

The Climate in Climate Economics*

Doris Folini

Institute for Atmospheric and Climate Science, ETHZ, Switzerland

Aleksandra Friedl

Department of Economics, University of Lausanne, Switzerland; ifo Institute – Leibniz Institute for Economic Research, University of Munich, Germany

Felix Kübler

Department for Banking and Finance, University of Zürich; Swiss Finance Institute (SFI), Switzerland

and

Simon Scheidegger

Department of Economics, University of Lausanne; Enterprise for Society (E4S), Switzerland

First version received June 2022; Editorial decision August 2023; Accepted January 2024 (Eds.)

To analyse climate change mitigation strategies, economists rely on simplified climate models—so-called climate emulators—that provide a realistic quantitative link between CO₂ emissions and global warming at low computational costs. In this paper, we propose a generic and transparent calibration and evaluation strategy for these climate emulators that are based on freely and easily accessible state-of-the-art benchmark data from climate sciences. We demonstrate that the appropriate choice of the free model parameters can be of key relevance for the predicted social cost of carbon. The key idea we put forward is to calibrate the simplified climate models to benchmark data from comprehensive global

Key words: Climate change, Social cost of carbon, Carbon taxes, Environmental policy, Deep learning, Integrated assessment models, DICE-2016

JEL codes: C61, E27, Q5, Q51, Q54, Q58

*We thank the managing editor Kurt Mitman and two anonymous referees for their extremely valuable comments and suggestions. We are very grateful to Pratyuksh Bansal, Yongyang Cai, Thomas Lontzek, Alena Miftakhova, Rick van der Ploeg, Karl Schmedders, Christian Träger, Takafumi Usui, Frank Venmans as well as seminar participants at the University of Lausanne, the University of Zurich, and various conferences and workshops for very useful conversations and comments. This work was supported by the Swiss National Science Foundation (SNF), under project ID “Can Economic Policy Mitigate Climate-Change?,” and the Swiss Platform for Advanced Scientific Computing (PASC), under project ID “Computing Equilibria in Heterogeneous Agent Macro Models on Contemporary HPC Platforms,” for research support, and the Swiss National Supercomputing Center (CSCS) under project ID 995. Simon Scheidegger acknowledges financial support from the Department of Economics at the University of Pennsylvania.

The editor in charge of this paper was Kurt Mitman.

climate models that took part in the coupled model intercomparison project, phase 5 (CMIP5). In particular, we propose to use four different test cases that are considered pivotal in the climate science literature: two highly idealized tests to separately calibrate and evaluate the carbon cycle and temperature response, an idealized test to quantify the transient climate response, and a final test to evaluate the performance for scenarios close to those arising from economic models, and that include exogenous forcing. As a concrete example, we re-calibrate the climate part of the widely used DICE-2016, fathoming the CMIP5 uncertainty range of model responses: the multi-model mean as well as extreme, but still permissible climate sensitivities and carbon cycle responses. We demonstrate that the functional form of the climate emulator of the DICE-2016 model is fit for purpose, despite its simplicity, but its carbon cycle and temperature equations are miscalibrated, leading to the conclusion that one may want to be skeptical about predictions derived from DICE-2016. We examine the importance of the calibration for the social cost of carbon in the context of a partial equilibrium setting where interest rates are exogenous, as well as the simple general equilibrium setting from DICE-2016. We find that the model uncertainty from different consistent calibrations of the climate system can change the social cost of carbon by a factor of 4 if one assumes a quadratic damage function. When calibrated to the multi-model mean, our model predicts similar values for the social cost of carbon as the original DICE-2016, but with a strongly reduced sensitivity to the discount rate and about 1 degree less long-term warming. The social cost of carbon in DICE-2016 is oversensitive to the discount rate, leading to extreme comparative statics responses to changes in preferences.

1. INTRODUCTION

Anthropogenic climate change and the associated economic damages constitute a substantial negative externality from CO₂ emissions. Economic policy can mitigate this externality and potentially lead to significant welfare gains across all economic agents (see, *e.g.* Nordhaus 2018; Kotlikoff *et al.* 2021a, 2021b, 2021c). In order to determine the optimal mitigation strategies, economists need to develop quantitative models that produce a realistic link between CO₂ emissions and global warming and that are informed by research in climate science as presented in the Intergovernmental Panel on Climate Change (IPCC)¹ reports, that is, the “state-of-the-art” in climate science.

The backbone of the IPCC reports are the coupled model intercomparison projects (CMIP5 and CMIP6; see Taylor *et al.* 2012; Eyring *et al.* 2016), which bundle the output from a collection of global climate models (GCMs) that run on pre-defined future scenarios, notably future greenhouse gas (GHG) concentrations. The GCMs come essentially in two flavors: Earth system models (ESMs), which cover biogeochemical processes like the carbon cycle and can take carbon emissions as input, and coupled atmosphere-ocean global climate models (AOGCMs), which lack biogeochemical processes and can take only GHG concentrations as input. One fundamental challenge is that the computational costs for ESMs (and for AOGCMs) are so significant that they are not suitable for studying the two-way feedback between the Earth system and human behavior.² Therefore, economic models focusing on this feedback have to rely on a much-simplified representation of the Earth system component. Simplified and computationally cheap-to-evaluate climate models come at different levels of complexity (and computational costs) and under different names: climate emulators (CEs), energy balance models, simple climate models, or ESMs of intermediate complexity. We avoid using this differentiation throughout this paper and, somewhat loosely, just use CE or the climate model.

There is a proliferation of different CEs used in the climate-change economics literature. The most popular emulators can be found in the dynamic integrated model of climate and the

1. For more details on IPCC, see, *e.g.* <https://www.ipcc.ch>.

2. See, *e.g.* Danabasoglu *et al.* (2020), who stated that the Community Earth System Model Version 2 (CESM2), one of the most popular ESMs, needs more than 3,000 CPU hours to simulate 1 year of the global climate.

economy (DICE) (Nordhaus 2018), in PAGE (Hope 2013), as well as that in Golosov *et al.* (2014). Unfortunately, as Calel and Stainforth (2017) pointed out, “*in failing to maintain clear links to the physical science literature, the climate components of these models have become opaque to the scientific community.*” Even worse, there is no consensus in the (economics) literature about what makes a “good” climate model, that is, which data the model is supposed to be matched to when choosing the model’s functional forms and parameters.

We close this gap by developing a comprehensive and transparent suite of tests that aims to answer the following key questions: (i) When a particular CE used in economics fit for purpose?, and (ii) Which calibration strategy should generally be applied so that the CE is in line with state-of-the-art climate science? We advocate the use of four test cases that are standard in climate science, notably in the context of CMIP (Eyring *et al.* 2016; Keller *et al.* 2018). Two of the said tests will be used for the calibration of the CE, whereas another two tests are used for evaluating the CE. For CEs as used by economists, these four tests can be regarded as comprehensive. However, the said test cases are typically not employed by economists to evaluate their emulators. We showcase our proposed strategy, including economic effects at an example from the seminal DICE model family: we re-calibrate one specific version—DICE-2016 (Nordhaus 2018)³—according to our battery of tests to match the output of climate science models and obtain “Calibrated DICE” (CDICE). In addition, we illustrate the economic importance of a proper calibration of the CE.

A key functionality of any CE is to translate anthropogenic emissions, as computed by the economic model, into a global mean temperature change. The task is typically split into two parts (see, *e.g.* DICE Nordhaus 2018): a “carbon cycle” which translates anthropogenic emissions in the wake of human economic activity into changes in the atmospheric CO₂ concentration, and a temperature model which translates changes in atmospheric CO₂ concentrations into global mean temperature changes. We propose first to calibrate both parts of the CE independently, using one test case for each calibration. Each of the two test cases comes, from a climate science perspective, with a plausible range of outcomes. To capture this range, we suggest for each test case one intermediate and two extreme calibration targets, to be detailed below. This leaves us with a total of nine differently calibrated CEs, three calibrations for the carbon cycle times three calibrations for the temperature part. Finally, we advocate evaluating the performance of these differently calibrated CEs against two additional climate science benchmarks, also detailed below, that are closer to real-world applications. The four tests and the associated CMIP5-based benchmark data are well consolidated and explored in the literature.⁴

The first test we use targets the calibration of the carbon cycle. It uses a highly idealized setup; an instantaneous emission of 100 GtC to the present-day atmosphere is followed in time. The CE is calibrated such that the (steadily decreasing) fraction of emitted carbon remaining in the atmosphere aligns with benchmark data adopted from Joos *et al.* (2013). We choose three specific calibration targets—one intermediate and two extremes—to mirror the range of carbon cycle responses considered plausible from a climate science point of view. As an intermediate target, we use the average response of all models in Joos *et al.* (2013); their multi-model mean

3. Below, we use the abbreviation DICE-2016 and DICE interchangeably.

4. The CE calibrations and test cases used here are all based on CMIP5 data, which is publicly available under <http://iacweb.ethz.ch/staff/beyerleu/cmip5/> and <https://esgf-node.llnl.gov/search/cmip5/>. Meanwhile, data from the sixth generation of the CMIP6 are available, and one may wonder whether switching from the CMIP5 data to the newer CMIP6 data makes a difference. Dedicated studies comparing CMIP5 and CMIP6 simulation data suggest that this is not the case for those aspects of climate and its modeling that we are interested in here. We discuss this point in detail in Section 3.4.

(MMM).⁵ As extreme calibration targets, we use two specific models from the study by [Joos et al. \(2013\)](#), leaving either a very large (cf. their model “MESMO”) or small (cf. their model “LOVECLIM”) fraction of emitted carbon in the atmosphere in the long run (cf. [Joos et al. 2013](#)). In the following, to make the different calibration targets transparent where necessary, we add the name of the corresponding target model to the general name CDICE.

The second test we use targets the calibration of the temperature equations. Being highly idealized, it follows the evolution of the global mean temperature after an instantaneous quadrupling of atmospheric CO₂ concentrations with respect to pre-industrial values. The CE is calibrated to meet benchmark data anchored in CMIP5 and adopted from [Geoffroy et al. \(2013\)](#). The range of warming displayed by the different CMIP5 models is considered plausible from a climate science point of view, strong or weak warming are plausible limiting cases ([Knutti et al. 2017](#)). We choose three specific calibration targets again to capture this range: the MMM, as well as two models that show a rather strong (model HadGEM2-ES, Hadley Center Global Environmental Model 2) or weak (model GISS-E2-R, Goddard Institute for Space Studies Model E) warming in the long run within CMIP5.

Any full-fledged quantitative economic treatment of climate change must take this “model uncertainty” seriously. Even in a simple representative agent framework, the attitude of the representative agent toward risk and uncertainty becomes a crucial determinant for optimal carbon policy (see, e.g. [Barnett et al. 2020, 2022](#)). To assess the effect of climate model uncertainty on economic models, the so-called equilibrium climate sensitivity (ECS) parameter is often varied within the range observed in CMIP5 models (see, e.g. [Hassler et al. 2018](#), or [Nordhaus 2018](#)). The ECS measures the long-term increase in global average temperature expected to occur after an instantaneous doubling of the atmospheric CO₂ concentration; different projected warmings imply a different ECS ([Geoffroy et al. 2013](#)). While being endogenously determined in AOGCMs, the ECS is an exogenous parameter determining long-run warming in CEs. The approach of just changing the ECS parameter to sample the uncertainty range is often a useful shortcut, but it misses part of the relevant temperature dynamics. We will show below that it makes a difference whether one changes only the ECS of the CE or, as we do, really calibrates the CE against CMIP5 model data whose ECS (high for HadGEM2-ES, low for GISS-E2-R) can then be associated with upper and lower bounds on the true costs of climate change and can also provide an average scenario (MMM) that can, perhaps, be viewed as “most likely.”

The third test evaluates the CE’s transient climate response (TCR), using the same setup as in climate science: starting in 1850 at pre-industrial levels, the atmospheric CO₂ concentration gradually increases at 1% per year until doubling after 70 years (quadrupling after 140 years). The TCR is defined as the warming after 70 years. It is not calibrated for but, rather, is an emergent property of both the CE and CMIP5 models—or any AOGCM or ESM in general. This test allows comparing the TCR of the CE with values from CMIP5. In fact, the TCR is a standard test in climate sciences and a compulsory test for any GCM to take part in CMIP (see [Eyring et al. 2016](#)). As the test takes CO₂ concentrations as input, not emissions, it addresses only the temperature equations of the CE.

The fourth test case is closer to what integrated assessment models (IAMs) are expected to cope with in real applications: the CMIP5 simulations for the historical period and future representative concentration pathways (RCPs) for three different forcing levels (RCP26, RCP45, RCP60, RCP85).⁶ Input data are either prescribed atmospheric CO₂ concentrations or CO₂

5. The multi-model mean is a debated, but regularly used, benchmark quantity in climate sciences (see, e.g. [Beusch et al. 2020](#); [Thao et al. 2021](#), and references therein).

6. The scenarios describe the results of different socio-economic narratives that produce particular concentration profiles of greenhouse gases, aerosols, and other climatically relevant forcing agents over the twenty-first century. The

emissions.⁷ In the latter case, the entire climate model is tested; in the former case, only its temperature equations. Within the CE, an assumption has to be made on non-CO₂ forcings, which form part of the CMIP5 data. The CE climate is evaluated against data from the CMIP5 archive.

To exemplify our calibration strategy, we apply it to the climate part of the widely used DICE-2016 model by Nordhaus (2018). As Dietz *et al.* (2021) pointed out, the most commonly used CEs in economics, including Nordhaus' widely used DICE-2016 model, seem seriously flawed in that they cannot reproduce the evolution of atmospheric CO₂ and temperature for the basic test case of an instantaneous carbon pulse to the present-day atmosphere (Joos *et al.* 2013). Based on this one test case, which mixes the response of the carbon cycle and the temperature equations, they conclude that “*economic models of climate change are out of line with state of the art in climate science*,” and they recommend that “*the climate modules in economic models be replaced*.” While we confirm the finding by Dietz *et al.* (2021) that the carbon cycle of DICE-2016 leaves too much CO₂ in the atmosphere and its temperature response to a sudden increase in CO₂ concentration is too slow, our study also demonstrates that their conclusion that the climate model must be replaced is incorrect. In particular, we show that the functional form of the CE in DICE-2016 is fit for purpose and can be re-calibrated to match results from our test cases with respect to the multi-model mean as well as concerning CMIP5 extreme cases. For the carbon cycle, no perfect calibration is possible as the functional form in DICE and the benchmark data from Joos *et al.* (2013) are different. By contrast, the temperature equations in DICE are formally identical to those in Geoffroy *et al.* (2013), thus the calibration target can be matched exactly.⁸ Our study goes beyond Dietz *et al.* (2021) in that we present a comprehensive set of tests to ascertain transparently whether a CE is fit for purpose and, if not, we also present means to re-calibrate a CE properly on the condition that its functional form is fit for purpose.

To examine the impact of different climate calibrations on the optimal price of carbon, we first consider a very simple partial equilibrium setup where we take the path of emissions, the growth rate, and the interest rate as exogenously given. As it is standard in economic models of climate change (see, *e.g.* Nordhaus 1979; Golosov *et al.* 2014), we assume that an aggregate damage function determines what fraction of output is destroyed as a function of the temperature. In these (deterministic) models, the optimal carbon tax equals the social cost of carbon (SCC), which can be computed as the marginal cost of carbon emissions, that is, as the sum of all future damages, discounted at the market interest rate, that results from an infinitesimal extra emission of CO₂ into the atmosphere. The ratio of growth rate and interest rate then constitutes the correct discount factor for these calculations and, as it is well documented in the literature, one of the most important determinants of the SCC. We show that for a given discount factor and emissions scenario, the model uncertainty from climate science can change the SCC by a factor of four if one assumes a quadratic damage function; one of the most common functional forms in the literature. This finding is (roughly) independent of discounting and emissions, but depends crucially on the functional form of the damage function. Simply adjusting the ECS in the equations of the climate emulator leads to significantly different estimates for the SCC than

RCP85 scenario, for instance, reflects a “no policy” narrative, in which total anthropogenic forcing reaches approximately 8.5 W/m² in the year 2100. Conversely, the RCP26 scenario involves aggressive decarbonization, causing radiative forcing to peak at approximately 3 W/m² around 2050 and to decline to approximately 2.6 W/m² at the end of the twenty-first century.

7. Meinshausen *et al.* (2011), <https://www.pik-potsdam.de/~mmalte/rcps/>.

8. The original formulation of DICE hides this similarity somewhat, as it absorbs the time-step of the climate part in the coefficients of the equations. We offer a more transparent formulation where the time-step is explicitly exposed, which allows us to directly use coefficients from Geoffroy *et al.* (2013) that are already calibrated to match the CMIP5 results.

using our full-blown calibrations of extreme climate models. In this simple partial equilibrium setting, the incorrect calibration of the CE in DICE-2016 leads to an SCC that is oversensitive to the discount rate. This high sensitivity has important consequences for a general equilibrium setting, where the interest rates are endogenous.

In order to investigate the economic consequences of our different calibrations in a general equilibrium framework, we use the economic model from DICE-2016 together with our newly calibrated versions of “CDICE” to compute the SCC and the optimal mitigation. We find that in our re-calibrated model, when updated with respect to the MMM, the SCC and the optimal mitigation are very similar to the values found in DICE-2016 if we take the exact same economic parameters.⁹ We point out that for the BAU scenario, the temperature increases in DICE-2016 are below those predicted by our extreme calibration, that is, CDICE-HadGEM2-ES. This finding implies that the DICE-2016 predictions are within the range of what is considered plausible in climate science. However, being in the plausible range is merely caused by the fact that two different flaws in the original calibration of the temperature equations and the carbon cycle in DICE-2016 partially offset each other. In particular, the calibration gives the wrong results for the optimal economic response to climate change. In the optimal mitigation scenario, carbon taxes turn out to be too low in the sense that the temperature response of DICE-2016 now falls outside of the CMIP5 range. As a result, the model with an optimal carbon policy predicts a higher temperature than CDICE-HadGEM2-ES.

As pointed out above, the SCC in the DICE-2016 calibration is oversensitive to the discount rate. The role of the rate of time preference of the social planner/representative agent for the SCC and optimal carbon taxes has been extensively documented and discussed in the literature (see, *e.g.* [Barrage 2018](#); [Hänsel and Quaas 2018](#)). While this parameter is obviously important for the SCC, we show that a large part of the quantitative results are driven by an incorrect calibration of the climate module in DICE-2016. Moreover, in a model with time-varying growth rates (as DICE-2016), it is not only the rate of time preference, but also the curvature of cardinal utility that determines the discount rates. We show that this implies that the DICE-2016 climate calibration does not produce reliable results for important comparative statics questions.

The remainder of the paper is organized as follows. In Section 2, we briefly connect our paper to the related literature. In Section 3, we detail our four test cases. Section 4 explains how to re-calibrate the parameters in the climate part of DICE to produce forecasts that are consistent with our test cases. Section 5 explores the implications of different CE calibrations for the SCC in a simple partial equilibrium model. Subsequently, in Section 6, we use the re-calibrated DICE model, that is, CDICE, to study the role of the climate calibration in the economic model of DICE-2016. Section 7 concludes. In addition, for the convenience of the reader, we provide a complete specification of the different variants of the DICE model (some of which we use in our numerical experiments below) in the [online Appendix B](#), including the DICE versions from 2007, 2016, as well as the re-calibration that we suggest in this paper, that is, CDICE.

2. LITERATURE

Starting with Nordhaus’ seminal work on climate change ([Nordhaus 1979](#)), a massive field of research was spawned, including Nordhaus’ own development of the DICE model (see, *e.g.* [Nordhaus 2012, 2013, 2018](#)). For example, [Anderson *et al.* \(2014\)](#) as well as [Miftakhova \(2018\)](#)

9. A significant difference is that the long-run temperature in DICE-2016 is substantially higher, that is, about 1°C, higher than in our calibration. This finding holds both for the business as usual (BAU) scenario without any mitigation and the optimally mitigated scenario.

conducted uncertainty quantification, whereas [Cai and Lontzek \(2019\)](#); [Lemoine and Traeger \(2014\)](#), and [Lontzek *et al.* \(2015\)](#) added stochastic tipping points on top of the DICE-2007 model. [Chang Hwang *et al.* \(2017\)](#) performed learning about equilibrium climate sensitivity with fat tails, whereas [Popp \(2004\)](#) studied endogenous growth in the DICE framework. [Botzen and van den Bergh \(2012\)](#) and [Michaelis and Wirths \(2020\)](#) studied DICE with an alternative damage function. The fact that DICE was criticized for a flawed climate model (in particular by [Dietz *et al.* \(2021\)](#)) puts the quantitative predictions of these contributions into a questionable position and motivates our study. We show below that the climate model in DICE-2016 can be re-calibrated to perform well in our tests.

[Golosov *et al.* \(2014\)](#) developed a simple CE that is at present widely used as an alternative to the climate module of DICE (see, *e.g.* [Hassler and Krusell 2018](#); [Hassler *et al.* 2019](#); [Kotlikoff *et al.* 2021a](#)). Due to its simplicity, the model is computationally much more tractable than DICE. However, the simplicity of this model implies that certain stylized facts of temperature change in response to CO₂ in the atmosphere cannot be captured adequately.¹⁰ This model, therefore, does not lend itself to our calibration strategy.¹¹

More directly related to our paper, there have been several important contributions about the calibration of CEs. [Miftakhova *et al.* \(2020\)](#) proposed a general emulation method for constructing low-dimensional approximations of complex dynamic climate models and developed an emulator for MAGICC to approximate the impact of emissions on global temperature. [Cai *et al.* \(2019\)](#) calibrated a spatial climate system to match four RCP scenarios and historical data. [Traeger \(2014\)](#), [Cai *et al.* \(2012\)](#), and [Cai and Lontzek \(2019\)](#) re-calibrated the DICE CE from a ten-year time-step (as in the DICE-2007 formulation) or five-year time step (as in the DICE-2016 formulation) to higher frequency data. In the climate-science literature, [Calel and Stainforth \(2017\)](#) discussed the physics of different climate models, including DICE. [Thompson \(2018\)](#) conducted an intercomparison study of models, including DICE to assess how they were designed from a physics point of view. [Geoffroy *et al.* \(2013\)](#) showed how two-layer energy balance models (formally identical to the temperature equations in DICE) can be used to match the results in CMIP5. [Joos *et al.* \(2013\)](#) quantified responses to emission pulses in a carbon cycle intercomparison project. Our analysis relies heavily on these two latter papers. In particular, we use [Geoffroy *et al.* \(2013\)](#)'s calibration of the temperature equation and take our carbon-cycle benchmarks from [Joos *et al.* \(2013\)](#). Our paper is closer to [Dietz *et al.* \(2021\)](#), and we compare our contributions in more detail in the following sections.

3. A COMPREHENSIVE FRAMEWORK TO CALIBRATE THE CLIMATE IN IAMs

Naively, one would think that existing historical data suffices for calibrating and evaluating CEs used in IAMs. However, it turns out that, broadly speaking, the signal-to-noise ratio in that sort of data is simply too low for a reliable calibration of CEs with their focus on forcing from CO₂. Put differently, in observed records of annual global mean temperature, the anthropogenic signal due to CO₂ emissions comes in combination with a range of other relevant effects. The change in global mean temperature over the last 150 years is not only due to

10. For example, since in [Golosov *et al.* \(2014\)](#), there is only one temperature equation that links CO₂ in the atmosphere directly to the current temperature, the temperature must decrease as soon as there are net-zero emissions and CO₂ concentration in the atmosphere decreases. We will show in Section 5.1 that the timing of warming plays an important role for the SCC.

11. Interestingly, [Dietz *et al.* \(2021\)](#) argued that although the model in [Golosov *et al.* \(2014\)](#) is very simple, it matches some important stylized facts that can be derived from CMIP5 data better than the original DICE-2016 calibration.

changes in CO₂. It also bears substantial imprints from other greenhouse gases, from aerosols of anthropogenic or natural origins, including volcanoes, and also from land use changes (Gambhir *et al.* 2017; Mengis and Damon Matthews 2020). Reliably disentangling the different contributions is challenging. To illustrate, Millar and Friedlingstein (2018) estimated that from 1880 to 2015, CO₂ contributed about 0.7 degrees to global warming, whereas non-CO₂ agents contributed about 0.3 degrees, with 5% to 95% uncertainty ranging from 0.5 to 1.5 degrees and −0.2 to +0.4 degrees, respectively (cf. their Figure 4). Hence, we need to base our suite of tests on historical observations, but also on the so-called third pillar of science: computational experiments.

As detailed in the introduction, the CMIP experiments are a logical choice for this purpose. We focus on CMIP5 in this paper—CMIP 6 is briefly discussed below. The range of models participating in CMIP then raises the issue that for any specific CMIP experiment (or CE test), one faces not a single benchmark, but a range of benchmarks. Given that the full range is considered plausible by climate scientists, it is important that any CE captures this model diversity in some way. In this light, a first and primary benchmark is the MMM,¹² the average of all cases reported in CMIP5. The MMM is complemented by extreme cases bracketing the CMIP5 range. Two extreme cases that bracket the plausible temperature response to CO₂ in the atmosphere in CMIP5 are the HadGEM2-ES model (strongly warming, ECS of 4.55 K) and the GISS-E2-R model (weakly warming, ECS of 2.15 K) (Geoffroy *et al.* 2013). Two extreme cases that bracket the amount of emitted carbon remaining in the atmosphere are MESMO and LOVECLIM, with about 55% and 30%, respectively, of the 100 GtC test pulse mass remaining in the atmosphere after 100 years, according to Figure 1a from Joos *et al.* (2013).

3.1. Choice of test cases

The choice of the four test cases outlined in the introduction, that is, two for the calibration, and two for the evaluation of the CE, was guided by the following considerations. First, all test cases should be widely used in the relevant climate-science literature and transparently documented. Thus, they can easily be reproduced and compared with published results. Second, some of the tests should be highly idealized to allow for the independent calibration of the carbon cycle and the temperature response, respectively. Third, we want some test cases to be close to the ultimately envisaged *real-world applications*, with gradual changes in CO₂ emissions, exogenous forcing, and associated temperature response. The proximity to real applications makes these tests somewhat like a gold standard, but also makes them more difficult to interpret than the highly idealized tests.

Taken together, the four test cases we propose to use may, loosely speaking, be regarded as *necessary and sufficient* to answer whether an emulator is fit for purpose. They allow the calibration and comprehensive evaluation of a CE like that in DICE, from its individual components to real-application cases. To this end, it is crucial that highly idealized settings, which are frequently used as benchmarks in the literature, are augmented with closer-to-reality tests. We advocate that the same four tests may be used to examine the performance of the climate parts of other IAMs featuring a carbon cycle and a set of temperature equations. Any climate model that fits the CMIP5 data for the four test cases will likely provide a good fit for all emissions paths arising in economic modeling.

12. For a recent application of the MMM in climate science, see, e.g. Beusch *et al.* (2020).

3.2. Technical setup of test cases

The first of four experiments in our test suite is used to calibrate the carbon cycle. Its highly idealized setup follows the temporal evolution of an instantaneous release of 100 GtC (GtC) to the present-day atmosphere. The said 100 GtC roughly corresponds to 10 years of present-day CO₂ emissions from fossil fuels (Quéré *et al.* 2018). The test requires two steps, which we outline here. For more detailed explanations of the test and its benchmark data, we refer to Joos *et al.* (2013). In a first step, the CE is used to simulate the evolution of the climate from its 1850 equilibrium to 2015, using as input historical carbon emissions from 1850 to 2015. The simulation is then continued from 2015 onward, with annual carbon emissions being iteratively determined to keep the atmospheric CO₂ concentration constant at the value of the year 2015. These emissions are then used in a second simulation that starts in 2015, but now with an additional, instantaneous pulse of 100 GtC to the atmosphere. Changes in atmospheric CO₂ are then taken as the difference between this second and the first simulation. The test is used to calibrate the time scales and rates at which the atmospheric CO₂ concentration decays in the wake of instantaneous carbon release to the atmosphere.¹³

The second test case in our proposed framework is used to calibrate the amplitude and time scales of a given emulator's temperature response, that is, how fast and how strongly the global mean temperature warms in response to a sudden quadrupling of CO₂.¹⁴ Starting from pre-industrial equilibrium conditions, that is, in the year 1850, with atmospheric CO₂ at 285 parts per million (ppm), the atmospheric CO₂ concentration is instantaneously quadrupled, and the change in temperature as a function of time is examined. The related benchmark data are anchored in CMIP5 and adopted from Geoffroy *et al.* (2013).

The third test case and its benchmark data come directly from CMIP5: starting in 1850 at pre-industrial levels, the atmospheric CO₂ concentration gradually increases from 285 ppm CO₂ at 1% per year until quadrupling after 140 years. It is used to evaluate the TCR of the CE in comparison with the CMIP5 values of TCR. The test is still relatively simple, as it involves the temperature part of the CE solely, but not the carbon cycle. In other words, it uses CO₂ concentrations as input, not emissions. It involves forcing only from CO₂. The change in CO₂ concentration is not an instantaneous step function, but is gradual in time. The test thus examines a situation where the temperature response time scales overlap with an additional time scale associated with the gradually increasing forcing from CO₂.

The fourth test case is close to what DICE, or some other IAM, is expected to cope with in real applications: the CMIP5 simulations for the historical period and future RCPs for four different forcing levels of 2.6 W/m², 4.5 W/m², 6.0 W/m², and 8.5 W/m² by 2100 (RCP26, RCP45, RCP60, RCP85). The input data consist of either prescribed atmospheric CO₂ concentrations or CO₂ emissions (Meinshausen *et al.* 2011, <http://www.pik-potsdam.de/~mmalte/rcps/>) as illustrated in the left and middle panel of Figure 1. The former case tests only the temperature response to gradually changing CO₂ concentrations. The latter case tests the full climate model of the CE, including the carbon cycle. In this test case, it is necessary to make an assumption on the non-CO₂, exogenous radiative forcing F_t^{EX} , which in the actual CMIP simulations is modeled in great detail. In the online Appendix C, we show how this choice affects the quantitative results of the full model. Here, we assume $F_t^{\text{EX}} = 0.3 \cdot F_t^{\text{CO}_2}$, that is, proportional to the forcing from CO₂ alone ($F_t^{\text{CO}_2}$) and resulting in a total radiative forcing, $F_t = F_t^{\text{CO}_2} + F_t^{\text{EX}}$, which is

13. Depending on the concrete formulation of the CE's carbon cycle, the decay of atmospheric CO₂ may result in increasing CO₂ concentrations in other carbon reservoirs, notably ocean or land.

14. Note that while the test described here applies to any emulator, it will in particular target equations (4) and (5) of the DICE-2016 model (cf. Section 4).

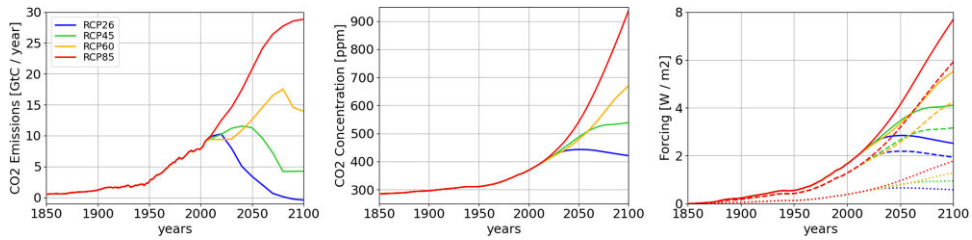


FIGURE 1

Illustration of CMIP5 historical and future evolution (RCP26, RCP45, RCP60, and RCP85 scenarios, shown in blue, green, yellow, and red), from 1850 to 2100, of prescribed CO₂ emissions (left, in GtC per year), alternatively prescribed CO₂ concentrations (middle, in ppm CO₂), and the forcing derived from prescribed CO₂ concentrations (right, in W/m²)

Notes: The total forcing F_t (solid lines) is decomposed into forcing from CO₂ ($F_t^{\text{CO}_2}$, dashed lines) and non-CO₂ forcing assumed as $F_t^{\text{EX}} = 0.3 \cdot F_t^{\text{CO}_2}$ (dotted lines)

30% larger than from CO₂ alone (cf. Section 4.1). This is in line with estimates for the radiative forcing in 2011 with respect to pre-industrial times (see, e.g. [Climate Change 2014 Synthesis Report IPCC 2014](#); [Gambhir et al. 2017](#)) and with estimates for the different RCPs in 2100: 26% (RCP26), 33% (RCP45), 32% (RCP60), 35% (RCP85), where the total forcing comes from the scenario and the CO₂ forcing is computed from concentrations given in [Meinshausen et al. \(2011\)](#); a base concentration of 285 ppm CO₂, and $F_{2\text{XCO}_2} = 3.68 \text{ W/m}^2$. An illustration of the temporal evolution of the different forcings is given in Figure 1, right panel. Note that estimates for the forcing from non-CO₂ GHGs by [MacDougall et al. \(2013\)](#) (their Figure 6) show more temporal structure than F_t^{EX} . Additionally, the assumption of strict proportionality between CO₂ and non-CO₂ forcings is subject to debate in climate sciences, especially in the context of strong mitigation scenarios (see, e.g. [Mengis and Damon Matthews 2020](#)). Benchmark data for the evolution of atmospheric temperature are taken directly from the CMIP5 archive. The proximity to real applications makes this test difficult to interpret, yet it is a key test that CEs in economic models should pass.

3.3. Metric for performance evaluation

The choice of a metric to assess the calibration and subsequent evaluation of a CE is not straightforward. The basic reason lies in the design of the CE, in its reduced functional form, and the limited number of free parameters as compared to ESMs, which capture only some aspects of the benchmark cases used for calibration and evaluation. This also implies that the choice of metric may depend on the concrete form of the CE, the simplicity or complexity of its carbon cycle, and temperature equations. If the CE explicitly covers further elements, like, for example, methane, then not only the metric but also the benchmarks for calibration and evaluation may need adaptation.

Focusing on simple CEs, like that in DICE, one prominent difference with respect to ESMs concerns the response time scales upon perturbation: an ESM features a whole range of response time scales, whereas the CE only a few, for example, two in the case of DICE-2016. Furthermore, an ESM explicitly covers a range of forcings; the CE explicitly covers only CO₂ via its carbon cycle, while any other forcings are included in a more or less ad hoc fashion. The temperature may affect the carbon cycle in an ESM, but not in the simple CE considered here. A comprehensive discussion of why to favor any particular metric in view of these issues is beyond the scope of the present paper. Therefore, we limit ourselves to a short statement of what we did and why.

We calibrate the carbon cycle of a CE via the response to an instantaneous 100 GtC pulse to the atmosphere. The comparison is performed against benchmark data from [Joos *et al.* \(2013\)](#), using the maximum norm over a time range of 100 years after the pulse. This choice of metric favors solutions, that is, calibrations that remain close to the benchmark data at all times over a time scale of interest in an economic context. The calibration of the temperature equations does not require any metric because their functional form is identical to the benchmark data from [Geoffroy *et al.* \(2013\)](#). Therefore, we can just adopt parameter values from the benchmark data.

We evaluate a calibrated CE first with regard to TCR, and we expose the CE to the corresponding test case from climate science, where TCR is defined as the change in global mean temperature after 70 years (and after 140 years) when the atmospheric CO₂ concentration steadily increases from 1850 onward at a rate of 1% per year. We measure the performance of the CE with respect to CMIP5 models by comparing the TCR. Finally, we evaluate the performance of the calibrated CE with regard to climate projections and historical simulation data from CMIP5. Here, we explicitly renounce using a specific metric for mainly two reasons: the CMIP5 models span a range of plausible climates, and they explicitly include non-CO₂ forcings that depend on the projection (RCP26, RCP45, or RCP85) and are included in the CE only in an ad hoc fashion. We consider the CE fit for purpose if a visual inspection shows that its climate in terms of atmospheric CO₂ concentration and the global mean temperature falls within the range of CMIP5 models.

3.4. *CMIP5 serves the purpose, no need for CMIP6*

Throughout this study, we use data from CMIP5 or studies relying on such data, notably [Geoffroy *et al.* \(2013\)](#) or [Joos *et al.* \(2013\)](#). We argue in the following that for the purpose at hand—calibration and testing of a simple CE—the CMIP5 database is sufficient; there is no need to consider the more recent CMIP6 data.

Starting with the calibration of the CE's carbon cycle, we note that the benchmark data from [Joos *et al.* \(2013\)](#) combine models that already span a rather wide range. Data from a more recent study ([Jeltsch-Thömmes and Joos 2020](#)) examining the same test case find a slightly faster decay of atmospheric CO₂ in the wake of the 100 GtC pulse (compare Figure 1 of both papers), yet well within the range of [Joos *et al.* \(2013\)](#). Similarly robust is the benchmark data from [Geoffroy *et al.* \(2013\)](#), used for calibrating the temperature response of the CEs. Using the same two-layer energy balance model and fitting procedure as in [Geoffroy *et al.* \(2013\)](#), but applied to 4xCO₂ data from eight CMIP6 models, [Jackson *et al.* \(2022\)](#) found parameter values in line with [Geoffroy *et al.* \(2013\)](#), the MMM $\pm 1\sigma$ range overlap (compare Tables 3 and 4 in [Geoffroy *et al.* \(2013\)](#) with Table S2 in the supplementary material in [Jackson *et al.* \(2022\)](#)). One may worry that the larger range of the ECS reported in CMIP6, notably the higher ECS of some models (*e.g.* [Zelinka *et al.* \(2020\)](#)), implies a too conservative ECS range in CMIP5 (bracketed here by the models GISS-E2-R and HadGEM2-ES). However, [Tokarska *et al.* \(2020\)](#) showed that the CMIP6 range is reduced considerably, becoming largely consistent again with CMIP5 if observational data are used to constrain the models.

Likewise, we see no necessity to use CMIP6 instead of CMIP5 when testing the calibrated CE against two more test cases. Warming in the “1% per year atmospheric CO₂ increase” experiment is slightly stronger in CMIP6 as compared to CMIP5, the MMM temperature across eleven models being around 10% higher after 140 years ([Fredriksen *et al.* 2023](#), Supplementary Figure S7), while the range of CMIP5 and CMIP6 models strongly overlaps. [Arora *et al.* \(2020\)](#) examined fully coupled simulations of the “1% year atmospheric CO₂” test case, finding a MMM warming after 140 years (quadrupling of atmospheric CO₂ concentration) by 4.87° C in CMIP5 and 4.74° C in CMIP6, that is, a difference of less than 3%.

Testing the CE against future scenarios, we note that a subset of the shared socioeconomic pathways (SSPs) in CMIP6 were designed similarly to the RCPs of CMIP5 in order to provide continuity. Comparing the simulated temperature evolution for corresponding RCPs and SSPs using reduced complexity models, [Nicholls *et al.* \(2020\)](#) found that global mean temperature evolution essentially falls on top of each other (their Figure 3). [Tebaldi *et al.* \(2021\)](#) found that global mean temperature change in the CMIP5 RCPs falls within the range of corresponding SSPs in CMIP6. They elaborated on a number of reasons for the slightly wider range in CMIP6 (including differences, for example, in assumed future aerosol forcing, or methane emissions) and more generally note that CMIP5 RCPs followed a more “middle-of-the-road” pollution policy path. We refer the interested reader to their original work or the paper by [Fredriksen *et al.* \(2023\)](#) for more detailed information on CMIP5 versus CMIP6. With regard to our purpose here—the calibration and testing of a simple CE—the above publications corroborate that no essential changes happened between CMIP5 and CMIP6. We take this as an argument to use the well-prepared CMIP5-based data in [Joos *et al.* \(2013\)](#) and [Geoffroy *et al.* \(2013\)](#) to calibrate a CE’s carbon cycle and temperature response, respectively. Subsequent further testing of the calibrated CE may then be done against either CMIP5 RCP or CMIP6 SSP data.

4. CDICE—RE-CALIBRATING THE CLIMATE OF DICE

So far, we put forward our suite of four test cases: one test to calibrate the carbon cycle of the CE, one test case to calibrate the temperature equation of the CE, and two test cases to examine emerging properties of the calibrated CE, notably its performance under gradually changing conditions in terms of CO₂ and non-CO₂ forcings. We claim that, taken together, these tests are necessary and sufficient to transparently evaluate and, if needed, re-calibrate the CE of an IAM, which is used in an economic context.

In this section, we illustrate our claim with the example of the climate module of the seminal DICE-2016 model. We choose DICE-2016 as an example because it is widely used in the IAM modeling community and because it is known to be flawed ([Dietz *et al.* \(2021\)](#)). We re-examine these flaws in the light of our four test cases and present an improved version, CDICE, that has the same functional form as DICE-2016 but is fully re-calibrated. We go beyond the work of [Dietz *et al.* \(2021\)](#) in three essential points: (i) while [Dietz *et al.* \(2021\)](#) use one test, the 100 GtC pulse, to evaluate the CE, we stress that it is crucial to separately examine each part of the CE, carbon cycle, and temperature equations; (ii) we provide corresponding test setups and show how they can be used to calibrate, and not only to examine a CE; (iii) we demonstrate that even a simple CE like that in DICE is fit for purpose if properly calibrated in this way.

This section is structured as follows: First, we lay down in Section 4.1 the fundamental building blocks of DICE’s CE, its carbon cycle, and temperature equations, along with the commonly used calibration of DICE-2016. Next, we demonstrate in Sections 4.2, 4.3, 4.4, and 4.5 how our battery of tests can be used to systematically test and re-calibrate DICE-2016. Finally, a critical synthesis of our learning is provided in Section 4.6. A summary of all CDICE model parameters is given in Table 1 in [online Appendix A](#).

4.1. Model equations

The climate model in DICE-2016 translates carbon emissions from the model’s economic side into atmospheric CO₂ concentrations (*i.e.* the carbon cycle; cf. Section 4.1) and, subsequently, into a global mean temperature change (*i.e.* the temperature equations; cf. Section 4.1), which then feeds back into the part of the model that represents the economic side of the IAM. Thus, the carbon cycle affects temperature, but not vice versa.

The carbon cycle is a linear three-box model, where the three carbon reservoirs represent the atmosphere (AT), the upper ocean (UO), and the lower ocean (LO), with the respective carbon masses $M = (M^{\text{AT}}, M^{\text{UO}}, M^{\text{LO}})$. Carbon can be exchanged between the atmosphere and upper ocean and between the upper and lower ocean, but not directly between the atmosphere and lower ocean. The global mean temperature is modeled via a system of two ordinary differential equations that couple two heat reservoirs, that is, the atmosphere, including the upper ocean, and the lower ocean, $T = (T^{\text{AT}}, T^{\text{LO}})$. The climate is anchored at an assumed pre-industrial equilibrium state, M_{EQ} and T_{EQ} . Climate change is quantified as deviations from this equilibrium state. In an economic context, time integration in DICE-2016 typically does not start at pre-industrial times, but at present day, for example, in the year 2015. A corresponding initial state M_{INI} and T_{INI} must be determined so as to be in line with carbon emissions since pre-industrial times and the equations of the CE (cf. Section 4.1).

4.1.1. Carbon cycle. The carbon cycle model (see, *e.g.* Keeling 1973) formally reads

$$M_{t+1} = (I + \Delta t \cdot B) \cdot M_t + \Delta t \cdot E_t, \quad (1)$$

with I being the identity matrix, Δt being the time-step in years, and $M_t = (M_t^{\text{AT}}, M_t^{\text{UO}}, M_t^{\text{LO}})$ representing the carbon mass at time t in three reservoirs. The carbon emissions per year to the atmosphere, as well as to the ocean,¹⁵ are specified via E_t . The time-constant matrix B ,

$$B = \begin{pmatrix} b_{11} & b_{21} & b_{31} \\ b_{12} & b_{22} & b_{32} \\ b_{13} & b_{23} & b_{33} \end{pmatrix}, \quad (2)$$

describes the mass transfer among reservoirs and has units “mass fraction per time-step.” Assuming that mass conservation holds, that is,

$$\sum_i b_{ji} = 0 \quad \text{for } j = 1, 2, 3, \quad (3)$$

and that there is no direct mass transfer between AT and LO (implying that $b_{13} = b_{31} = 0$), leaves four free parameters in matrix B that are used to calibrate the carbon cycle. In DICE, the said parameters are chosen as the transfer coefficients from AT to UO (b_{12}) and from UO to LO (b_{23}), as well as the equilibrium carbon mass ratios at pre-industrial times between the reservoirs, $r_1 = M_{\text{EQ}}^{\text{AT}}/M_{\text{EQ}}^{\text{UO}}$ and $r_2 = M_{\text{EQ}}^{\text{UO}}/M_{\text{EQ}}^{\text{LO}}$.

The remaining matrix entries b_{ij} are then given by $b_{11} = -b_{12}$; $b_{21} = b_{12} \cdot r_1$; $b_{22} = -b_{21} - b_{23}$; $b_{32} = b_{23} \cdot r_2$; $b_{33} = -b_{32}$. Parameter values for CDICE are given in Table 1. In DICE-2016, the numerical values for the free parameters are given by $M_{\text{EQ}} = (588, 360, 1,720)$, $r_1 = 1.633$, $r_2 = 0.209$, as well as $b_{12} = 0.12$ and $b_{23} = 0.007$, with a time-step of $\Delta t = 5$ years being hard-wired (absorbed) into the coefficients b_{ij} . To alleviate this limitation of being bound to a time-step of fixed size, the b_{ij} in our formulation are specified in units of 1 year and are then explicitly multiplied with a time-step. The latter thus can be freely chosen, for example, as the time-step prescribed by the respective economic model.

15. This is an interesting option in the context of carbon dioxide removal techniques (see, *e.g.* Rickels *et al.* 2018).

TABLE 1

Values of free parameters in the carbon cycle, original DICE-2016 (with $\Delta_t = 5$ years) and our proposed, different calibrations of CDICE

Model	b_{12}	b_{23}	M_{EQ}
DICE-2016, 5yr	0.12	0.007	(588, 360, 1,720)
CDICE	0.054	0.0082	(607, 489, 1,281)
CDICE-MESMO	0.059	0.0080	(607, 305, 865)
CDICE-LOVECLIM	0.067	0.0095	(607, 600, 1,385)

TABLE 2

Values of the free parameters of the temperature equations for different versions of DICE: original DICE-2016 with $\Delta_t = 5$ years and re-calibrated (CDICE)

Model	c_1	c_3	c_4	ECS	F_{2XCO_2}	λ
DICE-2016, 5yr	0.1005	0.088	0.025	3.1	3.6813	1.19
CDICE	0.137	0.73	0.00689	3.25	3.45	1.06
CDICE-HadGEM2-ES	0.154	0.55	0.00671	4.55	2.95	0.65
CDICE-GISS-E2-R	0.213	1.16	0.00921	2.15	3.65	1.70

Notes: Also given are versions mirroring CMIP5 models with extreme ECS from [Geoffroy et al. \(2013\)](#), which are used for comparison purposes. Note that $\lambda = F_{2XCO_2}/ECS$ is a derived quantity.

4.1.2. Temperature. The two-layer energy balance model in DICE-2016 reads as

$$T_{t+1}^{AT} = T_t^{AT} + \Delta_t \cdot c_1 (F_t - \lambda T_t^{AT} - c_3 (T_t^{AT} - T_t^{OC})), \quad (4)$$

$$T_{t+1}^{OC} = T_t^{OC} + \Delta_t \cdot c_4 (T_t^{AT} - T_t^{OC}), \quad (5)$$

and thus formally corresponds to that described in [Geoffroy et al. \(2013\)](#). T_t^{AT} and T_t^{OC} denote the temperature change with respect to pre-industrial times of the upper layer (atmosphere and upper ocean) as well as the lower layer (deep ocean), respectively, at time-step t . The free parameters c_1 , c_3 , c_4 , and λ in equations (4) and (5) are used for calibration. From a physics perspective, they may be interpreted as a heat exchange coefficient between the upper and lower layers, $\gamma = c_3$, effective heat capacities of the upper and lower layers, $C = 1/c_1$ and $C_0 = \gamma/c_4$, and the radiative feedback parameter $\lambda = F_{2XCO_2}/T_{2XCO_2}$, that is, the ratio of the forcing from a doubling of CO_2 to the associated temperature change, that is, the ECS. Finally, F_t denotes the total radiative forcing from CO_2 , $F_t^{CO_2}$, and other exogenous factors, F_t^{EX} , such as GHGs other than CO_2 and also aerosols

$$F_t = F_{2XCO_2} \frac{\log(M_t^{AT}/M_{EQ}^{AT})}{\log(2)} + F_t^{EX}. \quad (6)$$

In, DICE-2016, F_t^{EX} is assumed to change linearly with time from 0.5 in the year 2015 to 1.0 in 2100. Parameter values for CDICE are given in Table 2. Parameter values in DICE-2016 are $c_1 = 0.1005$, $c_3 = 0.088$, $c_4 = 0.025$, and $\lambda = 3.6813/3.1 = 1.1875$, with a time-step of $\Delta_t = 5$ years formally hard wired (absorbed) into c_1 and c_4 . To lift this issue, we provide, here again, a generic formulation of the temperature equations such that the time-step can be chosen freely in discrete values of 1 year, that is, 1 year, 2 years, and so forth.

4.1.3. Initialization at present day. In the context of economic models, the time integration typically does not start at pre-industrial times but rather at the present day. Corresponding initial values for the carbon and temperature reservoirs used in CDICE for 2015 are $M_{INI} = (851, 628, 1,323)$ and $T_{INI} = (1.10, 0.27)$ for the MMM case (see Section 4.5 for more

TABLE 3

Initial conditions for the year 2015 as used in DICE-2016 and in CDICE, as well as for extreme calibrations of CDICE with respect to the carbon cycle and/or temperature

Model	M_{INI}	T_{INI}
DICE-2016, 5yr	(851, 460, 1,740)	(0.85, 0.0068)
CDICE	(851, 628, 1,323)	(1.10, 0.27)
CDICE-HadGEM2-ES	(851, 628, 1,323)	(1.10, 0.27)
CDICE-GISS-E2-R	(851, 628, 1,323)	(1.10, 0.27)
CDICE-MESMO	(851, 403, 894)	(1.10, 0.27)
CDICE-LOVECLIM	(850, 770, 1,444)	(1.10, 0.27)
CDICE-HadGEM2-ES-MESMO	(851, 403, 894)	(1.10, 0.27)
CDICE-HadGEM2-ES-LOVECLIM	(850, 770, 1,444)	(1.10, 0.27)
CDICE-GISS-E2-R-MESMO	(851, 403, 894)	(1.10, 0.27)
CDICE-GISS-E2-R-LOVECLIM	(850, 770, 1,444)	(1.10, 0.27)

Notes: The atmospheric CO₂ concentration is always 400 ppm (851 GtC). All other values result from integrating the differently calibrated CEs from 1850 to the present day using the CMIP5 setup as described in the fourth test case.

details regarding the MMM and the alternative calibrations). Values for extreme calibrations of CDICE are also given in Table 3.

These initial values should meet three criteria: the atmospheric carbon mass and temperature change should be consistent with observed changes from pre-industrial to present times, and all values should be internally consistent in that they can be obtained by integrating the CE in time from its assumed pre-industrial equilibrium to the present. Regarding the first criterion, the 851 Gt atmospheric carbon corresponds to 400 ppm CO₂, in line with measured concentrations in 2015.¹⁶ The 1.1 K change in global mean temperature since pre-industrial times corresponds to the 2015 centered 11-year average (years 2010 to 2020) observation-based value.¹⁷ We assert the third criterion by integrating the CE from its 1850 equilibrium state to the present day. We use the historical carbon emissions from CMIP5. We slightly adjust the non-CO₂ forcing, by default, at 30% of the CO₂ forcing (see Section 3.2) to meet the warming target of 1.1 K by 2015. Such an adjustment is legitimate given that from a climate science perspective, this forcing comes with substantial uncertainty (see, *e.g.* Mengis and Damon Matthews 2020). From the internally consistent time series of M_t and T_t obtained in this way, we pick that time (year) with the desired atmospheric CO₂ concentration, for example, 400 ppm, as initial conditions, M_{INI} and T_{INI} .

The original initialization values used in DICE-2016 are $M_{INI} = (851, 460, 1,740)$ and $T_{INI} = (0.85, 0.0068)$. The 0.85 K change in global mean temperature tends to be on the low side when compared to observation-based estimates.¹⁸

4.2. Atmospheric CO₂ after 100 GtC pulse to the atmosphere

To evaluate and re-calibrate the carbon cycle of DICE-2016, we use the idealized test case of an instantaneous 100 GtC pulse to the present-day atmosphere, as described in Section 3.2 and illustrated in Figure 2. Compared to benchmark data from Joos *et al.* (2013) (black and gray lines), the known deficiency of DICE-2016 (blue dashed; see, *e.g.* Dietz *et al.* 2021) is apparent: DICE-2016 overestimates the fraction of the pulse remaining in the atmosphere. Retaining

16. <https://www.co2levels.org>.

17. <http://berkeleyearth.org/global-temperature-report-for-2020>.

18. <http://berkeleyearth.org/global-temperature-report-for-2020>.

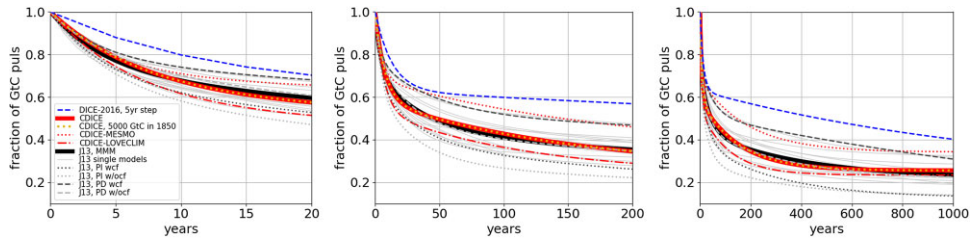


FIGURE 2

Fraction of an instantaneous 100 GtC pulse remaining in the atmosphere (y-axis) as a function of time (x-axis, three different time scales from left to right)

Notes: Shown are benchmark data adapted from Joos *et al.* (2013) (black and gray lines, J13) and three calibrations against this data: CDICE (red solid), which is calibrated against the MMM (black solid), and two extreme calibrations, CDICE-MESMO (red dotted) and CDICE-LOVECLIM (red dash-dotted), which essentially capture the range of benchmark models (thin gray lines). DICE-2016 (blue dashed) removes too little CO₂ from the atmosphere compared to the benchmark data. Absolute values of pulse height and equilibrium masses are irrelevant in DICE's carbon cycle model (CDICE, 5,000 GtC on 1850 equilibrium masses, orange dotted), which is in contrast to findings in J13 (dashed and dotted gray lines, see their paper for details).

the functional form of the carbon cycle, but re-calibrating the model against the MMM benchmark (black solid) using the maximum norm over the first 100 years after the pulse (see Section 3.3), the new best-fit version, CDICE (red solid), performs clearly better. Full agreement between CDICE and the MMM benchmark data cannot be reached because of the different functional forms of the carbon cycle.¹⁹ Extreme calibrations against MESMO and LOVECLIM (CDICE-MESMO, red dotted, and CDICE-LOVECLIM, dash-dotted) are also shown. They are representative of scientifically plausible, but extreme, carbon cycles that remove CO₂ very efficiently (or inefficiently) from the atmosphere, and which we use later on in our economic analysis.

Figure 2 also shows the functional limitation of the carbon cycle model in CDICE (or DICE in general). The fraction of the carbon pulse remaining in the atmosphere is sensitive to neither the pulse height nor the equilibrium state. The 100 GtC test case (red solid line) shows the same decay as a 5,000 GtC pulse to pre-industrial equilibrium (orange dashed line). This is in contrast with informed expectations (see, *e.g.* Joos *et al.* 2013): the pulse should decay more slowly if it is larger, if it is applied to present day instead of pre-industrial conditions, or if climate feedback processes are taken into account (dashed and dotted gray lines, PI and PD indicating pre-industrial and present day, respectively, adapted from Joos *et al.* 2013).

The test case used here is further suited to examine sensitivities to the parameter values of the carbon cycle (colored lines) in Figure 3, anchored at CDICE (see Table 2). As expected, the transfer coefficient b_{12} between the atmosphere and the upper ocean has the largest effect on short time scales. On intermediate time scales, up to about 50 years, assumptions about the equilibrium masses in the upper and lower ocean start to matter. The effect of the transfer coefficient b_{23} starts to have a clear effect only later on. The long-term evolution is equally dependent on b_{23} and the equilibrium masses in the upper and lower ocean.

In this context, the eigenvalues (EV) of matrix B are of interest, as they relate to the decay or the half-life time scales ($\tau = \Delta_t \cdot \ln(0.5) / \ln(EV)$) of an instantaneous perturbation (carbon emission). One EV equals 1 and corresponds to the equilibrium solution. The other two EVs are $(1 + g \pm h)/2$, with $g = 1 - b_{12} \cdot (1 + r_1) - b_{23} \cdot (1 + r_2)$ and $h = ((1 - g)^2 - 4f)^{1/2}$, where

19. The benchmark data in Joos *et al.* (2013) have three characteristic decay time scales, whereas DICE with its three carbon reservoirs has only two.

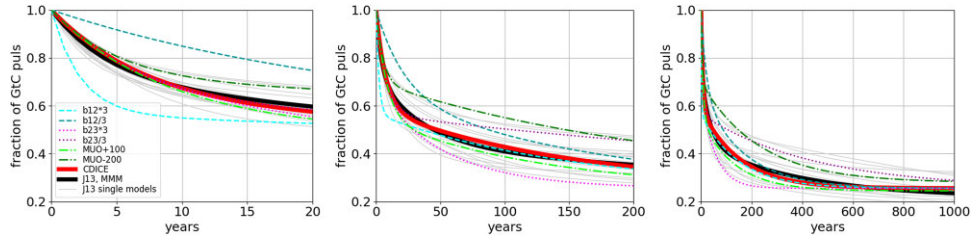


FIGURE 3

Parameter sensitivities (thin colored lines) of the carbon cycle, anchored at CDICE (red), illustrated at the example of an instantaneous 100 GtC pulse to the atmosphere

Notes: Shown is the fraction of the pulse remaining in the atmosphere (y-axis) as a function of time (x-axis, three different time scales from left to right). Benchmark data from Joos *et al.* (2013) are shown in black and gray.

$f = b_{12} \cdot b_{23} \cdot (1 + r_2(1 + r_1))$. Associated numerical values for DICE-2016 are 0.6796 and 0.9959, corresponding to half-life times of 9 and 851 years. For our CDICE calibration, we arrive at EVs of 0.8742 and 0.9933, corresponding to half-life times of 5 and 102 years. This highlights the difference between our carbon-cycle calibration and that in DICE-2016.

4.3. Temperature evolution upon quadrupling of CO_2 concentration

Our second test case, namely the temperature response to an instantaneous quadrupling of the atmospheric CO_2 concentration (cf. Section 3.2), allows for an evaluation and re-calibration of the temperature equations of the CE under consideration. In principle, a metric would again be needed for this purpose (see Section 3.3). However, as CDICE (or DICE in general) and the benchmark data from Geoffroy *et al.* (2013) share the same functional form (cf. Section 4.1), a re-calibration is trivial: we simply adopt the parameter values already fitted to the CMIP5 data by Geoffroy *et al.* (2013). Specifically, we obtain CDICE using their fit to the multi-model mean and two extreme models, that is, CDICE-HadGEM2-ES and CDICE-GISS-E2-R. As explained above, these extreme cases play an important role in our economic analysis below (cf. Sections 5 and 6). Table 2 summarizes the relevant coefficients for these alternative calibrations.

The test confirms previous findings that DICE-2016 warms too slowly compared to any CMIP5 model (dashed blue versus solid gray lines in Figure 4). All the CMIP5 models (solid gray lines) agree that warming occurs rapidly. Meanwhile, their level of warming differs by a factor of roughly two, in line with their differing ECS, given in Table 2. The overly slow warming in DICE-2016 finds confirmation via formal expressions for the characteristic fast and slow response time scales associated with equations (4) and (5) (see Geoffroy *et al.* 2013, their Table 1), which yield numerical values of 4 and 249 years for the CMIP5 multi-model mean in Geoffroy *et al.* (2013), but 40 and 219 years for DICE-2016.

The test discussed here is further suited to shed some light on the role of the different parameter values in the temperature equations of CDICE, as illustrated in Figure 5. How quickly the temperature responds to an increase in atmospheric CO_2 is mainly governed by the heat capacity of the upper ocean via the parameter c_1 (left panel, yellow and brown dashed lines), which, in turn, does not affect the mid- and long-term absolute warming (beyond roughly 20 years, middle and right panel). As expected, the very long-term warming is set by the ECS (right panel, pink and violet dash-dotted lines). The track toward long-term equilibrium across intermediate time scales and levels of warming depends on the energy exchange between the upper and lower ocean, via the choice of c_3 (solid green lines, after roughly four years), and on the heat capacity of the deep ocean, entering in the form of c_4 (cyan dotted lines, after roughly 20 years).

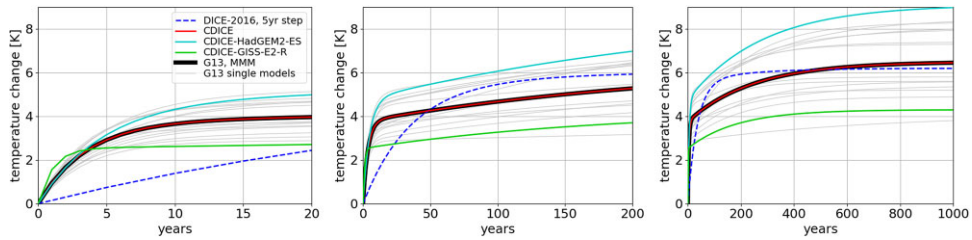


FIGURE 4

Temperature response to instantaneous quadrupling of atmospheric CO_2 with respect to pre-industrial values on time scales of 20 years (left), 200 years (middle), and 1,000 years (right)

Notes: Parameters in CDICE (red solid) can be chosen such as to exactly reproduce the calibration target, the CMIP5 multi-model mean (black solid, G13-MMM). Also shown are two extreme ECS calibrations (solid green and light blue, CDICE-HadGEM2-ES and CDICE-GISS-E2-R). In DICE-2016 (dashed blue, 5-year time-step), temperature increase is too slow compared to any CMIP5 model (thin gray lines, G13, from Geoffroy *et al.* 2013).

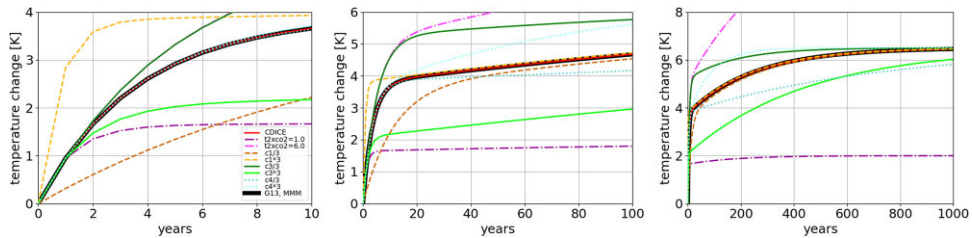


FIGURE 5

Parameter sensitivity of the temperature equations in DICE, illustrated at the example of an instantaneous quadrupling of the CO_2 concentration with respect to pre-industrial values

Notes: Shown is warming (y-axis) as a function of time (x-axis) for different time scales (from left to right). Parameters are varied one by one with respect to the CMIP5 multi-model mean from Geoffroy *et al.* (2013) (black solid, G13-MMM) and its adaptation in CDICE (red solid).

In summary, the test case demonstrates that the temperature equations in DICE-2016 can be re-calibrated to reach general agreement with corresponding results from CMIP5. Specific CMIP5 models or the multi-model mean as given in Geoffroy *et al.* (2013) can be recovered precisely by appropriate parameter choice.

4.4. Temperature evolution as atmospheric CO_2 increases at 1% per year

The purpose of the third test is to evaluate a CE's TCR. In contrast, the first two test cases primarily aim at calibrating a CE. The TCR is of interest as a measure of how the (model) climate responds to steadily changing conditions. From the point of view of a CE, one may also say that the TCR examines the CE's performance upon a series of perturbations; not a single one as in the tests used for calibration of the CE. Although still highly idealized, this test is thus closer to reality. Consequently, TCR is often considered a better measure to evaluate the performance of a climate model than ECS, which assumes the climate system to have equilibrated. A reasonable TCR is a necessary condition for any CE but, as we shall see, not a sufficient one.

Of the test results shown in Figure 6, two aspects deserve highlighting. First, all variants of DICE (shown in different color coding) are compatible with the CMIP5 data (black and gray lines). The temperature evolution of DICE-2016 (blue dashed lines) now lies within the range of CMIP5 models (solid gray lines). However, the warming is still slightly slower than the CMIP5 multi-model mean (black solid) during the first few decades. CDICE (solid red) warms slightly

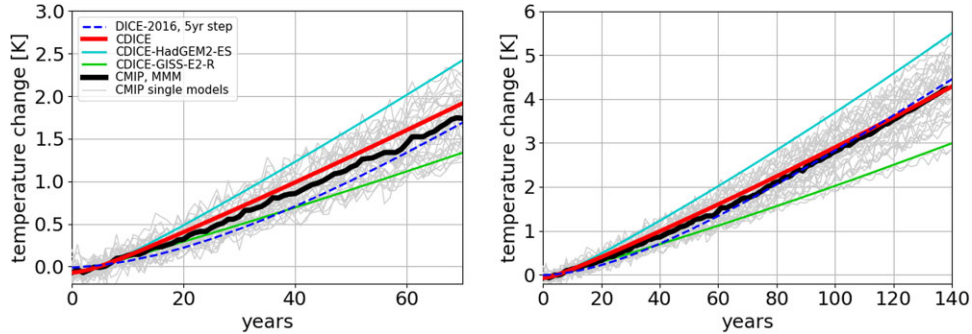


FIGURE 6

Temperature response (y-axis) to transient, 1% per year, increase of the CO_2 concentration as a function of years (x-axis) from 1850 onward

Notes: The CO_2 concentration doubles within 70 years (left panel), and associated warming levels correspond, by definition, to the TCR of the model. CO_2 quadruples within 140 years (right panel), and TCR roughly doubles. DICE-2016 (dashed blue), our new version CDICE (solid red), as well as its extreme scaling variants (solid cyan and green), are all compatible with CMIP5 models and their TCR (individual models and multi-model mean, solid gray and black, respectively).

faster than the CMIP5 multi-model mean. Variants of CDICE with extreme ECS (green and cyan lines, see Table 2 for parameters) indeed bracket the individual CMIP5 models. Second, the transient warming of the models remains clearly below their ECS, even after the quadrupling of CO_2 after 140 years. The TCR of the models is between 1.3 K and 2.3 K across CMIP5 models (gray lines in Figure 6). Associated values for ECS are roughly 50% to 100% larger, ranging between 2.05 K and 4.55 K (see Table 2). Upon the quadrupling of CO_2 after 140 years, ECS still exceeds the transient temperature change by about 40% to 70%.

Contrasting the above findings with those from Section 4.3 demonstrates that the model performance is highly sensitive to the concrete test. In particular, the striking under-performance of DICE-2016 in the idealized CO_2 quadrupling test case is hardly detectable if CO_2 increases gradually at a rate of 1% per year instead. The decent performance in terms of TCR may seem to come as a relief given the ample amount of existing literature using DICE-2016. This impression is, however, deceptive, as will be further elaborated on in Section 6.

4.5. CMIP5 historical and RCP evolution as simulated by DICE

Our fourth test was chosen to be as close as possible to real applications, that is, a necessary gold standard test any CE should pass: the CMIP5 simulations for historical and future representative concentration pathways (RCP26, RCP45, RCP60, and RCP85, from top to bottom in Figure 7). Compared to the other three tests, this test features some additional hurdles. First, the entire CE is examined, that is, the carbon cycle combined with the temperature equations. Second, strong mitigation scenarios are considered that test a CE's ability to capture a regime transition from, basically, emission-dominated to response-dominated. Third, non- CO_2 forcings must be taken into account. They form part of the CMIP5 simulations and the real-world alike and thus must be incorporated and examined in the context of any CE. To facilitate the interpretation of results, we separately examine in Figure 7 CO_2 concentrations due to changing carbon emissions (left column, performance of carbon cycle), the temperature in response to changing CO_2 concentrations (middle column, performance of temperature equations), and finally, temperature as a function of carbon emissions (right column, performance of full CE).

The CMIP5 data (gray and black lines in Figure 7) provide a benchmark range: depending on the model, warming by 2100 is between 1 K and 2.5 K for RCP26 and between

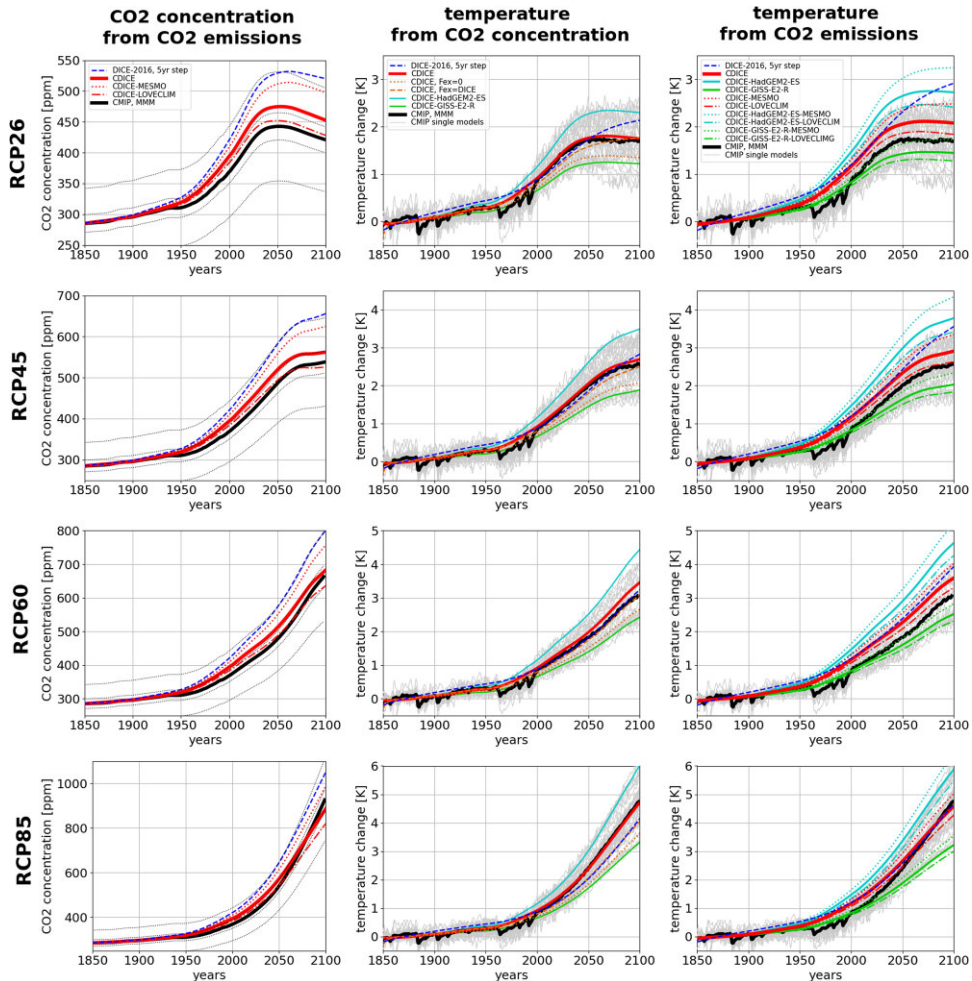


FIGURE 7

Comparison of DICE with historical and future (years 1850 to 2100, scenarios RCP26, RCP45, RCP60, and RCP85, from top to bottom) CMIP5 data

Notes: Shown are atmospheric CO₂ concentrations (left) as prescribed in CMIP5 (solid black, with dotted lines indicating $\pm 5\%$ and $\pm 20\%$ ranges) and as computed from CO₂ emissions with DICE-2016 (dashed blue), CDICE (solid red), as well as CDICE-MESMO and CDICE-LOVECLIM (dotted and dash-dotted red). Also shown is temperature evolution based on prescribed CO₂ concentrations (middle) and based on emissions (right) for DICE-2016, CDICE, and variants of CDICE.

3 K and over 6 K for RCP85. The range exists, although the models feed on prescribed concentrations, not emissions. A subset of CMIP5 models recomputed the RCP85 scenario based on prescribed emissions. By 2100, their CO₂ concentrations lie in a range from 800 ppm to 1,150 ppm, or within about 20% of the CMIP5 prescribed value of 935 ppm, and warming ranges roughly from 2.7 K to 6.5 K (Figure 12.36 in the [Intergovernmental Panel on Climate Change and Intergovernmental Panel on Climate Change 2015](#)). The episodic cooling apparent in Figure 7, which is due to significant volcanic eruptions such as Mt. Pinatubo in 1991, is a reminder that the temperature evolution in CMIP5 encapsulates a whole range of forcings. The decreasing CO₂ concentrations and temperatures in RCP26

are of particular interest as they allow us to benchmark the CE in the context of mitigation scenarios.

Looking at the carbon cycle (left column in Figure 7), we find that CO₂ concentrations modeled by the DICE family are mostly within 20% of those prescribed by CMIP5 (black dotted lines), thus within the range of emission-based CMIP5 simulations (see above). Concentrations are always the highest for DICE-2016, followed by CDICE-MESMO, CDICE, and CDICE-LOVECLIM, in line with calibration (CDICE family) and evaluation (DICE-2016) against the 100 GtC pulse test (see Section 4.2). Looking at the different scenarios, the strong mitigation scenario RCP26 has concentrations prescribed in CMIP5 (black solid) that are lower than in any model of the DICE family. The situation changes as one moves to RCP45 and RCP85, where toward the end of the century (the year 2100), concentrations in CDICE are even lower than those prescribed in CMIP5. The behavior may indicate a dependence of the carbon cycle on the scenario (strong mitigation versus continued strong emission) either on the side of the DICE model family or on the side of CMIP5. A dedicated analysis is beyond the scope of the present paper.

Turning to the temperature response to CO₂ concentrations as prescribed in CMIP5 (middle column), the curves for the different versions of CDICE echo corresponding calibrations (see Section 4.3). In particular, CDICE (red solid) falls on top of the CMIP5 multi-model mean (black solid) independent of the RCP scenario considered. This is remarkable in several respects. The temperature equations in CDICE were calibrated against one perturbation (instantaneous quadrupling of CO₂), whereas CO₂ keeps constantly changing here. The temperature evolution is well captured even in RCP26 after 2050, when CO₂ concentrations decrease again. Finally, the data by Geoffroy *et al.* (2013), used to calibrate CDICE, incorporates only a subset of CMIP5 models.²⁰ DICE-2016 (blue dashed line) features relatively high temperatures under strong mitigation (RCP26) and rather low temperatures in the strong forcing RCP85 scenario. One reason could be that the slow temperature response of DICE-2016 to an instantaneous increase in CO₂ concentrations turns into a lagging response upon gradually changing CO₂. Another reason could be the non-CO₂ forcing (see Section 4.1), which may be too strong in the RCP26 scenario, but too weak in the RCP85 scenario. For CDICE, the (assumed) non-CO₂ forcing (dotted orange line) contributes around 0.5- to 1-degree warming by 2100.

The performance of the entire CE (Figure 7, right column) encapsulates the features just described. Some combinations of the independently calibrated carbon cycle and temperature equations fall outside the range of CMIP5 models (thin gray lines). For example, CDICE-HadGEM2-ES-MESMO and CDICE-HadGEM2-ES are too warm in RCP26, whereas CDICE-GISS-E2-R-LOVECLIME tends to be on the cool side in RCP85. The relative importance of the carbon cycle and temperature response depends on the forcing scenario. Upon strong forcing, like RCP85, temperatures differ mainly because of the different ECS calibrations (cyan, red, and green curves). In contrast, the calibration of the carbon cycle plays a minor role (dotted, solid, and dash-dotted curves). For RCP26, CEs that combine high ECS with efficient removal of carbon emissions (CDICE-HadGEM2-ES-LOVECLIM, cyan dash-dotted) have a similar temperature evolution as CEs combining inefficient carbon removal with an intermediate ECS (CDICE-MESMO, red dotted). In Section 6, we examine the degree to which this degeneracy persists within an economic context.

In summary, CDICE falls within CMIP5 models for all future scenarios. This is not the case for some variants of CDICE that combine extreme calibrations of the carbon cycle and temperature equations. DICE-2016 performs better for the strong forcing scenario RCP85 than

20. Only a subset of all CMIP5 data was available when Geoffroy *et al.* (2013) performed their analysis.

for the strong mitigation scenario RCP26. Concrete assumptions about the non-CO₂ forcing term, notably its strength, are demonstrated to matter for global mean temperature change.

4.6. *Synthesis and lessons learned from the climate test cases*

The results presented in this section show that it is important to independently calibrate the two basic building blocks of any CE, that is, the carbon cycle and the temperature equations. In this paper, we provide two corresponding, highly idealized test cases that are based on published data from CMIP5. We advocate that the calibrated CE should, subsequently, be exposed to two additional tests: one to quantify the CE's transient climate response and the other to evaluate the CE's performance for realistic future scenarios, under strong forcing as well as under strong mitigation. Associated tests are described that rely only on publicly accessible data from CMIP5. We illustrated that DICE-2016, although failing both calibration tests, passes the two additional tests, possibly due to compensating errors. Additionally, we demonstrated that the functional form of the DICE family is fit for purpose: a re-calibrated version, CDICE, passes all four tests.

The CMIP5 test cases highlight the existence of a target range for both the carbon cycle and the temperature equations. Full-fledged ESMs do not agree on global mean temperature change, even if this change is only due to prescribed gradual changes in atmospheric CO₂ concentration, as in the 1% per year CO₂ increase test case. Variants of CDICE using calibrations for the temperature response that are extreme yet in line with CMIP5 have been shown to bracket the range of CMIP5 models for all four RCPs examined. We advocate exploiting this bracketing behavior in the context of economic studies as a measure of the uncertainty arising from the climate part. The same applies with regard to the carbon cycle. We caution, however, that combining extreme calibrations for both the carbon cycle and the temperature equations may result in an overly extreme climate of the CE.

Non-CO₂ forcings have been shown to play a prominent role and, as such, constitute another source of uncertainty for any CE. Their assumed linear increase with time in DICE-2016 has the advantage of being simple. Climate literature rather advocates that non-CO₂ forcings amount to about one-third of the forcing from CO₂ (cf. Section 3.2); the form adopted in CDICE. It is not obvious whether any of the two forms is clearly superior with regard to the envisaged applications; studying the interplay of climate, society, economy, and technology. Net-zero targets, mitigation scenarios, or scenarios presenting even more heavy use of fossil fuels may affect the mix of non-CO₂ forcing agents and associated lifetimes. A more detailed description of such developments is not desirable as it would necessitate including non-CO₂ agents, reservoirs, and processes, thereby violating the deliberately simple approach to climate taken in DICE. Retaining a simple form, but testing for associated sensitivities, seems a more promising avenue.

Focusing again on CO₂ forcings alone, the temperature part of CDICE has been shown to perform very well with respect to any benchmark data for prescribed CO₂ concentrations. Working with carbon emissions has turned out to be more challenging. It is difficult for CDICE to reproduce both high-emission and mitigation scenarios, like RCP85 and RCP26 (cf. Figure 7, left column) equally well. The point is relevant given that DICE should be applicable to a broad range of socio-economic scenarios, from fossil fuel dominated to mitigation and net-zero emissions. We speculate that the issue is rooted at least in part in the availability of only two time scales (of 5 and 102 years, see Section 4.1) related to the three reservoir carbon cycle.

A reservoir-based carbon cycle model has the advantage that the total amount of carbon is preserved. In particular, carbon leaving the atmospheric reservoir is still present in another reservoir, notably the ocean, in the case of DICE. From there, it may continue to play a role for the climate system at some point in time. This capability of a reservoir-based approach is especially interesting with regard to strong mitigation scenarios, where the amount of carbon currently

being stored in the (upper) ocean is likely to affect the carbon uptake capability of the ocean as emissions decline (Ridge and McKinley 2021). For the same reason, it seems questionable whether even simpler climate models without any reservoirs—an example of such a model could be the approximately linear relationship between cumulative carbon emissions and global mean temperature—would be able to properly cope with strong mitigation scenarios.

In summary, the emission-based global mean temperature evolution from 1850 to 2100 as modeled by either DICE-2016 or CDICE lies, despite the models' simple functional form, well within the range of CMIP5 results for most RCPs. For DICE-2016, this finding is noteworthy as the model, in contrast to CDICE, clearly fails more idealized tests (instantaneous quadrupling of CO₂ and 100 GtC pulse). Moreover, it is deceiving, as compensating errors play an important role in DICE-2016. Noteworthy also is the challenge posed to either DICE-2016 or CDICE by the strong mitigation scenario RCP26: CDICE struggles, and DICE-2016 clearly fails and clearly warms far too much toward 2100. As already noted by Traeger (2014), such strong mitigation scenarios excessively challenge the models' functional form.

5. THE SOCIAL COST OF CARBON IN PARTIAL EQUILIBRIUM

In standard economic models, climate change is treated as an externality, and, going back all the way to the work of Pigou (1920), it is well known that Pareto efficiency can be restored by adding the marginal value of the externality to the market price of carbon. To reach this goal, IAMs compute the net present value of all future damages caused by the emission of an additional unit of carbon at some date t . The effect of these extra emissions on damages depends on the calibration of the damage function and the path of emissions. Computing the net present value of the damages requires knowledge of all future interest rates. A change in the market price of carbon will change the path of emissions and of interest rates. Consequently, one has to impose an economic model with many assumptions to obtain an accurate estimate of the SCC. However, one can gain important insights by examining the future path of damages caused by a pulse of emissions for a given RCP emissions scenario. One can then assume constant interest and growth rates and make general statements about the SCC. As we will show in Section 5.1, in the BAU case of the DICE-2016 model, the effects of the climate model on emissions and the interest rates are indeed tiny (of course, the effects on optimal abatement and mitigated emissions are very large). The social cost of carbon is defined for each date t . As it turns out, if one assumes constant interest rates, the SCC relative to output is relatively constant across the next 50 years, and we, therefore, focus on the somewhat arbitrary date 2020 for reporting the SCC in this section.

In this, and the subsequent Section 6, we assume that exogenous forcing, F_t^{EX} , changes linearly with time from 0.5 in the year 2015 to 1.0 in 2100, that is, as specified in DICE-2016 (cf. equation (6)). This measure allows us to compare the impact of our CDICE calibration with DICE-2016 consistently. However, to be fully in line with state-of-the-art climate science, it might be better to assume $F_t^{\text{EX}} = 0.3 \cdot F_t^{\text{CO}_2}$ (see, e.g. Climate Change 2014 Synthesis Report IPCC 2014; Gambhir *et al.* 2017). In the online Appendix C, we show how such change in the full model affects our quantitative results.

Following the majority of the economic literature on climate change, we assume that the temperature influences the economic activity via a damage function $\Omega(\cdot)$ that multiplies total output $Y_t^{\text{Gross}}(\cdot)$ at time t . Given a constant interest rate r , and a constant growth rate of output, g , the present value of damages at some future date t is given by $(\frac{1+g}{1+r})^t \Omega(T_{AT,t})$. In this section, we take $\frac{1+r}{1+g}$ as given, and refer to it simply as the “ g -adjusted interest rate.” Note that output growth can come from population growth as well as total factor productivity (TFP) growth, and

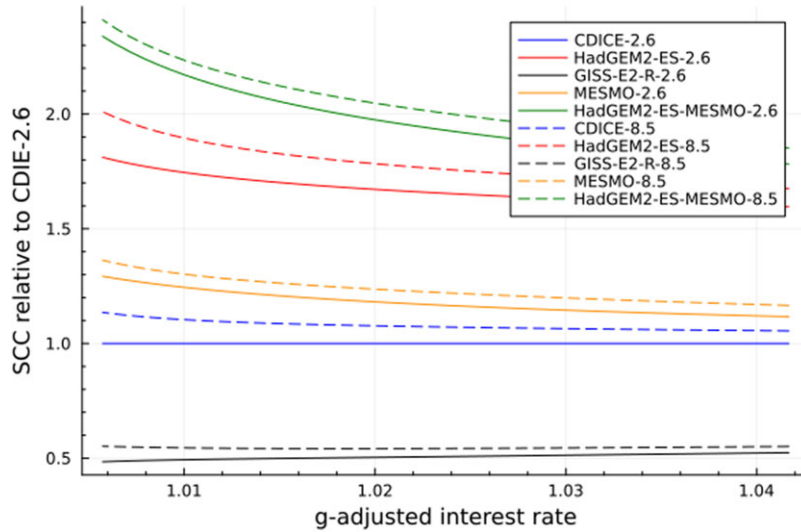


FIGURE 8

SCC for different CDICE calibrations, emission scenarios, and a quadratic damage function

it is likely not to be constant over the next 300 years. Real risk-free rates and returns to capital certainly vary over the business cycle, but also might change in the long term (see, *e.g.* [Bauer and Rudebusch 2021](#)).

The damage functions used in DICE-2016, but also in [Hänsel *et al.* \(2020\)](#) and [Howard and Sterner \(2017\)](#) among others, all assume the following quadratic functional form:

$$\Omega(T_{AT,t}) = \psi_2 \cdot (T_{AT,t})^2. \quad (7)$$

Since, in this section, we only consider the SCC relative to the CDICE calibration (both the temperature equations as well as the carbon cycle calibrated to the MMM), the value of the parameter ψ_2 is irrelevant. Of course, ψ_2 matters for the absolute value of the SCC and for optimal abatement, but not for ratios of the SCC for different calibrations. However, as we show, the functional form is very important, and higher-order terms (as, for instance, discussed in [Weitzman 2012](#)) can make a significant difference in the effect of climate model uncertainty on the SCC.

5.1. Different climate calibrations and the SCC

Figure 8 displays the SCC under two different RCP emission scenarios, that is, RCP85 with very high emissions and RCP26 with very low emissions, and four different calibrations of our CE. In all the cases in the figure, the social cost of carbon is relative to the CDICE MMM social cost of carbon for RCP26. We consider the MMM for both the temperature equations and the carbon cycle (denoted by CDICE). Furthermore, we also look at HadGEM2-ES and GISS-E2-R as alternative temperature calibrations, as well as MESMO as an alternative carbon-cycle calibration. As we vary the growth-adjusted interest rate from 0.5% to 4%, the SCC for the RCP 2.6 scenario with the basic CDICE model (both temperature equation and carbon cycle calibrated to the MMM) is always normalized to one. All models considered in this figure are a calibration of CDICE; for simplicity, the term CDICE is often omitted in the legends in Figures 8 and 9.

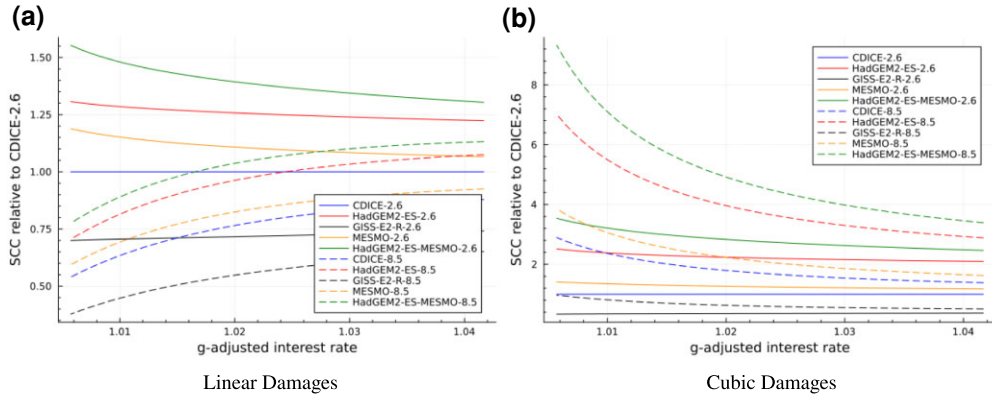


FIGURE 9

SCC for alternative damage functions: (a) linear damages and (b) cubic damages

There are four key takeaways from this stylized exercise. First, model uncertainty for the temperature equation can lead to three to five times higher SCC for very pessimistic scenarios as opposed to the most optimistic scenario. Second, the effect of an extreme carbon cycle is much smaller compared to the temperature block (it increases the SCC by about 20% relative to the MMM), but the combined effect of a pessimistic carbon cycle and a pessimistic temperature scenario can be very large. Third, the relative uncertainty introduced by climate model uncertainty does not change very much with discounting, at least for g-adjusted interest rates above 2%. Fourth, the SCC does not change very much with an alternative to the emissions scenario. It is uniformly higher when the emissions are very large, but relative to the effect of the climate model uncertainty and the discount factor, the changes are small.

It is important to note that the last two observations depend crucially on the assumption that damages are quadratic (cf. equation (7)). Typically, for any strictly convex cost function, one would expect that marginal costs increase in quantity and that, therefore, the SCC is substantially higher for RCP85 than for RCP26. However, the marginal costs of emissions also include the non-linear effects of emissions on the temperature. The concavity in the forcing equation (6) counters the convexity of damages, and thus, the overall shape of the cost function is an unknown a priori.

This point is further illustrated in Figure 9, where we show the analog of Figure 8 for a linear damage function (left panel) as well as a cubic damage function of the form $\Omega(T_{AT,t}) = \psi_3 \cdot (T_{AT,t})^3$ (right panel).

For linear damages, the marginal costs decrease in emissions. For high-emission scenarios, the SCC is lower than for low-emission scenarios. The overall effect of climate model uncertainty is significantly smaller than in the quadratic cost case. For the cubic damages, on the other hand, marginal costs clearly increase. For high-emission scenarios, the SCC is much higher. The effect of model uncertainty, in this case, is huge. Depending on the assumed interest rate, the SCC for the most pessimistic climate scenario can be almost eight times the SCC for the most optimistic scenario.

5.2. The effect of miscalibration

How important are the exact details of our calibration for the SCC? What happens, for instance, if, instead of using the CDICE-HadGem2-ES calibration, we just change the ECS parameter

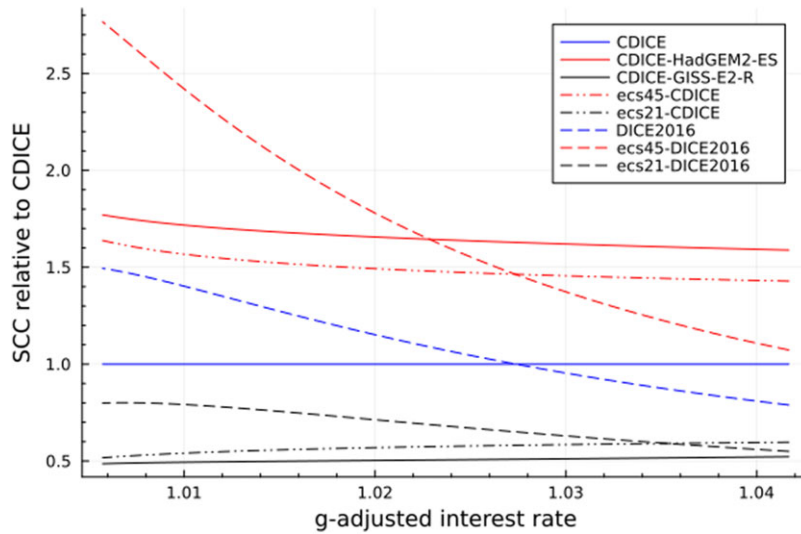


FIGURE 10

SCC for different calibrations. RCP8.5 and quadratic damage function

in equation (4) to 2.15 or to 4.55 without changing any other parameters? How does the SCC for the original DICE-2016 calibration differ from the SCC in our calibration? Figure 10 shows how the SCC changes with various calibrations of the CE. We focus here on the RCP 8.5 scenario with quadratic damages and compute the SCC at $t = 5$ (the year 2020) relative to the CDICE calibration. Figure 10 shows the SCC for CDICE-HadGEM2-ES and CDICE-GISS-E2-R as alternative temperature calibrations and compares them to simply setting the ECS to the associated values (4.55 in CDICE-HadGEM2-ES and 2.15 in CDICE-GISS-E2-R), but leaving the CDICE calibration unchanged otherwise (cf. the dot-dashed lines labeled “ecs45-CDICE” and “ecs21-CDICE”). This is how uncertainty about the ECS is typically treated in economic models of climate change (see, e.g. Nordhaus 2018 or Hassler and Krusell 2018). We observe that even within the CDICE calibration, just changing the ECS is not a very good proxy for the extreme warming scenarios in CMIP5. In particular, the SCC for an ECS of 2.15 is about 15% too large relative to GISS-E2-R, and the SCC for an ECS of 4.55 is about 10% too small relative to CDICE-HadGEM2-ES. This indicates that for a quantitative treatment of climate model uncertainty, it is not sufficient to simply vary the ECS parameter in the CE. As explained above, the ECS is not a free parameter in Earth system models, but emerges as one characteristic of simulated climate change describing the very long-run behavior of temperature. For the social cost of carbon, this very long run is largely irrelevant, and the speed of warming is often more important. Modeling the climate uncertainty by simply varying the ECS parameters in the climate emulator incorrectly reduces the uncertainty about the SCC by a significant amount.

Figure 10 also shows the ECS for the DICE-2016 calibration, the MMM, as well as calibrations with an ECS of 2.15 and with 4.55. The differences with respect to CDICE are apparent: For intermediate values of the discount factor (say 0.97 or so), the SCC for CDICE and DICE-2016 is very similar, both for the MMM and for the calibrations with ECS of 2.15 and 4.55, respectively. In the case of a discount factor of 0.96, DICE-2016 underestimates the SCC substantially and is about 20% lower than for CDICE. For high values of this parameter, DICE-2016 produces much higher values for the SCC, for a value of 0.99, about 30% larger than CDICE.

6. THE SOCIAL COST OF CARBON AND OPTIMAL ABATEMENT IN THE DICE ECONOMY

In this section, we present the optimal solutions for the adjusted DICE-2016 model, which features both the calibration of the economic part as presented in DICE-2016 (Nordhaus 2018) merged with the re-calibrated CDICE climate part (cf. Section 4). We will, in the following, also refer to this complete IAM as CDICE. A comprehensive summary of the complete calibration of CDICE is available in the [online Appendix B](#). We will refer to the original calibration of the DICE model as DICE-2016.²¹

The key non-climate part of the DICE-2016 model consists of a single, infinitely lived, representative consumer and a single firm. As it is standard in economics the equilibrium allocation can be described as the solution to a planner's problem (see, *e.g.* Golosov *et al.* 2014 for more details, and Kotlikoff *et al.* 2021b for a critique of this approach). The planner/representative agent maximizes a time-separable utility function over (per capita) consumption $(\frac{C_t}{L_t})_{t=0}^{\infty}$ with a constant intertemporal elasticity of substitution (IES), $\psi > 0$, and a time preference parameter, $0 < \beta < 1$.²² The optimal value, V_0 is given by the following expression:

$$V_0 = \max_{\{C_t, \mu_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t \frac{\left(\frac{C_t}{L_t}\right)^{1-1/\psi} - 1}{1 - 1/\psi} L_t \quad (8)$$

$$\text{s.t. } K_{t+1} = (1 - \Theta(\mu_t) - \Omega(T_{AT,t})) K_t^\alpha (A_t L_t)^{1-\alpha} + (1 - \delta)K_t - C_t \quad (9)$$

Equations (1), (4), (5)

$$0 \leq K_{t+1} \quad (10)$$

$$0 \leq \mu_t \leq 1 \quad (11)$$

$$\text{where } E_t = \sigma_t Y_t^{\text{Gross}} (1 - \mu_t) + E_t^{\text{Land}}. \quad (12)$$

The CO₂ emissions, denoted by E_t , consist of non-industrial emissions, E_t^{Land} , as well as industrial emissions that are modeled as a fraction of output, $\sigma_t Y_t^{\text{Gross}}$, with σ_t being emission intensity and $\mu_t \geq 0$ being mitigation. Output is produced in a Cobb–Douglas technology with capital, K_t , and labor, L_t . Mitigation is costly and decreases output at a rate $\Theta(\mu_t)$. Higher temperatures decrease output at a damage $\Omega(T_t^{\text{AT}})$. For a detailed specification and parametrization of all the equations including exogenous variables and damages, please refer to the [online Appendix B](#).

First, the planner solves the maximization problem stated above without understanding that higher mitigation μ_t leads to lower damages from a temperature increase. In the BAU scenario, she only chooses an investment path, and mitigation is set to zero. Then, she solves the problem optimally by choosing mitigation as well as an investment path. In both cases, the SCC is the marginal cost of atmospheric carbon in terms of the numeraire good. Following the literature (see, *e.g.* Traeger 2014; Cai and Lontzek 2019), we can write the SCC as the planner's marginal rate of substitution between the atmospheric carbon concentration and the capital stock.

21. Note that the calibration for the original DICE-2016 stems from the publicly available GAMS code specification. For details, see <http://www.econ.yale.edu/~nordhaus/homepage/homepage/DICE2016R-091916ap.gms>.

22. In the [online Appendix B](#), we list all exogenous variables, equations, and parameters of DICE-2016 and CDICE.

Thus, we have:²³

$$SCC_t = -\frac{\partial V_t / \partial M_{AT,t}}{\partial V_t / \partial K_t}. \quad (13)$$

The optimal carbon tax (CT_t) is the tax that equates the private and the social cost of carbon. Nordhaus (2018), among others, defined the optimal carbon tax as a function of mitigation μ_t . The social planner chooses the mitigation μ_t , which is equivalent to choosing the carbon tax in units [USD/tC]²⁴, that is,

$$CT_t = \frac{\theta_{1,t}\theta_2\mu_t^{\theta_2-1}}{\sigma_t}, \quad (14)$$

where $\theta_{1,t}$ is the abatement cost, and σ_t is emission intensity.²⁵

By definition, the SCC is equal to the optimal carbon tax if $\mu_t < 1$.

In the following, we first consider the CDICE model under different CMIP5 calibrations for the carbon cycle and the temperature equation. This is to illustrate that different reasonable calibrations for the CE can have large effects on the optimal abatement. Second, we compare CDICE to DICE-2016 to show some economic consequences of the incorrect calibration in DICE-2016. We do so in two steps. First, we compare the predictions of the two calibrations with the exact economic calibration from DICE-2016. Second, we also investigate the effects of differences in the IES-parameter ψ , which can have large effects on future g-adjusted interest rates.

6.1. CDICE—economic consequences of climate model uncertainty

We consider the economic model of DICE-2016 combined with our nine different reasonable calibrations for carbon-cycle and temperature equations. We explore the economic consequences of the large uncertainty in our climate model.²⁶

We take as a starting date ($t = 0$) the year 2015. In addition, the initial conditions are listed in Table 3.

We start our analysis by examining the predictions of the model in a scenario without mitigation. The first observation is that in this framework, the emissions are not sensitive to the climate calibration used. Figure 11(a) shows the path of emissions for our nine versions of CDICE and compares it to the emissions in RCP 85 and RCP 6. The assumed climate parameters have an effect on BAU emissions via differences in damages (higher temperatures imply higher damages that, in turn, imply less carbon emissions in this model), but the effect is quantitatively insignificant. One reason for this behavior is that the damages only become large once emissions have started to decline substantially. The fraction by which emissions are reduced by these damages

23. We mention here for the sake of accuracy that in DICE-2016, the cost of backstop is given in 2010 thousand USD per ton of CO₂ in 2015, whereas the mass of carbon in the atmosphere is measured in tons of carbon. Thus to make the social cost of carbon formula numerically correct with respect to units of measurement, one needs to adjust it by a factor of 3.66, which is a conversion rate between carbon and CO₂. The SCC is measured in 2010-US dollars per ton of carbon.

24. One needs to multiply the carbon intensity σ_t in the denominator by 1,000 because we define the carbon intensity in the units of 1,000 GtC.

25. A detailed description of all parameters and exogenous parameters of the model can be found in the [online Appendix D](#).

26. All results we report in the following were obtained by using “deep equilibrium nets” by Azinovic *et al.* (2022), a method for computing global solutions to high-dimensional dynamic economic models. Additional implementation details, as well as the level of accuracy for our computations, are provided in the [online Appendix D](#).

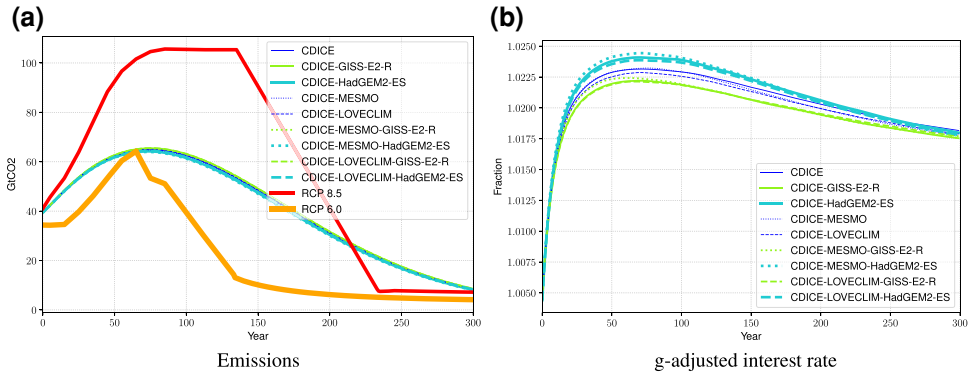


FIGURE 11

(a) BAU emissions as well as (b) growth-adjusted interest rate for different climate calibrations

Notes: Year zero on the graph corresponds to the starting year 2015.

might be significant, but the absolute value of the reduction is small. Figure 11(a) also shows that the emissions from DICE-16 lie somewhere between RCP6 and RCP85. The lessons from our detailed analysis of these emission scenarios above should also carry over to the DICE-2016 BAU emissions.

Figure 11(b) shows that the path of the growth-adjusted interest rate is also more or less the same across different climate calibrations. There is some variation that can be explained by differences in future damages that translate into differences in the interest rate. Higher expected future damages imply a lower capital stock and a (slightly) higher interest rate today. The variation over time, which is much more substantial, is mostly explained by the assumption in DICE-2016 that TFP growth, as well as population growth, flattens over time. The DICE-2016 calibration starts with an initial population growth of 1.4% and an initial TFP growth of 2.25%, with population growth declining fast over the next 50–100 years and TFP growth declining substantially slower. The way the time-varying growth rates translate into the time-varying interest rate depends crucially on the preferences of the representative agent. We will return to this issue in Section 6.3. Note for now that the initial g -adjusted interest rate is close to zero, whereas after 50 years, it is around 2%, and then finally converges to 1.5% in the very long run.

Figure 12 shows the effect of different climate calibrations on damages as a percentage of GDP and the SCC as a percentage of GDP. As is to be expected from our discussions in Section 5.1, our three different calibrations for the carbon cycle imply three vastly different predictions for CO₂ concentrations. Across different temperature equations, for a fixed carbon cycle, this then results in nine vastly different paths for temperature. As already pointed out in Section 5.1, the effects of the carbon cycle calibration on temperatures and damages are not nearly as large as the effect of the calibrations of the temperature equation, and the combination of an extreme carbon cycle and an extreme temperature equations yields extreme damages. This is true both for CDICE-MESMO-HadGEM2-ES as well as for CDICE-LOVECLIM-GISS-E2-R. The range of plausible damages in 300 years goes from around three percent of GDP to over 17.5% of GDP; almost a factor of six. As expected this has a very large effect on optimal abatement. The relative SCC depicted over the next 50 years in Figure 12(b) is fully consistent with our results in Section 5.1. In this section, we discussed only the SCC at a fixed date $t = 5$, whereas we examine it here over the next 30 periods. The SCC initially falls over time, and this effect is caused by the stark rise in the g -adjusted interest rate discussed above. After that, it stays

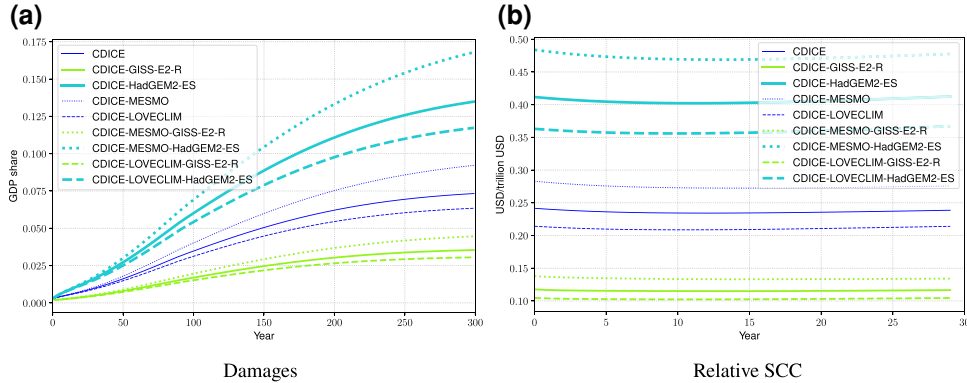


FIGURE 12

This figure shows the (a) evolution damages and the (b) relative SCC over time for nine CDICE calibrations in the business as usual case

Notes: Year zero on the graph corresponds to the starting year 2015.

relatively flat, and the relative SCC between the different temperature calibrations is similar to that described in Section 5.1.

The large differences in damages and the SCC across calibrations imply large differences in the optimal solution. We now assume that the social planner optimally chooses mitigation, μ_t , in the optimization-problem (8). Figure 13(a) shows the optimal abatement choices for the nine different climate calibrations. Differences in the social cost of carbon translate mechanically into differences in optimal abatement via equation (13). Since the exogenous parameters in equation (13) are time-varying, optimal mitigation is time-varying. The differences in optimal mitigation between the different climate calibrations are huge over the next 80 years; about a factor of three. By definition, differences in abatement lead to differences in emissions. Figure 13(b) shows BAU emissions for our nine cases compared to the emission scenarios in RCP 2.6 and in RCP 4.5. It is interesting to note that in all cases, optimal emissions remain far above RCP 2.6 emissions. This is a consequence of the specification of the damage function in DICE-2016, which we use here, and the fact that the model does not allow for carbon removal, which becomes an active part of future policy in RCP 2.6. Emissions in the MMM case closely mirror optimal emissions in RCP 4.5.

The fact that optimal emissions stay far above RCP26, which is the strong mitigation scenario from IPCC, readily implies that the optimal temperature rises far above the Paris limit. Figure 14(a) shows that in the MMM calibration, temperature rises (relative to pre-industrial levels) by 3 K and stays at that level for hundreds of years. In our most pessimistic climate calibration, the optimal temperature rises by over 4 K. As pointed out before, this depends crucially on the choice of the damage function, which we take directly from DICE-2016 for comparison purposes. As Nordhaus (2008) pointed out, “the economic impact of climate change . . . is the thorniest issue in climate-change economics.” See Hänsel et al. (2020) or Carleton et al. (2020) for more realistic treatments of damages.

Naturally, the choice of the damage function leads to relatively low SCC in levels. Figure 14(b) depicts the social cost of carbon in dollars over the next 50 years. Note that it increases only because of GDP growth. The relative SCC is almost identical to that in the BAU case depicted in Figure 12(b) above. Recall that the optimal tax on carbon is equal to the SCC. Thus, the model’s optimal tax lies, for the MMM case, slightly below what is typically discussed for economic policy purposes.

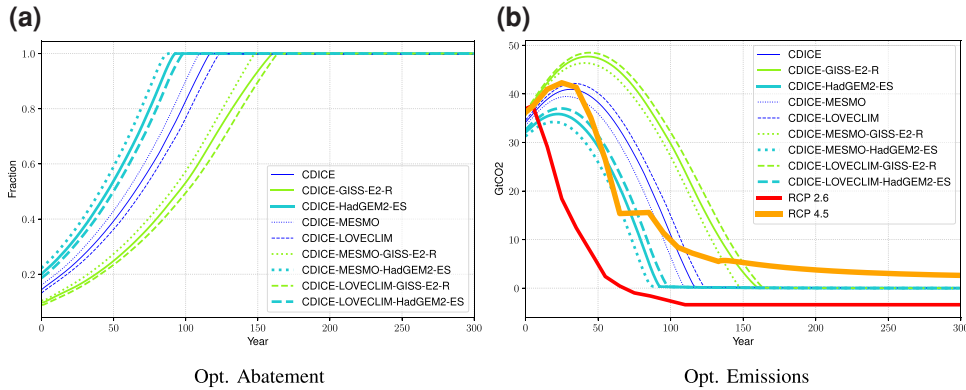


FIGURE 13
This figure shows (a) optimal abatement and (b) emissions over time

Notes: Year zero on the graph corresponds to the starting year 2015.

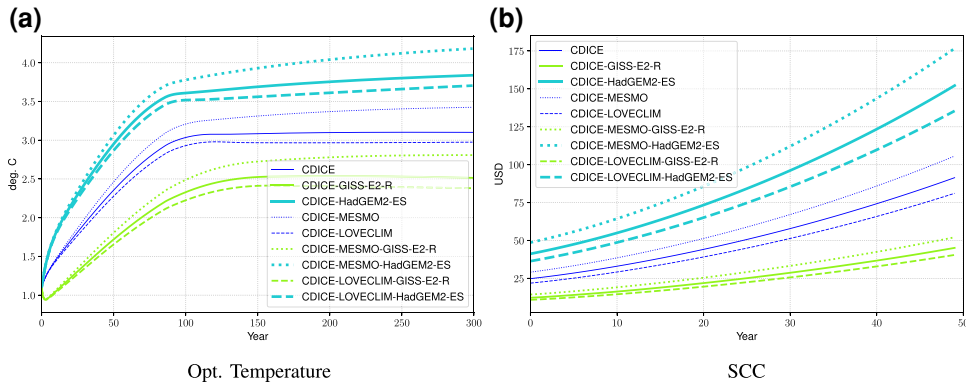


FIGURE 14
Evolution of (a) optimal temperature and (b) SCC with optimal abatement

Notes: Year zero on the graph corresponds to the starting year 2015.

6.2. The economic consequences of miscalibrated climate

As pointed out above, the calibration of the climate part of DICE-2016 has two serious flaws. Both the temperature equations and the carbon cycle are miscalibrated. As mentioned in Section 4.2, the carbon cycle in DICE-2016 overstates the fraction of CO₂ emissions that end up in the atmosphere relative to the MMM in Joos *et al.* (2013) and the extreme CDICE-MESMO case. We can confirm this by inspecting the BAU results. Compared to all three carbon calibrations in CDICE, DICE-2016 predicts way too much carbon in the atmosphere (cf. Figure 15(a)). In fact, the amount of carbon in the atmosphere is larger than for the CDICE-MESMO calibration (which corresponds to the extreme case in Joos *et al.* 2013) at any point in time.

The effect is less pronounced if one considers the overall temperature response. In addition to a mistake in the carbon cycle, the DICE-2016 calibration also makes a mistake in the temperature equation. As Figure 15(b) illustrates, in the short run (for the first 50 years or so), CDICE predicts more warming than the DICE-2016, despite the fact that there is far less CO₂ in the atmosphere. Even for an ECS of 4.5, DICE-2016 initially warms slower than the CDICE MMM calibration. The combined effect of temperature- and carbon-cycle miscalibration is crucial here; if one

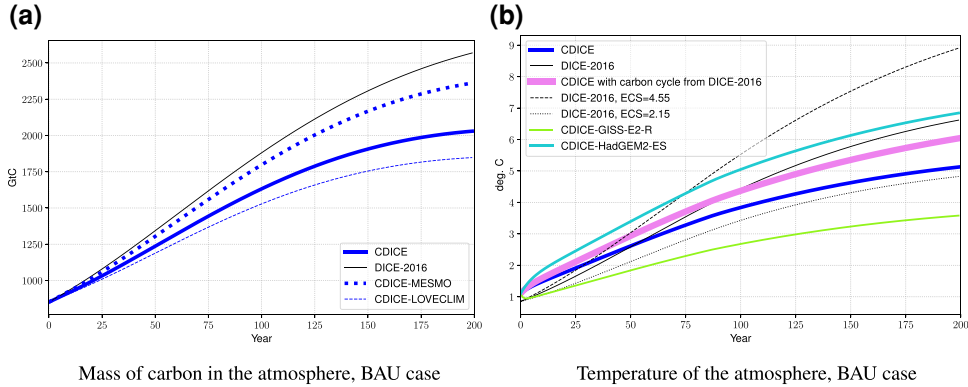


FIGURE 15

(a) Mass of carbon and (b) atmospheric temperature, BAU case for the CDICE and DICE-2016 calibrations

Notes: Year zero on the graph corresponds to the starting year 2015.

takes the CDICE MMM temperature equations together with the DICE-2016 carbon cycle, the temperature evolution is comparable to DICE-2016. The most notable difference is the initially overly slow warming in the DICE-2016 temperature equations.

Note that in the very long run, the dynamic system, defined by equations (1), (4), and (5), that governs climate in our simple model will converge to a new steady-state after all emissions are zero. In this steady-state, relative masses of carbon are given by the vector $M_{EQ} / \sum_{l=1}^3 M_{EQ,l}$ and it can be easily seen that the fraction of emitted CO_2 that remains in the atmosphere is about 0.26 for our CDICE calibration and about 0.22 for DICE-2016. The long-run effect on temperature is determined by the ECS, which is chosen to be 3.25 in CDICE and 3.1 in DICE-2016. Overall, the very long-run behavior of the two calibrations will be similar, with more warming in CDICE than in DICE-2016. Nevertheless, the miscalibration in DICE-2016 implies a consistently higher mass of carbon in the atmosphere over the next 500 years and a higher temperature after about 100 years. The important lesson is that parameters that determine the very long run of the climate system are more or less irrelevant to temperature over the next 300 years and, therefore, for the SCC that determines optimal policy today.

So far, we are repeating statements that we already made in Section 4. The crucial question is what the implications of this miscalibration are for the optimal policy. Figure 16(a) depicts the optimal levels of abatement for the different calibrations. Abatement is presented as a share of industrial emissions that are mitigated. Hence, its numerical values are bound to the interval $[0, 1]$, where 0 corresponds to the absence of any abatement, whereas 1 implies using the full mitigation capacities. The figure shows that the DICE-2016 calibrations imply a little more optimal abatement than the respective CDICE calibrations. The differences are so small because initial warming lies below the CDICE case and, with discounting, this is important for abatement over the next decade. This finding again highlights the relevance of both parts of the climate model, of the carbon cycle and the temperature response, and of caution against the potential for compensating errors.

From this figure, one might be tempted to argue that for policy recommendations it does not make much of a difference whether one uses the CDICE calibration or the DICE-2016 calibration. However, it makes a large difference for optimal temperature as Figure 16(b) demonstrates. The DICE-2016 (with the ECS of 3.1) atmospheric temperature is now outside of the CMIP5 range. It can be seen that after about 90 years, the black line (corresponding to DICE-2016)

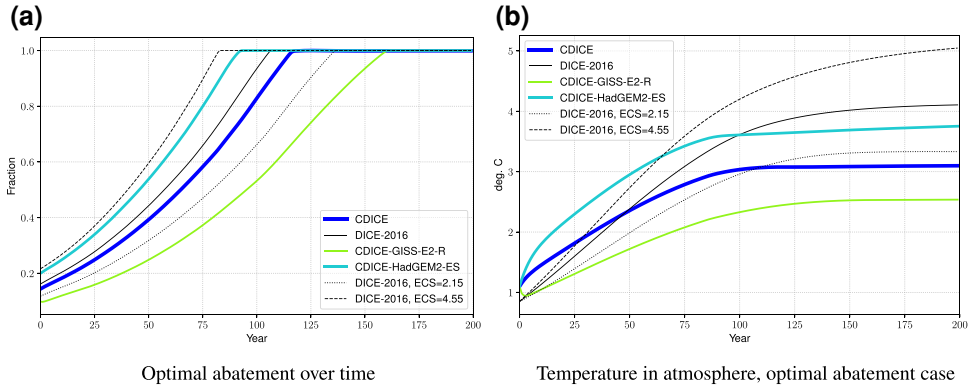


FIGURE 16

Optimal abatement in CDICE and DICE-2016: (a) optimal abatement over time and (b) temperature in atmosphere, optimal abatement case

Notes: Year zero on the graph corresponds to the starting year 2015.

crosses the predicted temperature under the optimal mitigation in DICE-2016, and it is above the temperature predicted by the extreme climate sensitivity scenario CDICE-HadGEM2-ES. In the BAU scenario, DICE-2016 falls well within the CMIP5 range, whereas in the economic model with mitigation, this is no longer the case. This issue arises because, in the BAU scenario, DICE-2016 is “helped” by offsetting errors in the climate equations and the carbon cycle. In the case of optimal mitigation, the flaw in the carbon cycle dominates and leads to so little mitigation that the temperature rises by more than 4 K. This is already implicit in the BAU SCC. Despite the fact that DICE-2016 leads to much larger damages and much higher temperatures than CDICE, the SCC is almost identical. This implies (to first order) the same optimal carbon tax for CDICE and DICE-2016, resulting in far higher temperatures in DICE-2016.

Although long-run temperatures differ so much, the reason why abatement and the SCC are comparable between CDICE and DICE-2016 is simply that these temperature differences only become significant in 75 years and, as shown in Figure 11(b), the g -adjusted interest rate is relatively high at that point in time. The damages that occur are heavily discounted. If one takes a different perspective and, instead of maximizing the utility of a social planner, restricts long-run warming to lie below 2 K, the differences between the two climate model calibrations become starker. In fact, for the DICE-2016 calibration, it is impossible to keep the temperature below 2 K. Full abatement that reduces industrial emissions to zero does not suffice to keep global warming below 2 K. This is very different in CDICE, where putting an immediate stop to industrial emissions will keep warming below 1.7 K. Somewhat more realistically, if one considers the RCP 2.6 emissions profile (that includes carbon removal), DICE-2016 warms by more than 2.5 K (as can already be seen in Figure 7) and stays above 2.3 K beyond 2300 while CDICE warms by 2.1 K, but warming is below 1.8 K by 2300.

Clearly, discounting is crucial for the small differences in abatement between the two calibrations. The interest rate is endogenous, as it depends on consumption growth, but also on the preferences of the representative agent.

6.3. The role of discounting

There is a large literature (see, *e.g.* Stern 2007; Hänsel and Quaas 2018) about the “right” discount rate for the social planner, that is, about the correct value of the parameter $\beta = \frac{1}{1+\rho}$ in the

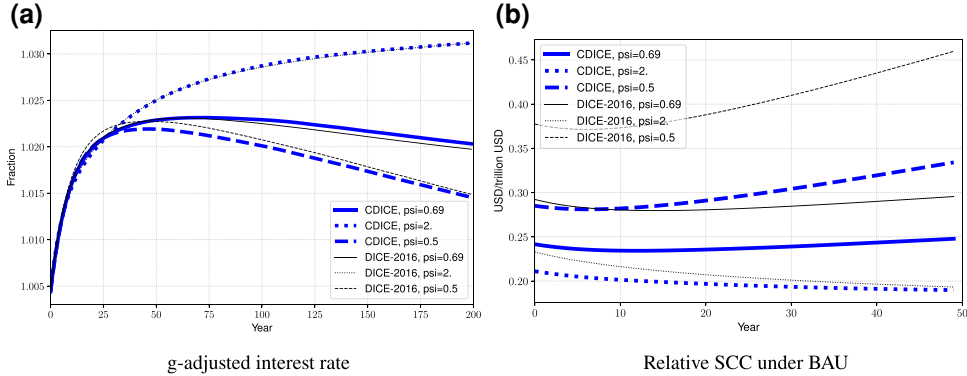


FIGURE 17

(a) Interest rate and (b) SCC under different values for ψ

Notes: Year zero on the graph corresponds to the starting year 2015.

planner's utility function. It is sometimes argued that this parameter can be pinned down by the observed average rates of return to capital (or interest rates). In a growth economy, this is not quite correct as the curvature of the planner's felicity function, ψ , and the time preference β are not jointly identified from average interest rates. The fact that in DICE-2016, the SCC is so sensitive to the growth-adjusted interest rate (documented in Section 5) implies that changing these two preference parameters simultaneously while keeping average interest rates fixed has large effects on the SCC and on optimal abatement in the DICE-2016 calibration. For simplicity, we take the baseline per-capita consumption growth rate to be 2%, which gives an interest rate of 4% in the DICE-2016 calibration (the long-run average growth rate in DICE-2016 is assumed to be zero, so we pick the 2% somewhat arbitrary to be consistent with the historical average and common long-run assumptions).

Figure 17(a) shows the paths of growth-adjusted interest rate for different values of the IES, ψ , in the business as usual case. G-adjusted interest rates are quite similar for the first 40 years, but there is significant divergence in the long run. After 200 years, the difference in g-adjusted interest rates between $\psi = 0.5$ and $\psi = 2$ is almost a factor of 2. As Figure 17(b) shows, this naturally translates into a large difference in the social cost of carbon across different specifications of the IES. The figure shows the SCC relative to output. The absolute SCC increases over time in all cases because of output growth. Because of the incorrect timing in the climate calibration of DICE-2016, the effect of the parameter ψ on interest rates is amplified, and the differences in the SCC between low and high ψ are larger than 50%. While a value of $\psi = 2$ might be considered an unrealistically high value of the IES, $\psi = 0.5$ is certainly well within the range of the IES which is typically assumed.²⁷

For the $\psi = 0.5$ case, the relative social cost of carbon initially decreases slightly and then strongly increases over time. This is obviously caused by the fact that g-adjusted interest rates decrease over time. In contrast, for a high value of the IES, $\psi = 2$, g-adjusted interest rates start decreasing significantly in about 40 years, leading to decreases in the SCC over time.

27. There are large differences in the IES typically assumed in the macro-literature (one or larger) to those typically found in household studies. There it is found that the IES varies significantly across households, ranging from 0.3 to 1. See, e.g. Calvet *et al.* (2021).

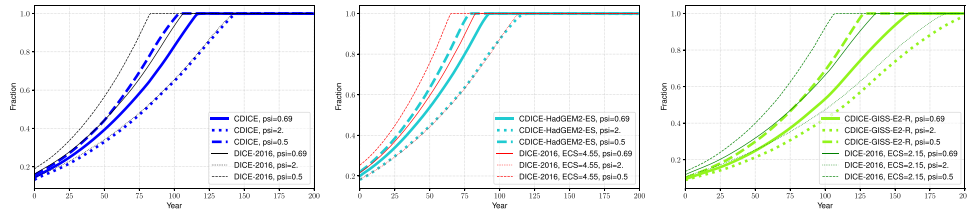


FIGURE 18

Optimal abatement for different values of ψ and different climate calibrations

Notes: The left panel shows results for MMM, the middle panel for high ECS calibrations, and the right panel for low ECS calibrations. Year zero on the graph corresponds to the starting year 2015.

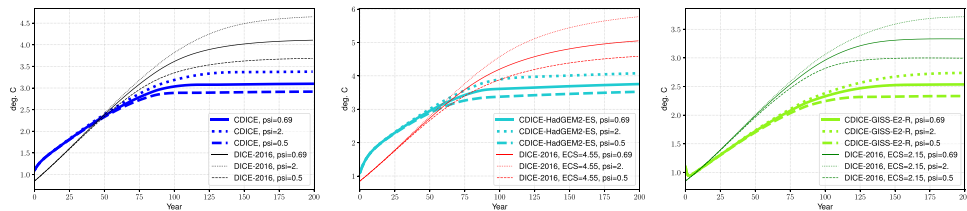


FIGURE 19

Optimal temperatures for different values of ψ and different climate calibrations

Notes: The left panel shows results for MMM, the middle panel for high ECS calibrations, and the right panel for low ECS calibrations. Year zero on the graph corresponds to the starting year 2015.

These differences in the BAU SCC obviously imply large differences in optimal policies. Figure 18 shows optimal abatement for the three values of ψ , comparing the DICE-2016 calibration to CDICE. The three panels in the figure show different ECS calibrations for DICE-2016 and HadGEM2-ES and GISS-E2-R for CDICE. Differences in abatement are very large between the different ψ -cases. In particular, in DICE-2016, for large ψ , the use of carbon is extended by about 50 years relative to the low IES case. This is true independently of the assumption on the ECS. Since the timing in the DICE-2016 calibration is wrong, the effects of future lower interest rates are much more pronounced than for CDICE. In fact, for a high value of ψ (higher future interest rates), the optimal policy in CDICE and DICE-2016 are quite similar; the differences become large as expected future interest rates become small. The range of plausible values for ψ adds another model uncertainty to economic models of climate change, and it turns out that the miscalibration in DICE-2016 amplifies the effect. In more interesting models with uncertainty and Epstein–Zin utility (see, *e.g.* Cai and Lontzek 2019), the use of the DICE-2016 climate calibration is likely to lead to serious problems.

As above, optimal long-run temperatures in DICE-2016 are far above the values in CDICE. Now there is an even larger spread caused by different values for ψ . Figure 19 shows future temperature under optimal abatement for different values of ψ and for the three different specifications of ECS. For the case of a high ECS, the optimal warming under DICE-2016 is now quite dramatic if one assumes a high value of ψ . In 200 years, optimal warming is almost 6 degrees; far above anything considered acceptable by most climate scientists. Even for the MMM calibration, a high ψ implies warming by more than 4.5 K. For this high IES case, warming prescribed by DICE-2016 falls significantly out of the CMIP5 range.

Future interest rates and growth rates matter a lot for optimal mitigation and optimal interest rates. Even if one settles on a path for TFP growth, different reasonable preference specifications lead to large differences in interest rates. The fact that the DICE-2016 calibration of the

climate emulator contains two flaws (discussed in detail above) implies an unrealistically large sensitivity to discounting, and differences in future interest rates translate into an unrealistically large difference in the SCC.

The results presented here clearly highlight the need for reliable climate calibrations and show the advantages of our approach. The fact that DICE-2016 lies within the range of CMIP5 predictions in some cases does not ensure that it can be used as a reliable tool within a calibrated economic model.

7. CONCLUSION

IAMs provide a quantitative framework that takes into account economic and environmental stocks and flows as well as their interaction and which allows researchers to investigate climate change and possible mitigation pathways (Hassler *et al.* 2016). IAMs differ very much in terms of structure, complexity, level of detail, and possible solution methods (Weyant 2017; Hare *et al.* 2018; Doukas *et al.* 2019). Models like the one by Clarke *et al.* (2009) have very rich climate representation, whereas Bosetti *et al.* (2007) provided a detailed economic model. Other models provided a detailed representation of the energy sector. Models like Nordhaus (2018), Golosov *et al.* (2014), Traeger (2019), Hambel *et al.* (2021), and Hassler and Krusell (2018) have a more parsimonious structure and exhibit a higher level of aggregation, but allow for monetary estimates both for the cost of curbing climate change and for the economic benefits that it brings, which is a necessary feature from a policy-making perspective. Having a very detailed model with the possibility of a cost-benefit analysis would be the preferable option. However, computational costs for such kinds of models are extremely high. Thus, there is a strong demand for IAMs, on the one hand, for capturing and reflecting relevant processes both in the economy and climate. On the other hand, IAMs are necessary for quantitative estimation (uncertainty, risks, tipping points, endogenous growth), so they should be parsimonious enough to have reasonable computational complexity. The perfect candidate to satisfy both requirements could be the DICE model (Nordhaus 2018). However, it has been strongly criticized by climate scientists and recently by economists for representing climate not in accordance with the recent climate science advancements (Dietz *et al.* 2021).

Our paper's three main contributions are as follows. First, we develop a series of tests to evaluate the quality of CEs used in economic modeling. These tests include one test that is similar to Dietz *et al.* (2021), but overall they are more comprehensive in that they specifically test both, the temperature equations and the carbon cycle of CEs.

Second, we update the parsimonious climate representation of climate in DICE with a new calibration of coefficients that aligns it with the CMIP5 benchmark, making the DICE model both simple and transparent yet making it realistically represent climate. The functional form of the climate part in DICE is retained. This allows authors who have the climate module of DICE "hard-wired" in their economic model to obtain more reliable results by simply changing parameters.

Third, we explain that, although the DICE-2016 predictions in a BAU scenario fall within the range of CMIP5, this is caused by two compensating errors in the calibration. We demonstrate that mitigation scenarios are highly sensitive to such compensating errors between the carbon cycle and the temperature response of the model. We advocate that this result is generic and not specific to DICE-2016 and that the proposed battery of tests enables the identification of such issues in climate models used in climate economics. Regarding DICE-2016, our conclusion is similar to that in Dietz *et al.* (2021); the calibration of DICE-2016 is not suited for reliable policy analysis.

An important aspect for the calibration and testing of a CE, is what physics is covered by either the test data or the functional form of the CE—as, for example, the box-model formulation used here or an impulse response function (IRF) formulation frequently seen in the literature. This is a vast question that may easily fill a paper of its own, but it is useful to point out a few aspects that come to mind in the context of the present study. We first note that a box model, as used here for the carbon cycle, preserves the total carbon mass in the system, apart from carbon added to the system via human activity. A corollary is that such a box description “knows” when its reservoirs (boxes) are out of (pre-industrial) equilibrium, factoring in any resulting associated fluxes among reservoirs automatically. This is in contrast to IRF approaches, as used, for example, in [Joos *et al.* \(2013\)](#), where the temporal evolution of a perturbation to a reservoir (as a carbon emission pulse to the atmosphere) is described as a function of time but without taking into account where the part of the perturbation leaving that reservoir (*e.g.* the atmosphere) goes. Consequently, an IRF formulation cannot be aware of whether or how strongly out of equilibrium different reservoirs (atmosphere, ocean, land, etc.) are. Coefficients in either formulation, box, or IRF, can be made aware (calibrated) such as to take into account feedbacks affecting the exchange among reservoirs (*e.g.* [Stern and Johansson \(2017\)](#) for box models, or [Joos *et al.* 2001](#) for IRFs). On the side of the carbon cycle, such feedbacks include carbon concentration and carbon-climate feedback, that is, carbon fluxes among reservoirs are sensitive to changes in atmospheric carbon concentration or global mean temperature that have already occurred relative to pre-industrial times (see, *e.g.* [Arora *et al.* 2013, 2020](#)). Other effects present in the full climate system are more difficult to capture via simple climate emulators. An example here is feedback due to two-dimensional patterns of ocean surface temperature (see, *e.g.* [Andrews *et al.* 2022](#)). By design, such feedbacks are beyond the scope of simple zero-dimensional CEs as the one studied here. Nevertheless, such simple CEs are of obvious interest for economic applications.

Supplementary Data

[Supplementary data](#) are available at *Review of Economic Studies* online.

Data availability statement

The data underlying this article are publicly available on Zenodo, at <https://doi.org/10.5281/zenodo.10276218>.

REFERENCES

- ANDERSON, B., BORGONOVO, E., GALEOTTI, M., *et al.* (2014), “Uncertainty in Climate Change Modeling: Can Global Sensitivity Analysis Be of Help?”, *Risk Analysis*, **34**, 271–293. ISSN 02724332. doi:[10.1111/risa.12117](https://doi.org/10.1111/risa.12117)
- ANDREWS, T., BODAS-SALCEDO, A., GREGORY, J. M., *et al.* (2022), “On the Effect of Historical SST Patterns on Radiative Feedback”, *Journal of Geophysical Research (Atmospheres)*, **127**, e2022JD036675. doi:[10.1029/2022JD036675](https://doi.org/10.1029/2022JD036675)
- ARORA, V. K., BOER, G. J., FRIEDLINGSTEIN, P., *et al.* (2013), “Carbon-Concentration and Carbon-Climate Feedbacks in CMIP5 Earth System Models”, *Journal of Climate*, **26**, 5289–5314. doi:[10.1175/JCLI-D-12-00494.1](https://doi.org/10.1175/JCLI-D-12-00494.1)
- ARORA, V. K., KATAVOUTA, A., WILLIAMS, R. G., *et al.* (2020), “Carbon-Concentration and Carbon-Climate Feedbacks in CMIP6 Models and Their Comparison to CMIP5 Models”, *Biogeosciences*, **17**, 4173–4222. doi:[10.5194/bg-17-4173-2020](https://doi.org/10.5194/bg-17-4173-2020)
- AZINOVIC, M., GAEGAUF, L. and SCHEIDEGGER, S. (2022), “Deep Equilibrium Nets”, *International Economic Review*, **63**, 1471–1525. doi:[10.1111/iere.12575](https://doi.org/10.1111/iere.12575)
- BARNETT, M., BROCK, W. and HANSEN, L. P. (2020), “Pricing Uncertainty Induced by Climate Change”, *The Review of Financial Studies*, **33**, 1024–1066.
- BARNETT, M., BROCK, W. and PETER HANSEN, L. (2022), “How Should Climate Change Uncertainty Impact Social Valuation and Policy?” (Working Paper).
- BARRAGE, L. (2018), “Be Careful What You Calibrate for: Social Discounting in General Equilibrium”, *Journal of Public Economics*, **160**, 33–49.
- BAUER, M. D. and RUDEBUSCH, G. D. (2021), “The Rising Cost of Climate Change: Evidence from the Bond Market”, *The Review of Economics and Statistics*, (105), 1–45.

- BEUSCH, L., GUDMUNDSSON, L. and SENEVIRATNE, S. I. (2020), “Emulating Earth System Model Temperatures with Mesmer: From Global Mean Temperature Trajectories to Grid-Point-Level Realizations on Land”, *Earth System Dynamics*, **11**, 139–159. doi:10.5194/esd-11-139-2020
- BOSETTI, V., ENI, F., MATTEI, E., *et al.* (2007), “The WITCH Model: Structure, Baseline, Solutions” (Working Paper). <http://dx.doi.org/10.2139/ssrn.960746>.
- BOTZEN, W. J. W. and VAN DEN BERGH, J. C. J. M. (2012), “How Sensitive Is Nordhaus to Weitzman? Climate Policy in DICE with an Alternative Damage Function”, *Economics Letters*, **117**, 372–374. ISSN 01651765. doi:10.1016/j.econlet.2012.05.032
- CAI, Y., BROCK, W., XEPAPADEAS, A., *et al.* (2019), “Climate Policy Under Spatial Heat Transport: Cooperative and Noncooperative Regional Outcomes” arXiv preprint arXiv:1909.04009.
- CAI, Y., JUDD, K. L. and LONTZEK, T. S. (2012), “Continuous-Time Methods for Integrated Assessment Models” (Working Papers 18365, NBER: National Bureau of Economic Research, Inc.) <https://ideas.repec.org/p/nbr/nberwo/18365.html>.
- CAI, Y. and LONTZEK, T. (2019), “The Social Cost of Carbon with Economic and Climate Risks”, *Journal of Political Economy*, **127**, 2684–2734. ISSN 1537534X. doi:10.1086/701890
- CALEL, R. and STAINFORTH, D. A. (2017), “On the Physics of Three Integrated Assessment Models”, *Bulletin of the American Meteorological Society*, **98**, 1199–1216. ISSN 00030007. doi:10.1175/BAMS-D-16-0034.1
- CALVET, L. E., CAMPBELL, J. Y., GOMES, F., *et al.* (2021), “The Cross-Section of Household Preferences” (Technical Report, National Bureau of Economic Research).
- CARLETON, T. A., JINA, A., DELGADO, M. T., *et al.* (2020), “Valuing the Global Mortality Consequences of Climate Change Accounting for Adaptation Costs and Benefits” (Technical Report, National Bureau of Economic Research).
- CLARKE, L., EDMONDS, J., KREY, V., *et al.* (2009), “International Climate Policy Architectures: Overview of the EMF 22 International Scenarios”, *Energy Economics*, **31**, S64–S81. doi:10.1016/j.eneco.2009.10.013
- Climate Change 2014 Synthesis Report IPCC (2014), “Climate Change 2014 Synthesis Report Summary Chapter for Policymakers” IPCC, 31. ISSN 1476–4687.
- DANABASOGLU, G., LAMARQUE, J. F., BACMEISTER, J., *et al.* (2020), “The Community Earth System Model Version 2 (CESM2)”, *Journal of Advances in Modeling Earth Systems*, **12**, e01916. doi:10.1029/2019MS001916
- DIETZ, S., VAN DER PLOEG, F., REZAI, A., *et al.* (2021), “Are Economists Getting Climate Dynamics Right and Does It Matter?”, *Journal of the Association of Environmental and Resource Economists*, **8**, 895–921. doi:10.1086/713977
- DOUKAS, H., FLAMOS, A. and LIEU, J. (2019), *Understanding Risks and Uncertainties in Energy and Climate Policy* (Cham: Springer). ISBN 9783030031510. doi:10.1007/978-3-030-03152-7
- EYRING, V., BONY, S., MEEHL, G. A., *et al.* (2016), “Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization”, *Geoscientific Model Development*, **9**, 1937–1958. doi:10.5194/gmd-9-1937-2016
- FREDRIKSEN, H.-B., SMITH, C. J., MODAK, A., *et al.* (2023), “21st Century Scenario Forcing Increases More for CMIP6 Than CMIP5 Models”, *Geophysical Research Letters*, **50**, e2023GL102916. doi:10.1029/2023GL102916
- GAMBHIR, A., NAPP, T. and HAWKES, A. (2017), “The Contribution of Non-CO2 Greenhouse Gas Mitigation to Achieving Long-Term Temperature Goals”, *Energies*, **10**, 1–23. ISSN 1996-1073. doi:10.3390/en10050602
- GEOFFROY, O., SAINT-MARTIN, D., BELLON, G., *et al.* (2013), “Transient Climate Response in a Two-Layer Energy–Balance Model. Part II: Representation of the Efficacy of Deep-Ocean Heat Uptake and Validation for CMIP5 AOGCMs”, *Journal of Climate*, **26**, 1859–1876. ISSN 08948755. doi:10.1175/JCLI-D-12-00196.1
- GOLOSOV, M., HASSLER, J., KRUSELL, P., *et al.* (2014), “Optimal Taxes on Fossil Fuel in General Equilibrium”, *Econometrica*, **82**, 41–88. ISSN 0012-9682. doi:10.3982/ECTA10217
- HAMBEL, C., KRAFT, H. and SCHWARTZ, E. (2021), “Optimal Carbon Abatement in a Stochastic Equilibrium Model with Climate Change”, *European Economic Review*, **132**, 103642. ISSN 00142921. doi:10.1016/j.euroecorev.2020.103642
- HÄNSEL, M. C., DRUPP, M. A., JOHANSSON, D. J. A., *et al.* (2020), “Climate Economics Support for the UN Climate Targets”, *Nature Climate Change*, **10**, 781–789. ISSN 17586798. doi:10.1038/s41558-020-0833-x
- HÄNSEL, M. C. and QUAAS, M. F. (2018), “Intertemporal Distribution, Sufficiency, and the Social Cost of Carbon”, *Ecological Economics*, **146**, 520–535. ISSN 09218009. doi:10.1016/j.ecolecon.2017.11.024
- HARE, B., BRECHA, R. and SCHAEFFER, M. (2018), “Integrated Assessment Models : What Are They and How Do They Arrive at Their Conclusions?”, *Climate Analytics*, 1–12.
- HASSLER, J. and KRUSELL, P. (2018), *Environmental Macroeconomics: The Case of Climate Change* (Vol. 4) (Oxford: Elsevier B.V.). doi:10.1016/bs.hesenv.2018.08.003
- HASSLER, J., KRUSELL, P. and OLOVSSON, C. (2018), “The Consequences of Uncertainty: Climate Sensitivity and Economic Sensitivity to the Climate”, *Annual Review of Economics*, **10**, 189–205. ISSN 19411391. doi:10.1146/annurev-economics-080217-053229
- HASSLER, J., KRUSELL, P. and SMITH, A. A. (2016), “Environmental Macroeconomics”, *Handbook of Macroeconomics*, **2**, 1893–2008. ISSN 15740048. doi:10.1016/bs.hesmac.2016.04.007
- HASSLER, J., OLOVSSON, C. and REITER, M. (2019), “Integrated Assessment in a Multi-region World with Multiple Energy Sources and Endogenous Technical Change” (Technical Report, FRBSF). <http://theflatearthsociety.org/home/index.php/about-the-society/>.
- HOPE, C. (2013), “Critical Issues for the Calculation of the Social Cost of CO2: Why the Estimates from PAGE09 are Higher Than Those from PAGE2002”, *Climatic Change*, **117**, 531–543. doi:10.1007/s10584-012-0633-z

- HOWARD, P. H. and STERNER, T. (2017), “Few and Not So Far Between: A Meta-Analysis of Climate Damage Estimates”, *Environmental and Resource Economics*, **68**, 197–225. ISSN 15731502. doi:[10.1007/s10640-017-0166-z](https://doi.org/10.1007/s10640-017-0166-z)
- HWANG, I. C., REYNÈS, F. and TOL, R. S. J. (2017), “The Effect of Learning on Climate Policy under Fat-Tailed Risk”, *Resource and Energy Economics*, **48**, 1–18. ISSN 09287655. doi:[10.1016/j.reseneeco.2017.01.001](https://doi.org/10.1016/j.reseneeco.2017.01.001)
- Intergovernmental Panel on Climate Change and Intergovernmental Panel on Climate Change (2015), “Assessing Transformation Pathways”, In *Climate Change 2014 Mitigation of Climate Change*, (Cambridge: Cambridge University Press) 413–510. doi: [10.1017/cbo9781107415416.012](https://doi.org/10.1017/cbo9781107415416.012).
- JACKSON, L. S., MAYCOCK, A. C., ANDREWS, T., *et al.* (2022), “Errors in Simple Climate Model Emulations of Past and Future Global Temperature Change”, *Geophysical Research Letters*, **49**, e98808. doi:[10.1029/2022GL098808](https://doi.org/10.1029/2022GL098808)
- JELTSCH-THÖMMES, A. and JOOS, F. (2020), “Modeling the Evolution of Pulse-Like Perturbations in Atmospheric Carbon and Carbon Isotopes: The Role of Weathering-Sedimentation Imbalances”, *Climate of the Past*, **16**, 423–451. doi:[10.5194/cp-16-423-2020](https://doi.org/10.5194/cp-16-423-2020)
- JOOS, F., PRENTICE, I. C., SITCH, S., *et al.* (2001), “Global Warming Feedbacks on Terrestrial Carbon Uptake Under the Intergovernmental Panel on Climate Change (IPCC) Emission Scenarios”, *Global Biogeochemical Cycles*, **15**, 891–907. doi:[10.1029/2000GB001375](https://doi.org/10.1029/2000GB001375)
- JOOS, F., ROTH, R., FUGLESTVEDT, J. S., *et al.* (2013), “Carbon Dioxide and Climate Impulse Response Functions for the Computation of Greenhouse Gas Metrics: A Multi-Model Analysis”, *Atmospheric Chemistry and Physics*, **13**, 2793–2825.
- KEELING, C. D. (1973), “Industrial Production of Carbon Dioxide from Fossil Fuels and Limestone”, *Tellus*, **25**, 174–198. doi:[10.1111/j.2153-3490.1973.tb01604.x](https://doi.org/10.1111/j.2153-3490.1973.tb01604.x)
- KELLER, D. P., LENTON, A., SCOTT, V., *et al.* (2018), “The Carbon Dioxide Removal Model Intercomparison Project (CDRMI): Rationale and Experimental Protocol for cmip6”, *Geoscientific Model Development*, **11**, 1133–1160. doi:[10.5194/gmd-11-1133-2018](https://doi.org/10.5194/gmd-11-1133-2018)
- KNUTTI, R., RUGENSTEIN, M. A. A. and HEGERL, G. C. (2017), “Beyond Equilibrium Climate Sensitivity”, *Nature Geoscience*, **10**, 727–736. doi:[10.1038/ngeo3017](https://doi.org/10.1038/ngeo3017)
- KOTLIKOFF, L., KUBLER, F. and POLBIN, A. (2021a), “Pareto-Improving Carbon-Risk Taxation”, *Economic Policy*, **36**, 551–589. ISSN 0266-4658. doi:[10.1093/epolic/eiab008](https://doi.org/10.1093/epolic/eiab008)
- KOTLIKOFF, L., KUBLER, F., POLBIN, A., *et al.* (2021b), “Making Carbon Taxation a Global Win-Win”, No Brainers and Low-Hanging Fruit in National Climate Policy.
- KOTLIKOFF, L. J., KUBLER, F., POLBIN, A., *et al.* (2021c), “Can Today’s and Tomorrow’s World Uniformly Gain from Carbon Taxation?” (Working Paper 29224, National Bureau of Economic Research). <http://www.nber.org/papers/w29224>.
- LE QUÉRÉ, C., ANDREW, R. M., FRIEDLINGSTEIN, P., *et al.* (2018), “Global Carbon Budget 2018”, *Earth System Science Data*, **10**, 2141–2194. doi:[10.5194/essd-10-2141-2018](https://doi.org/10.5194/essd-10-2141-2018)
- LEMOINE, D. and TRAEGER, C. (2014), “Watch Your Step: Optimal Policy in a Tipping Climate”, *American Economic Journal: Economic Policy*, **6**, 137–166. ISSN 19457731. doi:[10.1257/pol.6.1.137](https://doi.org/10.1257/pol.6.1.137)
- LONTZEK, T. S., CAI, Y., JUDD, K. L., *et al.* (2015), “Stochastic Integrated Assessment of Climate Tipping Points Indicates the Need for Strict Climate Policy”, *Nature Climate Change*, **5**, 441–444.
- MACDOUGALL, A. H., EBY, M. and WEAVER, A. J. (2013), “If Anthropogenic CO₂ Emissions Cease, Will Atmospheric CO₂ Concentration Continue to Increase?”, *Journal of Climate*, **26**, 9563–9576. doi:[10.1175/JCLI-D-12-00751.1](https://doi.org/10.1175/JCLI-D-12-00751.1)
- MEINSHAUSEN, M., SMITH, S. J., CALVIN, K., *et al.* (2011), “The RCP Greenhouse Gas Concentrations and Their Extensions from 1765 to 2300”, *Climatic Change*, **109**, 213–241. doi:[10.1007/s10584-011-0156-z](https://doi.org/10.1007/s10584-011-0156-z)
- MENGIS, N. and MATTHEWS, H. D. (2020), “Non-CO₂ Forcing Changes Will Likely Decrease the Remaining Carbon Budget for 1.5 C”, *npj Climate and Atmospheric Science*, **3**, 1–7.
- MICHAELIS, P. and WIRTHS, H. (2020), “DICE-RD: An Implementation of Rate-Related Damages in the DICE Model”, *Environmental Economics and Policy Studies*, **22**, 555–584. ISSN 1867383X. doi:[10.1007/s10018-020-00269-4](https://doi.org/10.1007/s10018-020-00269-4)
- MIFTAKHOVA, A. (2018), “Global Sensitivity Analysis in Integrated Assessment Modeling”, *SSRN Electronic Journal*, **2018**, 1–43. ISSN 1556-5068. doi:[10.2139/ssrn.3215873](https://doi.org/10.2139/ssrn.3215873)
- MIFTAKHOVA, A., JUDD, K. L., LONTZEK, T. S., *et al.* (2020), “Statistical Approximation of High-Dimensional Climate Models”, *Journal of Econometrics*, **214**, 67–80.
- MILLAR, R. J. and FRIEDLINGSTEIN, P. (2018), “The Utility of the Historical Record for Assessing the Transient Climate Response to Cumulative Emissions”, *Philosophical Transactions of the Royal Society of London Series A*, **376**, 20160449. doi:[10.1098/rsta.2016.0449](https://doi.org/10.1098/rsta.2016.0449)
- NICHOLLS, Z. R. J., MEINSHAUSEN, M., LEWIS, J., *et al.* (2020), “Reduced Complexity Model Intercomparison Project Phase 1: Introduction and Evaluation of Global-Mean Temperature Response”, *Geoscientific Model Development*, **13**, 5175–5190. doi:[10.5194/gmd-13-5175-2020](https://doi.org/10.5194/gmd-13-5175-2020)
- NORDHAUS, W. D. (1979), *Efficient Use of Energy Resources* (New Haven, CT: Yale University Press).
- NORDHAUS, W. D. (2008), *A Question of Balance: Weighing the Options on Global Warming Policies* (New Haven, CT: Yale University Press).
- NORDHAUS, W. D. (2012), “Estimates of the Social Cost of Carbon: Background and Results from the Rice-2011 Model”, *SSRN Electronic Journal*, **1**, 273–312. ISSN 0898-2937. doi:[10.2139/ssrn.1945844](https://doi.org/10.2139/ssrn.1945844)

- NORDHAUS, W. (2013), *Integrated Economic and Climate Modeling* (Vol. 1) (Amsterdam: Elsevier). doi:[10.1016/B978-0-444-59568-3.00016-X](https://doi.org/10.1016/B978-0-444-59568-3.00016-X)
- NORDHAUS, W. (2018), "Evolution of Modeling of the Economics of Global Warming: Changes in the DICE Model, 1992–2017", *Climatic Change*, **148**, 623–640. ISSN 15731480. doi:[10.1007/s10584-018-2218-y](https://doi.org/10.1007/s10584-018-2218-y)
- NORDHAUS, W. (2018), "Projections and Uncertainties about Climate Change in an Era of Minimal Climate Policies", *American Economic Journal: Economic Policy*, **10**, 333–360. ISSN 1945774X. doi:[10.1257/pol.20170046](https://doi.org/10.1257/pol.20170046)
- PIGOU, A. C. (1920), *Wealth and Welfare*, London, 1912. Later Editions, *The Economics of Welfare*.
- POPP, D. (2004), "ENTICE: Endogenous Technological Change in the DICE Model of Global Warming", *Journal of Environmental Economics and Management*, **48**, 742–768. ISSN 00950696. doi:[10.1016/j.jeem.2003.09.002](https://doi.org/10.1016/j.jeem.2003.09.002)
- RICKELS, W., REITH, F., KELLER, D., *et al.* (2018), "Integrated Assessment of Carbon Dioxide Removal", *Earth's Future*, **6**, 565–582. doi:[10.1002/2017EF000724](https://doi.org/10.1002/2017EF000724)
- RIDGE, S. M. and MCKINLEY, G. A. (2021), "Ocean Carbon Uptake under Aggressive Emission Mitigation", *Biogeosciences*, **18**, 2711–2725. doi:[10.5194/bg-18-2711-2021](https://doi.org/10.5194/bg-18-2711-2021)
- STERN, N. (2007), *The Economics of Climate Change: The Stern Review* (London: Cambridge University Press).
- STERNER, E. O. and JOH ANSSON, D. J. A. (2017), "The Effect of Climate-Carbon Cycle Feedbacks on Emission Metrics", *Environmental Research Letters*, **12**, 034019. doi:[10.1088/1748-9326/aa61dc](https://doi.org/10.1088/1748-9326/aa61dc)
- TAYLOR, K. E., STOUFFER, R. J. and MEEHL, G. A. (2012), "An Overview of CMIP5 and the Experiment Design", *Bulletin of the American Meteorological Society*, **93**, 485–498. doi:[10.1175/BAMS-D-11-00094.1](https://doi.org/10.1175/BAMS-D-11-00094.1)
- TEBALDI, C., DEBEIRE, K., EYRING, V., *et al.* (2021), "Climate Model Projections from the Scenario Model Inter-comparison Project (ScenarioMIP) of CMIP6", *Earth System Dynamics*, **12**, 253–293. doi:[10.5194/esd-12-253-2021](https://doi.org/10.5194/esd-12-253-2021)
- THAO, S., GARVIK, M., MARIETHOZ, G., *et al.* (2021), "Combining Global Climate Models Using Graph Cuts" (Technical Report).
- THOMPSON, T. M. (2018), "Modeling the Climate and Carbon Systems to Estimate the Social Cost of Carbon", *Wiley Interdisciplinary Reviews: Climate Change*, **9**, 1–12. ISSN 17577799. doi:[10.1002/wcc.532](https://doi.org/10.1002/wcc.532)
- TOKARSKA, K. B., STOLPE, M. B. and SIPPEL, S. (2020), "Past Warming Trend Constrains Future Warming in cmip6 Models", *Science Advances*, **6**, eaaz9549–eaaz9549. doi:[10.1126/sciadv.aaz9549](https://doi.org/10.1126/sciadv.aaz9549)
- TRAEGER, C. P. (2014), "A 4-States DICE: Quantitatively Addressing Uncertainty Effects in Climate Change", *Environmental and Resource Economics*, **59**, 1–37. ISSN 09246460. doi:[10.1007/s10640-014-9776-x](https://doi.org/10.1007/s10640-014-9776-x)
- TRAEGER, C. P. (2019), "ACE – Analytic Climate Economy (with Temperature and Uncertainty)", *SSRN Electronic Journal*. ISSN 1556-5068. doi:[10.2139/ssrn.3307622](https://doi.org/10.2139/ssrn.3307622)
- WEITZMAN, M. L. (2012), "GHG Targets as Insurance Against Catastrophic Climate Damages", *Journal of Public Economic Theory*, **14**, 221–244.
- WEYANT, J. (2017), "Some Contributions of Integrated Assessment Models of Global Climate Change", *Review of Environmental Economics and Policy*, **11**, 115–137. ISSN 17506824. doi:[10.1093/reep/rew018](https://doi.org/10.1093/reep/rew018)
- ZELINKA, M. D., MYERS, T. A., MCCOY, D. T., *et al.* (2020), "Causes of Higher Climate Sensitivity in CMIP6 Models", *Geophysical Research Letters*, **47**, 1–12. ISSN 19448007. doi:[10.1029/2019GL085782](https://doi.org/10.1029/2019GL085782)