Application the Cloud Feedback Model Intercomparison Project (CFMIP) to become a CMIP6-Endorsed MIP

Mark Webb, Chris Bretherton, Sandrine Bony, Jen Kay, Steve Klein, Pier Siebesma,

Bjorn Stevens, George Tselioudis, Masahiro Watanabe,

Peter Good, Timothy Andrews, Roger Marchand, Robin Chadwick and Hervé Douville

Updated 3rd June 2015

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

CFMIP started in 2003 and its first phase (CFMIP-1) organised an intercomparison based on perpetual July SST forced Cess style +2K experiments and 2xCO_2 equilibrium mixed-layer model experiments containing ISCCP simulator in parallel with CMIP3 (McAvaney and Le Treut, 2003). Results from 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, and featured prominently in the fourth and fifth IPCC assessments.

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 to understand cloud-climate feedback processes. CFMIP-2 organized further experiments as part of CMIP5, introducing seasonally varying SST perturbation experiments for the first time, as well as fixed SST CO_2 forcing experiments to examine cloud adjustments, and idealized 'aquaplanet' experiments to establish the contributions of land and zonally asymmetric circulations to cloud feedback uncertainties (Bony et al., 2011). CFMIP-2 also introduced satellite simulators to CMIP via the CFMIP Observation Simulator Package (COSP), 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. Additionally CFMIP-2 introduced into CMIP5 process diagnostics such as temperature and humidity budget tendency terms and high frequency 'cfSites' outputs at 120 locations around the globe. CFMIP also organized a joint project with the GEWEX Global Atmospheric System Study (GASS) called CGILS (the CFMIP-GASS Intercomparison of LES and SCMs) to develop cloud feedback intercomparison cases to assess the physical credibility of cloud feedbacks in climate models by comparing Single Column Models (SCM) versions of GCMs with high resolution Large Eddy Simulations (LES) models. Additionally CFMIP-2 developed the CFMIP-OBS data portal and the CFMIP diagnostic codes repository (see http://www.cfmip.net for more details).

Early studies arising from CFMIP-2 include numerous model evaluation studies using COSP, studies attributing cloud feedbacks and cloud adjustments to different cloud types, and the finding that idealized 'aquaplanet' experiments without land or Walker circulations are able to capture the essential differences between models' global cloud feedbacks and cloud adjustments. 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. CGILS has demonstrated a consensus in the responses of LES models to climate forcings and identified a number of shortcomings in the physical representations of cloud feedbacks in climate models. 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 CO_2 forcing and climate change. Studies using CFMIP-2 outputs from CMIP5 remain ongoing and many further results are expected to feed into future assessments of the representation of clouds and cloud feedbacks in climate models. For a list of publications arising from CFMIP-2, please refer to the CFMIP publications page at http://www.cfmip.net.

Given the previous record of CFMIP activities and the case outlined below we would like to request that CFMIP be endorsed as a CMIP6 project to continue support for community activities in this important area of research. We provide information on our plans for CFMIP-3 structured according to the provided criteria below.

Name of MIP: The Cloud Feedback Model Intercomparison Project (CFMIP)

Co-chairs: Mark Webb mark.webb@metoffice.gov.uk, Chris Bretherton breth@washington.edu

Members of the Scientific Steering Committee: Mark Webb (Met Office), Chris Bretherton (U. Washington), Sandrine Bony (IPSL), Jen Kay (CIRES), Steve Klein (PCMDI), Pier Siebesma (KNMI), Bjorn Stevens (MPG), George Tselioudis (NASA GISS), Masahiro Watanabe (U. Tokyo)

Link to website: http://www.cfmip.net

Goal of the MIP and a brief overview: The primary goal of CFMIP is to inform improved assessments of climate change cloud feedbacks. However, the CFMIP approach is increasingly being used to understand other aspects of climate response, such as 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 to be achieved by:

- Ongoing organized coordinated 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. cfSites) as well as process-based investigation of cloud maintenance and feedback mechanisms (e.g. cfSites, budget tendency terms, etc.)
- Ongoing development and improvement of COSP and CFMIP-OBS infrastructure.
- Ongoing collaboration with the cloud process modelling community (via GASS collaboration) on CGILS and via new efforts to develop a hierarchy of experiments linking GCMs with cloud resolving models (CRMs) and Large Eddy Simulation (LES) models run on large domains (e.g. via the IMPULSE project consortium).
- Organising annual meetings to provide a focus for community activities relevant to CFMIP and also to the broader community working to understand changes in clouds, circulation and precipitation which impact regional projections of climate change. (These two communities are increasingly becoming connected because the experiments designed for CFMIP are also useful in addressing a broader range of questions not directly related to clouds.)

References:

- Andrews, T., (2014), Using an AGCM to diagnose historical effective radiative forcing and mechanisms of recent decadal climate change. *J. Climate*, 27, 1193–1209, doi:10.1175/JCLI-D-13-00336.1.
- Bony, S., Webb, M., Bretherton, C. S., Klein, S. A., Siebesma, P., Tselioudis, G., & Zhang, M. (2011). CFMIP: Towards a better evaluation and understanding of clouds and cloud feedbacks in CMIP5 models. Clivar Exchanges, 56(2), 20-22.
- Good, P., Andrews, T., Bouttes, N., Chadwick, R., Gregory, J. M., Lowe, J. A. (2014). The nonlinMIP intercomparison project: physical basis, experimental design and analysis principles. In preparation; (attached)
- McAvaney BJ, Le Treut H (2003) The cloud feedback intercomparison project: (CFMIP). In: CLIVAR Exchanges—supplementary contributions. 26: March 2003.
- Skinner, C.B., M. Ashfaq, and N.S. Diffenbaugh (2012). Influence of twenty-first-century atmospheric and sea surface temperature forcing on West African climate. J. Climate, 25, 527-542.

- Stevens B., Bony S., Frierson, D.M, Jakob, C., Kageyama, M., Pincus, R, Shepherd, T., Sherwood, S., Siebesma, A. P., Sobel, A., Watanabe, M., Webb, M.J. (2014). Clouds, Circulation and Climate Sensitivity: A Grand Science Challenge. World Climate Research Programme Report No. 8/2014

We argue below the CFMIP and its proposed experiments meet the requirements laid out by the CMIP panel, as outlined below.

1. CFMIP and its experiments directly address the key science questions of CMIP6. The question that CFMIP most directly addresses is `How does the Earth system respond to forcing?' The CFMIP emphasis on understanding cloud feedbacks makes CFMIP highly relevant to this question. The next most relevant question is `What are the origins and consequences of systematic model biases?' CFMIP has a strong model evaluation component via the use of satellite simulators, process diagnosis and comparisons with LES, and a proven track record in investigating the link between errors in cloud processes and cloud feedbacks. CFMIP is also relevant to the question `How can we assess future climate changes given climate variability, climate predictability, and uncertainties in scenarios?' CFMIP will continue to supplement fully coupled CMIP experiments with idealised experiments that focus on basic understanding of the dominant uncertainties associated with cloud feedbacks. This will continue to support work which relates variability on observable timescales (e. g. seasonal to decadal) to longer term climate change responses (e.g. via 'emergent constraints'). For example the amipPiForcing experiment proposed below will support studies relating cloud variability and feedbacks on observable timescales to long term cloud feedbacks (Andrews, 2014).

Note also that the WCRP Grand Challenge on Clouds, Circulation and Climate is led by two CFMIP committee members (Bony and Stevens), and has three additional CFMIP committee members on its steering committee (Webb, Siebesma, Watanabe), including one of the CFMIP co-chairs. This puts CFMIP in an excellent position to directly address the questions arising from the WCRP Grand Challenge.

2. CFMIP builds on and connects to the shared CMIP DECK and CMIP6 historical experiments. The AMIP experiment is the control simulation for the CFMIP amip4K, amip4xCO2 and amipFuture experiments which were proposed by CFMIP for CMIP5 and which we would like to see continued in CMIP6 as Tier I experiments. The proposed Tier II experiments also connect to the AMIP DECK experiment; the AMIP preindustrial forcing experiment and amip minus 4K experiments also use the DECK AMIP experiment as a control. The abrupt +/- 4% solar constant experiments build on and contrast with the DECK abrupt4xCO2 experiment, as do the abrupt4xCO2 and abrupt0.5CO2 experiments. Additionally the atmosphere-only timeslice experiments build on the abrupt4xCO2 experiment, decomposing the regional response of each model's abrupt4xCO2 run into separate responses to each aspect of forcing and warming. Additionally CFMIP will propose additional process diagnostics and simulator outputs for the CMIP6 historical experiment, which will allow process based comparisons with the AMIP experiments to assess the impact of coupled SST errors on the simulation of clouds and regional precipitation patterns in the CMIP6 models.

3. CFMIP will continue to follow the CMIP modeling infrastructure standards and conventions, in terms of experimental design, data format and documentation. CFMIP-2 experiments were organized as part of CMIP5 and the CFMIP co-chairs have demonstrated the ability to follow all of the relevant standards in experimental protocols, in specification of diagnostic output requests, data formats and documentation. We commit to continuing in this spirit for CFMIP experiments which are coordinated through CMIP6.

4. All experiments are tiered, well‑defined, and useful in a multi‑model context and don't overlap with other CMIP6 experiments.

These are outlined below, and detailed specifications are provided in the accompanying spreadsheet. They are tiered into Tiers I and II. Additionally we give guidance on other experiments currently under development which we may propose as additional Tier II experiments in the future. Alternatively these additional experiments may be coordinated outside of CMIP.

These experiments are we believe useful in the multi-model context because the common purpose that they share is a focus on understanding the inter-model uncertainty/spread in cloud adjustments and cloud feedbacks as well as that in regional precipitation and circulation change and non-linear change. Investigation of inter-model requires multi-model analysis and hence all of these experiments are useful (and in fact require) a multi-model context. The usefulness of the Tier I experiments to a number of climate researchers has already been demonstrated by the large number of publications produced using CFMIP-2 experiments.

We have checked for overlaps with other CMIP6 experiments and are confident that links with other MIPS (e.g. nonLinMIP, GeoMIP, SolarMIP, RFMIP and PMIP) are based on complementary but non-overlapping experiments.

Summary of proposed experiments

Tier I Science questions, activities and experiments

1.1 Continuation of CFMIP-2 experiments - Lead coordinator: Mark Webb (Met Office)

Science Question: 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 CMIP5/CFMIP-2 experiments and diagnostic outputs have enabled considerable progress on these questions but participation by a larger fraction of modelling groups is required in CMIP6 for a more comprehensive assessment of the uncertainties across the full multi-model ensemble. Our proposal is essentially to retain the CFMIP-2/CMIP5 experiments in Tier I for CMIP6. The experiments to be retained are amip4K, amip4xCO2, amipFuture, aquaControl, aqua4xCO2 and aqua4K. These build on the amip DECK experiment. As the output requirements for the DECK are not yet finalised, it is possible that the DECK AMIP experiment will not contain all of the output diagnostics required for CFMIP. For this reason we also request an additional CFMIP AMIP experiment including the full set of CFMIP diagnostics, both for model evaluation and for interpretation of feedbacks and adjustments in conjunction with other Tier I CFMIP experiments. If all of the proposed CFMIP diagnostics are included in the DECK experiment, this additional CFMIP AMIP experiment will not be required.

Tier II Science questions, activities and experiments

2.1 Abrupt +/-4% Solar Forced AOGCM experiments - Lead coodinators: Chris Bretherton (UW), Roger Marchand (UW), Bjorn Stevens (MPI)

Science Question: How do responses in the climate system due to changes in solar forcing differ from changes due to CO_2, and is the response sensitive to the sign of the solar forcing?

Rapid adjustments in clouds and precipitation are now recognized as significant components of models' responses to CO_2 forcing. While they 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 cloud and precipitation adjustment, due to a smaller atmospheric absorption for a given top of atmosphere forcing. Solar forcing also has a weaker impact on the stratosphere than CO_2, potentially resulting in different upper tropospheric meridional temperature gradients and storm track responses.

A +4% solar experiment would be equivalent to the abrupt4xCO2 experiment but would increase the solar constant abruptly by 4 percent, resulting in a radiative forcing of a similar magnitude to that due to CO_2 quadrupling. This would provide a useful complement to the DECK abrupt4xCO2 experiment, and would support our understanding of regional responses of the coupled system with and without CO_2 adjustments. A complementary -4% abrupt solar forcing experiment 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 relate to past climates.

2.2 Abrupt2xCO2 and abrupt0.5xCO2 Experiments (nonLinMIP) - Lead Coordinator Peter Good (Met Office Hadley Centre)

Science Question: To what extent is regional-scale climate change per CO_2 doubling state-dependent (nonlinear), and why? How does the balance of mechanisms differ for high-forcing compared to low-forcing scenarios or paleoclimate simulations?

To address this question we propose two new experiments for Tier II, abrupt2xCO2 and abrupt0.5xCO2, to explore global and regional-scale nonlinear responses, highlighting different behavior under business-as-usual scenarios, mitigation scenarios and paleoclimate simulations. Additional experiments may be proposed for Tier II in the future, or coordinated via CFMIP outside of CMIP6. These include 100-year extensions to abrupt4xCO2 and abrupt2xCO2; a 1% ramp-down from the end of the 1pctCO2 experiment; an abrupt step-down to 1xCO2 from year 100 of the abrupt4xCO2. These would be used to explore longer-timescale responses, quantify nonlinear mechanisms more precisely and understand the reversibility of climate change.

2.3 amipMinus4K Experiment: Lead Coordinator: Mark Webb (Met Office)

Science Question: Are cloud feedbacks symmetric when subject to climate cooling rather than warming, and if not, why not?

An amipMinus4K experiment would take a similar form to the amip4K experiment, except that the sea surface temperatures would be uniformly reduced by 4K. This will be used to investigate asymmetric responses of clouds to a cooling climate in an idealized experiment, providing a link to PMIP. This experiment also complements the abrupt0.5xCO2 and the -4% solar experiments in that one can identify asymmetries in the warming/cooling response with and without interactions with the ocean. This experiment has been proposed for CFMIP following discussions with PMIP representatives (Pacale Braconnot, Masa Kageyama, and Masakazu Yoshimori).

2.4 Feedbacks in AMIP experiments: Lead Coordinator: Tim Andrews (Met Office)

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

Experiment and rationale: The previous CFMIP design was unable to diagnose time-dependent feedbacks that potentially undermine the simple linear forcing-feedback paradigm and which may be relevant to the gap between observed and modeled estimates of climate sensitivity. To address this we propose an additional experiment called ‘amipPiForcing’ (amip pre-industrial forcing), which is exactly the same as the standard amip run (i.e. SSTs and sea-ice) but run for the period 1870-present with constant pre-industrial forcings (i.e. all anthropogenic and natural forcing boundary conditions identical to the piControl run). Since the forcing constituents do not change in this experiment it readily allows a simple diagnosis of the simulated atmospheric feedbacks to observed SST changes, which can then be compared to feedbacks representative of long term change and climate sensitivity (e.g. from abrupt4xCO2 or amip4K). This has an advantage over the alternative approach of first estimating the forcing and adjustments (e.g. from RFMIP) and removing them from the amip experiment since the approach here only requires a single experiment (rather than pairs) which reduces the noise. The experiment has the additional benefit, by differencing with the standard amip run, of providing detailed information on the transient effective radiative forcing and adjustments in models relative to pre-industrial for the standard AMIP period. The inclusion of CFMIP process diagnostics not available in the RFMIP experiments will also enable a deeper understanding of the factors underlying forcing and feedback differences in the present and future climate.

2.5 Timeslice experiments for understanding regional climate responses to CO_2 forcing. Co-ordinators: Rob Chadwick (Met Office) and Hervé Douville (CNRM)

Science questions:

- How do regional climate responses (of e.g. precipitation) in a coupled model arise from the combination of responses to different aspects of CO_2 forcing and warming (uniform SST warming, pattern SST warming, direct CO_2 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?

We propose a set of 6 3~~2~~0-year atmosphere-only timeslice experiments to decompose the regional responses of each model's abrupt4xCO2 run into separate responses to each aspect of forcing and warming (uniform SST warming, pattern SST change, increased CO_2, plant physiological effect). As well as allowing regional responses in each individual model to be better understood, this set of experiments should prove especially useful for understanding the causes of model uncertainty in regional climate change.

The experiments are: 1) sstPi – –An AGCM experiment with monthly-varying SSTs, sea-ice and atmospheric constituents taken from 30 years of each model's own piControl run (using the 30 years of piControl that are parallel to years 111-140 of its abrupt4xCO2 run). ~~the same as amip but with monthly-varying SSTs and sea-ice from years 101-120 of each model's own control run rather than observed fields~~; 2) sstPiUniform~~4K~~ – the same as sstPi but with SSTs uniformly increased. The magnitude of the uniform increase is taken from each model's global, climatological annual mean SST change between abrupt4xCO2 minus piControl (using the mean of years 111-140 of abrupt4xCO2, and the parallel 30-year section of piControl). ~~by 4K~~; 3) sstPi4xCO2 – the same as sstPi but CO_2 as seen by the radiation scheme is quadrupled; 4) sstPi4xCO2Veg – the same as sstPi4xCO2 but with the plant physiological response also able to respond to the increased CO_2; 5) sstPiFuture –As sstPi, but with monthly-varying SSTs and sea-ice taken from years 111-140 of each model's own abrupt4xCO2 experiment instead of from piControl. ~~the same as sstPi but a seasonally varying monthly mean climatology of the SST pattern anomaly taken from years 91-140 of each model's own abrupt4xCO2 minus piControl is scaled to have a global mean increase of 4K and applied~~; 6) sstPiTot – the same as sstPiFuture but also with $4xCO_2$ including the plant effect. sstPiTot is used to establish whether a timeslice experiment can adequately recreate the coupled $abrupt4xCO_2$ response in each model, and then forms the basis for a decomposition using the other experiments.

We also propose an additional amip based experiment, amipTot: the 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.~~ the same as sstPiTot but with the SST pattern anomaly climatology from sstPiFuture added instead to the observed background SSTs and sea-ice (as for other amip experiments~~). Comparison of amipTot and sstPiTot should help to illuminate the impact of SST biases on regional climate responses in each model, and how this contributes to inter-model uncertainty.

2.6 Atmosphere-only experiments for understanding the role of cloud-radiative effects in the large-scale atmospheric circulation in current and perturbed climates. Co-ordinators: Sandrine Bony (IPSL) and Bjorn Stevens (MPI).

Science questions:

- How do cloud-radiative effects impact the structure, the strength and the variability of the general atmospheric circulation in the present-day climate?
- How much do cloud-radiative feedbacks contribute to the spread of circulation and precipitation responses in climate change?
- Can we identify robust aspects of the climate response to global warming that do not depend on cloud-radiative feedbacks?

It is increasingly recognized that clouds, and cloud-radiative effects in particular, play a critical role in the general circulation of the atmosphere (ITCZ, MJO, storm tracks, hurricanes) and its response to global warming. A better assessment of this role would greatly help 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 circulation is expected to improve our ability to answer two of the four science questions raised by the WCRP Grand Challenge on Clouds, Circulation and Climate Sensitivity: what controls the position, the strength, and the variability of the storm tracks and of the tropical rainbelts?

These questions provided the scientific motivation for the Clouds On/Off Klima Intercomparison Experiment (COOKIE) project proposed by the european consortium EUCLIPSE and CFMIP in 2012. The COOKIE experiments, which have been run by 4 to 8 climate models (depending on the experiment), consisted in switching 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-sea contrasts. The change in circulation between On and Off experiments was resulting from both effects, obscuring a bit the mechanisms through which the atmospheric cloud-radiative effects interact with the circulation for given surface boundary conditions. As the LW cloud-radiative effects are felt mostly within the troposphere (and represent most of the LW+SW cloud-radiative heating) while the SW effects are felt mostly at the surface, we could better isolate the role of tropospheric cloud-radiative effects on the circulation by running atmosphere-only experiments in which clouds are made transparent to radiation only in the LW.

We propose in Tier II a set of simple experiments similar to the amip, amip4K, aquaControl and aqua4K experiments of CMIP5/CFMIP2 (and Tier 1 of CMIP6) but in which cloud-radiative effects are switched off in the LW part of the radiation code. These experiments will be referred to as offlwamip, offlwamip4K, offlwaquaControl and offlwaqua4K. The analysis of idealized (aqua-planet) 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 dynamicists (e.g. DynVar). The comparison of the inter-model spread of simulations between AMIP and offlwAMIP experiments for present-day and globally warmer climates will help identify which aspects of the spread depend on the representation of cloud-radiative effects, and which aspects do not, thus better highlighting other sources of spread.

Additional CFMIP experiments under consideration for the future

We also propose to use these CMIP6 experiments as the foundation for further experiments planned in the context of the Grand Challenge on Clouds, Circulation and Climate Sensitivity. These will include for example sensitivity experiments to assess the impacts of different physical processes on cloud feedbacks and regional circulation/precipitation responses, and others designed to test specifically proposed cloud feedback mechanisms. Additional experiments further idealizing the aquaplanet framework to a non-rotating rotationally symmetric case are also under development. These will be proposed as additional Tier II experiments at a future time, or coordinated by CFMIP outside of CMIP6.

5. Unless a Tier 1 experiment differs only slightly from another well –established experiment, it must already have been performed by more than one modeling group. All of the proposed Tier I experiments were previously included in CMIP5 and so are well established and already performed by multiple groups.

6. A sufficient number of modelling centers (~8) are committed to performing all of CFMIP's Tier 1 experiments and providing all the requested diagnostics needed to answer at least one of its science questions. Fourteen modeling groups have so far agreed to participate in CFMIP as part of CMIP6, implying that they are prepared to perform the Tier I experiments. These are ACCESS (Australia), BCC (China), CanESM (Canada), CESM (USA), CNRM (France), FGOALS (China), GFDL (USA), IPSL-ESM (France), MIROC6-GCM (Japan) NICAM (Japan), MPI-ESM (Germany), MRI (Japan) and UKESM (United Kingdom).

7. The MIP presents an analysis plan describing how it will use all proposed experiments, any relevant observations, and specially requested model output to evaluate the models and address its science questions. Our analysis plan is outlined below.

We commit to contributing to the creation of the CMIP6 data request and to analyzing the data, as we did for CMIP5. This will include making proposals for an updated COSP request in CMIP6 (see the proposal from the COSP PMC), and also additional improvements to the CFMIP diagnostic specifications relating to temperature and humidity budget increments, 3D radiative fluxes, inclusion of aerosol diagnostics across CFMIP experiments, and the introduction of additional locations in the cfSites specification.

We also commit to identifying observations needed for model evaluation and improved process understanding, and to contributing directly to making such datasets available as part of obs4MIPs. For example the CFMIP community has up to now played a central role in providing versions of CloudSat and CALIPSO datasets designed for direct comparison with CMIP5 data through the CFMIP-OBS website (see http://climserv.ipsl.polytechnique.fr/cfmip-obs/) and part of this work has recently involved publishing this data via the ESG and linking into obs4MIPS (see for example references to CFMIP-OBS on the obs4MIPS website at https://www.earthsystemcog.org/projects/obs4mips/aboutus). This work will continue.

CFMIP analysis activities are ongoing and the CFMIP community is ready to analyse CMIP6 data at any time. We would like modelling groups to perform the proposed CFMIP/CMIP6 experiments at the same time or shortly after their DECK experiments. Subsequent CFMIP experiments which are not included in CMIP6 will build on the proposed DECK and CMIP6/CFMIP experiments and some will start as soon as CMIP6 DECK experiments start to become available. We envisage a succession of CFMIP related intercomparisons addressing different questions arising from the GC spanning the duration of CMIP6.

We commit to scientifically analyze, evaluate and exploit the proposed experiments, and have identified leads within CFMIP for different aspects of this activity. An overview of the proposed evaluation/analysis of the CMIP DECK and CMIP6 experiments follows:

- CFMIP will continue to exploit the CMIP DECK and CMIP6 experiments to understand and evaluate cloud processes and 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 applied to an expanding number of models, including the new generation of models currently under development. These activities will include evaluation of clouds using additional simulators (see proposal regarding COSP below), investigation of cloud processes and cloud feedback/adjustment mechanisms using process outputs (cfSites, budget tendency terms, etc). The inclusion of COSP and budget tendency terms in additional DECK experiments (e.g. abupt4xCO2 and some scenario experiments, also see proposal for COSP below) will enable the CFMIP approach to be applied to a wider range of experimental configurations. (Lead coordinator Mark Webb).
- Analysis of the +/-4% solar model runs would 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 abrupt4xCO2 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 coordinator Chris Bretherton).
- Analysis of nonlinear climate processes will primarily involve comparing the abrupt4xCO2, abrupt2xCO2 and abrupt0.5xCO2 experiments over the same timescale (Good et al., 2014). (Lead coordinator Peter Good).
- Analysis of amipPiForcing has already been done in detail for a single model in Andrews (2014). We propose to use this has a starting point for a multi-model analysis. (Lead coordinator Timothy Andrews).
- An overview analysis of regional responses and model uncertainty in the timeslice and amipTot experiments will be carried out by the co-ordinators, in collaboration with members of contributing modeling 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 timeslice and amipTot experiments have already been successfully run with HadGEM2 (Met Office), and are currently in the planning stage for CNRM. (Lead coordinator Robin Chadwick).
- When analyzed together with the amip4K experiment, the amipMinus4K experiment allows one to exploit the CFMIP process diagnostics 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 coordinator Mark Webb).

8. The MIP has completed the MIP template questionnaire. We have done this.

9. The MIP contributes a paper on its experimental design to the CMIP6 Special Issue. We agree to do this.

10. The MIP considers reporting on the results by co−authoring a paper with the modelling groups. We agree to do this. Separate papers will be prepared for each of the experiment groups proposed.

Answers to other questions in the MIP template questionnaire

All model output archived by CMIP6-Endorsed MIPs is expected to be made available under the same terms as CMIP output. Most modeling groups currently release their CMIP data for unrestricted use. If you object to open access to the output from your experiments, please explain the rationale. We have no objection to this.

List of output and process diagnostics for the CMIP DECK/CMIP6 data request. Please see the accompanying spreadsheet and outline below.

Any proposed contributions and recommendations for model diagnostics and performance metrics, observations/reanalysis data products, tools, code or scripts. We have provided a database of performance metrics and codes at the CFMIP Diagnostics Code Repository and a set of observational data for comparison with CFMIP outputs at the CFMIP-OBS site. Both are accessible via the CFMIP website http://www.cfmip.net. We welcome additional contributions to both of these databases.

Any proposed changes from CMIP5 in NetCDF metadata (controlled vocabularies), file names, and data archive (ESGF) search terms. None expected.

Explanation of any proposed changes (relative to CMIP5) that will be required in CF, CMOR, and/or ESGF. None expected.

CFMIP Recommended Outputs For CMIP6 DECK experiments and CFMIP experiments.

CFMIP recommends a set of diagnostic outputs for the CMIP6 DECK and CFMIP experiments which are based on those from CFMIP-2, with some modifications. These are detailed in the accompanying spreadsheet CMIP6DataRequestCompilationCFMIP_20150331.xls, and are summarized below. The recommendations are in two parts. The first part describes updates to the CFMIP process diagnostics compared to those which were requested in CMIP5, in terms of additional variables and the experiments in which they are requested. This set was drawn up by the CFMIP committee and ratified by the modeling groups following a presentation at the 2014 CFMIP meeting. The second part describes recommendations for COSP outputs in the DECK and CMIP6 Historical experiments which were drawn up by the COSP Project Management Committee (PMC). Please refer to the request scoping worksheet in the accompanying spreadsheet for a summary of which outputs are requested in which DECK experiments CFMIP experiments.

For participation in CFMIP it is required that modeling groups commit to performing all of the Tier I experiments, and sufficient diagnostic outputs to answer at least one scientific question. Since a number of the science questions of CFMIP (e.g. those pertaining to precipitation responses) require no diagnostic outputs beyond the standard ‘Amon’ outputs from CMIP5, a modeling group may qualify for participation in CFMIP even if they run the Tier I experiments without CFMIP simulators or process outputs. Such a submission would be useful, in the main for the precipitation analysis aspects of CFMIP. However we strongly recommend that participating groups additionally submit as many of the COSP and process outputs as they are practically able to, to support investigations of the full range of scientific questions of CFMIP in CMIP6.

Proposed updates to CFMIP process outputs for the CMIP DECK, CMIP6 Historical and CMIP6 CFMIP experiments.

CFMIP Committee

March, 2015

The diagnostic request for CMIP5/CFMIP2 is summarised and motivated in the CFMIP-2 proposal document [Bony et al., 2009], and documented in detail in the CMIP5 Standard Output documentation at http://cmip-pcmdi.llnl.gov/cmip5/output_req.html in excel spreadsheet format (Worksheet 'CFMIP output' indicates which tables appear in which experiments and for which periods, which other worksheets such as cfMon, cfDay etc indicate the variables in each table). Our view is that the CFMIP-2 diagnostics set is fundamentally sound and forms a suitable basis for the process diagnostics in the DECK, CMIP6 Historical and CMIP6 CFMIP experiments. Thus, we present this proposal as changes with respect to the CMIP5/CFMIP-2 protocol in the accompanying spreadsheet, which includes a request scoping worksheet indicating which outputs are requested in which experiments, including the CMIP6 DECK + CMIP6 Historical experiments and the CFMIP experiments proposed within CMIP6. In the sections below we present and motivate the specific requested changes.

In this section we cite a number of peer reviewed publications. Please refer to http://www.cfmip.net -> CFMIP Publications for full references.

cfSites Outputs: The CFMIP cfSites outputs were requested in CMIP5 for 120 locations in the amip, amip4K, amipFuture and amip4xCO2 experiments, and for 73 locations along the Greenwich meridian in the aquaplanet experiments. These outputs have so far been used to evaluate the models with in-situ measurements (e.g. Nuijens et al. (submitted), Guichard et al. (in prep), Neggers et al. (submitted) and to examine cloud feedbacks on short timescales such as over the diurnal cycle (Webb et al. 2015). For CMIP6 we have dispensed with the cfSites outputs in the aquaplanet experiments, and in amipFuture, retaining them in amip, amip4K and amip4xCO2 only. At the request of the US CLIVAR ETOS WG we have added St. Helena to the list in light of upcoming field work/additional radiosondes from these islands, increasing the total number of locations to 121. (Ascension island was also requested, but this was already present).

Temperature and humidity tendency terms: CFMIP-2 requested cloud, temperature and humidity tendency terms. In CMIP6 we have omitted the cloud condensate tendency terms because no publications have arisen from those saved in CMIP5. The temperature and humidity tendency terms from CMIP5 have been widely used however. Temperature and humidity tendency terms have been demonstrably useful for understanding the roles of different parts of the model physics in cloud feedbacks, adjustments, and present-day variability (Williams et al 2013, Webb and Lock 2013, Kamae and Watanabe 2012, Demoto et al 2013, Sherwood et al 2014, Ogura et al 2014, Brient et al. (submitted), Xavier et al. (submitted)). They have also been used to understand regional warming patterns such as polar amplification in coupled models (e.g. Yoshimori et al 2013,2014). For CMIP6 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. A shortcoming of the CMIP5 protocol was that we were unable to interpret the physical feedback mechanisms in coupled model experiments due to lack of process diagnostics. For this reason we are additionally requesting these budget terms in the DECK abrupt4xCO2 experiment and the pre-industrial control.

Additional daily diagnostics: So-called ‘clustering’ approaches 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., submitted). We have 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).

Proposal of request of COSP diagnostics for CMIP/DECK, CMIP6 Historical and CMIP6 CFMIP experiments

COSP Project Management Committee

March, 2015

1 Introduction

The initial design for CMIP6 has recently been published [Meehl et al., 2014]. It includes a set of 'DECK + CMIP6 Historical' experiments to be run by modelling groups whenever they develop a new model version:

- AMIP (1979-end)
- Pre-industrial control
- 1% yr^{-1} CO_2 increase up to $4xCO_2$
- Abrupt $4xCO_2$
- CMIP6 Historical run

These experiments are called the CMIP Diagnostic, Evaluation and Characterization of Klima (DECK) plus CMIP6 Historical experiments. In this document, we present the proposal of the list of COSP diagnostics to be requested to for the DECK + CMIP6 Historical experiments and additionally the CMIP6 CFMIP experiments. This proposal is the outcome of discussions by the COSP Project Management Committee (PMC) and the CFMIP Committee. The COSP diagnostic request for CMIP5/CFMIP2 is summarised and motivated in the CFMIP-2 proposal document [Bony et al., 2009], and documented in detail in the CMIP5 Standard Output documentation at http://cmip-pcmdi.llnl.gov/cmip5/output_req.html in excel spreadsheet format (Worksheet 'CFMIP output' indicates which tables appear in which experiments and for which periods, which other worksheets such as cfMon, cfDay etc indicate the variables in each table). Our view is that the CFMIP-2 diagnostics set fundamentally sound and forms a suitable basis for the COSP request for the DECK, CMIP6 Historical and CMIP6 CFMIP experiments, subject to some modifications. Thus, we present this proposal as changes with respect to the CMIP5/CFMIP-2 protocol in the accompanying spreadsheet. The request scoping sheet also shows which outputs are requested in which experiments. We have tried to address the concerns raised in the CMIP5 survey by simplifying the technical difficulty of the requests (sometimes at the expense of extra data) and basing the requests upon a frozen well-tested and already-released version of COSP (v1.4). In the sections below we present and motivate the specific requested changes.

2 Description of proposed changes

2.1 Change #1: Replacement of curtain data by full 3D fields, and deletion of cfOff table (proposed by Alejandro Bodas-Salcedo)

In CFMIP-2, the production of data along the A-train track ("curtain" data, table cf3hr offline) involved a substantial amount of post-processing. A second post-processing step required the gridding and time-averaging of these data to produce the monthly means requested in the cfOff table. This proved quite difficult for many modelling centres. Although not from the ESG archive, this type of data has been used in several model evaluation papers [Bony et al., 2009; Bodas-Salcedo et al., 2008; Field et al., 2011; Williams et al., 2013] involving case-study comparison of models with along-track observations from CloudSat and Calipso. We believe that by simplifying the request, the modelling centres will find easier to contribute these data. Hence, we propose to drop the orbital sampling, i.e. to request globally-complete fields on a standard lat/lon grid. Given this change, the calculation of monthly-averages from gridded 3-hourly data is trivial, and therefore we propose to delete the cfOff table.

2.2 Change #2: New table cfMonExtra. Add CloudSat and CALIPSO CFADs to cfMonExtra (proposed by Alejandro Bodas-Salcedo and Mark Webb)

Optimisations to the code in COSP v1.4 mean that it is now practical to run the CloudSat simulator inline in models and so for longer periods. We propose the introduction of a new table cfMonExtra for the inclusion of monthly mean COSP diagnostics used for model evaluation in the AMIP DECK experiment, but which we don't consider appropriate for coupled or climate change experiments. In this new table we include Cloud Frequency/Altitude Diagram (CFAD) diagnostics for CloudSat and CALIPSO for the entire AMIP integration. CFADs for CloudSat and CALIPSO have appeared in a number of published studies [e.g. Nam et al., 2014; Franklin et al., 2013; Bodas-Salcedo et al., 2011; Bodas-Salcedo et al., 2012; Nam and Quaas, 2012; Nam and Quaas, 2013; Kay et al., 2012; Kodama et al., 2012; Marchand et al., 2009; Abel and Boutle, 2012] and their inclusion as monthly means in the AMIP DECK experiment will make them available for analysts in a more convenient form than the higher frequency outputs currently requested in CMIP5.

2.3 Change #3: Standard monthly COSP and daily COSP 2D outputs in all of the DECK, CMIP6 Historical and CMIP6 CFMIP Experiments (proposed by Mark Webb and Steve Klein)

Many of the standard monthly COSP and daily COSP 2D have been shown to be valuable in the CMIP5 experiments, not only for cloud evaluation [e.g. Franklin et al., 2013; Bodas-Salcedo et al., 2012; Nam and Quaas, 2013; Lacagnina and Selten, 2014; Bodas-Salcedo et al., 2014; Klein et al., 2013; Cesana and Chepfer, 2012; Tsushima et al., 2013] but also in quantifying the contributions of different cloud types to cloud feedbacks and forcing adjustments in climate change experiments [e.g. Tsushima et al., 2013; Zelinka et al., 2012a; Zelinka et al., 2012a; Zelinka et al., 2013; Zelinka et al., 2014]. We propose to include these in all of the DECK, CMIP6 Historical and CMIP6 CFMIP experiments as standard for the entire length of the runs, to support evaluation of cloud, cloud feedbacks and cloud adjustments and to investigate trends in the observational record.

2.4 Change #4: Move PARASOL reflectance to cfMonExtra (proposed by Robert Pincus)

Top-of-atmosphere reflectance measurements from PARASOL were part of the standard request for CMIP5. They have been used in some applications [e.g. Nam et al. 2012] but have not been widely exploited. The proposal is to move them from the cfMon to cfMonExtra tables to reduce the number of integrations for which they are requested and to focus on model evaluation applications.

2.5 Change #5: Add MISR CTH-COD to cfMonExtra. Add MISR CTH-COD and ISCCP CTP-OD histograms to cf3hr (proposed by Roger Marchand)

Histograms of cloud-top-height (or cloud-top-pressure) and optical-depth produced by ISCCP have been widely used in the evaluation of climate models, often in combination with the ISCCP-simulator now part of COSP. Because top-of-atmosphere outgoing longwave fluxes are related to cloud-top-height and outgoing shortwave fluxes are related to cloud-optical-depth this framework provides a way to evaluate the distribution of model clouds in a way that is closely related to their radiative impact. Similar histograms of cloud-top-height and optical-depth are being produced from observations by the Multiangle Imaging Spectro-Radiometer (MISR). While similar, the cloud-top-height in the MISR dataset is obtained using a stereo-imaging technique that his purely geometric and insensitivity to the calibration of the MISR cameras. This technique 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 ISCCP provides greater sensitivity to optically-thin high-level clouds. In addition, 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]. We therefore recommend using both ISCCP and MISR observations and instrument-simulators in the evaluation of climate model, and such an analysis is underway using a few CFMIP5 models that have run the MISR simulator [Hillman et al. 2014]. While monthly data are useful for the broad evaluation of models on monthly or longer time scales, the acquisition of high frequency (Three hourly) data will enable analysis of events that are not well resolved with monthly data, including the diurnal cycle, the Madden-Julian Oscillation (MJO) and various synoptic states or weather patterns, such as frontal passages. We recognize that this represents a large increase in data-volume compared with monthly averages and propose collection of this three hourly data only for a period of about 1 year.

2.6 Change #6: Add MODIS cloud fractions (total, liquid, ice) to cfMonExtra (proposed by Robert Pincus)

The partitioning between liquid and ice phase has significant impacts on the energy and hydrologic impacts of clouds. As models move towards predicting more details of the aerosol distributions, including the ice nucleation ability, evaluation of the phase partitioning on the global scale will become more important. Evaluation to date has been based primarily on polarization measurements from active and passive sensors [e.g. Doutriaux-Boucher and Quaas, 2004; Komurcu et al., 2014] and height-resolved partitioning estimates from the CALIPSO sensor are requested below. Cloud phase estimates from the MODIS simulator were not available in CFMIP2 but may prove a useful complement by virtue of greater geographic sampling and longer time records.

2.7 Change #7: MODIS COT-particle size histograms by phase in cfMonExtra, cfDayExtra, cf3hr (proposed by Robert Pincus)

The joint distribution of optical thickness and particle size provides a window on the microphysical processes within clouds [Nakajima et al., 1991] and is influenced by direct and some indirect effects of aerosols on cloud optical properties [Han et al. 2002]. As models move towards predicting more details of the aerosol properties and cloud-aerosol interactions the assessment of these processes becomes more pressing.

Estimate of particle size from MODIS have been difficult to use for model evaluation to date because of observational artefacts not treated by the MODIS simulator. These artefacts are reduced by the use of observations at wavelengths with greater absorption by condensed water (e.g. by exploiting reflectance at 3.7 μm instead of 2.1 μm). The MODIS simulator and accompanying data for CFMIP3 will use measurements at 3.7 μm to infer particle size. This will also act to make output from the MODIS simulator roughly consistent with the PATMOS-X observations in the same way that distributions of optical thickness from the MODIS, MISR, and ISCCP simulators are nearly equivalent.

2.8 Change #8: add CALIPSO ice and liquid 3D cloud fractions to cfMonExtra (proposed by Hélène Chepfer)

Changes in cloud optical depth associated with cloud phase feedbacks can dominate the changes in high-latitude clouds in future climate projections [e.g. Senior and Mitchell, 1993]. Cloud phase identification capabilities have been recently added to the CALIPSO simulator in COSP, and a compatible observational dataset has been produced [Cesana and Chepfer, 2013]. We propose to include these in the AMIP DECK experiment to support the evaluation of the simulation of cloud phase.

2.9 Change #9: CALIPSO total cloud fraction and PARASOL reflectance to cfDayExtra (proposed by Hélène Chepfer and Dimitra Konsta)

The multi-sensor A-train observations (CALIPSO-GOCCP and MODIS, PARASOL) allow to make the correlations between the different cloud variables at the instantaneous time scale, and at high resolution. The use of the high-frequency relationships between different variables allows for process-oriented model evaluation. These diagnostics will help test the realism of the co-variation of key cloud properties that control cloud feedbacks in models. Konsta at al. (2014) have used these diagnostics in a pilot analysis.

3 References Using Satellite Simulators for the Evaluation of Clouds in Models (partial)

Bodas-Salcedo, A. et al., 2008: Evaluating cloud systems in the Met Office global forecast model using simulated CloudSat radar reflectivities, *J. Geophys. Res.*, 113, D00A13. DOI: 10.1029/2007JD009620.

Bodas-Salcedo, A., et al., 2011: Satellite simulation software for model assessment. *Bull. Am. Meteorol. Soc.*, 92. DOI: 10.1175/2011BAMS2856.1.

Bodas-Salcedo, A., et al., 2012: The surface downwelling solar radiation surplus over the Southern Ocean in the Met Office model: the role of midlatitude cyclone clouds, *J. Climate*, 25. DOI: 10.1175/JCLI-D-11-00702.1.

Bodas-Salcedo, A., et al., 2014: Origins of the Solar Radiation Biases over the Southern Ocean in CFMIP2 Models, J. Climate, 27. DOI: 10.1175/JCLI-D-13-00169.1.

Cesana, G., and Chepfer, H., 2012: How well do climate models simulate cloud vertical structure? A comparison between CALIPSO-GOCCP satellite observations and CMIP5 models, *Geophys. Res. Let.* DOI: 10.1029/2012GL053153.

Cesana, G., and Chepfer, H., 2013: Evaluation of the cloud thermodynamic phase in a climate model using CALIPSO-GOCCP, *J. Geophys. Res.*, 118, 7922–7937. DOI: 10.1002/jgrd.50376.

Doutriaux-Boucher, M., and J. Quaas, 2004: Evaluation of cloud thermodynamic phase parametrizations in the LMDZ GCM by using POLDER satellite data, *Geophys. Res. Lett.*, 31, L06126. DOI: 10.1029/2003GL019095.

Field, P. R., et al., 2011: Using model analysis and satellite data to assess cloud and precipitation in midlatitude cyclones, *Q. J. R. Meteorol. Soc.*, 137, 1501-1515. DOI: 10.1002/qj.858.

Franklin, C. N., et al., 2013: Evaluation of clouds in ACCESS using the satellite simulator package COSP: Global, seasonal, and regional cloud properties, *J. Geophys. Res.*. DOI: 10.1029/2012JD018469.

Hillman, B., R. Marchand, T. P. Ackerman, A.Bodas-Salcedo, J. Cole, J.-C. Golaz, J. E. Kay, 2014: Comparing Cloud Biases in CMIP5: Insights Using MISR and ISCCP Observations and Satellite Simulators, *in preparation*.

Kay, J. E., et al., Exposing Global Cloud Biases in the Community Atmosphere Model (CAM) Using Satellite Observations and Their Corresponding Instrument Simulators, *J. Climate*, 25, 2012. DOI: 10.1175/JCLI-D-11-00469.1.

Klein, S. A. et al., 2013: Are climate model simulations of clouds improving? An evaluation using the ISCCP simulator, *J. Geophys. Res.*, 118. DOI: 10.1002/jgrd.50141.

Kodama, C., et al., 2012: An assessment of the cloud signals simulated by NICAM using ISCCP, CALIPSO, and CloudSat satellite simulators, *J. Geophys. Res*, 117. DOI: 10.1029/2011JD017317.

Komurcu, M., T. Storelvmo, I. Tan, U. Lohmann, Y. Yun, J. E. Penner, Y. Wang, X. Liu, and T. Takemura ,2014: Intercomparison of the cloud water phase among global climate models, J. Geophys. Res., 119, 3372–3400. DOI:10.1002/2013JD021119.

Konsta, D., J-L. Dufresne, H. Chepfer, A. Idelkadi and G. Cesana, 2014: Evaluation of clouds simulated by the LMDZ5 GCM using A-train satellite observations (CALIPSO, PARASOL, CERES). *Climate Dynamics*, under review.

Lacagnina, C., and Selten, F., 2014: Evaluation of clouds and radiative fluxes in the EC-Earth general circulation model, *Clim. Dyn.* DOI: 10.1007/s00382-014-2093-9.
Marchand, R., et al., 2009: A comparison of simulated cloud radar output from the multiscale modeling framework global climate model with CloudSat cloud radar observations, *J. Geophys. Res., 114*, D00A20. DOI: 10.1029/2008JD009790.
Marchand, R., T. Ackerman, M. Smyth, and W. B. Rossow ,2010: A review of cloud top height and optical depth histograms from MISR, ISCCP, and MODIS, *J. Geophys. Res.*, 115, D16206. DOI:10.1029/2009JD013422.
Nam, C., S. Bony, J.-L. Dufresne, and H. Chepfer, 2012: The "too few, too bright" tropical low-cloud problem in CMIP5 models, *Geophys. Res. Lett.*, 39. DOI:10.1029/2012GL053421.
Nam, C. C. W., and Quaas, J., 2012: Evaluation of Clouds and Precipitation in the ECHAM5 General Circulation Model Using CALIPSO and CloudSat Satellite Data. I, *J. Climate*, 25, 4975-4992. DOI:10.1175/JCLI-D-11-00347.1.
Nam, C. C. W. and Quaas, J., 2013: Geographically versus dynamically defined boundary layer cloud regimes and their use to evaluate general circulation model cloud parameterizations, *Geophys. Res. Let.* DOI: 10.1002/grl.50945.
Nam, C. W. W., et al., 2014: Evaluation of boundary layer cloud parametrizations in the ECHAM5 general circulation model using CALIPSO and CloudSat satellite data. *JAMES*. DOI: 10.1002/2013MS000277.
Tsushima, Y. et al., 2013: Quantitative evaluation of the seasonal variations in climate model cloud regimes, *Clim. Dyn.*, 41. DOI: 10.1007/s00382-012-1609-4.
Williams, K. D., et al., 2013: The Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models, *J. Climate*, 26, 3258-3274. DOI: 10.1175/JCLI-D-12-00429.1.
Zelinka, M. D., et al., 2012a: Computing and Partitioning Cloud Feedbacks Using Cloud Property Histograms. Part I: Cloud Radiative Kernels. *J. Climate*. DOI: 10.1175/JCLI-D-11-00248.1.
Zelinka, M. D., et al., 2012b: Computing and Partitioning Cloud Feedbacks Using Cloud Property Histograms. Part II: Attribution to Changes in Cloud Amount, Altitude, and Optical Depth. *J. Climate*. DOI: 10.1175/JCLI-D-11-00249.
Zelinka, M. D., et al., 2013: Contributions of Different Cloud Types to Feedbacks and Rapid Adjustments in CMIP5, *J. Climate*. DOI: 10.1175/JCLI-D-12-00555.1.
Zelinka, M. D., et al., 2014: Quantifying Components of Aerosol-Cloud-Radiation Interactions in Climate Models. *J. Geophys. Res*. DOI: 10.1002/2014JD021710.

4 Other References

Abel, S. J., and Boutle, I. A., 2012: An improved representation of the raindrop size distribution for single-moment microphysics schemes, *Q. J. R. Meteorol. Soc.*, 138, 2151-2162. DOI: 10.1002/qj.1949.
Bony, S. et al., 2009: The Cloud Feedback Model Intercomparison Project : Summary of Activities and Recommendations for Advancing Assessments of Cloud-Climate Feedbacks, (http://www.cfmip.net -> CFMIP Strategy and Plans -> CFMIP2_experiments_March20th2009.pdf).
Han, Q, W B. Rossow, J Zeng, R Welch, 2002: Three Different Behaviors of Liquid Water Path of Water Clouds in Aerosol–Cloud Interactions. *J. Atmos. Sci.*, **59**, 726–735. DOI: 10.1175/1520-0469(2002)059<0726:TDBOLW>2.0.CO;2.
Meehl, G. et al., 2014: Climate Model Intercomparisons: Preparing for the Next Phase, *Eos Trans. AGU*, 95(9), 77. DOI: 10.1002/2014EO090001.
Nakajima, T, M D. King, J D. Spinhirne, L F. Radke, 1991: Determination of the Optical Thickness and Effective Particle Radius of Clouds from Reflected Solar Radiation Measurements. Part II: Marine Stratocumulus Observations, *J. Atmos. Sci.*, **48**, 728–751. DOI: 10.1175/1520-0469(1991)048<0728:DOTOTA>2.0.CO;2.
Senior, C. A., and Mitchell, J. F. B., 1993: Carbon Dioxide and Climate. The Impact of Cloud Parameterization. *J. Climate,* 6, 393–418. DOI: 10.1175/1520-0442(1993)006<0393:CDACTI>2.0.CO;2.

Acknowledgements

Thanks to D. Konsta for providing text for Section 2.9.

The nonlinMIP intercomparison project: physical basis, experimental design and analysis principles

P. Good[1], T. Andrews[1], N. Bouttes[2], R. Chadwick[1], J. M. Gregory[2,1], J. A. Lowe[1]

[1]{ Met Office Hadley Centre, Exeter, United Kingdom}
[2]{ NCAS-Climate, University of Reading, Reading, United Kingdom}
Correspondence to: P. Good (peter.good@metoffice.gov.uk)

Abstract

nonlinMIP aims to quantify and understand, at regional scales, climate responses that are non-linear under CO2 forcing (mechanisms for which doubling the CO2 forcing does not double the response). Non-linear responses can be large at regional scales, with important implications for understanding mechanisms and for GCM emulation techniques (e.g. energy balance models and pattern-scaling methods). However, these processes are hard to explore using traditional experiments, explaining why they have had little attention in previous studies. Some single model studies have established novel analysis principles and some physical mechanisms. There is now a need to explore robustness and uncertainty in such mechanisms across a range of models.

nonlinMIP addresses this using a simple, small set of CO2-forced experiments that are able to separate linear and non-linear mechanisms cleanly, with good signal/noise – while being demonstrably traceable to realistic transient scenarios. The design builds on the CMIP5 and CMIP6 DECK protocols, and is centred around a suite of abruptCO2 experiments, with a ramp-up-ramp-down experiment to test traceability to gradual forcing scenarios. The understanding gained will help interpret the spread in policy-relevant scenario projections.

Here we outline the basic physical principles behind nonlinMIP, and the method of establishing traceability from abruptCO2 to gradual forcing experiments, before detailing the experimental design and finally some analysis principles. The discussion on traceability of abruptCO2 to transient experiments is also relevant to the abrupt4xCO2 experiment in the CMIP5 and CMIP6 DECK protocols.

5 Introduction

Climate impacts assessments require, at regional scales, understanding of physical mechanisms of climate change in GCM projections. Also required is the ability to emulate (using fast simplified climate models) GCM behaviour for a much larger range of policy-relevant scenarios than may be evaluated using GCMs directly. These two requirements may be combined into a single question: what is the simplest conceptual framework that has quantitative predictive power and captures the key mechanisms behind GCM scenario projections?

Often, a pragmatic choice has been to assume some form of linearity. In studies of the global energy balance, linearity is often assumed in the form of a constant climate feedback parameter. This parameter may be used to quantify feedbacks in different models (e.g. Zelinka et al., 2013) or, in emulation methods, to parameterise global energy balance models (e.g. Huntingford and Cox, 2000). In understanding or emulating regional patterns of climate change, it is often assumed that regional climate change is roughly proportional to global mean warming. In emulation work, this is termed 'pattern scaling' (Mitchell, 2003;Santer et al., 1990;Tebaldi and Arblaster, 2014), but this assumption may also be applied either explicitly or implicitly in understanding mechanisms. Sometimes, patterns of change per K of global warming are quantified; often, physical mechanisms are studied for a single period of a single forcing scenario (implicitly assuming that the understanding is relevant for other periods or scenarios).

While these approximations appear to work well under some circumstances, significant limitations are increasingly being revealed in such assumptions. These are of two types: different timescales of response, and non-linear responses. In discussing this, a complication arises in that different linearity assumptions exist. Henceforth we define 'linear' as meaning 'consistent with linear systems theory' - i.e. responses that are linear in model forcing (i.e. where doubling the forcing doubles the response; this is different from assuming that pairs of responses are linearly related to each other – as in pattern scaling).

Even in a linear system (where responses are linear in forcing), the relationship between two system outputs (e.g. between global-mean temperature and regional sea surface temperature - SST) will in general be non-linear. This is due to different timescales of response in different locations and/or variables. Examples include lagged surface ocean warming due to a connection with the deeper ocean (Chadwick et al., 2013;Held et al., 2010;Williams et al., 2008;Manabe et al., 1990;Andrews and Ringer, 2014) or the direct response of precipitation to forcings (Andrews et al., 2010;Allen and Ingram, 2002;Mitchell et al., 1987). One (generally false) assumption of pattern scaling, then, is that regional climate responds over the same timescale as global-mean temperature. Different timescales of response are especially important in understanding and predicting behaviour under mitigation and geoengineering scenarios (or over very long timescales).

Non-linear system responses (e.g. Schaller et al., 2013) are more complex to quantify, understand and predict than those of linear systems. Some examples have been known for some time, such as changing feedbacks through retreating snow/sea-ice (Colman and McAvaney, 2009;Jonko et al., 2013), or the Atlantic Meridional Overturning Circulation. More recently, substantial non-linear precipitation responses have been demonstrated in spatial patterns of regional precipitation change in two Hadley Centre climate models with different atmospheric formulations (Good et al., 2012;Chadwick and Good, 2013). This is largely due to simultaneous changes in pairs of known robust pseudo-linear mechanisms (Chadwick and Good, 2013). Non-linearity has also been demonstrated in the response under idealised geoengineering scenarios, of ocean heat uptake, sea-level rise, and regional climate pattterns, with different behaviour found when forcings are decreasing than when they are increasing (Bouttes et al., 2013;Schaller et al., 2014).

Investigation of these mechanisms at regional scales has been constrained by the type of GCM experiment typically analysed. Most previous analyses (e.g. Solomon et al., 2007) have used results from transient forcing experiments, where forcing changes steadily through the experiment. There are three main problems with this approach. First, information about different timescales of response is masked. This is because the GCM response at any given time in a transient forcing experiment is a mixture of different timescales of response (Good et al., 2013;Held et al., 2010;Li and Jarvis, 2009), including short-timescale responses (e.g. ocean mixed layer response from forcing change over the previous few years) through long-timescale behaviour (including deeper ocean responses from forcing changes multiple decades to centuries earlier). Secondly, in transient forcing experiments, non-linear behaviour is hard to separate from linear mechanisms. For example, in an experiment where CO2 is increased by 1% per year for 140 years ('1pctCO2'), we might find different spatial patterns at year 70 (at 2xCO2) than at year 140 (at 4xCO2). This could be due to nonlinear mechanisms (due to the different forcing level and associated different climate state). However, it could also be due to linear mechanisms: year 140 follows 140 years of forcing increase, so includes responses over longer response timescales than at year 70 (only 70 years of forcing increase). Thirdly, signal/noise ratios of regional climate change can be relatively poor in such experiments.

These three issues may be addressed by the use of idealised abruptCO2 GCM experiments (Forster et al., 2012;Zelinka et al., 2013;Jonko et al., 2013;Good et al., 2013;Good et al., 2012;Chadwick and Good, 2013;Chadwick et al., 2013;Bouttes et al., 2013;Gregory et al., 2004): an experiment where CO2 forcing is abruptly changed, then held constant. In abrupt CO2 experiments, responses over different timescales are separated from each other. Further, responses at different forcing levels may be directly compared, e.g. by comparing the response in abrupt2xCO2 and abrupt4xCO2 experiments over the same timescale - both have identical forcing time histories, apart from the larger forcing magnitude in abrupt4xCO2. Thirdly, high signal/noise is possible: averages may be taken over periods of 100 years or more (after the initial ocean mixed layer adjustment, change is gradual in such experiments). Recent work (Good et al., 2011;Good et al., 2012;Good et al., 2013;Zelinka et al., 2013) has established that these experiments contain global and regional-scale information quantitatively traceable to more policy-relevant transient experiments - and equivalently, that they form the basis for fast simple climate model projections traceable to the GCMs.

The CMIP5 abrupt4xCO2 experiments have thus been used widely: including quantifying GCM forcing and feedback behaviour (Gregory et al., 2004;Zelinka et al., 2013), and for traceable emulation of GCM projections of global-mean temperature and heat uptake (Good et al., 2013;Stott et al., 2013). Abrupt4xCO2 is also part of the CMIP6 DECK protocol (Meehl et al., 2014).

NonlinMIP extends the CMIP5 and CMIP6 DECK designs to explore non-linear responses (via additional abruptCO2 experiments at different forcing levels. It also explores responses over slightly longer timescales (extending the CMIP5 abrupt4xCO2 experiment by 100 years).

6 Relating abruptCO2 to gradual forcing scenarios: the step-response model

In using the highly-idealised abruptCO2 experiments, it is essential that their physical relevance (traceability) to more realistic gradual forcing experiments is determined. Some GCMs could respond unrealistically to the abrupt forcing change. A key tool here is the step-response model (described below). This response-function method aims to predict the GCM response to any given transient-forcing experiment, using the GCM response to an abruptCO2 experiment. Such a prediction may be compared with the GCM transient-forcing simulation, as part of a traceability assessment (discussed in detail in section 5).

Once some confidence is established in traceability of the abruptCO2 experiments to transient-forcing scenarios, the step-response model has other roles: to explore the implications, for different forcing scenarios, of physical understanding gleaned from abruptCO2 experiments; to help separate linear and nonlinear mechanisms (section 5); and potentially as a basis for GCM emulation. The method description below also serves to illustrate the assumptions of linear system theory.

The step-response model represents the evolution of radiative forcing in a scenario experiment by a series of step changes in radiative forcing (with one step taken at the beginning of each year). The method makes two linear assumptions. First, the response to each annual forcing step is estimated by linearly scaling the response in a CO_2 step experiment according to the magnitude of radiative forcing change. Second, the response y_i at year i of a scenario experiment is estimated as a sum of responses to all previous annual forcing changes (see Figure 1 of Good et al., 2013 for an illustration):

$$y_i = \sum_{j=0}^{i} w_{i-j} x_j \qquad (1a)$$

where x_j is the response of the same variable in year j of the CO_2 step experiment. w_{i-j} scales down the response from the step experiment (x_j) to match the annual step change in radiative forcing from year i to year j of the scenario (denoted ΔF_{i-j}):

$$w_{i-j} = \frac{\Delta F_{i-j}}{\Delta F_s} \quad (1b)$$

where ΔF_s is the radiative forcing change in the CO_2 step experiment. All quantities are expressed as anomalies with respect to a constant-forcing control experiment.

This approach can in principle be applied at any spatial scale for any variable for which the assumptions are plausible (e.g. Chadwick et al., 2013).

7 Linear and non-linear mechanisms, and the relevance of abruptCO2 experiments

Here we discuss further, with examples, the distinction between linear and nonlinear mechanisms, when they are important, and the relevance of abruptCO2 experiments.

7.1 Linear mechanisms: different timescales of response

Even in a linear system, regional climate change per K of global warming will evolve during a scenario simulation. This happens because different parts of the climate system have different timescales of response to forcing change.

This may be due to different effective heat capacities. For example, the ocean mixed layer responds much faster than the deeper ocean, simply due to a thinner column of water (Li and Jarvis, 2009). However, some areas of the ocean surface (e.g. the Southern Ocean and south-east subtropical Pacific) show lagged warming, due to a greater connection (via upwelling or mixing) with the deeper ocean (e.g. Manabe et al., 1990;Williams et al., 2008). The dynamics of the ocean circulation and vegetation may also have their own inherent timescales (e.g. vegetation change may lag global warming by years to hundreds of years, Jones et al., 2009). At the other extreme, some responses to CO2 forcing are much faster than global warming: such as the direct response of global mean precipitation to forcings (Allen and Ingram, 2002;e.g. Andrews et al., 2010;Mitchell et al., 1987) and the physiological response of vegetation to CO2 (Field et al., 1995).

In a linear system, patterns of change per K of global warming are sensitive to the forcing history. For example in Figure 1, a scenario is illustrated where forcing is ramped up, then stabilized. Three periods are highlighted, which may have different patterns of change per K of global warming, due to different forcing histories: at the leftmost point, faster responses will be relatively more important, whereas at the right, the slower responses have had some time to catch up. This is illustrated in Figure 2 for sea-level rise. The blue curves show that for RCP2.6, global-mean warming ceases after 2050, while sea-level rise continues at roughly the same rate throughout the century. This is largely because deep ocean heat uptake is much slower than ocean mixed-layer warming.

By design, abruptCO2 experiments separate different timescales of GCM response to forcing change. This is used, for example, (Gregory et al., 2004) to estimate radiative forcing and feedback parameters for GCMs: plotting radiative flux anomalies against global mean warming can separate 'fast' and 'slow' responses (see e.g. Figure 3).

7.2 Non-linear responses

Nonlinear mechanisms arise for a variety of reasons. Often, however, it is useful to describe them as state-dependent feedbacks. For example, the snow-albedo feedback becomes small at high or low snow depth. Sometimes, nonlinear mechanisms may be better viewed as simultaneous changes in pairs of properties. For example, convective precipitation is broadly a product of moisture content and dynamics (Chadwick and Good, 2013;Chadwick et al., 2012). Both moisture content and atmospheric dynamics respond to CO2 forcing, so in general we might expect convective precipitation to have a nonlinear response to CO2 forcing. Of course, more complex nonlinear responses exist, such as for the Atlantic Meridional Overturning Circulation.

In contrast to linear mechanisms, nonlinear mechanisms are sensitive to the magnitude of forcing. For example, the two points highlighted in Figure 4 may have different patterns of change per K of global warming, due to nonlinear mechanisms.

An example is given in Figure 5, which shows the albedo feedback declining with increased global temperature, due to declining snow and ice cover, and the remaining snow and ice being in areas of lower solar insolation (Colman and McAvaney, 2009).

AbruptCO2 experiments may be used to separate nonlinear from linear mechanisms. This can be done by comparing the responses at the same timescale in different different abruptCO2 experiments. Figure 6 compares abrupt2xCO2 and abrupt4xCO2 experiments over years 50-149. A 'doubling difference' is defined, measuring the difference in response to the first and second CO2 doublings. In most current simple climate models (e.g. Meinshausen et al., 2011), the radiative forcing from each successive CO2 doubling is assumed identical (because forcing is approximately linear in log[CO2], Myhre et al., 1998). With this assumption, a linear system would have zero doubling difference everywhere. Therefore, the doubling difference is used as a measure of nonlinearity. The question of which abruptCO2 experiments to compare, and over which timescale, is discussed in section 5.

In some GCMs, the forcing per CO2 doubling has been shown to vary with CO2 (Colman and McAvaney, 2009;Jonko et al., 2013). However, this variation depends on the specific definition of forcing used (Jonko et al., 2013). Currently this is folded into our definition of nonlinearity. If a robust definition of this forcing variation becomes available in future, it could be used to scale out any difference in forcing between pairs of abruptCO2 experiments, to calculate an 'adjusted doubling difference'.

As an example, Figure 7 maps the response to abrupt2xCO2 and abrupt4xCO2, and the doubling difference, for precipitation in HadGEM2-ES over the ocean (taken from Chadwick and Good). The nonlinearities are large - comparable in magnitude to the responses to abrupt2xCO2, albeit with a different spatial pattern.

8 Experimental design

nonlinMIP is composed of a set of abruptCO2 experiments (the primary tools), plus a CO2-forced transient experiment. These build on the CMIP5 and CMIP6 DECK protocols (the required runs from these are detailed in Table 1). The additional nonlinMIP runs (Table 2) are assigned three priority levels. Three options for participation are: 1) only the 'essential' simulation; 2) all 'high priority' plus the 'essential' simulations; or, preferably, 3) all simulations. The experiments in Table 1 are required in all cases. All experiments must be initialized from the same year of a pre-industrial control experiment, except for abrupt4xto1x (see Table 2). A typical analysis procedure is outlined in section 5.

The nonlinMIP design is presently limited to CO2 forcing, although the same principles could be applied to other forcings.

9 Basic analysis principles

This section outlines the general principles behind analysis of nonlinMIP results. The primary idea is to find where the step-response model (section 2) breaks: since the step-response model is based on a linear assumption, this amounts to detecting non-linear responses.

The aim is to focus subsequent analysis. If non-linearities in a quantity of interest are found to be small, then analysis may focus on understanding different timescales of response from a single abruptCO2 experiment: linearity means that the physical response (over a useful range of CO2 concentrations) is captured by a single abruptCO2 experiment. This represents a considerable simplification. If, on the other hand, non-linearities are found to be important, the focus shifts to understanding the different responses in different abruptCO2 experiments. The choice of which abruptCO2 experiments to focus on, and over which timescales, is discussed below.

9.1 First step: check basic traceability of abrupt4xCO2 to the transient-forced response near 4xCO2

This is to confirm that the abruptCO2 experiments contain realistic physical responses in the variables of interest (as previously done for global-mean temperature and heat uptake for a range of CMIP5 models (Good et al., 2013), and for other global-mean quantities for HadCM3 (Good et al., 2011). This also, rules out the most pathological non-linearities (e.g. if the response to an abrupt CO2 change in a given GCM was unrealistic).

The linear step-response model should first be used with the abrupt4xCO2 response, to predict the response near year 140 of the 1pctCO2 experiment (i.e. near 4xCO2). This prediction is then compared with the actual GCM 1pctCO2 result. This should first be done for global mean temperature: this assessment has been performed for a range of CMIP5 models (Good et al., 2013; see Figure 8), giving an idea of the level of accuracy expected. If the abruptCO2 response is fundamentally unrealistic, it is likely to show up in the global temperature change. This approach may then be repeated for spatial patterns of warming, and then for the quantities of interest. Abrupt4xCO2 is used here as it has larger signal/noise than abrupt2xCO2, yet is representative of forcing levels in a business-as-usual scenario by 2100. However, the tests may also be repeated using abrupt2xCO2 – but compared with year 70 of the 1pctCO2 experiment (i.e. at 2xCO2).

The step-response model emulation under these conditions should perform well for most cases: the state at year 140 of the 1pctCO2 experiment is very similar to that of abrupt4xCO2 (same forcing, similar global-mean temperature), so errors from non-linear mechanisms should be minimal. If large errors are found, this may imply caution about the use of abruptCO2 experiments for these variables, or perhaps point to novel non-linear mechanisms that may be understood by further analysis.

9.2 Second step: detecting nonlinear responses

Having established some level of confidence in the abruptCO2 physical response, the second step is to look for nonlinear responses. This first involves repeating the tests from step 1 above, but for different parts of the 1pctCO2 and 1pctCO2 ramp-down experiments, and using different abruptCO2 experiments for the step-response model.

An example is given in Figure 9 (but for different transient-forcing experiments). This shows results for global-mean precipitation in the HadCM3 GCM (Good et al., 2012). Here, the step-response model prediction using abrupt4xCO2 (red curves) only works where a transient-forced experiment is near to 4xCO2. Similarly, the prediction using abrupt2xCO2 (blue curves) works only near 2xCO2. Otherwise, quite large errors are seen, and the predictions with abrupt2xCO2 and abrupt4xCO2 are quite different from each other. This implies that there are large non-linearities in the precipitation response in this GCM, and that they may be studied by comparing the responses in the abrupt2xCO2 and abrupt4xCO2 experiments.

Having identified some non-linear response, and highlighted two or more abruptCO2 experiments to compare (in the previous example, abrupt2xCO2 and abrupt4xCO2), the non-linear mechanisms may be studied in detail by comparing the responses in the different abruptCO2 experiments over the same timescale (e.g. via the doubling difference, as in Figures 6,7). This allows (Good et al., 2012;Chadwick and Good, 2013) non-linear mechanisms to be separated from linear mechanisms (not possible in a transient-forcing experiment).

10 Conclusions

This paper outlines the basic physical principles behind the nonlinMIP design, and the method of establishing traceability from abruptCO2 to gradual forcing experiments, before detailing the experimental design and finally some general analysis principles that should apply to most studies based on this dataset.

Acknowledgements

This work was supported by the Joint UK DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). Nathaelle Bouttes received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013), ERC Grant Agreement 247220, project "Seachange."

References

Allen, M. R., and Ingram, W. J.: Constraints on future changes in climate and the hydrologic cycle, Nature, 419, 224-+, 10.1038/nature01092, 2002.

Andrews, T., Forster, P. M., Boucher, O., Bellouin, N., and Jones, A.: Precipitation, radiative forcing and global temperature change, Geophysical Research Letters, 37, Artn L14701
Doi 10.1029/2010gl043991, 2010.

Andrews, T., and Ringer, M. A.: Cloud feedbacks, rapid adjustments, and the forcing-response relationship in a transient co2 reversibility scenario, Journal of Climate, 27, 1799-1818, Doi 10.1175/Jcli-D-13-00421.1, 2014.

Bouttes, N., Gregory, J. M., and Lowe, J. A.: The reversibility of sea level rise, Journal of Climate, 26, 2502-2513, Doi 10.1175/Jcli-D-12-00285.1, 2013.

Chadwick, R., Boutle, I., and Martin, G.: Spatial patterns of precipitation change in cmip5: Why the rich don't get richer., Journal of Climate, accepted, 2012.

Chadwick, R., and Good, P.: Understanding non-linear tropical precipitation responses to co2 forcing, Geophysical Research Letters, 40, 10.1002/grl.50932, 2013.

Chadwick, R., Wu, P. L., Good, P., and Andrews, T.: Asymmetries in tropical rainfall and circulation patterns in idealised co2 removal experiments, Climate Dynamics, 40, 295-316, DOI 10.1007/s00382-012-1287-2, 2013.

Colman, R., and McAvaney, B.: Climate feedbacks under a very broad range of forcing, Geophysical Research Letters, 36, L01702
10.1029/2008gl036268, 2009.

Field, C. B., Jackson, R. B., and Mooney, H. A.: Stomatal responses to increased co2 - implications from the plant to the global-scale, Plant Cell Environ, 18, 1214-1225, DOI 10.1111/j.1365-3040.1995.tb00630.x, 1995.

Forster, P. M., Andrews, T., Good, P., Gregory, J. M., Jackson, L., and Zelinka, M. D.: Evaluating adjusted forcing and model spread for historical and future scenarios in the cmip5 generation of climate models, Journal of Geophysical Research-Atmospheres (accepted pending minor revisions), 2012.

Good, P., Gregory, J. M., and Lowe, J. A.: A step-response simple climate model to reconstruct and interpret aogcm projections, Geophysical Research Letters, 38, Artn L01703
Doi 10.1029/2010gl045208, 2011.

Good, P., Ingram, W., Lambert, F. H., Lowe, J. A., Gregory, J. M., Webb, M. J., Ringer, M. A., and Wu, P. L.: A step-response approach for predicting and understanding non-linear precipitation changes, Climate Dynamics, 39, 2789-2803, DOI 10.1007/s00382-012-1571-1, 2012.

Good, P., Gregory, J. M., Lowe, J. A., and Andrews, T.: Abrupt co2 experiments as tools for predicting and understanding cmip5 representative concentration pathway projections, Climate Dynamics, 40, 1041-1053, DOI 10.1007/s00382-012-1410-4, 2013.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., and Williams, K. D.: A new method for diagnosing radiative forcing and climate sensitivity, Geophysical Research Letters, 31, L03205
10.1029/2003gl018747, 2004.

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F. R., and Vallis, G. K.: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing, Journal of Climate, 23, 2418-2427, Doi 10.1175/2009jcli3466.1, 2010.

Huntingford, C., and Cox, P. M.: An analogue model to derive additional climate change scenarios from existing gcm simulations, Climate Dynamics, 16, 575-586, 2000.

IPCC: Summary for policymakers, in: Climate change 2013: The physical science basis. Contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change, edited by: Stocker, T. F., Qin, D., Plattner, G. K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Jones, C., Lowe, J., Liddicoat, S., and Betts, R.: Committed terrestrial ecosystem changes due to climate change, Nat Geosci, 2, 484-487, Doi 10.1038/Ngeo555, 2009.

Jonko, A. K., Shell, K. M., Sanderson, B. M., and Danabasoglu, G.: Climate feedbacks in ccsm3 under changing co2 forcing. Part ii: Variation of climate feedbacks and sensitivity with forcing, Journal of Climate, 26, 2784-2795, Doi 10.1175/Jcli-D-12-00479.1, 2013.

Li, S., and Jarvis, A.: Long run surface temperature dynamics of an a-ogcm: The hadcm3 4xco(2) forcing experiment revisited, Climate Dynamics, 33, 817-825, 10.1007/s00382-009-0581-0, 2009.

Manabe, S., Bryan, K., and Spelman, M. J.: Transient-response of a global ocean atmosphere model to a doubling of atmospheric carbon-dioxide, J Phys Oceanogr, 20, 722-749, Doi 10.1175/1520-0485(1990)020<0722:Troago>2.0.Co;2, 1990.

Meehl, G. A., Moss, R., Taylor, K. E., Eyring, V., Stouffer, R. J., Bony, S., and Stevens, B.: Climate model intercomparisons: Preparing for the next phase, Eos Trans. AGU, 95, 77, 2014.

Meinshausen, M., Raper, S. C. B., and Wigley, T. M. L.: Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, magicc6-part 1: Model description and calibration, Atmos Chem Phys, 11, 1417-1456, DOI 10.5194/acp-11-1417-2011, 2011.

Mitchell, J. F. B., Wilson, C. A., and Cunnington, W. M.: On co2 climate sensitivity and model dependence of results, Q J Roy Meteor Soc, 113, 293-322, 1987.
Mitchell, T. D.: Pattern scaling - an examination of the accuracy of the technique for describing future climates, Climatic Change, 60, 217-242, 2003.
Myhre, G., Highwood, E. J., Shine, K. P., and Stordal, F.: New estimates of radiative forcing due to well mixed greenhouse gases, Geophysical Research Letters, 25, 2715-2718, 1998.
Santer, B., Wigley, T., Schlesinger, M., and Mitchell, J. F. B.: Developing climate scenarios from equilibrium gcm results, Report No. 47, Max Planck Institute for Meteorology, Hamburg, 1990.
Schaller, N., Cermak, J., Wild, M., and Knutti, R.: The sensitivity of the modeled energy budget and hydrological cycle to co2 and solar forcing, Earth Syst Dynam, 4, 253-266, DOI 10.5194/esd-4-253-2013, 2013.
Schaller, N., Sedláček, N. J., and Knutti, R.: The asymmetry of the climate system's response to solar forcing changes and its implications for geoengineering scenarios, Journal of Geophysical Research: Atmospheres, 10, 5171–5184, 2014.
Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L.: Contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
Stott, P., Good, P., Jones, G., Gillett, N., and Hawkins, E.: The upper end of climate model temperature projections is inconsistent with past warming, Environ Res Lett, 8, Artn 014024
Doi 10.1088/1748-9326/8/1/014024, 2013.
Tebaldi, C., and Arblaster, J. M.: Pattern scaling: Its strengths and limitations, and an update on the latest model simulations, Climatic Change, 122, 459-471, DOI 10.1007/s10584-013-1032-9, 2014.
Williams, K. D., Ingram, W. J., and Gregory, J. M.: Time variation of effective climate sensitivity in gcms, Journal of Climate, 21, 5076-5090, Doi 10.1175/2008jcli2371.1, 2008.
Zelinka, M. D., Klein, S. A., Taylor, K. E., Andrews, T., Webb, M. J., Gregory, J. M., and Forster, P. M.: Contributions of different cloud types to feedbacks and rapid adjustments in cmip5, Journal of Climate, 26, 5007-5027, Doi 10.1175/Jcli-D-12-00555.1, 2013.

Table 1. List of CMIP5/CMIP6 DECK experiments required by nonlinMIP.

Experiment	Description	Role
piControl	Pre-industrial control experiment	
Abrupt4xCO2	CO2 abruptly quadrupled, then held constant for 150 years.	Separate different timescales of response.
1pctCO2	CO2 increased at 1% per year for 140 years (i.e. as CMIP5 1pctCO2 experiment), then decreased by 1% per year for 140 years (i.e. returning to pre-industrial conditions).	To test traceability of the abruptCO2 experiments to more realistic transient-forcing conditions. Adding the ramp-down phase explores physics relevant to mitigation and geo-engineering scenarios.

Table 2. NonlinMIP experimental design. Three options are: only the 'essential' simulation; all 'high priority' plus the 'essential' simulations; or, preferably, all simulations. The experiments in Table 1 are required in all cases.

Experiment (priority)	Description	Role
Abrupt2xCO2 (essential)	As abrupt4xCO2 (see Table 1), but at double pre-industrial CO2 concentration.	To diagnose non-linear responses (in combination with abrupt4xCO2). Assess climate response and (if appropriate) make climate projections with the step-response model at forcing levels more relevant to mid- or low-forcing scenarios.
1pctCO2 ramp-down (high priority)	Initialised from the end of 1pctCO2. CO2 is decreased by 1% per year for 140 years (i.e. returning to pre-industrial conditions).	To test traceability of the abruptCO2 experiments to more realistic transient-forcing conditions. Adding the ramp-down phase explores a much wider range of physical responses, providing a sterner test of traceability. Relevant also to mitigation and geo-engineering scenarios, and offers a sterner test of.
Extend both abrupt2xCO2 and abrupt4xCO2 by 100 years (high priority)		Allow traceability tests (via the step-response model) against most of the 1pctCO2 ramp-up-ramp-down experiment. Explore longer timescale responses than in CMIP5 experiment. Permit improved signal/noise in diagnosing some regional-scale non-linear responses Provide a baseline control for the abrupt4xto1x experiment.
Abrupt4xto1x (medium priority)	Initialised from year 100 of abrupt4xCO2, CO2 is abruptly returned to pre-industrial levels, then held constant for 150 years.	Quantify non-linearities over a larger range of CO2 (quantifies responses at 1xCO2). Assess non-linearities that may be associated with the direction of forcing change.
Abrupt8xCO2 (medium priority)	As abrupt4xCO2, but at 8x pre-industrial CO2 concentration. Only 150 years required here.	Quantify non-linearities over a larger range of CO2.

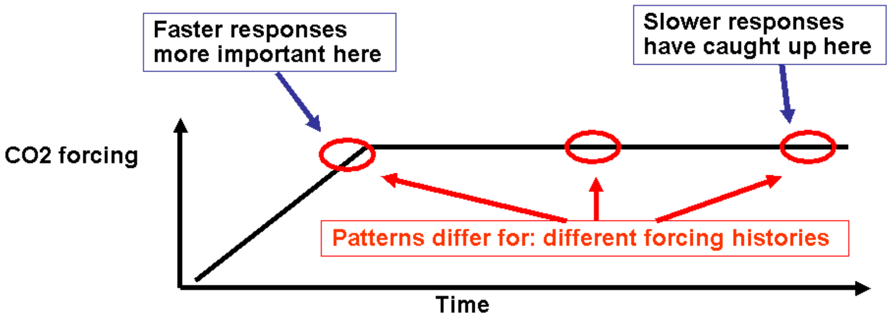

Figure 1. Schematic illustrating a situation where linear mechanisms can cause climate patterns to evolve. This represents a scenario where forcing (black line) is ramped up, then stabilised.

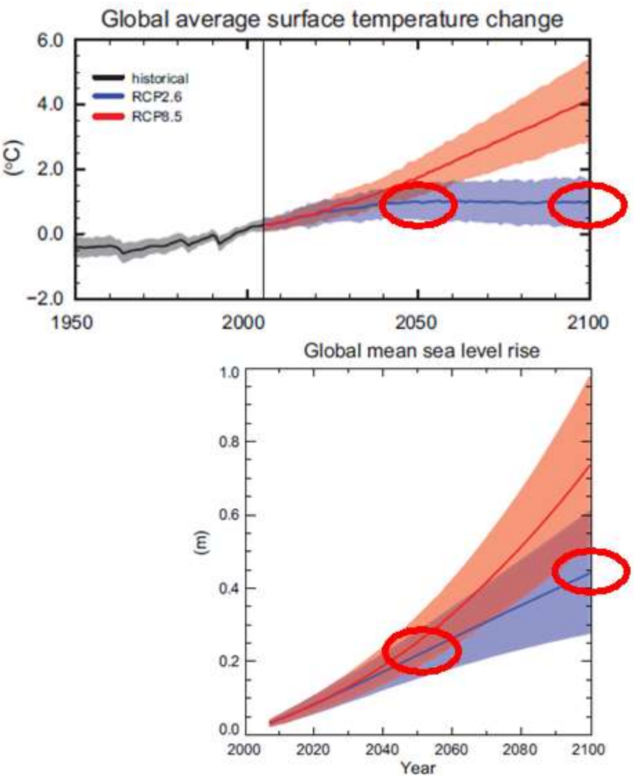

Figure 2. Adapted (red ovals overlaid) from the IPCC Fifth Assessment Report (IPCC, 2013), Figures SPM.7 and SPM.9. Global mean warming (top) and global mean sea level rise (bottom), relative to 1986-2005, for rcp8.5 (red) and rcp2.6 (blue).

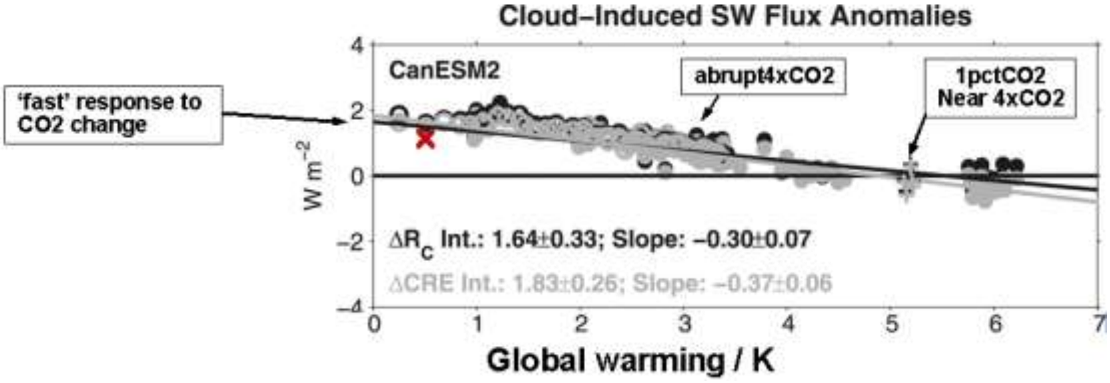

Figure 3. Illustrating a method (Gregory et al., 2004) for separating 'fast' and 'slow' responses to radiative forcing change. Figure adapted (labels in rectangles overlaid) from Zelinka et al. (2013). Global-mean cloud-induced SW flux anomalies against global warming, for the CanESM2 model (black & grey represent two methods of calculating cloud-induced fluxes). This also illustrates one test of traceability of abrupt4xCO2 to 1pctCO2 responses: the linear fit to the abrupt4xCO2 response (straight lines) passes through the 1pctCO2 response near 4xCO2 (i.e. near year 140 of that experiment).

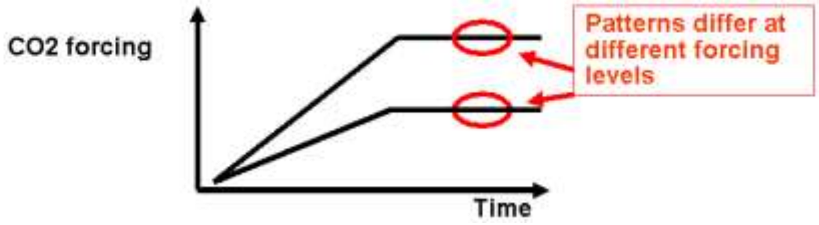

Figure 4. Schematic illustrating the point that nonlinear mechanisms can cause climate patterns to differ at different forcing (and hence global temperature) levels.

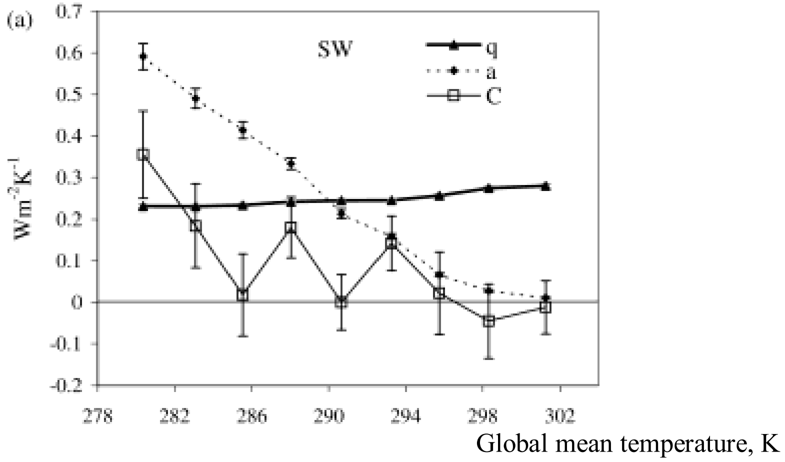

Figure 5. Albedo feedback (dotted line) strength (y-axis) decreasing with global mean temperature (x-axis, K) in a climate model (figure from Colman and McAvaney, 2009).

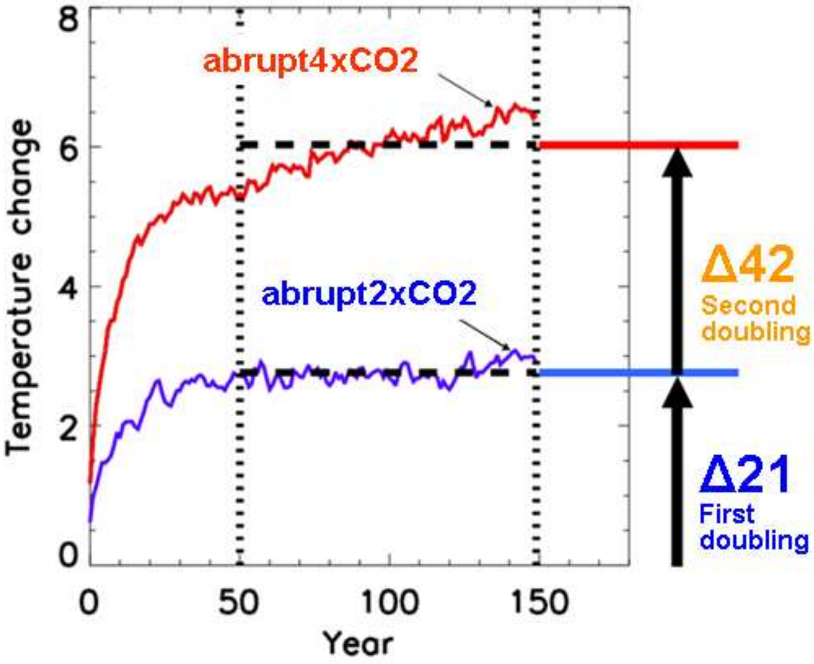

Figure 6. Defining the 'doubling difference'. Doubling difference = Δ42 – Δ21 (the difference in response between the first and second CO2 doublings. This is defined for a specific timescale after the abrupt CO2 change – in this example, it is the mean over years 50-149.

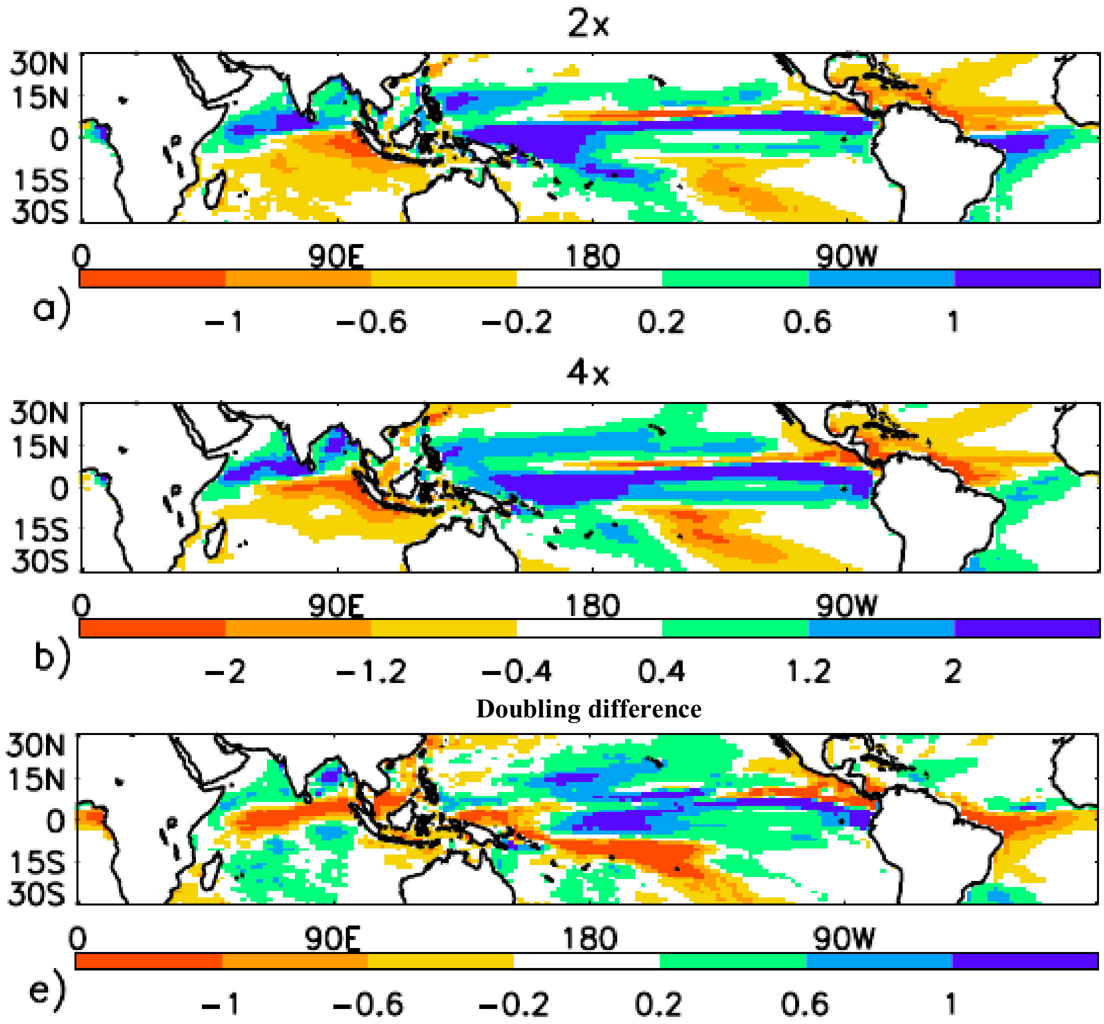

Figure 7. Non-linear regional precipitation responses over the ocean in HadGEM2-ES (figure from Chadwick and Good, 2013). Precipitation change (mm/day) averaged over years 50-149 for (top) abrupt2xCO2 and (middle) abrupt4xCO2, and the doubling difference (bottom). Note that the top and bottom panels have the same scale.

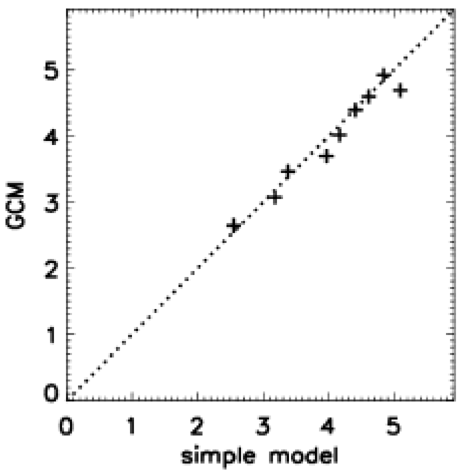

Figure 8. Checking basic traceability of abrupt4xCO2 to a transient forcing experiment (1pctCO2) (figure from Good et al., 2013). Global-mean warming (K) averaged over years 120-139 of 1pctCO2 for (y-axis) the GCM simulation and (x-axis) the reconstruction from abrupt4xCO2 using the step-response method.

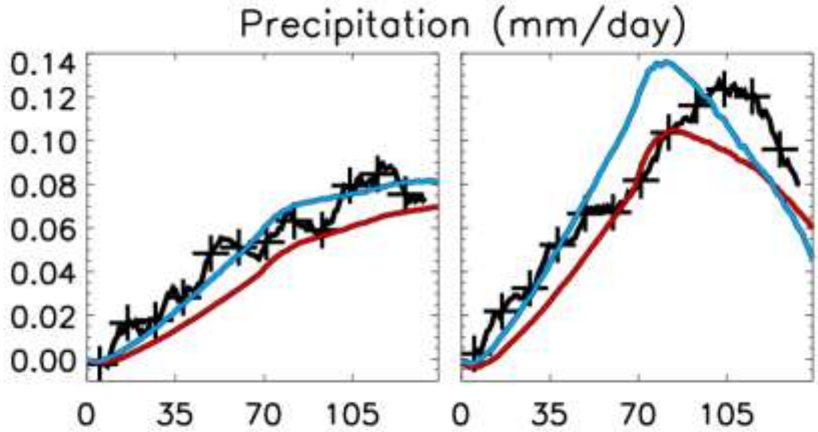

Figure 9. Finding nonlinear responses in transient forcing experiments. (figure from Good et al., 2012). Left: where CO2 is increased by 1% per year, then stabilised at 2x pre-industrial levels. Right: where CO2 is increased by 2% per year for 70 years, then decreased by 2% per year for 70 years. Black: GCM. Red: step-response model using the abrupt4xCO2 response. Blue: the abrupt2xCO2 response.