterned SST anomaly is applied on top of the monthly-varying amip SSTs. This anomaly is a monthly climatology, taken from each model's own abrupt4xCO2 run minus piControl (using the mean of years 111-140 of abrupt4xCO2, and the parallel 30-year section of piControl). CO2 is quadrupled, and the increase in CO2 is seen by both the radiation scheme and vegetation.</td>
<td>Atmos-only</td>
<td>1979</td>
<td>36</td>
</tr>
<tr>
<td>Dataset</td>
<td>Years</td>
<td>Observables</td>
<td>Applications</td>
<td>References</td>
</tr>
<tr>
<td>---------------</td>
<td>---------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------</td>
</tr>
</tbody>
</table>

References


Chadwick, R., Which Aspects of CO$_2$ forcing and SST Warming Cause Most Uncertainty in Projections of Tropical Rainfall? 21


Harrop, B.E. and Hartmann, D.L. The role of cloud radiative heating in determining the location of the ITCZ in aqua planet simulations. Journal of Climate, 2016.


Oueslati, B., Bony, S., Risi, C. and Dufresne, J.-L.: Interpreting the inter-model spread in regional precipitation projections in the tropics of surface evaporation and cloud radiative effects. Climate Dynamics, (Published Online), 2016.


Voigt, A., Bony, S., Dufresne, J.-L. and Stevens, B.: The radiative impact of clouds on the shift of the inter-tropical

Voigt, A. and Shaw, T. A.: Circulation response to warming shaped by radiative changes of clouds and water vapour, Nature

Webb, M., C. Senior, S. Bony, and J. J. Morcrette, Combining ERBE and ISCCP data to assess clouds in the Hadley Centre,

Webb, M. J., C. A. Senior, D. M. H. Sexton, W. J. Ingram, K. D. Williams, M. A. Ringer, B. J. McAvaney, Colman, R.,
Soden, B.J., Gudgel, R., Knutson, T., Emori, S., Ogura, T., Tsushima, Y., Andronova, N., Li, B., Musat, I., Bony, S. and
Taylor, K. E.: On the contribution of local feedback mechanisms to the range of climate sensitivity in two GCM ensembles.

Webb, M. J. and Lock, A. P.: Coupling between subtropical cloud feedback and the local hydrological cycle in a climate

Webb, M.J., Lock, A.P., Bodas-Salcedo, A., Bony, S., Cole, J.N.S., Koshiro, T., Kawai, H., Lacagnina, C., Selten, F.M.,
Roehring, R. and Stevens, B.: The diurnal cycle of marine cloud feedback in climate models. Climate Dynamics 44.5-6, 1419-
1436, 2015a.

Roehring, R., Shin, Y., Mauritsen, T., Sherwood, S.C., Vial, J., Watanabe, M., Woelfle, M.D. and Zhao, M: The impact of
parametrized convection on cloud feedback. Phil. Trans. R. Soc. A, 373, 2054, 20140414, 2015b.


Williams, K. D., Ingram, W. J., and Gregory, J. M.: Time variation of effective climate sensitivity in GCMs. Journal of
Climate, 21, 5076-5090, 2008.

Williams, K.D., Harris, C.M., Bodas-Salcedo, A., Camp, J., Comer, R.E., Copsey, D., Fereday, D., Graham, T., Hill, R.,
Hinton, T. and Hyder, P.. The Met Office global coupled model 2.0 (GC2) configuration. Geoscientific Model Development,
8, 1509-1524, 2015.

Williamson, D. L., Blackburn, M. Hoskins, B. J., Nakajima, K., Ohfuchi, W., Takahashi, Y. O., Hayashi, Y.-Y., Nakamura,
Ishiwatari, M., McGregor, J. L., Borth, H., Wirth, V., Frank, H., Bechtold, P., Wedi, N., P., Tomita, H., Satoh, M., Zhao,

Williams, K.D., Harris, C.M., Bodas-Salcedo, A., Camp, J., Coner, R.E., Copsey, D., Fereday, D., Graham, T., Hill, R.,
Hinton, T. and Hyder, P.: The Met Office global coupled model 2.0 (GC2) configuration. Geoscientific Model Development,
8, 1509-1524, 2015.


10.1175/JCLI-D-11-00249, 2012b.

Zelinka, M.D., Klein, S.A. and Hartmann, D.L: Computing and Partitioning Cloud Feedbacks Using Cloud Property
Histograms. Part II: Attribution to Changes in Cloud Amount, Altitude, and Optical Depth., J. Climate, .25, 3736-3754, DOI:
10.1175/JCLI-D-11-00249, 2012b.


<table>
<thead>
<tr>
<th>Page</th>
<th>Action</th>
<th>User</th>
<th>Date/Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Deleted</td>
<td>mark.webb</td>
<td>28/09/2016 09:34:00</td>
</tr>
<tr>
<td>1</td>
<td>Inserted</td>
<td>mark.webb</td>
<td>12/10/2016 09:50:00</td>
</tr>
<tr>
<td>1</td>
<td>Inserted</td>
<td>mark.webb</td>
<td>31/08/2016 13:34:00</td>
</tr>
<tr>
<td>2</td>
<td>Inserted</td>
<td>mark.webb</td>
<td>12/10/2016 10:13:00</td>
</tr>
<tr>
<td>1</td>
<td>Inserted</td>
<td>mark.webb</td>
<td>12/10/2016 09:50:00</td>
</tr>
<tr>
<td>1</td>
<td>Deleted</td>
<td>mark.webb</td>
<td>31/08/2016 13:34:00</td>
</tr>
<tr>
<td>2</td>
<td>Inserted</td>
<td>mark.webb</td>
<td>30/08/2016 14:06:00</td>
</tr>
<tr>
<td>1</td>
<td>Deleted</td>
<td>mark.webb</td>
<td>30/08/2016 14:06:00</td>
</tr>
<tr>
<td>2</td>
<td>Deleted</td>
<td>mark.webb</td>
<td>30/08/2016 14:22:00</td>
</tr>
<tr>
<td>2</td>
<td>Inserted</td>
<td>mark.webb</td>
<td>30/08/2016 14:07:00</td>
</tr>
<tr>
<td>2</td>
<td>Deleted</td>
<td>mark.webb</td>
<td>30/08/2016 14:08:00</td>
</tr>
<tr>
<td>2</td>
<td>Deleted</td>
<td>mark.webb</td>
<td>30/08/2016 14:08:00</td>
</tr>
<tr>
<td>2</td>
<td>Inserted</td>
<td>mark.webb</td>
<td>12/10/2016 10:13:00</td>
</tr>
<tr>
<td>2</td>
<td>Deleted</td>
<td>mark.webb</td>
<td>30/08/2016 14:12:00</td>
</tr>
<tr>
<td>2</td>
<td>Deleted</td>
<td>mark.webb</td>
<td>30/08/2016 14:22:00</td>
</tr>
<tr>
<td>2</td>
<td>Deleted</td>
<td>mark.webb</td>
<td>30/08/2016 15:25:00</td>
</tr>
<tr>
<td>2</td>
<td>Inserted</td>
<td>mark.webb</td>
<td>15/09/2016 15:24:00</td>
</tr>
</tbody>
</table>

The CFMIP approach is also increasingly being used to understand other aspects of climate change, and so a second objective has now been introduced, to improve understanding of circulation, regional-scale precipitation, and non-linear changes.

Bretherton et al., 2015 and in comprehensive climate models (e.g. Webb and Lock 2013; Brient and Bony 2013; Ringer et al., 2014; Medeiros et al., 2015; Bretherton et al., 2015; Andrews and Gregory, 2016; Andrews et al., 2015; Brient et al., 2015; Tsushima et al., 2015;
The primary objective of CFMIP is to inform future assessments of cloud feedbacks through improved understanding of cloud-climate feedback mechanisms and better evaluation of cloud processes and cloud feedbacks in climate models. However, the CFMIP approach is also increasingly being used to understand other aspects of climate change, and so a second objective has been introduced, to improve understanding of circulation, regional-scale precipitation, and non-linear changes.

The primary goal of CFMIP is to inform improved assessments of cloud feedbacks on climate change. However, the CFMIP approach is increasingly being used to understand other aspects of climate response, such as regional circulation and precipitation changes, and non-linear changes.

This paper describes and documents the CFMIP contribution to the current phase on the Coupled Model Intercomparison Project (CMIP6, Eyring et al., 2016). It is anticipated that CFMIP-3 will eventually be broader than what is described here, for instance including studies with process models, but for the purposes of this document CFMIP-3 should be considered to be synonymous with the CFMIP contribution to CMIP6.
It is anticipated that CFMIP-3 will eventually be broader than what is described here, for instance including studies with process models, but for the purposes of this document CFMIP-3 should be considered to be synonymous with the CMIP contribution to CMIP6.

Most of the CFMIP-3 experiments are based on CO$_2$ concentration forced amip, piControl and abrupt-4xCO2 CMIP DECK (Diagnostic, Evaluation and Characterization of Klima) experiments (Eyring et al., 2016). Unless otherwise specified below, the CFMIP-3 experiments should be configured consistently with the DECK experiments on which they are based, using consistent model formulation, and forcings and boundary conditions as specified by Eyring et al., 2016.

with atmosphere-only general circulation models (AGCMs)
e.g. Briant and Bony, 2012; Webb and Lock, 2013; Briant and Bony, 2013; Ringer et al., 2014; Bretherton et al., 2014; Lacagnina et al., 2014; Gordon and Klein, 2014; Chepfer et al., 2014; Sherwood et al., 2014; Medeiros et al., 2015; Briant et al., 2015; Tsushima et al., 2015; Bellomo and Clement, 2015; Dal Gesso at al., 2015; Webb et al., 2015a, Webb et al., 2015b, Ceppi et al., 2016

simulator outputs (e.g. Gordon and Klein, 2014; Chepfer et al., 2014; Tsushima et al., 2015, Ceppi et al., 2016) and

e.g. Webb and Lock, 2013; Sherwood et al., 2014; Briant et al., 2015; Webb et al., 2015a; Dal Gesso at al., 2015

various quantities

, including precipitation
Because for the it is not possible to tune the models to reproduce a known answer, these, as it is not possible to tune them to observations in the same way as is for more realistic configurations under
Care should be taken to ensure that SSTs are increased in any inland bodies of water and near coastal edges, for example by linearly interpolating the provided warming pattern dataset to fill in missing data before re-gridding to the target resolution.

In models which employ a fixed lower threshold near freezing for the SST used in the calculation of the surface fluxes, this should ideally also be reduced by 4K.

, for example in interpreting differing cloud feedbacks between future CO₂ forced experiments and those representing the Last Glacial Maximum, as highlighted by Yoshimori et al., 2009

; Voigt and Shaw, 2015
In this configuration, the models will have a shortwave cloud feedback but no longwave cloud feedback. We note that the presence of clouds does affect the shortwave radiative heating of the atmosphere, although this is a much smaller effect than its longwave equivalent (e.g. Pendergrass and Hartmann, 2014).

while retaining those in the shortwave (Fermepin and Bony, 2014)

Care should also be taken to remove the effects of cloud on any longwave cooling used in other model schemes (e.g. turbulent mixing) if these are calculated independently of the radiation scheme.

An alternative method (proposed by Aiko Voigt) was also considered, in which clear-sky heating rates would be applied in the atmosphere while retaining the all-sky fluxes at the surface. Although this approach would potentially isolate the effects of cloud heating in the atmosphere more cleanly than the Iwoff experiments proposed here, it is yet to be demonstrated in a pilot study, and is considered more technically difficult to implement than the Iwoff experiments, which are very similar to those piloted by Fermepin and Bony, 2014.

When changing the solar constant, the shape of the spectral solar irradiance distribution should remain consistent with that in the piControl experiment.

; Meraner et al., 2013

These are the same as the DECK abrupt4xCO2 experiment except that CO₂ concentrations are doubled and halved respectively relative to the preindustrial control. These experiments are
because this requires the removal of the time varying forcing.

The experiment therefore complements the alternative approach of diagnosing

Time
in the *amip* experiment could alternatively be diagnosed by subtracting a time-varying radiative forcing diagnosed from RFMIP experiments. However, the *amip-piForcing* approach has the benefit of diagnosing the time-varying feedbacks over the full 1870-present period rather than the last 36 years, and does so with reference to a single experiment, which reduces noise compared to that which would be present with a double difference of the amip experiment and two RFMIP experiments.

Also, this first requires estimating the forcing and adjustments (e.g. from RFMIP) and removing them from the standard *amip* experiment, since the approach here extends the time-period of the *amip* simulation and only requires a single experiment (rather than pairs) which reduces the noise.

These are forced with monthly- and annually-varying monthly mean SSTs and sea ice, which reproduce regional precipitation patterns more accurately than is possible using climatological SST forcing (Skinner et al., 2012).
The time slice experiments can be combined in various ways to isolate the climate response to each individual aspect of forcing and warming. For example the response to SST pattern change is given by taking the difference between $a4SST$ and $piSST-pxK$, and the plant physiological response is found by taking the difference between $piSST-4xCO2$ and $piSST-4xCO2-rad$.

we have dispensed with the cfSites outputs in the aquaplanet and amip-future4K experiments because these have been less widely used compared to those from the other experiments.

Kamae and Watanabe 2012;
and cloud adjustments (e.g. Kamae and Watanabe 2012; Ogura et al., 2014)
We have dispensed with the cloud water tendency terms because these have been less widely used than the temperature and humidity tendencies.

https://www.earthsystemcog.org/projects/wip/CMIP6DataRequest
(Please note that in the full data request these variable groups are in many cases split into a number of sub-tables. As noted above, the formal data request provides the definitive specification of the model outputs.)

COSP is available via the CFMIP website (https://www.earthsystemcog.org/projects/cfmip). Version 1.4 is a stable code release that was made available well in advance of CMIP6 at the request of the modelling groups. Small updates are required to enable some new diagnostics requested by CFMIP3/CMIP6, most notably joint histograms of particle size and optical thickness from the MODIS simulator; with these updates the code is known as version 1.4.1. Modeling centers are encouraged to update to COSP 1.4.1 to provide these new diagnostics but may provide results from COSP 1.4.
Developed over the last few years, COSP 2 substantially revises the infrastructure for integrating satellite simulators in climate models. COSP 2 makes many fewer inherent assumptions about the model representation of clouds than do previous versions but contains an optional interface allowing it to be used as a drop-in replacement for COSP 1.4 or COSP 1.4.1. At the time of this writing, COSP 2 is undergoing final testing in two climate models. Availability of the final version will be announced on the CFMIP website and modelling groups are free to adopt it for use in CFMIP at that time.

COSP 1.4, available via the CFMIP website ( ), is the official version to be used for CMIP6. This is a stable release that was made available well in advance of CMIP6 at the request of the modelling groups. Version 2 of COSP is under active development. At the time of writing, COSP 2 is in beta testing and does not have a stable release, and so is not currently permitted for production of CMIP6 data. COSP-2 may be permitted for use in CMIP6 along with COSP 1.4 in the future; if and when this happens details will be posted on the CFMIP website.

The CFMIP approach is also increasingly being used to understand other aspects of climate change, such as circulation, regional-scale precipitation and non-linear changes.

We are grateful to Florent Brient, Hideo Shiogama, Aiko Voigt, Mark Ringer and two anonymous referees for helpful comments on the manuscript.
The lead coordinators are responsible for encouraging analysis of the relevant experiments as broadly as possible across the scientific community. While they may lead some analysis themselves, they do not have any first claim on analysing or publishing the results. All interested investigators are encouraged to exploit the data from these experiments. While investigators may wish to liaise with the lead coordinators to avoid duplicating work that others are doing, this is not a requirement.

Ozone values are provided up to 0.28hPa (about 60km altitude in mid-latitudes). For models with tops above this level, a high top ozone dataset is also provided, which is available via the DOI [http://doi.org/10.5065/D64X5653](http://doi.org/10.5065/D64X5653). The ozone climatologies provided uses pressure as a vertical coordinate. Most models use a sigma or hybrid vertical coordinate in pressure or altitude, which will mean that the pressure on a given model level varies in time, near the surface at the very least. Although the ozone climatology can be interpolated to the pressure of each model level as it varies in time within the model, for simplicity we recommend interpolating the ozone dataset onto the model vertical grid before the experiment is performed, and then specifying ozone values which are constant in time on each model level. This vertical interpolation will require a zonally symmetric climatology of pressure on model levels which is as consistent as possible with that expected in the aqua-control experiment. This could for example be produced by initially running a test version of the aqua-control experiment with an ozone climatology taken from a more realistic model configuration such as the AMIP DECK experiment.
We have retained the SST forcing based on the CMIP3 coupled models because we consider it more important to be able to compare CMIP5 and CMIP6 models forced with the same SST pattern than to use a pattern which is consistent with, say, the CMIP5 coupled response.


