Reduced Complexity Model Intercomparison Project (Phase 1)

Zebedee R. J. Nicholls[1,2], Malte Meinshausen[1,2,3], Jared Lewis[1], Robert Gieseke[4], Dietmar Dommenget[5], Kalyn Dorheim[6], Chen-Shuo Fan[5], Jan S. Fuglestvedt[7], Thomas Gasser[8], Ulrich Golüke[9], Philip Goodwin[10], Corinne Hartin[6], Austin P. Hope[11], Elmar Kriegler[3], Nicholas J. Leach[12], Davide Marchegiani[5], Laura A. McBride[13], Yann Quilcaille[8], Joeri Rogelj[8,14], Ross J. Salawitch[11,13,15], Bjørn H. Samset[7], Marit Sandstad[7], Alexey N. Shiklomanov[6], Ragnhild B. Skeie[7], Christopher J. Smith[8,16], Steve Smith[6], Katsumasa Tanaka[17,18], Junichi Tsutsui[19], and Zhiang Xie[5]

[1]Australian–German Climate and Energy College, The University of Melbourne, Parkville, Victoria, Australia
[2]School of Earth Sciences, The University of Melbourne, Parkville, Victoria, Australia
[3]Potsdam Institute for Climate Impact Research (PIK), Member of the Leibniz Association, Potsdam, Germany
[4]Unaffiliated
[5]Monash University, School of Earth, Atmosphere and Environment, Clayton, Victoria 3800, Australia
[6]Joint Global Change Research Institute, Pacific Northwest National Laboratory, College Park, MD, USA
[7]CICERO Center for International Climate Research, Oslo, Norway
[8]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
[9]BI Norwegian Business School, Nydalsveien 37, 0484 Oslo, Norway
[10]School of Ocean and Earth Science, University of Southampton, Southampton, UK
[11]Department of Atmospheric and Oceanic Sciences, University of Maryland-College Park, College Park, 20740, USA
[12]Department of Physics, Atmospheric Oceanic and Planetary Physics, University of Oxford, United Kingdom
[13]Department of Chemistry and Biochemistry, University of Maryland-College Park, College Park, 20740, USA
[14]Grantham Institute for Climate Change and the Environment, Imperial College London, UK
[15]Earth System Science Interdisciplinary Center, University of Maryland-College Park, College Park, 20740, USA
[16]Priestley International Centre for Climate, University of Leeds, UK
[17]National Institute for Environmental Studies (NIES), Tsukuba, Japan
[18]Laboratoire des Sciences du Climat et de l'Environnement (LSCE), Commissariat à l'énergie atomique et aux énergies alternatives (CEA), Gif sur Yvette, France
[19]Central Research Institute of Electric Power Industry, Abiko, Japan

Correspondence: Zebedee Nicholls (zebedee.nicholls@climate-energy-college.org)

Abstract.

Reduced complexity climate models (RCMs) are critical in the policy and decision making space, and are directly used within multiple Intergovernmental Panel on Climate Change (IPCC) reports to complement the results of more comprehensive Earth System Models. To date, evaluation of RCMs has been limited to a few independent studies. Here we propose a systematic evaluation of RCMs in the form of the Reduced Complexity Model Intercomparison Project (RCMIP). We have performed Phase 1 of RCMIP with two scientific themes: examining how RCMs compare to observations and how RCMs compare to results from more complex climate models such as those participating in the Sixth Coupled Model Intercomparison Project (CMIP6). We also present our standardised data formats, experiment protocols and output specifications. So far 15 models have participated and submitted results for over 50 experiments. We present illustrative figures comparing model output with historic global surface air temperature (GSAT) observations, showing probabilistic projections, demonstrating different calibrations

with CMIP model output as well as temperature change against cumulative emissions, and exploring differences between CMIP5's Representative Concentration Pathways (RCPs) and CMIP6's SSP-based (Shared Socioeconomic Pathways based) scenarios. Further research on these and other questions can build on the open data and open source processing code provided with this paper.

1 Introduction

Sufficient computing power to enable running our most comprehensive, physically complete climate models for every application of interest is not available. Thus, for many applications less computationally demanding approaches are used. One common approach is the use of reduced complexity climate models (RCMs), also known as simple climate models (SCMs).

RCMs are designed to be computationally efficient tools, allowing for exploratory research and have smaller spatial and temporal resolution than complex models. Typically, they describe highly parameterised macro properties of the climate system. Usually this means that they simulate the climate system on a global-mean, annual-mean scale although some RCMs have slightly higher spatial and/or temporal resolutions. As a result of their highly parameterised approach, RCMs can be on the order of a million or more times faster than more complex models (in terms of simulated model years per unit CPU time).

The computational efficiency of RCMs means that they can be used where computational constraints would otherwise be limiting. For example, some applications of Integrated Assessment Models (IAMs) require iterative climate simulations. As a result, hundreds to thousands of climate realisations must be integrated by the IAM for a single scenario to be produced. RCMs also enable the exploration of interacting uncertainties from multiple parts of the climate system or the constraining of unknown parameters by combining multiple lines of evidence in an internally consistent setup. In the context of the assessment reports of the Intergovernmental Panel on Climate Change (IPCC), a prominent example is the climate assessment of socioeconomic scenarios by IPCC Working Group 3 (WGIII). Hundreds of emission scenarios were assessed in the IPCC's Fifth Assessment Report (AR5, see Clarke et al. (2014)) as well as its more recent Special Report on Global Warming of 1.5°C (SR1.5, see Rogelj et al. (2018); Huppmann et al. (2018)). (Scenario data is available at https://secure.iiasa.ac.at/web-apps/ene/AR5DB and https://data.ene.iiasa.ac.at/iamc-1.5c-explorer/ for AR5 and SR1.5 respectively, both databases are hosted by the IIASA Energy Program). For the IPCC's forthcoming Sixth Assessment (AR6), it is anticipated that the number of scenarios will be in the several hundreds to a thousand (for example, see the full set of scenarios based on the SSPs at https://tntcat.iiasa.ac.at/SspDb). Both the number of scenarios and the tight timelines of the IPCC assessments render it infeasible to use the world's most comprehensive models to estimate the climate implications of these IAM scenarios.

There are two key modes of use which are relevant for the assessment of a large number of IAM scenarios. The first is 'emulation' mode, where the RCMs are run in a setup which has been calibrated to reproduce the behaviour of a Coupled Model Intercomparison Project (CMIP) (Eyring et al., 2016; Taylor et al., 2012) model as closely as possible over a range of scenarios. The second is 'probabilistic' mode, where the RCMs are run with a parameter ensemble which captures the uncertainty in estimates of specific Earth system quantities, be it observations of historical global mean temperature increase, radiative forcing, ocean heat uptake, or cumulative land or ocean carbon uptake. Probabilistic climate projections are derived by running parametric ensembles of RCM simulations which capture the range of responses consistent with our understanding of the climate system (Meinshausen et al., 2009; Smith et al., 2018; Goodwin, 2016). The resulting ensemble is designed to capture the likelihood that different warming levels are reached under a specific emissions scenario (e.g. 50% and 66%) based on the combined available evidence hence is quite different from an ensemble emulating multiple model outputs, which have been produced independently with no relative relationship or probabilities in mind. The two approaches, emulation of complex models and historically constrained probabilistic mode, can also be combined, e.g. where historical constraints are very weak. For example, the MAGICC6 probabilistic setup used in AR5 (Clarke et al., 2014) used randomly drawn emulations for the carbon cycle response whilst using a probabilistic parameter ensemble for the climate response to radiative forcing (Meinshausen et al., 2009).

RCMs also play the role of 'integrators of knowledge', examining the combined response of multiple interacting components of the climate system. The most comprehensive RCMs will include (highly parameterised) representations of the carbon cycle, permafrost, non-CO_2 gas cycles, aerosol chemistry, temperature response to radiative forcing, ocean heat uptake, sea-level rise and all their interactions and feedbacks. More complex models cannot include as many interactive components without the computational cost quickly becoming prohibitive for running multiple century-long simulations. As a result, RCMs are able to examine the implications of the Earth System's feedbacks and interactions in a way which cannot be done with other techniques.

1.1 Evaluation of reduced complexity climate models

The validity of the RCM approach rests on the premise that RCMs are able to replicate the behaviour of the Earth system and response characteristics of our most complete models. Over time, multiple independent efforts have been made to evaluate this ability. In 1997, an IPCC Technical Paper (Houghton et al., 1997), investigated the simple climate models used in the IPCC Second Assessment Report and compared their performance with idealised Atmosphere-Ocean General Circulation Model (AOGCM) results. Later, Van Vuuren et al. (2011) compared the climate components used in IAMs, such as DICE (Nordhaus, 2014), FUND (Waldhoff et al., 2011) and the RCM MAGICC (version 4 at the time (Wigley and Raper, 2001)), which is used in several IAMs. They focused on five CO_2-only experiments to quantify the differences in the behaviour of the RCMs used by each IAM. Harmsen et al. (2015) extended the work of van Vuuren et al. (2011) to consider the impact of non-CO_2 climate drivers in the RCPs. Recently, Schwarber et al. (2019) proposed a series of impulse tests for simple climate models in order to isolate differences in model behaviour under idealised conditions.

Building on these efforts, an ongoing comprehensive evaluation and assessment of RCMs requires an established protocol. The Reduced Complexity Model Intercomparison Project (RCMIP) proposed here provides such a protocol (also see rcmip.org). We aim for RCMIP to provide a focal point for further development and an experimental design which allows models to be readily compared and contrasted. We believe that a comprehensive, systematic effort will result in a number of benefits seen in other MIPs (Carlson and Eyring, 2017) including building a community of reduced complexity modellers, facilitating comparison of model behaviour, improving understanding of their strengths and limitations, and ultimately also improving RCMs.

RCMIP focuses on RCMs and is not one of the official CMIP6 (Eyring et al., 2016) endorsed intercomparison projects that are designed for Earth System Models. However, RCMIP does replicate selected experimental designs of many of the CMIP-endorsed MIPs, particularly the DECK simulations (Eyring et al., 2016), ScenarioMIP (O'Neill et al., 2016), AerChemMIP (Collins et al., 2017), C4MIP (Jones et al., 2016), ZECMIP (Jones et al., 2019), DAMIP (Gillett et al., 2016) and PMIP4 (Kageyama et al., 2018). Hence whilst RCMIP is not a CMIP6 endorsed intercomparison, its design is closely related in the hope that its results may be useful beyond the RCM community.

In what follows, we describe RCMIP Phase 1. In section 2, we detail the domain of RCMIP Phase 1 and its scientific objectives. In sections 3 and 4, we described the simulations performed and outputs requested from each model. In section 5 we present sample results from RCMIP Phase 1, before presenting possible extensions to RCMIP Phase 1 and conclusions in sections 6 and 7.

2 Science themes

In the RCMIP community call (available at rcmip.org) RCMs were broadly defined as follows: "[...] RCMIP is aimed at reduced complexity, simple climate models and small emulators that are not part of the intermediate complexity EMIC or complex GCM/ESM categories." In practice, we encouraged and encourage any group in the scientific community who identifies with the label of RCM to participate in RCMIP, see Table 1 for an overview of the models which participated in RCMIP Phase 1.

RCMIP Phase 1 focuses on evaluation of RCMs. Specifically, comparing them against observations of the Earth System and the output of more complex models from CMIP5 and CMIP6 within two scientific themes.

Theme 1: To what extent can reduced complexity models reproduce observed ranges of key climate change indicators (e.g. surface warming, ocean heat uptake, land carbon uptake)?

The first theme focuses on evaluating models against observations. Before using any model, one important question to ask is whether it can reproduce observations of the climate's recent evolution. For RCMs, the key observation is changes in air and ocean temperatures (Morice et al., 2012; Cowtan and Way, 2014). Beyond this, RCMs should also be evaluated against observed changes in ocean heat uptake (Zanna et al., 2019; von Schuckmann et al., 2020) and estimates of carbon content in the air, land and oceans (Friedlingstein et al., 2019).

These comparisons evaluate the extent to which the model's approximations cause its response to deviate from observational data. However, most RCMs can be calibrated, i.e. have their parameters adjusted, such that they reproduce our best-estimate (typically median) observations. Hence, where available, we also evaluate the extent to which RCMs can be configured to reproduce the range of available observational estimates too. The handling of such observational estimates, particularly their uncertainties, is a complex topic in and of itself. In RCMIP we rely on published estimates and make basic assumptions about how their uncertainty estimates should be compared to model output ranges, each of which we detail when the comparison is performed.

Given the limited amount of observations available and the ease of calibration of RCMs, comparing only with observations leaves us with little understanding of how RCMs perform in scenarios apart from a historic one in which anthropogenic emissions are heating the climate. Recognising that there are a range of possible futures, it is vital to also assess RCMs in other scenarios. Prominent examples include stabilising or falling anthropogenic emissions, strong mitigation of non-CO_2 climate forcers and scenarios with CO_2 removal. The limited observational set motivates RCMIP's second theme: evaluation against more complex models.

Theme 2: To what extent can reduced complexity models emulate the response of more complex models?

Whilst the response of more comprehensive models may not represent the behaviour of the actual Earth System, they are the best available representation of the Earth System's physical processes. By evaluating RCMs against more complex models, we can quantify the extent to which the simplifications made in RCMs limit their ability to capture physically-based model responses. For example, the extent to which the approximation of a constant climate feedback limits an RCM's ability to replicate ESMs' longer-term response under either higher forcing or lower overshoot scenarios (Rohrschneider et al., 2019).

In combination, these two research themes examine how well the reduced complexity approach can a) reproduce historical observations of the climate and b) respond to scenarios other than the recent past in a way which is consistent with our best understanding of the Earth system's physical and biogeochemical processes.

3 Simulation design

RCMIP Phase 1 includes over 50 experiments. To help modelling groups prioritise model runs and ensure comparibility of core experiments three tiers of model runs and output variables were defined. Ideally at least all Tier 1 scenarios and variables for a default model version should be submitted. The following describes the simulation design, model runs as well as data sources and format of RCMIP.

3.1 Model configuration

RCMs are usually highly flexible. Their response to anthropogenic and natural drivers strongly depends on the configuration in which they are run (i.e. their parameter values). To mitigate this as a cause of difference between models in RCMIP Phase 1, we have requested that all models provide one set of simulations in which their equilibrium climate sensitivity is equal to 3°C. While this does not define the entirety of a model's behaviour, it removes a major cause of difference between model output which is not related to model structure. On top of these 3°C climate sensitivity simulations, we have also invited groups to submit other default configurations, where each participating modelling group is free to choose their own defaults. In practice, these defaults are typically a group's most likely parameter values given their own expert judgement. Finally, where available, we have also requested probabilistic output i.e. output which quantifies the probable range of a number of output variables rather than a single timeseries for each output variable (see section 1).

3.2 RCM drivers

Depending on the experiment in RCMIP, the drivers of the RCMs will vary e.g. the RCMs might run with prescribed CO_2 concentrations and calculate consistent CO_2 emissions or the opposite i.e. run with prescribed CO_2 emissions and calculate consistent CO_2 concentrations. Below we describe each of the different setups used in RCMIP. However, a model did not need to be able to run in all of these ways to participate in RCMIP Phase 1.

3.2.1 Concentration driven

The concentration driven setup can strictly better be described as 'well-mixed greenhouse gas concentration' driven. Here, 'well-mixed greenhouse gases' refers to CO_2, CH_4, N_2O, hydrofluorocarbons (HFCs), perfluorocarbons (PFCs) and hydrochlorofluorocarbons (HCFCs). Depending on the experiment, these simulations are also supplemented by aerosol emissions and natural effective radiative forcing (specifically solar and volcanic forcings). For models which do not include the aerosol emissions to effective radiative forcing step, prescribed aerosol effective radiative forcing can instead be used.

This setup mirrors the majority of experiments performed in CMIP5 and CMIP6 such as the historical, RCP/SSP-based scenario and one percent per year rise in atmospheric CO_2 concentration (1pctCO2) experiments. The key difference between the RCMIP experiments and the CMIP experiments is that some RCMs include more anthropogenic drivers than CMIP models. Specifically, CMIP models do not include the full range of HFC, PFC and HCFC species, instead using equivalent concentrations (Meinshausen et al., 2017, 2019). In addition, some CMIP models will not include the effect of aerosol precursors such as nitrates, ammonia and organic carbon (McCoy et al., 2017).

3.2.2 CO_2 emissions driven

In the CO_2 emissions driven setup CO_2 emissions are amended with concentrations of non-CO_2 well-mixed greenhouse gases. Like the concentration-driven setup, these simulations are also supplemented by aerosol emissions (or aerosol effective radiative forcing) and natural effective radiative forcings.

This setup mirrors the CO_2 emissions driven experiments performed in CMIP5 and CMIP6 such as the esm-hist, esm-ssp/rcp and esm-1pctCO2 experiments. As above, a cause of difference between CMIP and RCMIP simulations is the number of climate drivers that are explicitly modelled.

3.2.3 Emissions driven

The emissions driven or rather 'well-mixed greenhouse gas emissions' driven setup is, like the concentration-driven and CO_2 emissions driven setups, supplemented by aerosol emissions (or aerosol effective radiative forcing) and natural effective radiative forcings.

These experiments have no obvious equivalent within the CMIP protocol. However, for many climate policy applications they are the most relevant set of experiments, given that anthropogenic emissions and reduction targets are what climate policy is directly concerned with (rather than atmospheric concentrations of GHGs). In addition, these experiments are of particular interest to the Integrated Assessment Modelling Consortium (IAMC) community and their contribution in IPCC WGIII because they require climate assessment of socioeconomic scenarios that are described in terms of their corresponding emissions, not concentrations.

3.3 Experimental design

RCMIP's experimental design focuses on a limited set of the CMIP6 experiment protocol (Eyring et al., 2016) plus some CMIP5 experiments (Taylor et al., 2012). We then complement this CMIP-based set with other experiments of interest to the RCM and IAMC communities.

Systematic intercomparison projects such as RCMIP require the definition of a clear input and output data handling framework (see Section 4 for output specifications). Historically, comparing RCMs required learning how to set up, configure and run multiple RCMs in order to produce results. This required significant time and hence, as previously discussed, has only been attempted in standalone cases with a limited number of models (Houghton et al., 1997; van Vuuren et al., 2011; Harmsen et al., 2015; Schwarber et al., 2019). With a common framework, once a model has participated in RCMIP, it is simpler to run it again in different experiments and provide output in a common, standardised format. This allows researchers to design, run and analyse experiments with far less effort than was previously required. As a result, it becomes feasible to do more regular and targeted assessment of RCMs. This capacity improves our knowledge of RCMs, our understanding of the implications of their quantitative results and our ability to develop and improve them.

Our input protocol is designed to be easy to use and hence easily able to be extended within future RCMIP phases or in separate research. The full set of RCMIP experiments is described in Supplementary Table S3 and available at rcmip.org.

3.3.1 Input format

All input data is provided in a text-based format based on the specifications used by the IAMC community (Gidden and Huppmann, 2019). The computational simplicity of RCMs means that their input specifications are relatively lightweight and hence using an uncompressed, text-based input format is possible. Further, the format is explicit about associated metadata and ensures metadata remains attached to the timeseries. As the IAMC community is a major user of RCMs, as well as being the source of input data for many experiments run with RCMs, using their data format ensures that data can be shared easily and assessment of IAM emissions scenarios can be performed with minimal data handling overhead.

The inputs are formatted as text files with comma separated values (CSV), with each row of the CSV file being a timeseries (see rcmip.org). This format is also often referred to as 'wide' although this term is imprecise (Wickham, 2014). The columns provide metadata about the timeseries, specifically the timeseries' variable, units, region, model and scenario. Other columns provide the values for each timestep within the timeseries.

Being simplified models, RCMs typically do not take gridded input. Hence we use a selection of highly aggregated socioeconomic regions, which once again follow IAMC conventions (Gidden and Huppmann, 2019). RCMIP's variables and units are described in Section 4.1. The regions used in RCMIP are described in Table S2. Scenarios are discussed in section 3.3.3 and summarised in Table S3.

One complication of using the IAMC format is that the 'model' column is reserved for the name of the integrated assessment model which produced the scenario. To enhance compatibility with the IAMC format, we don't use the 'model' column. Instead, as described in Section 4, we use the separate 'climate_model' column to store metadata about the climate model which provided the timeseries.

In general, we follow the naming conventions provided by the CMIP6 protocol (Eyring et al., 2016). These typically specify CO_2-emissions driven runs by prefixing the scenario name with 'esm-', with all other scenarios being concentration-driven. Where it is not possible to follow CMIP6 naming schemes, we use our own custom conventions. For example, full greenhouse gas emissions driven runs are typically not performed in CMIP6 because of computational cost. RCMIP's convention is to denote all greenhouse gas emissions driven by prefixing the scenario name with 'esm-' as well as suffixing the name with '-allGHG' (e.g. 'esm-ssp245-allGHG'). In addition, RCMIP includes a number of CMIP5 experiments, which sometimes have the same name as their CMIP6 counterpart (e.g. 'historical'). Where such a clash exists, we append the CMIP5 experiment with '-cmip5' to distinguish the two (e.g. 'historical-cmip5'). Finally, if an experiment is not a CMIP6-style experiment then we cannot use a CMIP6 name for it. In such cases, we choose our own name and describe it within Table S3.

3.3.2 Idealised experiments

The first group of experiments in RCMIP is idealised experiments. They focus on examining model response in highly idealised experiments. These experiments provide an easy point of comparison with output from other models, particularly CMIP output, as well as information about basic model behaviour and dynamics which can be useful for understanding the differences between models.

RCMIP's Tier 1 idealised experiments are: piControl, esm-piControl, 1pctCO2, 1pctCO2-4xext, abrupt-4xCO2, abrupt-2xCO2 and abrupt-0p5xCO2 (Table S3). The piControl and esm-piControl control experiments serve as a useful check of model type. Most RCMs are perturbation models and hence do not include any internal variability, so will simply return constant values in their control experiments. Deviations from constant values in the control experiments quickly reveals those models with more complexity. Apart from esm-piControl, all of the Tier 1 experiments are concentration driven.

After the control experiments, the other Tier 1 experiments examine the models' responses to idealised, CO_2-only concentration changes. They reveal differences in model response to forcing, particularly whether the RCM response to forcing includes non-linearities. In addition, these experiments also provide a direct comparison with CMIP experiments (i.e. more complex model behaviour) and are a key benchmark when examining an RCM's ability to emulate more complex models.

The idealised Tier 2 experiments add idealised CO_2 removal experiments, which complement the typically rising/abruptly changing Tier 1 experiments. Idealised Tier 3 experiments examine the carbon cycle response in more detail with idealised emissions driven experiments as well as experiments in which the carbon cycle is only coupled to the climate system radiatively or biogeochemically (the '1pctCO2-rad' and '1pctCO2-bgc' experiments (Jones et al., 2016)). In concentration-driven experiments, RCMs report emissions (often referred to as 'inverse emissions') and carbon cycle behaviour consistent with the prescribed CO_2 pathway. For brevity, we do not go through all Tier 2 and 3 experiments in detail here, further information can be found in Table S3.

3.3.3 Scenario experiments

In addition to the idealised experiments, RCMIP also includes a number of scenario based experiments. These examine model responses to historical transient forcing as well as a range of future scenarios. The historical experiments provide a way to compare RCM output against observational data records, and are complementary to the idealised experiments which provide a cleaner assessment of model response to forcing. The future scenarios probe RCM responses to a range of possible climate futures, both continued warming as well as stabilisation or overshoots in forcing. The variety of scenarios is a key test of model behaviour, evaluating them over a range of conditions rather than only over the historical period. Direct comparison with CMIP output then provides information about the extent to which the simplifications involved in RCM modelling are able to reproduce the response of our most advanced, physically-based models.

RCMIP's Tier 1 scenario experiments are: historical, ssp119, ssp585, esm-hist, esm-ssp119, esm-ssp585, esm-hist-allGHG, esm-ssp119-allGHG and esm-ssp585-allGHG. We focus on simulations (historical plus future) which cover the highest forcing (ssp585) and lowest forcing (ssp119) scenarios from the CMIP6 ScenarioMIP exercise (O'Neill et al., 2016). These quickly reveal differences in model projections over the widest available scenario range which can also be compared to CMIP6 output.

The Tier 2 experiments expand the CMIP6 scenario set to include the full range of ScenarioMIP concentration-driven experiments (O'Neill et al., 2016), which examine scenarios between the two extremes of ssp585 and ssp119, as well as the CMIP5 historical experiments. The CMIP5 experiments are particularly useful as they provide a direct comparison between CMIP5 and CMIP6, something which has only been done to a limited extent with more complex models (Wyser et al., 2019). Finally, the Tier 3 experiments add the remaining emissions-driven ScenarioMIP experiments, the rest of the CMIP5 scenario experiments (the so-called 'RCPs') and detection and attribution experiments (Gillett et al., 2016) designed to examine the response to specific climate forcers over both the historical period and under a middle of the road emissions scenario (ssp245).

3.3.4 Data sources

CMIP6 emissions projections follow Gidden et al. (2019) and are available at https://tntcat.iiasa.ac.at/SspDb/dsd?Action=htmlpage&page=60 (hosted by IIASA). Where well-mixed greenhouse gas emissions are missing, we use inverse emissions based on the CMIP6 concentrations from MAGICC7.0.0 (Meinshausen et al., 2019). Where regional emissions information is missing, we use the downscaling procedure described in Meinshausen et al. (2019). The emissions extensions also follow the convention described in Meinshausen et al. (2019).

For CMIP6 historical emissions (year 1850-2014), we have used data sources which match the harmonisation used for the CMIP6 emissions projections. This ensures consistency with CMIP6, although it means that we do not always use the latest data sources. CMIP6 historical anthropogenic emissions for CO_2, CH_4, BC, CO, NH_3, NOx, OC, SO_2 and non-methane volatile organic compounds (NMVOCs) come from CEDS (Hoesly et al., 2018). Biomass burning emissions data for CH_4, BC, CO, NH_3, NOx, OC, SO_2 and NMVOCs come from UVA (van Marle et al., 2017). The biomass burning emissions are a blend of both anthropogenic and natural emissions, which could lead to some inconsistency between RCMs as they make different assumptions about the particular anthropogenic/natural emissions split. CO_2 global land-use emissions are taken from the Global Carbon Budget 2016 (Quéré et al., 2016). Emissions of N_2O and the regional breakdown of CO_2 land-use emissions come from PRIMAP-hist Version 1.0 (Gütschow et al., 2016, see https://doi.org/10.5880/PIK.2016.003). Where required, historical emissions were extended back to 1750 by assuming a constant relative rate of decline based on the period 1850-1860 (noting that historical emissions are somewhat uncertain, we require consistent emissions inputs in Phase 1, uncertainty in historical emissions will be explored in future research).

CMIP6 concentrations follow Meinshausen et al. (2019). CMIP6 radiative forcings follow the data provided at https://doi.org/10.5281/zenodo.3515339). CMIP5 emissions, concentrations and radiative forcings follow Meinshausen et al. (2011b) and are taken from http://www.pik-potsdam.de/~mmalte/rcps/.

4 Output specifications

RCMIP Phase 1's submission template (see rcmip.org or https://doi.org/10.5281/zenodo.3593570) is composed of two parts. The first part is the data submission and is identical to the input format (see Section 3.3.1). This allows for simplified analysis with the same tools we used to develop the input protocols and exchange with the IAMC community as they can analyse the data using existing tools such as pyam (Gidden and Huppmann, 2019). The second part is model metadata. This includes the model's name, version number, brief description, literature reference and other diagnostics (see Section 4.2). We also request a configuration label, which uniquely identifies the configuration in which the model was run to produce the given results.

Given the typical temporal resolution of RCMs, we request all output be reported with an annual timestep. In addition, to facilitate use of the output, participating modelling groups agree to have their submitted data made available under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) license. All input and output data, as well as all code required to produce this paper, is available at gitlab.com/rcmip/rcmip and archived at https://doi.org/10.5281/zenodo.3593569.

4.1 Variables

RCMIP has a large variable request (26 Tier 1 variables, 344 Tier 2 variables and 13 Tier 3 variables), reflecting the large number of climate components included in RCMs. Here we discuss the Tier 1 variables. Tier 2 and 3 variables, which go into more detail for various parts of the climate system, are described in Supplementary Table S4.

The Tier 1 variables focus on key steps in the cause-effect chain from emissions to warming. We request emissions of black carbon, CH_4, carbon monoxide, CO_2, N_2O, NH_3, nitrous oxides, organic carbon, sulphates and non-methane volatile organic compounds. These cover the major greenhouse gases plus aerosol pre-cursor emissions. In the case of emissions driven runs, these emissions are prescribed hence we only request that these variables are reported as outputs where the modelling groups have had to alter them (e.g. their model includes internal land-use calculations which cannot be exogenously overridden). In the case of concentration-driven runs, we request emissions compatible with the prescribed concentration pathway (where these can be derived). We also request cumulative emissions of CO_2 given their strong relationship with peak warming (Allen et al., 2009; Matthews et al., 2009; Meinshausen et al., 2009; Zickfeld et al., 2009).

In Tier 1, we only request atmospheric concentrations of CO_2 and CH_4. Many models are capable of reporting much more detail than this, and we encourage them to report this detail, however some models only focus on a limited set of concentrations hence we restrict our Tier 1 variables.

In addition to concentrations, we request total, anthropogenic, CO_2 and aerosol effective radiative forcing and radiative forcing. These forcing variables are key indicators of the long-term drivers of climate change within each model as well as being a key metric for the IAMC community. Effective radiative forcing and radiative forcing are defined following Myhre et al. (2013). In contrast to radiative forcing, effective radiative forcing includes rapid adjustments beyond stratospheric temperature adjustments thus is a better indicator of long-term climate change.

Finally in Tier 1, we request output of total climate system heat uptake, ocean heat uptake, surface air temperature change and surface ocean temperature change. These variables are most directly comparable to available observations and CMIP output, with surface temperature also being highly policy-relevant. Focusing on these key variables allows us to discern major differences between RCMs, with Tier 2 and 3 variables then providing further points of comparison at a finer level of detail.

4.1.1 Probabilistic outputs

To reduce the total data volume, we request that groups provide only a limited set of percentiles from reporting probabilistic outputs, rather than every run which makes up the probabilistic ensemble. The 10th, 50th (median) and 90th percentiles are Tier 1, with the 5th, 17th, 33rd, 67th, 83rd and 95th percentiles being Tier 2. When calculating these percentiles, groups must take care to calculate derived quantities (e.g. Effective Climate Sensitivity) from each run in the probabilistic ensemble first and then calculate the percentiles in a second step. Doing the reverse (calculating percentiles first, then derived quantities from percentiles) will not necessarily lead to the same answer.

4.2 Diagnostics

On top of the variable request, we ask for one other diagnostic. This is the equilibrium climate sensitivity, defined as ‘the equilibrium warming following an instantaneous doubling of atmospheric CO_2 concentrations’. Unlike more complex models, RCMs typically have analytically tractable equilibrium climate sensitivities. This means we do not need to include ten thousand year long simulations, which would allow the models to reach true equilibrium. In contrast to the equilibrium climate sensitivity, the more commonly used effective climate sensitivity, derived using the Gregory method (Gregory, 2004), underestimates warming at true equilibrium in many models (Rohrschneider et al., 2019).

5 Illustrative results

15 models have participated in RCMIP Phase 1 (see Table 1 for an overview and links to key description papers). This is a promising start, demonstrating that the protocol is accessible to a wide range of modelling teams. We encourage any other interested groups to join further phases of the project.

The groups which have participated have submitted a number of results. We provide a brief overview of these here to give an initial assessment of the diversity of models which have submitted results to date. However, this is not intended as a comprehensive comparison or evaluation.

Firstly, we present a comparison of model best-estimates against observational best estimates (Figure 1). Such comparisons are a natural starting point for evaluation of all RCMs. We see that all the RCMs are able to capture the approximately 1 °C of warming seen in the historical observations compared to a pre-industrial reference period (Richardson et al., 2016; Rogelj et al., 2019). We also see that all the RCMs include some representation of the impact of volcanic eruptions, most notably the drop in global-mean temperatures after the eruption of Mount Agung in 1963. The exception is the CO_2-only model, GREB, which lacks the volcanic and aerosol induced cooling signals of the 19th and 20th Centuries.

Another way to evaluate RCMs is to compare their probabilistic results to observational best estimates as well as uncertainties (Figure 2). Such comparisons are vital to understanding the limits of projected probabilistic ranges and their dependence on model structure. Here we see large differences in probabilistic projections despite the similarities in the models' historical simulations. Determining the underlying causes of such differences requires investigation into and understanding of how the probabilistic distributions are created.

RCMIP also facilitates a comparison of model calibrations and CMIP output (Figure 3). Each RCM is calibrated to a different number of CMIP models (some RCMs provide no calibrations at all) because there is no common resource of calibration data. Instead, the CMIP models to which each RCM is calibrated depends on each RCM development team's capability and the time at which they last accessed the CMIP archives.

Examining multiple emulation setups (Figures S1 - S24), RCMs can reproduce the temperature response of CMIP models to idealised forcing changes to within a root-mean square error of 0.2°C (Table 2). In scenario-based experiments, it appears to be harder for RCMs to emulate CMIP output than in idealised experiments. We suggest two key explanations. The first is that effective radiative forcing cannot be easily diagnosed in SSP scenarios hence it is hard to know how best to force the RCM during calibration. The second is that the forcing in these scenarios includes periods of increase, sudden decrease due to volcanoes as well as longer term stabilisation rather than the simpler changes seen in the idealised experiments. Fitting all three of these regimes is a more difficult challenge than fitting the the idealised experiments alone.

We also present plots of the relationship between surface air temperature change and cumulative CO_2 emissions from the 1pctCO2 and 1pctCO2-4xext experiments (Figure 4). These can be used to derive the transient climate response to emissions (Matthews, 2018), a key metric in the calculation of our remaining carbon budget (Rogelj et al., 2019). The illustrative results here demonstrate a range of relationships between these two key variables, from weakly sub-linear to weakly super-linear (see further discussion in Nicholls et al. (2020)).

Finally, we present initial results from running both CMIP5 and CMIP6 generation scenarios ('RCP' and 'SSP-based' scenarios respectively) with the same models (Figure 5). In the small selection of models which have submitted all RCP, SSP-based scenario pairs, the SSP-based scenarios are 0.21°C (standard deviation 0.10°C across the models' default setups) warmer than their corresponding RCPs (Figure 5(b)). This difference is driven by the $0.42 \pm 0.26\ \mathrm{Wm}^{-2}$ larger effective radiative forcing in the SSP-based scenarios (Figure 5(d)), which itself is driven by the larger CO_2 effective radiative forcing in the SSP-based scenarios (Figure 5(f)). As noted previously, these are only initial results, not a comprehensive evaluation and should be treated as such. Nonetheless, they agree with other work (Wyser et al., 2020) which suggests that even when run with the same model (in a concentration-driven setup), the SSP-based scenarios result in (non-trivially) warmer projections than the RCPs.

6 Extensions

RCMIP Phase 1 provides proof of concept of the RCMIP approach to RCM evaluation, comparison and examination. The RCMIP Phase 1 protocol focuses on model evaluation hence is limited to experiments which are directly comparable to observations and CMIP output. In this section we present a number of ways in which further research and phases of RCMIP could build on the work presented in this paper.

The first is a deeper evaluation of the results submitted to RCMIP Phase 1. Here we have only presented illustrative results, however these can be evaluated and investigated in far more detail. For example, quantifying the degree to which different RCMs agree with observations, carefully considering how to handle observational uncertainties, natural variability (which many RCMs cannot capture) and model tuning.

Secondly, there is a wide range of RCMs available in the literature. This variety can be confusing, especially to those who are not intimately involved in developing the models. An overview of the different models, their structure and relationship to one another would help reduce the confusion and provide clarity about the implications of using one model over another.

The third suggested extension is an investigation into how different RCMs reach equilibrium in response to a step change in forcing. In RCMIP Phase 1, we only specified the equilibrium climate sensitivity value but temperature response is potentially further defined by linear and nonlinear feedbacks on different timescales. Further phases could investigate whether model structure is a driver of difference between model output or whether these differences are largely controlled by differences in parameter values.

Fourthly, emulation results have generally only been submitted for a limited set of experiments (see Supplementary Table S1 and Supplementary Figures S1 - S24). Hence it is still not clear whether the emulation performance seen in idealised experiments also carries over to scenarios, particularly the SSP-based scenarios. As the number of available CMIP6 results continues to grow, this area is ripe for investigation and will lead to improved understanding of the limits of the reduced complexity approach. A common resource for RCM calibration would greatly aid this effort because CMIP6 data handling requires specialist big data handling skills.

Fifthly, while RCMIP Phase 1 allows us to evaluate the differences between RCMs, the root causes of these differences may not be clear. This can be addressed by extending RCMIP to include experiments which specifically diagnose the reasons for differences between models e.g. simple pulse emissions of different species or prescribed step changes in atmospheric greenhouse gas concentrations. Such experiments could build on existing research (van Vuuren et al., 2011; Schwarber et al., 2019) and would allow even more comprehensive examination and understanding of RCM behaviour.

Following this, there is clearly some variation in probabilistic projections. However, what is not yet known is the extent to which variations in model structure, calibration data and calibration technique drive such differences. Investigating these questions would help understand the limits of probabilistic projections and their uncertainties. Experiments could involve constraining two different models with the same constraining technique and data, constraining a single model with two different techniques but the same data or constraining a single model with a single technique but two different datasets.

Next, the current experiments can be extended to examine the behaviour of models' gas cycles, particularly their interactions and feedbacks with other components of the climate system. This will require custom experiments but is important for understanding the behaviour of these emissions driven runs. Such experiments are particularly important for the carbon cycle, which is strongly coupled to other parts of the climate system. It should be noted that, for ESMs, the suggestion of extra experiments is limited by human and computational constraints. This constraint does not apply to RCMs because of their computational efficiency: adding extra RCM experiments adds relatively little technical burden.

One final suggestion for future research is the importance of the choice of reference period. Within the reference period, all model results and observations will be artificially brought together, narrowing uncertainty and disagreement within this period (Hawkins and Sutton, 2016). This can alter conclusions as the reference period will become less important for any fitting algorithm (because of the artificial agreement), placing more weight on other periods. Developing a method to rebase both the mean and variance of model and observational results onto other reference periods would allow the impact of the reference period choice to be explored in a more systematic fashion.

7 Conclusions

RCMs are used in many applications, particularly where computational constraints prevent other techniques from being used. Due to their importance in climate policy assessments, in carbon budget calculations, as well as applicability to a wide range of scientific questions understanding the behaviour and output from RCMs is highly relevant and requires continuous updating with the latest science. Here we have presented the Reduced Complexity Model Intercomparison Project (RCMIP), an effort to facilitate the evaluation and understanding of RCMs in a systematic, standardised and detailed way. We hope this can greatly improve ease of use of and familiarity with RCMs.

We have performed RCMIP Phase 1, which provides an initial database of experiments conducted with 15 participating models from the RCM community. RCMIP Phase 1 focused on basic evaluation and benchmarking of RCMs, providing some key starting points for all users of RCMs to examine when considering their model of choice. Here we have only presented illustrative results and further analysis is warranted to quantify the differences in behaviour (and associated uncertainty) between the different RCMs. Further work will examine the results from Phase 1 and RCMs in more detail, improving evaluation, comparison and understanding of the implications of differences between models.

RCMIP aims to fill a gap in our understanding of RCM behaviour, in particular, in how different RCMs perform relative to each other as well as in absolute terms. This gap is particularly important to fill given the widespread use of RCMs throughout the integrated assessment modelling community and in large-scale climate science assessments. We welcome requests, suggestions and further involvement from throughout the climate modelling research community. With our efforts, we hope to increase understanding of and confidence in RCMs, particularly for their many users at the science-policy interface.

Code and data availability. RCMIP input timeseries and results data along with processing scripts as used in this submission are available from the RCMIP GitLab repository at https://gitlab.com/rcmip/rcmip and archived by Zenodo (https://doi.org/10.5281/zenodo.3593569).

The ACC2 model code is available upon request.

The implementation of the AR5IR model used in this study is available in the OpenSCM repository: https://github.com/openscm/openscm/blob/ar5ir-notebooks/notebooks/ar5ir_rcmip.ipynb

The model version of ESCIMO used to produce the RCMIP runs can be downloaded from http://www.2052.info/wp-content/uploads/2019/12/mo191107%202%20ESCIMO-rcimpfrom%20mo160911%202100%20ESCIMO.vpm. The vpm extension allows you to view, examine and run the model, but not save it. The original model with full documentation is available from http://www.2052.info/escimo/.

FaIR is developed on GitHub at https://github.com/OMS-NetZero/FAIR and v1.5 used in this study is archived at Zenodo (Smith et al., 2019).

The GREB model source code used is available, upon request, on Bitbucket: https://bitbucket.org/rcmipgreb/greb-official/src/official-rcmip/. The last stable versions are available on GitHub at https://github.com/christianstassen/greb-official/releases.

The Held two layer model implementation used in this study is available in the OpenSCM repository: https://github.com/openscm/openscm/blob/ar5ir-notebooks/notebooks/held_two_layer_rcmip.ipynb

Hector is developed on GitHub at https://github.com/JGCRI/hector. The exact version of Hector used for these simulations can be found at https://github.com/JGCRI/hector/releases/tag/rcmip-tier1. The scripts for the RCMIP runs are available at https://github.com/ashiklom/hector-rcmip.

MAGICC's Python wrapper is archived at Zenodo (https://doi.org/10.5281/zenodo.1111815) and developed on GitHub at https://github.com/openclimatedata/pymagicc/.

OSCAR v3 is available on GitHub at https://github.com/tgasser/OSCAR.

WASP's code for the version used in this study is available from the supplementary material of Goodwin (2018): https://doi.org/10.1029/2018EF000889. See also the WASP website at http://www.waspclimatemodel.info/download-wasp.

The other participating models are not yet available publicly for download or as open source. Please also refer to their respective model description papers for notes and code availability.

Author contributions. ZN and RG conceived the idea for RCMIP. ZN, MM and JL setup the RCMIP website (rcmip.org), produced the first draft of the protocol and derived the data format. All authors contributed to updating and improving the protocol. ACC2 results were provided by KT and EK. AR5IR and Held et al. two layer model were provided by ZN. CICERO-SCM results were provided by JF, BS, MS and RS. ESCIMO results were provided by UG. FaIR results were provided by CS. GIR results were provided by NL. GREB results were provided by DD, CF, DM and ZX. Hector results were provided by AS and KD. MAGICC results were provided by MM, JL and ZN. MCE results were provided by JT. OSCAR results were provided by TG and YQ. WASP results were provided by PG. ZN wrote, except for the model descriptions, the first manuscript draft, produced all the figures and led the manuscript writing process with support from RG. All authors contributed to writing and revising the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We acknowledge the World Climate Research Program (WCRP) Coupled Model Intercomparison Project (CMIP) and thank the climate modelling groups for producing and making available their model output. RCMIP could not go ahead without the outputs of CMIP6 nor without the huge effort which is put in by all the researchers involved in CMIP6 (some of whom are also involved in RCMIP).

We also thank the RCMIP Steering Committee, comprised of Maisa Corradi, Piers Forster, Jan Fuglestvedt, Malte Meinshausen, Joeri Rogelj and Steven Smith, for their support and guidance throught Phase 1. We look forward to their ongoing support in further phases.

ZN benefited from support provided by the ARC Centre of Excellence for Climate Extremes (CE170100023). RG acknowledges support by the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety (16_II_148_Global_A_IMPACT) while working at PIK in the beginning of RCMIP. KT is supported by the Integrated Research Program for Advancing Climate Models (TOUGOU Program), the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan.

References

Allen, M. R., Frame, D. J., Huntingford, C., Jones, C. D., Lowe, J. A., Meinshausen, M., and Meinshausen, N.: Warming caused by cumulative carbon emissions towards the trillionth tonne, Nature, 458, 1163, 2009.

Bruckner, T., Hooss, G., Füssel, H.-M., and Hasselmann, K.: Climate System Modeling in the Framework of the Tolerable Windows Approach: The ICLIPS Climate Model, Climatic Change, 56, 119–137, https://doi.org/10.1023/a:1021300924356, 2003.

Canty, T., Mascioli, N. R., Smarte, M. D., and Salawitch, R. J.: An empirical model of global climate – Part 1: A critical evaluation of volcanic cooling, Atmospheric Chemistry and Physics, 13, 3997–4031, https://doi.org/10.5194/acp-13-3997-2013, https://www.atmos-chem-phys.net/13/3997/2013/, 2013.

Carlson, D. and Eyring, V.: Contributions to Climate Science of the Coupled Model Intercomparison Project, https://public.wmo.int/en/resources/bulletin/contributions-climate-science-of-coupled-model-intercomparison-project, 2017.

Clarke, L., Jiang, K., Akimoto, K., Babiker, M., Blanford, G., Fisher-Vanden, K., Hourcade, J.-C., Krey, V., Kriegler, E., Löschel, A., and et al.: Assessing Transformation Pathways, p. 413–510, Cambridge University Press, 2014.

Collins, W. J., Lamarque, J.-F., Schulz, M., Boucher, O., Eyring, V., Hegglin, M. I., Maycock, A., Myhre, G., Prather, M., Shindell, D., and Smith, S. J.: AerChemMIP: quantifying the effects of chemistry and aerosols in CMIP6, Geoscientific Model Development, 10, 585–607, https://doi.org/10.5194/gmd-10-585-2017, 2017.

Cowtan, K. and Way, R. G.: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends, Quarterly Journal of the Royal Meteorological Society, 140, 1935–1944, https://doi.org/10.1002/qj.2297, 2014.

Dommenget, D., Nice, K., Bayr, T., Kasang, D., Stassen, C., and Rezny, M.: The Monash Simple Climate Model experiments (MSCM-DB v1.0): an interactive database of mean climate, climate change, and scenario simulations, Geoscientific Model Development, 12, 2155–2179, https://doi.org/10.5194/gmd-12-2155-2019, 2019.

Dorheim, K., Link, R., Hartin, C., Kravitz, B., and Snyder, A.: Calibrating simple climate models to individual Earth system models: Lessons learned from calibrating Hector, Under Review at Earth and Space Science.

Etminan, M., Myhre, G., Highwood, E. J., and Shine, K. P.: Radiative forcing of carbon dioxide, methane, and nitrous oxide: A significant revision of the methane radiative forcing, Geophysical Research Letters, 43, 12,614–12,623, https://doi.org/10.1002/2016gl071930, 2016.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geoscientific Model Development (Online), 9, 2016.

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., Peters, W., Pongratz, J., Sitch, S., Le Quéré, C., Bakker, D. C. E., Canadell, J. G., Ciais, P., Jackson, R. B., Anthoni, P., Barbero, L., Bastos, A., Bastrikov, V., Becker, M., Bopp, L., Buitenhuis, E., Chandra, N., Chevallier, F., Chini, L. P., Currie, K. I., Feely, R. A., Gehlen, M., Gilfillan, D., Gkritzalis, T., Goll, D. S., Gruber, N., Gutekunst, S., Harris, I., Haverd, V., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Joetzjer, E., Kaplan, J. O., Kato, E., Klein Goldewijk, K., Korsbakken, J. I., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Marland, G., McGuire, P. C., Melton, J. R., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-I., Neill, C., Omar, A. M., Ono, T., Peregon, A., Pierrot, D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Séférian, R., Schwinger, J., Smith, N., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Werf, G. R., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2019, Earth System Science Data, 11, 1783–1838, https://doi.org/10.5194/essd-11-1783-2019, https://www.earth-syst-sci-data.net/11/1783/2019/, 2019.

Gasser, T., Ciais, P., Boucher, O., Quilcaille, Y., Tortora, M., Bopp, L., and Hauglustaine, D.: The compact Earth system model OSCAR~v2.2: description and first results, Geoscientific Model Development, 10, 271–319, https://doi.org/10.5194/gmd-10-271-2017, 2017.

Gidden, M. and Huppmann, D.: pyam: a Python Package for the Analysis and Visualization of Models of the Interaction of Climate, Human, and Environmental Systems, Journal of Open Source Software, 4, 1095, https://doi.org/10.21105/joss.01095, 2019.

Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., van Vuuren, D. P., van den Berg, M., Feng, L., Klein, D., Calvin, K., Doelman, J. C., Frank, S., Fricko, O., Harmsen, M., Hasegawa, T., Havlik, P., Hilaire, J., Hoesly, R., Horing, J., Popp, A., Stehfest, E., and Takahashi, K.: Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century, Geoscientific Model Development, 12, 1443–1475, https://doi.org/10.5194/gmd-12-1443-2019, 2019.

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., Santer, B. D., Stone, D., and Tebaldi, C.: The Detection and Attribution Model Intercomparison Project (DAMIP~v1.0) contribution to CMIP6, Geoscientific Model Development, 9, 3685–3697, https://doi.org/10.5194/gmd-9-3685-2016, 2016.

Goodwin, P.: How historic simulation–observation discrepancy affects future warming projections in a very large model ensemble, Climate Dynamics, 47, 2219–2233, https://doi.org/10.1007/s00382-015-2960-z, 2016.

Goodwin, P.: On the Time Evolution of Climate Sensitivity and Future Warming, Earth's Future, 6, 1336–1348, https://doi.org/10.1029/2018ef000889, 2018.

Goodwin, P., Williams, R. G., and Ridgwell, A.: Sensitivity of climate to cumulative carbon emissions due to compensation of ocean heat and carbon uptake, Nature Geoscience, 8, 29–34, https://doi.org/10.1038/ngeo2304, 2014.

Goodwin, P., Williams, R. G., Roussenov, V. M., and Katavouta, A.: Climate Sensitivity From Both Physical and Carbon Cycle Feedbacks, Geophysical Research Letters, 46, 7554–7564, https://doi.org/10.1029/2019gl082887, 2019.

Gregory, J. M.: A new method for diagnosing radiative forcing and climate sensitivity, Geophysical Research Letters, 31, https://doi.org/10.1029/2003gl018747, 2004.

Gütschow, J., Jeffery, M. L., Gieseke, R., Gebel, R., Stevens, D., Krapp, M., and Rocha, M.: The PRIMAP-hist national historical emissions time series, Earth System Science Data, 8, 571–603, https://doi.org/10.5194/essd-8-571-2016, 2016.

Harmsen, M. J. H. M., van Vuuren, D. P., van den Berg, M., Hof, A. F., Hope, C., Krey, V., Lamarque, J.-F., Marcucci, A., Shindell, D. T., and Schaeffer, M.: How well do integrated assessment models represent non-CO2 radiative forcing?, Climatic Change, 133, 565–582, https://doi.org/10.1007/s10584-015-1485-0, 2015.

Hartin, C. A., Patel, P., Schwarber, A., Link, R. P., and Bond-Lamberty, B. P.: A simple object-oriented and open-source model for scientific and policy analyses of the global climate system – Hector v1.0, Geoscientific Model Development, 8, 939–955, https://doi.org/10.5194/gmd-8-939-2015, 2015.

Hawkins, E. and Sutton, R.: Connecting Climate Model Projections of Global Temperature Change with the Real World, Bulletin of the American Meteorological Society, 97, 963–980, https://doi.org/10.1175/bams-d-14-00154.1, 2016.

Held, I. M., Winton, M., Takahashi, K., Delworth, T., Zeng, F., 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, https://doi.org/10.1175/2009jcli3466.1, 2010.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-i., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R., and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geoscientific Model Development, 11, 369–408, https://doi.org/10.5194/gmd-11-369-2018, 2018.

Hooss, G., Voss, R., Hasselmann, K., Maier-Reimer, E., and Joos, F.: A nonlinear impulse response model of the coupled carbon cycle-climate system (NICCS), Climate Dynamics, 18, 189–202, https://doi.org/10.1007/s003820100170, 2001.

Hope, A. P., Canty, T. P., Salawitch, R. J., Tribett, W. R., and Bennett, B. F.: Forecasting Global Warming, pp. 51–114, Springer Climate, 2017.

Houghton, J. T., Meira Filho, L. G., Griggs, D. J., and Maskell, K.: An introduction to simple climate models used in the IPCC Second Assessment Report, Cambridge University Press Cambridge, http://large.stanford.edu/courses/2015/ph240/girard1/docs/houghton.pdf, 1997.

Huppmann, D., Rogelj, J., Kriegler, E., Krey, V., and Riahi, K.: A new scenario resource for integrated 1.5 °C research, Nature Climate Change, 8, 1027–1030, https://doi.org/10.1038/s41558-018-0317-4, 2018.

Jones, C. D., Arora, V., Friedlingstein, P., Bopp, L., Brovkin, V., Dunne, J., Graven, H., Hoffman, F., Ilyina, T., John, J. G., Jung, M., Kawamiya, M., Koven, C., Pongratz, J., Raddatz, T., Randerson, J. T., and Zaehle, S.: C4MIP – The Coupled Climate–Carbon Cycle Model Intercomparison Project: experimental protocol for CMIP6, Geoscientific Model Development, 9, 2853–2880, https://doi.org/10.5194/gmd-9-2853-2016, https://www.geosci-model-dev.net/9/2853/2016/, 2016.

Jones, C. D., Frölicher, T. L., Koven, C., MacDougall, A. H., Matthews, H. D., Zickfeld, K., Rogelj, J., Tokarska, K. B., Gillett, N. P., Ilyina, T., Meinshausen, M., Mengis, N., Séférian, R., Eby, M., and Burger, F. A.: The Zero Emissions Commitment Model Intercomparison Project (ZECMIP) contribution to C4MIP: quantifying committed climate changes following zero carbon emissions, Geoscientific Model Development, 12, 4375–4385, https://doi.org/10.5194/gmd-12-4375-2019, 2019.

Joos, F., Bruno, M., Fink, R., Siegenthaler, U., Stocker, T. F., Quéré, C. L., and Sarmiento, J. L.: An efficient and accurate representation of complex oceanic and biospheric models of anthropogenic carbon uptake, Tellus B: Chemical and Physical Meteorology, 48, 394–417, https://doi.org/10.3402/tellusb.v48i3.15921, 1996.

Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., Peterschmitt, J.-Y., Abe-Ouchi, A., Albani, S., Bartlein, P. J., Brierley, C., Crucifix, M., Dolan, A., Fernandez-Donado, L., Fischer, H., Hopcroft, P. O., Ivanovic, R. F., Lambert, F., Lunt, D. J., Mahowald, N. M., Peltier, W. R., Phipps, S. J., Roche, D. M., Schmidt, G. A., Tarasov, L., Valdes, P. J., Zhang, Q., and Zhou, T.: The PMIP4 contribution to CMIP6 – Part 1: Overview and over-arching analysis plan, Geoscientific Model Development, 11, 1033–1057, https://doi.org/10.5194/gmd-11-1033-2018, 2018.

Kriegler, E.: Imprecise probability analysis for integrated assessment of climate change, Doctoral thesis, Universität Potsdam, https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/index/index/docId/497, 2005.

Leach, N. J., Nicholls, Z., Jenkins, S., Smith, C. J., Lynch, J., Cain, M., Wu, B., Tsutsui, J., and Allen, M. R.: GIR v1.0.0: a generalised impulse-response model for climate uncertainty and future scenario exploration, Geoscientific Model Development Discussions, 2020, 1–29, https://doi.org/10.5194/gmd-2019-379, https://www.geosci-model-dev-discuss.net/gmd-2019-379/, 2020.

Matthews, H. D., Gillett, N. P., Stott, P. A., and Zickfeld, K.: The proportionality of global warming to cumulative carbon emissions, Nature, 459, 829, 2009.

Matthews, J. B. R. e.: Annex I: Glossary, In Press, Geneva, Switzerland, https://www.ipcc.ch/sr15/chapter/glossary/, 2018.

McCoy, D. T., Bender, F. A.-M., Mohrmann, J. K. C., Hartmann, D. L., Wood, R., and Grosvenor, D. P.: The global aerosol-cloud first indirect effect estimated using MODIS, MERRA, and AeroCom, Journal of Geophysical Research: Atmospheres, 122, 1779–1796, https://doi.org/10.1002/2016JD026141, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD026141, 2017.

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C. B., Frieler, K., Knutti, R., Frame, D. J., and Allen, M. R.: Greenhouse-gas emission targets for limiting global warming to 2 °C, Nature, 458, 1158–1162, https://doi.org/10.1038/nature08017, 2009.

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, Atmospheric Chemistry and Physics, 11, 1417–1456, https://doi.org/10.5194/acp-11-1417-2011, 2011a.

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J.-F., Matsumoto, K., Montzka, S. A., Raper, S. C. B., Riahi, K., Thomson, A., Velders, G. J. M., and van Vuuren, D. P.: The RCP greenhouse gas concentrations and their extensions from 1765 to 2300, Climatic Change, 109, 213–241, https://doi.org/10.1007/s10584-011-0156-z, 2011b.

Meinshausen, M., Vogel, E., Nauels, A., Lorbacher, K., Meinshausen, N., Etheridge, D. M., Fraser, P. J., Montzka, S. A., Rayner, P. J., Trudinger, C. M., Krummel, P. B., Beyerle, U., Canadell, J. G., Daniel, J. S., Enting, I. G., Law, R. M., Lunder, C. R., O'Doherty, S., Prinn, R. G., Reimann, S., Rubino, M., Velders, G. J. M., Vollmer, M. K., Wang, R. H. J., and Weiss, R.: Historical greenhouse gas concentrations for climate modelling (CMIP6), Geoscientific Model Development, 10, 2057–2116, https://doi.org/10.5194/gmd-10-2057-2017, https://www.geosci-model-dev.net/10/2057/2017/, 2017.

Meinshausen, M., Nicholls, Z., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P., Reimann, S., Smith, S. J., van den Berg, M., Velders, G. J. M., Vollmer, M., and Wang, H. J.: The SSP greenhouse gas concentrations and their extensions to 2500, https://doi.org/10.5194/gmd-2019-222, 2019.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, Journal of Geophysical Research: Atmospheres, 117, n/a–n/a, https://doi.org/10.1029/2011jd017187, 2012.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing, book section 8, p. 659–740, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.018, http://www.climatechange2013.org, 2013.

Nauels, A., Meinshausen, M., Mengel, M., Lorbacher, K., and Wigley, T. M. L.: Synthesizing long-term sea level rise projections – the MAGICC sea level model v2.0, Geoscientific Model Development, 10, 2495–2524, https://doi.org/10.5194/gmd-10-2495-2017, 2017.

Nicholls, Z., Gieseke, R., Lewis, J., Nauels, A., and Meinshausen, M.: Implications of non-linearities between cumulative CO2 emissions and CO2 -induced warming for assessing the remaining carbon budget, Environmental Research Letters, http://iopscience.iop.org/10.1088/1748-9326/ab83af, 2020.

Nordhaus, W.: Estimates of the Social Cost of Carbon: Concepts and Results from the DICE-2013R Model and Alternative Approaches, Journal of the Association of Environmental and Resource Economists, 1, 273–312, https://doi.org/10.1086/676035, 2014.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J.-F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geoscientific Model Development, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

Quéré, C. L., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I., Peters, G. P., Manning, A. C., Boden, T. A., Tans, P. P., Houghton, R. A., Keeling, R. F., Alin, S., Andrews, O. D., Anthoni, P., Barbero, L., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Currie, K., Delire, C., Doney, S. C., Friedlingstein, P., Gkritzalis, T., Harris, I., Hauck, J., Haverd, V., Hoppema, M., Goldewijk, K. K., Jain, A. K., Kato, E., Körtzinger, A., Landschützer, P., Lefèvre, N., Lenton, A., Lienert, S., Lombardozzi, D., Melton, J. R., Metzl, N., Millero, F., Monteiro, P. M. S., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-i., O'Brien, K., Olsen, A., Omar, A. M., Ono, T., Pierrot, D., Poulter, B., Rödenbeck, C., Salisbury, J., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Stocker, B. D., Sutton, A. J., Takahashi, T., Tian, H., Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., and Zaehle, S.: Global Carbon Budget 2016, Earth System Science Data, 8, 605–649, https://doi.org/10.5194/essd-8-605-2016, 2016.

Randers, J., Golüke, U., Wenstøp, F., and Wenstøp, S.: A user-friendly earth system model of low complexity: the ESCIMO system dynamics model of global warming towards 2100, Earth System Dynamics, 7, 831–850, https://doi.org/10.5194/esd-7-831-2016, 2016.

Richardson, M., Cowtan, K., Hawkins, E., and Stolpe, M. B.: Reconciled climate response estimates from climate models and the energy budget of Earth, Nature Climate Change, 6, 931–935, https://doi.org/10.1038/nclimate3066, 2016.

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshgi, H., Kobayashi, S., Kriegler, E., and et al.: Mitigation pathways compatible with 1.5°C in the context of sustainable development, p. 93–174, IPCC/WMO, http://www.ipcc.ch/report/sr15/, 2018.

Rogelj, J., Forster, P. M., Kriegler, E., Smith, C. J., and Séférian, R.: Estimating and tracking the remaining carbon budget for stringent climate targets, Nature, 571, 335–342, https://doi.org/10.1038/s41586-019-1368-z, 2019.

Rohrschneider, T., Stevens, B., and Mauritsen, T.: On simple representations of the climate response to external radiative forcing, Climate Dynamics, 53, 3131–3145, https://doi.org/10.1007/s00382-019-04686-4, 2019.

Schlesinger, M. E., Jiang, X., and Charlson, R. J.: Implication of Anthropogenic Atmospheric Sulphate for the Sensitivity of the Climate System, in: Climate Change and Energy Policy: Proceedings of the International Conference on Global Climate Change: Its Mitigation Through Improved Production and Use of Energy, edited by Rosen, L. and Glasser, R., pp. 75–108, American Institute of Physics, 1992.

Schwarber, A. K., Smith, S. J., Hartin, C. A., Vega-Westhoff, B. A., and Sriver, R.: Evaluating climate emulation: fundamental impulse testing of simple climate models, Earth System Dynamics, 10, 729–739, https://doi.org/10.5194/esd-10-729-2019, 2019.

Skeie, R. B., Fuglestvedt, J., Berntsen, T., Peters, G. P., Andrew, R., Allen, M., and Kallbekken, S.: Perspective has a strong effect on the calculation of historical contributions to global warming, Environmental Research Letters, 12, 024 022, https://doi.org/10.1088/1748-9326/aa5b0a, 2017.

Skeie, R. B., Berntsen, T., Aldrin, M., Holden, M., and Myhre, G.: Climate sensitivity estimates – sensitivity to radiative forcing time series and observational data, Earth System Dynamics, 9, 879–894, https://doi.org/10.5194/esd-9-879-2018, 2018.

Smith, C. J., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., and Regayre, L. A.: FAIR v1.3: a simple emissions-based impulse response and carbon cycle model, Geoscientific Model Development, 11, 2273–2297, https://doi.org/10.5194/gmd-11-2273-2018, 2018.

Smith, C. J., Gieseke, R., and Nicholls, Z.: OMS-NetZero/FAIR: RCMIP phase 1, https://doi.org/10.5281/ZENODO.3588880, https://zenodo.org/record/3588880, 2019.

Tanaka, K. and O'Neill, B. C.: The Paris Agreement zero-emissions goal is not always consistent with the 1.5 °C and 2 °C temperature targets, Nature Climate Change, 8, 319–324, https://doi.org/10.1038/s41558-018-0097-x, 2018.

Tanaka, K., Kriegler, E., Bruckner, T., Hooss, G., Knorr, W., Raddatz, T., and Tol, R.: Aggregated Carbon cycle, atmospheric chemistry and climate model (ACC2): description of forward and inverse mode, https://pure.mpg.de/rest/items/item_994422/component/file_994421/content, 2007.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bulletin of the American Meteorological Society, 93, 485–498, https://doi.org/10.1175/bams-d-11-00094.1, 2012.

Tsutsui, J.: Quantification of temperature response to CO2 forcing in atmosphere–ocean general circulation models, Climatic Change, 140, 287–305, https://doi.org/10.1007/s10584-016-1832-9, 2017.

Tsutsui, J.: Diagnosing Transient Response to CO2 Forcing in Coupled Atmosphere-Ocean Model Experiments Using a Climate Model Emulator, Geophysical Research Letters, 47, https://doi.org/10.1029/2019gl085844, 2020.

van Marle, M. J. E., Kloster, S., Magi, B. I., Marlon, J. R., Daniau, A.-L., Field, R. D., Arneth, A., Forrest, M., Hantson, S., Kehrwald, N. M., Knorr, W., Lasslop, G., Li, F., Mangeon, S., Yue, C., Kaiser, J. W., and van der Werf, G. R.: Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fire models (1750–2015), Geoscientific Model Development, 10, 3329–3357, https://doi.org/10.5194/gmd-10-3329-2017, 2017.

van Vuuren, D. P., Lowe, J., Stehfest, E., Gohar, L., Hof, A. F., Hope, C., Warren, R., Meinshausen, M., and Plattner, G.-K.: How well do integrated assessment models simulate climate change?, Climatic Change, 104, 255–285, https://doi.org/10.1007/s10584-009-9764-2, 2011.

Vega-Westhoff, B., Sriver, R. L., Hartin, C. A., Wong, T. E., and Keller, K.: Impacts of Observational Constraints Related to Sea Level on Estimates of Climate Sensitivity, Earth's Future, 7, 677–690, https://doi.org/10.1029/2018ef001082, 2019.

von Deimling, T. S., Meinshausen, M., Levermann, A., Huber, V., Frieler, K., Lawrence, D. M., and Brovkin, V.: Estimating the near-surface permafrost-carbon feedback on global warming, Biogeosciences, 9, 649–665, https://doi.org/10.5194/bg-9-649-2012, 2012.

von Schuckmann, K., Cheng, L., Palmer, M. D., Tassone, C., Aich, V., Adusumilli, S., Beltrami, H., Boyer, T., Cuesta-Valero, F. J., Desbruyères, D., Domingues, C., García-García, A., Gentine, P., Gilson, J., Gorfer, M., Haimberger, L., Ishii, M., Johnson, G. C., Killik, R., King, B. A., Kirchengast, G., Kolodziejczyk, N., Lyman, J., Marzeion, B., Mayer, M., Monier, M., Monselesan, D. P., Purkey, S., Roemmich, D., Schweiger, A., Seneviratne, S. I., Shepherd, A., Slater, D. A., Steiner, A. K., Straneo, F., Timmermans, M.-L., and Wijffels, S. E.: Heat stored in the Earth system: Where does the energy go? The GCOS Earth heat inventory team, Earth System Science Data Discussions, 2020, 1–45, https://doi.org/10.5194/essd-2019-255, https://www.earth-syst-sci-data-discuss.net/essd-2019-255/, 2020.

Waldhoff, S. T., Anthoff, D., Rose, S., and Tol, R. S.: The marginal damage costs of different greenhouse gases: An application of FUND, Economics Discussion Paper, 2011.

Wickham, H.: Tidy Data, Journal of Statistical Software, Articles, 59, 1–23, https://doi.org/10.18637/jss.v059.i10, https://www.jstatsoft.org/v059/i10, 2014.

Wigley, T. M. L. and Raper, S. C. B.: Interpretation of High Projections for Global-Mean Warming, Science, 293, 451–454, https://doi.org/10.1126/science.1061604, https://science.sciencemag.org/content/293/5529/451, 2001.

Wyser, K., van Noije, T., Yang, S., von Hardenberg, J., O'Donnell, D., and Döscher, R.: On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6, Geoscientific Model Development Discussions, 2019, 1–13, https://doi.org/10.5194/gmd-2019-282, 2019.

Wyser, K., Kjellström, E., Koenigk, T., Martins, H., and Döscher, R.: Warmer climate projections in EC-Earth3-Veg: the role of changes in the greenhouse gas concentrations from CMIP5 to CMIP6, Environmental Research Letters, 15, 054 020, https://doi.org/10.1088/1748-9326/ab81c2, https://doi.org/10.1088%2F1748-9326%2Fab81c2, 2020.

Zanna, L., Khatiwala, S., Gregory, J. M., Ison, J., and Heimbach, P.: Global reconstruction of historical ocean heat storage and transport, Proceedings of the National Academy of Sciences, 116, 1126–1131, https://doi.org/10.1073/pnas.1808838115, https://www.pnas.org/content/116/4/1126, 2019.

Zickfeld, K., Eby, M., Matthews, H. D., and Weaver, A. J.: Setting cumulative emissions targets to reduce the risk of dangerous climate change, Proceedings of the National Academy of Sciences, 106, 16 129–16 134, https://doi.org/10.1073/pnas.0805800106, 2009.

Table 1. Overview of the physical components of the models participating in RCMIP Phase 1.

Model (acronym used in figures)	Spatial resolution	Key references
ACC2 (ACC2-v4-2)	Global land/ocean	Tanaka and O'Neill (2018); Tanaka et al. (2007) (also Hooss et al. (2001); Bruckner et al. (2003); Kriegler (2005))
AR5IR (ar5ir-2box, ar5ir-3box)	Global	Myhre et al. (2013)
CICERO-SCM (CICERO-SCM)	Hemispheric	Skeie et al. (2017) (also Schlesinger et al. (1992); Joos et al. (1996); Etminan et al. (2016); Skeie et al. (2018))
EMGC (EMGC)	Global	Canty et al. (2013); Hope et al. (2017)
ESCIMO (ESCIMO)	Global	Randers et al. (2016)
FaIR (FaIR-v1-5)	Global	Smith et al. (2018); Etminan et al. (2016)
GIR (GIR)	Global	Leach et al. (2020)
GREB (GREB-v1-0-1)	96 x 48 grid	Dommenget et al. (2019)
Hector (hector|62381e71)	Global	Hartin et al. (2015); Dorheim et al. (Under Review at Earth and Space Science); Vega-Westhoff et al. (2019) (see also Kriegler (2005); Tanaka et al. (2007))
Held et. al two layer model (held-two-layer-uom)	Global	Rohrschneider et al. (2019); Held et al. (2010)
MAGICC (MAGICC-v7-1-0-beta)	Hemispheric land/ocean	Meinshausen et al. (2011a, 2019) (see also von Deimling et al. (2012); Nauels et al. (2017))
MCE (MCE-v1-1)	Global	Tsutsui (2017, 2020) (see also Joos et al. (1996); Hooss et al. (2001))
OSCAR (OSCAR-v3-0)	Global, with regionalized land carbon cycle	Gasser et al. (2017)
WASP (WASP-v2)	Global	Goodwin (2018); Goodwin et al. (2019) (see also Goodwin et al. (2014); Goodwin (2016))

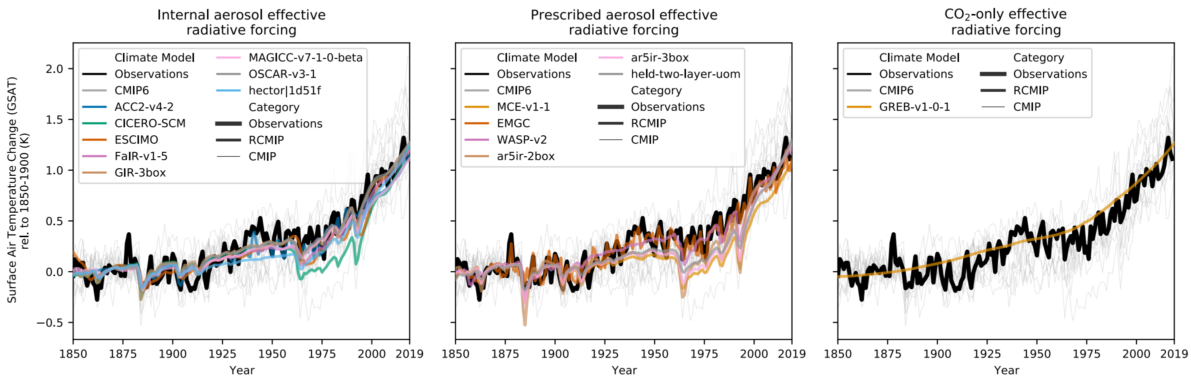

Figure 1. Historical global-mean annual mean surface air temperature (GSAT) simulations. Thick black line is observed GSAT (Richardson et al., 2016; Rogelj et al., 2019). Medium thickness lines are illustrative configurations for RCMIP models. Thin grey solid lines are CMIP6 models. In order to provide timeseries up until 2019, we have used data from the combination of historical and ssp585 simulations for RCMIP and CMIP6 models and rcp85 data for CMIP5 models.

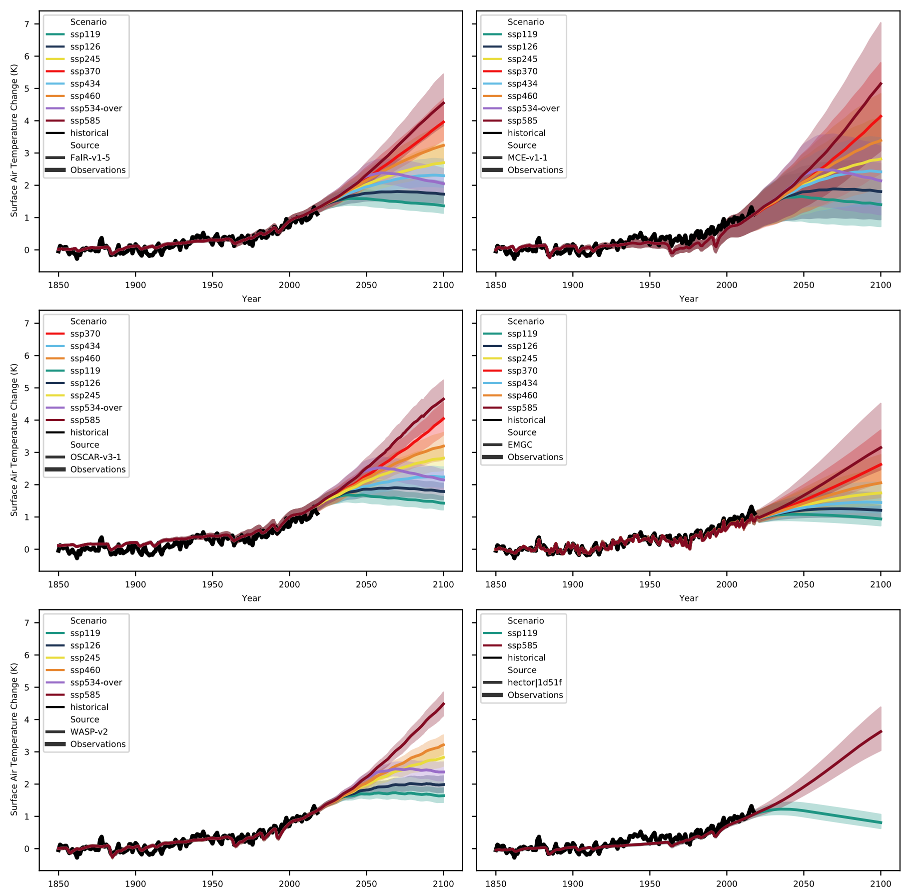

Figure 2. Probabilistic projections. Black line is observed GSAT (Richardson et al., 2016; Rogelj et al., 2019). Coloured lines are results for different RCMs for the SSP-based scenarios (ranges are 66% ranges). Note that not all groups have been able to perform all simulations.

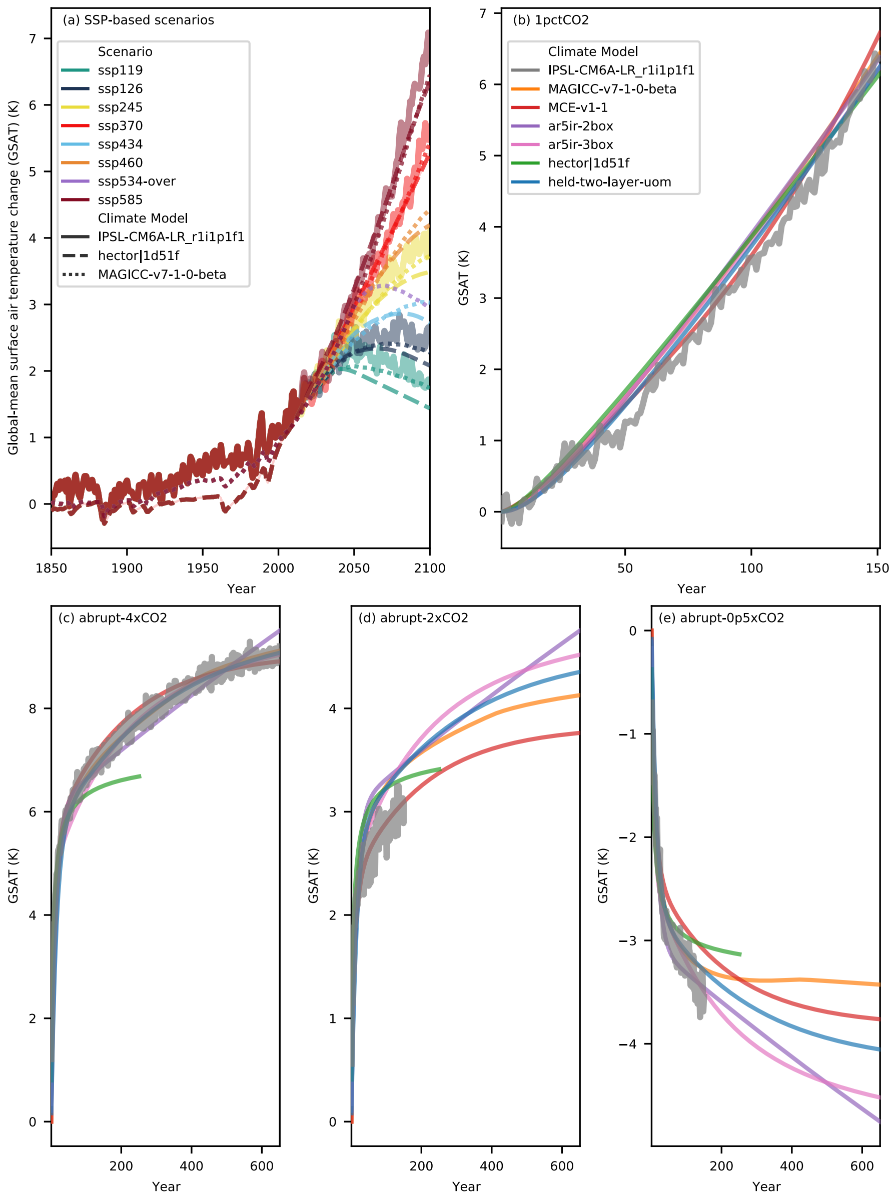

Figure 3. Emulation of CMIP6 models by RCMs. The thick transparent lines are the target CMIP6 model output (here from IPSL-CM6A-LR r1i1p1f1). The thin lines are emulations from different RCMs. Panel (a) shows results for scenario based experiments while panels (b) - (e) show results for idealised CO2-only experiments (note that panels (b) - (e) share the same legend). See the Supplementary Information for other target CMIP6 models.

Table 2. Model emulation scores over all emulated models and scenarios. Here we provide root-mean square errors over the SSPs plus four idealised CO2-only experiments (abrupt-2xCO2, abrupt-4xCO2, abrupt-0p5xCO2, 1pctCO2). As the models have not all provided emulations for the same set of models and scenarios, the model emulation scores are indicative only and are not a true, fair test of skill. For target model by target model emulation scores, see Table S1.

Model (number of emulated scenarios)	**Surface Air Temperature Change (GSAT aka tas) root-mean square error (indicative only)**
MAGICC-v7-1-0-beta (131)	0.21 K
MCE-v1-1 (44)	0.19 K
ar5ir-2box (36)	0.24 K
ar5ir-3box (36)	0.28 K
hector|1d51f (64)	0.28 K
held-two-layer-uom (34)	0.18 K

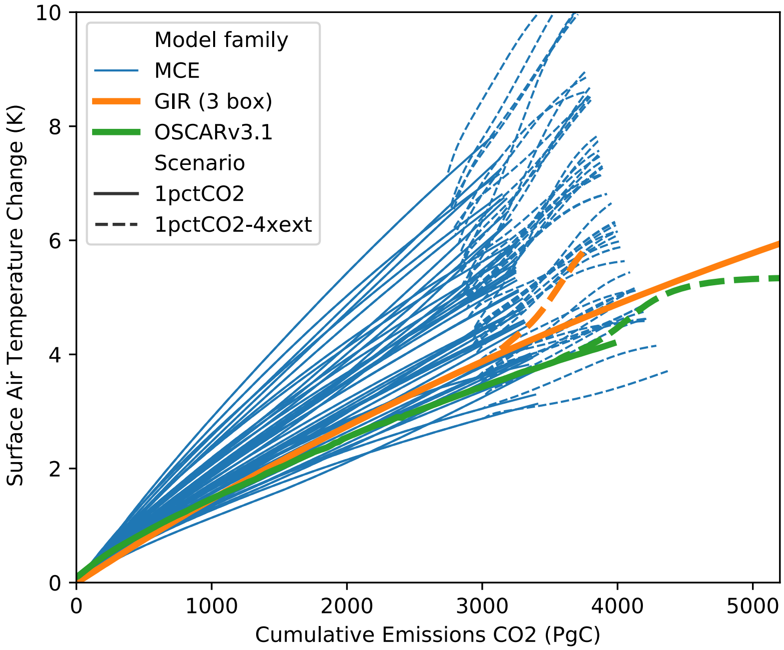

Figure 4. Surface air temperature change against cumulative CO_2 emissions in the 1pctCO2 and 1pctCO2-4xext experiments. Thin lines are used for the MCE model's family of emulation setups. Thick lines are used for the GIR (3 box) and OSCARv3.1 default setups (OSCARv3.1's probabilistic output is available but not shown).

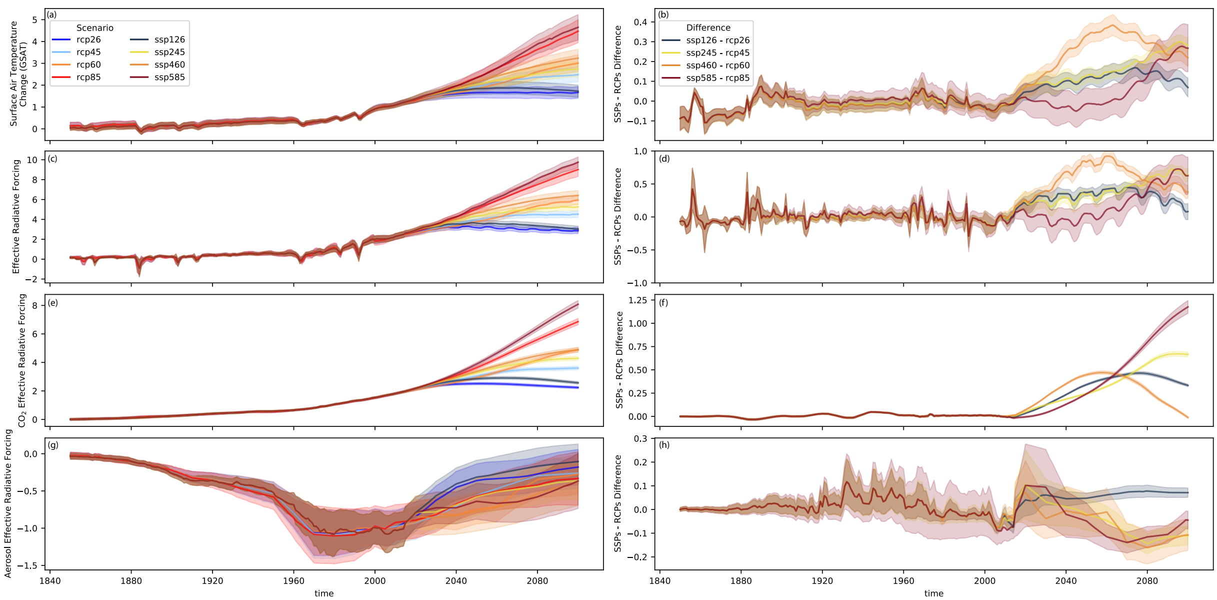

Figure 5. Output from the RCPs and SSP-based scenarios up until 2100. The left-hand column shows raw model output. The right-hand column shows the difference between scenarios for a given model's output. The shaded range shows one standard deviation about the median (solid lines). Output is shown for surface air temperature change (GSAT, (a) and (b)), effective radiative forcing ((c) and (d)), CO_2 effective radiative forcing ((e) and (f)) and aerosol effective radiative forcing ((g) and (h)). The results here are illustrative and provided only for those models which have done RCP, SSP-based scenario pairs.