nonlinMIP contribution to CMIP6: model intercomparison project for nonlinear mechanisms - physical basis, experimental design and analysis principles (v1.0)

P. Good[1], T. Andrews[1], R. Chadwick[1], J. L. Dufresne[3], J. M. Gregory[2,1], J. A. Lowe[1] , N. Schaller[4], H. Shiogama[5].

[1]Met Office Hadley Centre, Exeter, United Kingdom

[2]NCAS-Climate, University of Reading, Reading, United Kingdom

[3]Laboratoire de Météorologie Dynamique, Institut Pierre Simon Laplace, Paris,France

[4]Atmospheric, Oceanic and Planetary Physics, University of Oxford, Parks Road, Oxford OX1 3PU, United Kingdom

[5]Climate Risk Assessment Section, Centre for Global Environmental Research, National Institute for Environmental Studies, Tsukuba, Japan

Correspondence to: P. Good (peter.good@metoffice.gov.uk)

Abstract

nonlinMIP provides experiments that account for state-dependent regional and global climate responses. The experiments have two main applications: 1) to focus understanding of responses to CO_2 forcing on states relevant to specific policy or scientific questions (e.g. change under low-forcing scenarios, the benefits of mitigation, or from past cold climates to the present-day); or 2) to understand the state-dependence (nonlinearity) of climate change – i.e. why doubling the forcing may not double the response. State dependence (nonlinearity) of 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 (point 2 above), and more broadly, to focus work on understanding the response to CO_2 on climate states relevant to specific policy/science questions (point 1).

nonlinMIP addresses this using a simple, small set of CO_2-forced experiments that are able to separate linear and non-linear mechanisms cleanly, with a good signal/noise ratio – 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 instantaneous atmospheric CO_2 change experiments, with a ramp-up-ramp-down experiment to test traceability to gradual forcing scenarios. In all cases the models are intended to be used with CO_2 concentrations rather than CO_2 emissions as the input. 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 abruptCO_2 to gradual forcing experiments, before detailing the experimental design and finally some analysis principles. The test of traceability from abruptCO_2 to transient experiments is recommended as a standard analysis within the CMIP5 and CMIP6 DECK protocols.

1 Introduction

Robust climate impacts assessments require, at regional scales, understanding of physical mechanisms of climate change in GCM projections. A further, pragmatic requirement for impacts assessments is the ability to emulate (using fast but 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, for a given well defined model application, that has quantitative predictive power and captures the key mechanisms behind GCM scenario projections?

Often, a 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 explicitly that regional climate change is roughly proportional to global mean warming. In emulation work, this is termed 'pattern scaling' (Santer et al., 1990;Mitchell, 2003;Ishizaki et al., 2012;Tebaldi and Arblaster, 2014), but this assumption may also be applied implicitly in understanding mechanisms. Often, physical mechanisms are studied for a single period of a single forcing scenario or in a single high-forcing experiment such as abrupt4xCO2 (implicitly assuming that the understanding is relevant for other periods or scenarios). The use of pattern-scaling is prevalent in studies of climate impacts.

While these approximations appear to work well under many 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 regional climate change is proportional to global mean warming – 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 not be linear. This is due to different timescales of response in different locations and/or variables (section 3.1). Examples include lagged surface ocean warming due to a connection with the deeper ocean (Manabe et al., 1990;Williams et al., 2008;Held et al., 2010;Chadwick et al., 2013;Andrews and Ringer, 2014) or the direct response of precipitation to forcings (Mitchell et al., 1987;Allen and Ingram, 2002;Andrews et al., 2010;Bala et al., 2010;Bony et al., 2014). One (generally false, but potentially acceptable) 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 (section 3.2). Some examples have been known for some time, such as changing feedbacks through retreating snow/sea-ice or increasing water vapour (Hansen et al., 2005;Colman and McAvaney, 2009;Jonko et al., 2013;Meraner et al., 2013). Some paleoclimate evidence supports the idea that climate sensitivity increases with warming (Caballero and Huber, 2013;Shaffer et al., 2016), which is important for the risk of high-end global warming (Bloch-Johnson et al., 2015). The nonlinear behaviour of the Atlantic Meridional Overturning Circulation is another example (Hofmann and Rahmstorf, 2009;Ishizaki et al., 2012). 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). Regional warming has been shown to be different for a first and second CO2 doubling, with implications primarily for impact assessment models or studies combining linear energy balance models with pattern scaling (Good et al., 2015). 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;Bouttes et al., 2015).

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 (Li and Jarvis, 2009;Held et al., 2010;Good et al., 2013), 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: an experiment where CO2 forcing is instantaneously changed, then held constant. The simplified forcing in such experiments simplifies the understanding of physical mechanisms of response. In these abrupt CO2 experiments, responses over different timescales (fast and slow responses) 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., 2012;Good et al., 2013;Zelinka et al., 2013;Bouttes et al., 2015;Good et al., 2015) 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. In other studies (e.g. Frolicher et al., 2014), pulse experiments have been used to separate different timescales of response (where forcing is abruptly increased, then abruptly returned to the control state). We use abruptCO2 experiments because they offer greater signal/noise in the change signal (important for regional-scale studies); and also for consistency with the CMIP6 DECK abrupt4xCO2 experiment.

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 builds on 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.

2 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. We cannot a priori reject the possibility that some GCMs could respond unrealistically to the abrupt forcing change. A key tool here is the step-response model (described below). This (Hasselmann et al., 1993) is a response-function method, which 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 change in radiative forcing during year i-j of the scenario (denoted ΔF_{i-j}):

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

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).

3 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.

3.1 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 (Mitchell et al., 1987;Allen and Ingram, 2002;Andrews et al., 2010) 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. A key example is the different responses of global-mean warming and global-mean sea level rise under RCP2.6, as shown in Figures SPM.7 and SPM.9 of the IPCC Fifth Assessment Report (IPCC, 2013). Under RCP2.6, global-mean warming ceases after 2050, when radiative forcing is approximately stabilised (corresponding qualitatively to the period when the black line is horizontal in Figure 1). In contrast, sea-level rise continues at roughly the same rate throughout the century. Therefore, in RCP2.6, the sea-level rise per K of global warming increases after 2050. This is largely because the timescale of deep ocean heat uptake is much longer than that of ocean mixed-layer warming.

By design, abruptCO2 experiments separate GCM responses with different timescales (i.e. separating faster responses from slower responses): the response of a given variable in year Y of the experiment corresponds to the response of that variable over the timescale Y. 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. For example, the top-of-atmosphere outgoing shortwave flux shows a rapid initial change before the global mean temperature has had time to respond.

3.2 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 and sea-ice albedo feedbacks becomes small at high or low snow depth (Hall, 2004;Eisenman, 2012). Soil moisture–temperature feedbacks can also be state-dependent (Seneviratne et al., 2006;Seneviratne et al., 2010): feedback is small when soil moisture is saturated, or so low that moisture is tightly bound to the soil (in both regimes, evaporation is insensitive to change in soil moisture). 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 et al., 2012;Chadwick and Good, 2013;Bony et al., 2014;Oueslati et al., 2016). 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. In addition, the Clausius Clapeyron equation introduces some nonlinearity in the increase of specific humidity with warming. 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 2 may have different patterns of change per K of global warming, due to nonlinear mechanisms (in contrast, linear mechanisms would cause no difference in the patterns of change per K of global warming between the two points in Figure 2, because the two scenarios have the same forcing history apart from a constant scaling factor).

An example is the snow/ice albedo feedback, which tends to change in magnitude 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 abruptCO2 experiments. Figure 3 compares abrupt2xCO2 and abrupt4xCO2 experiments over years 50-149. A 'doubling difference' is defined (Good et al., 2015), 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'.

4 Experimental design

nonlinMIP is composed of a set of abruptCO2 experiments (the primary tools), plus a CO2-forced transient experiment. AbruptCO2 experiments are driven by changes in atmospheric CO2 concentration: CO2 is abruptly changed, then held constant. 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.

5 Basic analysis principles

This section outlines the applications and general principles behind analysis of nonlinMIP results. First, some general applications are introduced, before giving more detail on how one particular application (quantifying and understanding nonlinear change) may be analysed.

The addition of the abrupt2xCO2 experiment to the standard DECK abrupt4xCO2 permits quantifying and understanding climate change due to CO2 for three main applications:

1) under global warming approximately comparable to that envisaged by the Paris agreement. (quantified by abrupt2xCO2 – pre-industrial control)

2) climate change approximately comparable to that avoided by mitigation (quantified by abrupt4xCO2 - abrupt2xCO2).
3) nonlinear change (the difference between 2 and 1).

Applications 1 and 2 are expected to be of the widest interest to the community, as they could be analysed using the same methods as have already been used extensively to study the response in the CMIP5 abrupt4xCO2 experiment, but for climate states more relevant to the policy questions outlined in 1) and 2). Useful signal/noise should be possible because ~100 year means may be analysed (e.g. over years 50-149, where climate is relatively stable as it follows the initial ocean mixed layer warming). Application 3 is more specialised, and is discussed in more detail below.

The abrupt0.5xCO2 experiment permits analogous work, extending the relevance to colder past climates, and exploring one aspect of how past change may differ from future change. It also allows nonlinear mechanisms to be studied with greater signal/noise:

4) change under past cold climates (abrupt0.5xCO2 - piControl).
5) nonlinear change: as 3, but with larger signal/noise ([abrupt4xco2 - abrupt2xco2] – [piControl - abrupt0.5xCO2]).

In quantifying nonlinear change (applications 3 or 5 above), 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.

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

The test described here is recommended as a routine analysis of the CMIP6 DECK experiments (even if nonlinMIP experiments are not performed). The aim is to confirm whether 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), for regional-scale warming and ocean heat updake (Bouttes et al., 2015;Good et al., 2015) 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). Although this test has been done for a range of models and variables, traceability cannot be assumed to hold for all models and variables.

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 previously 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.

5.2 Second step: characterising 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 4 (but for different transient-forcing experiments). This shows results for global-mean precipitation in the HadCM3 GCM (Good et al., 2012), under an idealised simulation where forcing is ramped up at a constant rate for 70 years, then ramped down at the same rate for 70 years. Here, the step-response model prediction using abrupt4xCO2 (red curves) is only close to the actual GCM simulation (black) where the transient-forced simulation is near to 4xCO2 (i.e. near year 70). Similarly, the prediction using abrupt2xCO2 (blue curves) works only near 2xCO2 (near years 35 or 105). 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 global-mean 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 Figure 3). This allows (Good et al., 2012;Chadwick and Good, 2013;Good et al., 2015) non-linear mechanisms to be separated from linear mechanisms (not possible in a transient-forcing experiment). It is expected that analysis will focus on the 100-year period over years 40-139 of the experiments (the relatively stable period after the initial ocean mixed-layer warming).

In the same spirit as other CMIP5 and CMIP6 idealised experiments, nonlinMIP will help understand nonlinear mechanisms by isolating the signal of nonlinear mechanisms more effectively. This occurs in two ways: first, by using simplified forcing compared to the time-dependent, RCP projections (the latter feature multiple forcings of evolving strength). The simplified forcing means that alternative mechanisms (from different forcing agents or linear mechanisms) may be ruled out by design. Secondly, contamination of the signal from internal variability may be reduced, as averages of around 100 years are possible.

The magnitude of internal variability may also be estimated at the different levels of CO2 forcing. This could be used to help explore changes in variability with warming (Seneviratne et al., 2006;Screen, 2014), and to assess significance of any signal of nonlinear change in the time mean climate. Internal variability could be estimated from years 40-139 of the experiments (after the initial warming of the ocean mixed layer), after removing a fitted linear trend.

6 Conclusions

These experiments can help improve climate science and consequent policy advice in a number of ways. The focus is on understanding mechanisms (given the idealised nature of the experiments). A further application, however, is that energy balance models could be tuned to the different experiments, to explore the importance, for projections, of state-dependence of feedback parameters (Hansen et al., 2005;Colman and McAvaney, 2009;Caballero and Huber, 2013). Also, if certain regions are found to show strongly nonlinear behaviour in these experiments, this could help focus assessment of impact tools like pattern-scaling or time-shifting (e.g. Herger et al., 2015).

Of probably widest interest is the fact that the additional experiments will allow understanding work to focus on climate states more directly relevant to discrete policy/science questions (the benefits of mitigation; impacts of scenarios consistent with the Paris agreement; or understanding past cold climates; see start of section 5). These questions may show important differences, due to state-dependence (nonlinearity) of mechanisms, but for many cases the nature of the nonlinearity may not need to be assessed. A classical example is the snow-albedo feedback: the strength of this would be different in a warm versus a cold world (due to different baseline snow cover), but if the focus is on understanding the warm world, the first priority is to study experiments representative of the warm world (with the correct climate state).

There is also a need to quantify and understand, at regional scales, nonlinear mechanisms of climate change: that is, do the above science/policy questions give significantly different answers (e.g. different patterns of rainfall change), and why? This is difficult to do using transient model experiments alone, for two reasons: contamination due to different timescales of response, and noise from internal variability.

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.

7 Data availability

Results will be made available as part of the CFMIP project, within the sixth model intercomparison project, CMIP6.

Acknowledgements

This work was supported by the Joint UK DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). Jonathan Gregory received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013), ERC Grant Agreement 247220, project "Seachange." Hideo Shiogama was supported by the SOUSEI program from the Ministry of Education, Culture, Sports, Science and Technology of Japan and the Environment Research and Technology Development Fund (S-10) of the Ministry of the Environment of Japan. Nathalie Schaller was supported by the Swiss National Science Foundation.

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.

Bala, G., Caldeira, K., and Nemani, R.: Fast versus slow response in climate change: Implications for the global hydrological cycle, Climate Dynamics, 35, 423-434, 10.1007/s00382-009-0583-y, 2010.

Bloch-Johnson, J., Pierrehumbert, R. T., and Abbot, D. S.: Feedback temperature dependence determines the risk of high warming, Geophysical Research Letters, 42, 4973-4980, 10.1002/2015GL064240, 2015.

Bony, S., Bellon, G., Klocke, D., Sherwood, S., Fermepin, S., and Denvil, S.: Robust direct effect of carbon dioxide on tropical circulation and regional precipitation (vol 4, pg 447, 2013), Nat Geosci, 7, 547-547, 10.1038/NGEO2192, 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.

Bouttes, N., Good, P., Gregory, J. M., and Lowe, J. A.: Nonlinearity of ocean heat uptake during warming and cooling in the famous climate model, Geophysical Research Letters, 42, 2409-2416, 10.1002/2014GL062807, 2015.

Caballero, R., and Huber, M.: State-dependent climate sensitivity in past warm climates and its implications for future climate projections, Proceedings of the National Academy of Sciences of the United States of America, 110, 14162-14167, 10.1073/pnas.1303365110, 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.

Eisenman, I.: Factors controlling the bifurcation structure of sea ice retreat, Journal of Geophysical Research-Atmospheres, 117, Artn D01111

Doi 10.1029/2011jd016164, 2012.

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.

Frolicher, T. L., Winton, M., and Sarmiento, J. L.: Continued global warming after co2 emissions stoppage, Nat Clim Change, 4, 40-44, 10.1038/Nclimate2060, 2014.

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.

Good, P., Lowe, J. A., Andrews, T., Wiltshire, A., Chadwick, R., Ridley, J. K., Menary, M. B., Bouttes, N., Dufresne, J. L., Gregory, J. M., Schaller, N., and Shiogama, H.: Nonlinear regional warming with increasing co2 concentrations, Nat Clim Change, 5, 138-142, 10.1038/Nclimate2498, 2015.

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.

Hall, A.: The role of surface albedo feedback in climate, Journal of Climate, 17, 1550-1568, Doi 10.1175/1520-0442(2004)017<1550:Trosaf>2.0.Co;2, 2004.

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, M., Bauer, S., Bell, N., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., Hall, T., Jackman, C., Kelley, M., Kiang, N., Koch, D., Lean, J., Lerner, J., Lo, K., Menon, S., Miller, R., Minnis, P., Novakov, T., Oinas, V., Perlwitz, J., Rind, D., Romanou, A., Shindell, D., Stone, P., Sun, S., Tausnev, N., Thresher, D., Wielicki, B., Wong, T., Yao, M., and Zhang, S.: Efficacy of climate forcings, Journal of Geophysical Research-Atmospheres, 110, 45, D18104

10.1029/2005jd005776, 2005.

Hasselmann, K., Sausen, R., Maierreimer, E., and Voss, R.: On the cold start problem in transient simulations with coupled atmosphere-ocean models, Climate Dynamics, 9, 53-61, 1993.

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.

Herger, N., Sanderson, B. M., and Knutti, R.: Improved pattern scaling approaches for the use in climate impact studies, Geophysical Research Letters, 42, 3486-3494, 10.1002/2015GL063569, 2015.

Hofmann, M., and Rahmstorf, S.: On the stability of the atlantic meridional overturning circulation, Proceedings of the National Academy of Sciences of the United States of America, 106, 20584-20589, DOI 10.1073/pnas.0909146106, 2009.

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.

Ishizaki, Y., Shiogama, H., Emori, S., Yokohata, T., Nozawa, T., Ogura, T., Abe, M., Yoshimori, M., and Takahashi, K.: Temperature scaling pattern dependence on representative concentration pathway emission scenarios, Climatic Change, 112, 535-546, DOI 10.1007/s10584-012-0430-8, 2012.

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.

Meraner, K., Mauritsen, T., and Voigt, A.: Robust increase in equilibrium climate sensitivity under global warming, Geophysical Research Letters, 40, 5944-5948, 10.1002/2013GL058118, 2013.

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.

Oueslati, B., Bony, S., Risi, C., and Dufresne, J. L.: Interpreting the inter-model spread in regional precipitation projections in the tropics, Climate Dynamics, in press, doi 10.1007/s00382-016-2998-6, 2016.

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.

Screen, J. A.: Arctic amplification decreases temperature variance in northern mid- to high-latitudes, Nat Clim Change, 4, 577-582, 10.1038/Nclimate2268, 2014.

Seneviratne, S. I., Luthi, D., Litschi, M., and Schar, C.: Land-atmosphere coupling and climate change in europe, Nature, 443, 205-209, Doi 10.1038/Nature05095, 2006.

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., and Teuling, A. J.: Investigating soil moisture-climate interactions in a changing climate: A review, Earth-Sci Rev, 99, 125-161, DOI 10.1016/j.earscirev.2010.02.004, 2010.

Shaffer, G., Huber, M., Rondanelli, R., and Pedersen, J. O. P.: Deep time evidence for climate sensitivity increase with warming, Geophysical Research Letters, 43, 6538-6545, 10.1002/2016GL069243, 2016.

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, 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.
Abrupt0.5xCO2 (essential)	As abrupt4xCO2 (see Table 1), but at half pre-industrial CO2 concentration	To diagnose non-linear responses (in combination with abrupt4xCO2 and abrupt2xCO2). Offers greater signal/noise for regional precipitation change than if just abrupt2xCO2 was used. Also relevant to paleoclimate studies.
Extend both abrupt2xCO2 and abrupt4xCO2 by 100 years (high priority)		Permit improved signal/noise in diagnosing some regional-scale non-linear responses Explore longer timescale responses than in CMIP5 experiment. Permit step-response model scenario simulations from 1850-2100

		Allow traceability tests (via the step-response model) against most of the 1pctCO2 ramp-up-ramp-down experiment. Provide a baseline control for the abrupt4xto1x experiment.
1pctCO2 ramp-down (medium 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.
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.

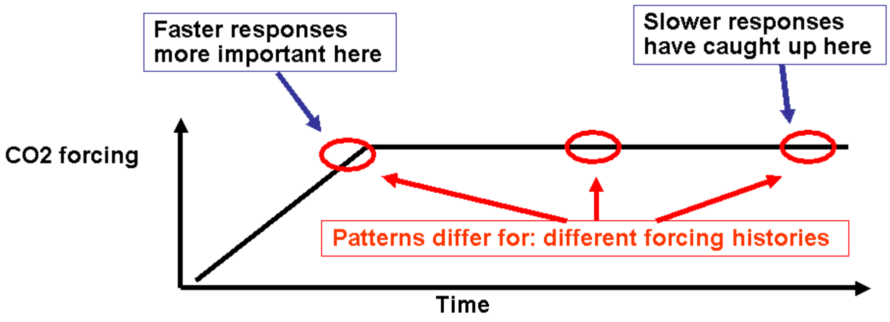

Figure 1. Schematic illustrating a situation where linear mechanisms can cause climate patterns to evolve. This represents a scenario where global-mean radiative forcing (black line) is ramped up, then stabilised. At the time indicated by the left red oval, responses with shorter timescales are relatively important, due to the recent increase in forcing. At the time marked by the right-hand oval, forcing has been stabilised for an extended period, so the responses with longer timescales (such as sea-level rise) have had more time to respond to the initial forcing increase.

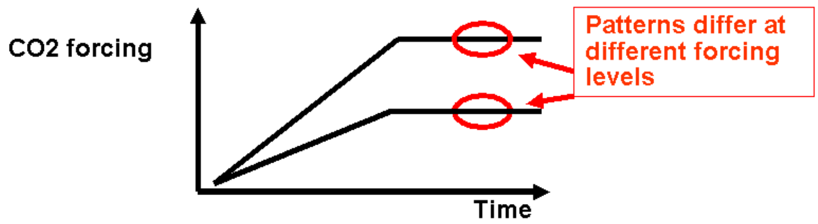

Figure 2. Schematic illustrating the point that nonlinear mechanisms can cause climate patterns to differ at different forcing (and hence global temperature) levels. This represents two different scenarios, whose forcing timeseries is identical apart from a constant scale factor (the higher forcing scenario has about twice the forcing of the lower scenario).

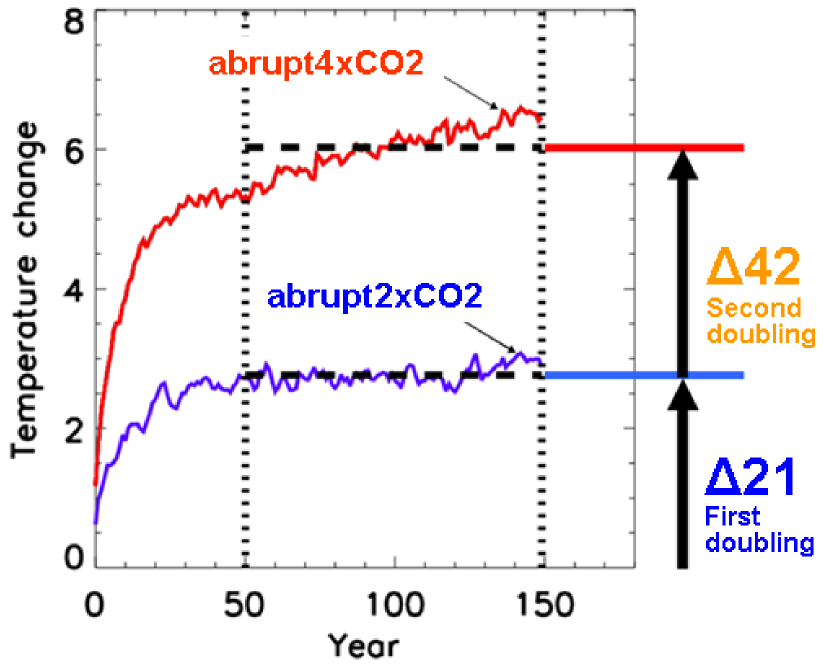

Figure 3. Defining the 'doubling difference'. The red and blue lines show illustrative time-series of a variable (in this example, global-mean temperature from HadGEM2-ES) from the abrupt4xCO2 and abrupt2xCO2 experiments. Doubling difference = Δ42 – Δ21 (the difference in response between the first and second CO2 doublings. This is defined for a specific timescale after the abrupt CO2 change – in this example, it is for means over years 50-149.

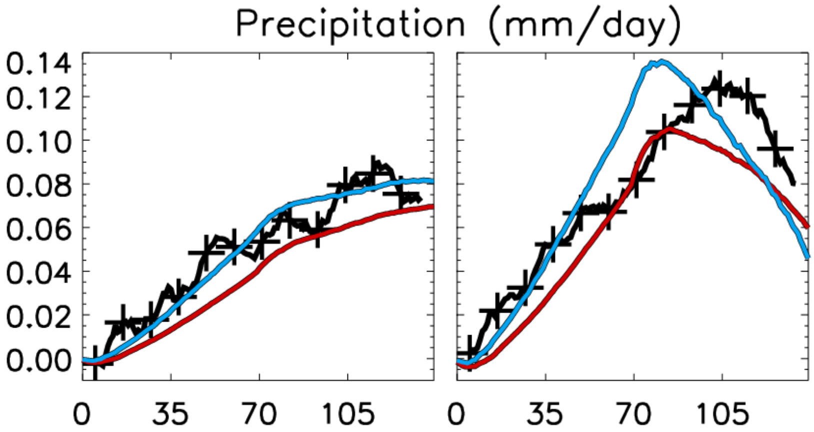

Figure 4. Finding nonlinear responses in transient forcing experiments. (figure from Good et al., 2012). Time-series of global-mean precipitation change under two experiments. Left: where CO2 is increased by 1% per year, then stabilised at 2x pre-industrial levels. Right: where CO2 is increased by 2% 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.