Statistical approximation of high-dimensional climate models

Alena Miftakhova [a,*], Kenneth L. Judd [b], Thomas S. Lontzek [c],
Karl Schmedders [a,d]

[a] *University of Zurich, Switzerland*
[b] *Hoover Institution, Stanford University, USA*
[c] *RWTH Aachen University, Germany*
[d] *International Institute for Management Development, Lausanne, Switzerland*

ARTICLE INFO

ABSTRACT

We propose a general emulation method for constructing low-dimensional approximations of complex dynamic climate models. Our method uses artificially designed uncorrelated CO_2 emissions scenarios, which are much better suited for the construction of an emulator than are conventional emissions scenarios. We apply our method to the climate model MAGICC to approximate the impact of emissions on global temperature. Comparing the temperature forecasts of MAGICC and our emulator, we show that the average relative out-of-sample forecast errors in the low-dimensional emulation models are below 2%. Our emulator offers an avenue to merge modern macroeconomic models with complex dynamic climate models.

1. Introduction

Integrated assessment models (IAMs) aim to merge dynamic models of the climate system with dynamic economic models to study their interactions and formulate policies related to limiting greenhouse gas emissions. It is currently intractable to merge the best climate models with modern macroeconomic models.[1] In this paper we address this problem by constructing a low-dimensional dynamic system that accurately represents the impact of world CO_2 emissions on world average global temperature. This is a valid reduction, because CO_2 disperses rapidly in the atmosphere and the major impact of climate change can be represented by average world temperature. The basic procedure for constructing a reduced model (often called an emulator) is to specify a set of emissions paths, use each one as input into a complex climate model, observe the resulting temperature paths, and use these simulated emissions and temperature data[2] to specify an approximating dynamic system. The structure has a pooled cross-sectional nature since each path is a time

---

* Correspondence to: Department of Business Administration, University of Zurich, Moussonstrasse 15, CH8044 Zurich, Switzerland.
  E-mail address: alena.miftakhova@business.uzh.ch (A. Miftakhova).

[1] The climate models rely on PDEs with an initial value representing the physical state of the climate at some initial time. Myopic economic models (called "recursive" models in the IAM literature) are also initial value problems, since information about the future does not affect current behavior in such models. It is therefore straightforward to merge myopic economic models with any climate model. Perfect foresight or rational expectations models are two-point boundary value problems with some states pinned down by initial conditions but other economic variables determined by transversality conditions at the terminal time. Merging even the smallest perfect foresight or rational expectations model with a large climate model would create an intractably large two-point boundary value problem.

[2] The word "data" in econometrics conventionally means observations of the real world. In the climate change literature, in turn, "data" can refer both to historical observations and to the simulated scenarios of the future climate, as in the IPCC Data Distribution Centre (IPCC-TGICA, 2007). Following this terminology, all references to data in the present paper refer to the simulated output of the climate model concerned.

series and multiple paths are used. Given the high computational costs of running complex climate models, one important criterion for the input emissions paths is their efficiency in "extracting" information from the system. We construct uncorrelated CO_2 emissions scenarios, and show that, when used as input for simulations, they prove more efficient than conventional scenarios. In the demonstrative case of emulating a climate model, forecast errors decrease by almost half when we use uncorrelated scenarios.

IAMs, deterministic or stochastic, should use as many state variables as required to ensure a realistic specification of the climate. One commonly used climate model is the Model for the Assessment of Greenhouse-gas Induced Climate Change (MAGICC), a reduced-complexity climate emulator (Meinshausen et al., 2011a). Since the computational complexity of solving dynamic models increases with their dimensionality, a dynamic system of the size of MAGICC is too large to be commonly applied in stochastic economic models with continuous state and control variables and large time horizons. Using our approximation approach, we construct an emulator of MAGICC that produces reliable predictions of temperature response to CO_2 emissions at a much lower cost.

Although statistical approximations seem alien to the physical nature of climate models, they are in fact closely connected to it. Studies show that statistical methods can produce models that adhere to basic climate physics. Kaufmann et al. (2013) test two time series models that link radiative forcing to global surface temperature as simulated by complex climate models and conclude that the model with a stochastic trend is supported more strongly by the data. They also show that this statistical model is consistent with the relation characterized by a globally averaged energy balance model, which is based on the physical relationship between radiative forcing and surface temperature. Pretis (2019) extends this analysis to a two-component (ocean–atmosphere) energy balance model by showing its equivalence to a cointegrated time series relation of the modeled climate variables.

Often, the methodology for emulating high-dimensional models assumes the preexistence of some input–output data sets from computer simulations. The general requirement for the input scenarios is that they are "carefully chosen" and "excite"[3] the emulated system sufficiently (Young and Ratto, 2011; Castelletti et al., 2012). The existing data sets of climate models' simulations contain the predictions of complex climate models in response to some commonly prescribed scenarios (e.g., IPCC-DDC 1998). A widely known example of such common scenarios is the Representative Concentration Pathways (RCPs) adopted in the most recent assessment report of the Intergovernmental Panel on Climate Change (IPCC) (van Vuuren et al., 2011). These scenarios are designed to provide a consistent base for scientific research, from single studies to large intercomparison projects. In this paper, however, we show that they are not a wise choice for the particular task of approximation.

In cases when additional simulations are not computationally expensive, researchers generate a large collection of emissions paths and the resulting temperature paths. For example, to emulate a climate model of intermediate complexity, Holden and Edwards (2010) construct an ensemble of possible future concentration profiles using the Latin hypercube method, and apply dimensionality reduction techniques to construct an emulator. A few studies consider statistical emulation an applicable data compression tool when storing the full data sets from climate models' runs proves too costly (Castruccio and Genton, 2016; Guinness and Hammerling, 2018). With the increasing volumes of generated climate data, the methodology enables the compression of data sets while preserving their distributional properties. For such experiments it is especially important to design the runs of the emulated complex models in such a way that the resulting input–output data sets are as informative as possible.

Unfortunately, the most detailed and complex climate models are costly to run: it can take several months to simulate a few hundred model years (Dringnei et al., 2008). This limits the collection of existing simulations. If the existing data is deemed insufficient to design a robust emulator, it is expensive to significantly increase the available data set. In a recent study, Castruccio et al. (2014) recognize the need to run more scenarios of the Community Climate System Model, version 3 (CCSM3) (Collins et al., 2006; Yeager et al., 2006). They design five new scenarios specifically for emulation and demonstrate that the trajectories of temperature and precipitation can be emulated using CO_2 concentrations paths. In choosing new scenarios, they do not follow any experimental design procedure, but make part of the scenarios similar to those readily available in multimodel experiments and let the other part induce rapid changes in CO_2 concentrations.

The key distinction of the present paper is our focus on the procedure for designing efficient input scenarios. We do not aim for our scenarios to resemble the existing runs of the complex climate models, but propose a task-driven approach to scenario creation. The computational cost of running complex climate models makes it imperative that input scenarios are chosen to maximize the information gained from these computations. This paper takes a mathematical view motivated by approximation theory. Intuitively, the input scenarios should be orthogonal in some sense. But when we apply principal component analysis (PCA) to the four conventional RCPs, we find that they jointly contain little more information than one scenario would. We, therefore, propose, instead, to use a family of orthogonal polynomials as a base for constructing the input scenarios for emulation. These scenarios obviously do not look like anything we expect will happen to the climate system; but this is not important for our task. And indeed, as intuition from approximation theory would indicate, we find that our four orthogonal polynomial input scenarios produce a significantly better emulator than the four standard RCPs. We test the usefulness of our emulation approach for stochastic IAMs by evaluating its ability to accurately simulate the distribution of temperatures in response to a stochastic emissions process. We find that it is very good at this task.

---

[3] The word "excite" here is borrowed from the emulation literature and essentially means to force or trigger a response from the emulated system.

Our emulator can be directly used to improve IAMs. Currently, most IAMs are deterministic, assuming that the future climate and economy are perfectly predicable. These perfect foresight models do not entail random variability within the model but attempt to incorporate uncertainty by applying Monte Carlo simulation techniques (see, for example, Mastrandrea and Schneider 2004 or Hope 2013). These deterministic models are solved many times, each time with their parameters being drawn from a distribution. The approximation of the stochastic framework is then presented by probabilistically averaging the results of all deterministic simulations. These models certainly produce valuable insights, for example regarding the range of possible model outcomes. However, they do not model economic and climate-related uncertainty in the decision-maker's problem, and therefore miss real features of decision-making under uncertainty. Crost and Traeger (2013) show that the Monte Carlo approach can produce misleading implications for policy making.

Even studies that account for some kind of uncertainty often point to the "curse of dimensionality" as an explanation for their simplified representations of the climate (EPA, 2010; Webster et al., 2012; Newbold et al., 2013; Jensen and Traeger, 2014). Moreover, stochastic IAMs often rely on very simple assumptions when formulating the actual stochastic processes.[4] However, recent advances in computational methods have described ways to solve high-dimensional economic models, even beyond 100 dimensions (Maliar and Maliar, 2015; Judd et al., 2011; Cai et al., 2015; Brumm and Scheidegger, 2017). Combining our approach to the construction of emulators with these new computational methods offers the potential to build more realistic models.

The remainder of the paper is organized as follows: in Section 2 we describe the methods used for approximation; we motivate the use of uncorrelated scenarios, explain the procedure of constructing them, and present the general model we estimate. Based on the results of the estimations, Section 3 states the specifications for recommended representations of the climate system and assess our emulator's performance on alternative scenarios and under different sets of initial parameters of the climate model. Section 4 concludes.

2. Approximating high-dimensional models

In this section we describe our approach to the construction of an emulator of a complex climate model. First, we briefly describe our source of temperature predictions, MAGICC. Second, we explain why conventional emissions scenarios are not a good source for the construction of an emulator. Next, we derive a set of uncorrelated emissions scenarios, which is ideally suited for the derivation of an emulator. And finally, we depict our general constructive approach.

2.1. MAGICC

MAGICC is a carbon cycle–climate model used to emulate the insights from large and complex Atmosphere–Ocean General Circulation Models (AOGCMs). The emulation task is to generate global temperature responses (the output variable) to various exogenous emissions scenarios (the input variable). In the climate research community, MAGICC is considered to be a reduced complexity climate model. Yet, it includes representations of the most essential physical and biological components of complex AOGCMs. Despite being a "simple" model, MAGICC performs exceptionally well in emulating the results of the large AOGCMs (see Meinshausen et al., 2011a). MAGICC was therefore used in the recent IPCC reports (IPCC, 2014) as the prime tool for evaluating carbon and climate responses to various emissions scenarios. MAGICC is publicly available and easy to operate. According to Meinshausen et al. (2011a), MAGICC is flexible enough to deliver accurate results when running scenarios outside of the original calibration space. For the purpose of the present paper, MAGICC is, therefore, best suited to generating reliable responses of global temperature to any emissions scenario—responses that we use as benchmarks for evaluating the accuracy of our statistical model.

We chose the model MAGICC because its computer implementation is very fast (a matter of seconds on many computers at the time of writing of this paper). This speed of implementation makes it feasible for us to compare the temperature forecasts of our emulator to those of MAGICC for 200 different alternative model settings (see Section 3.4). In addition, even for such a fast climate model an emulator may make it much easier to incorporate the model's insights into a complex macroeconomic model.

2.2. Conventional emissions scenarios

The four basic RCPs endorsed by the IPCC are the most common and ready-to-use scenarios and currently serve as a coherent base for integrated climate/economic modeling and model intercomparison projects (Taylor et al., 2012). Each RCP is a distinct pathway of radiative forcing named after the level of forcing it achieves in 2100. RCP2.6, RCP4.5, RCP6, and RCP8.5 thus imply forcing of 2.6 to 8.5 W/m^2 by the end of the century. Each scenario also specifies the associated levels of emissions and concentrations of greenhouse gases and other forcing agents. Fig. 1 displays the decadal CO_2 emissions paths implied by the RCPs and the corresponding temperature rises as predicted by MAGICC.[5] The lowest-forcing scenario is the only one that does not hit the critical value of warming the atmosphere to 2 ℃ above the preindustrial level; the most aggressive scenario among the four – RCP8.5 – implies warming of above 4 ℃.

---

[4] Cai et al. (2015) and Lontzek et al. (2015), among others, discuss the implications of structural assumptions in IAMs regarding modeling the risk of abrupt and irreversible catastrophic climate events.

[5] The RCP scenarios are specified on a decadal scale; MAGICC uses their linear interpolation to produce the corresponding annual paths of temperature anomalies.
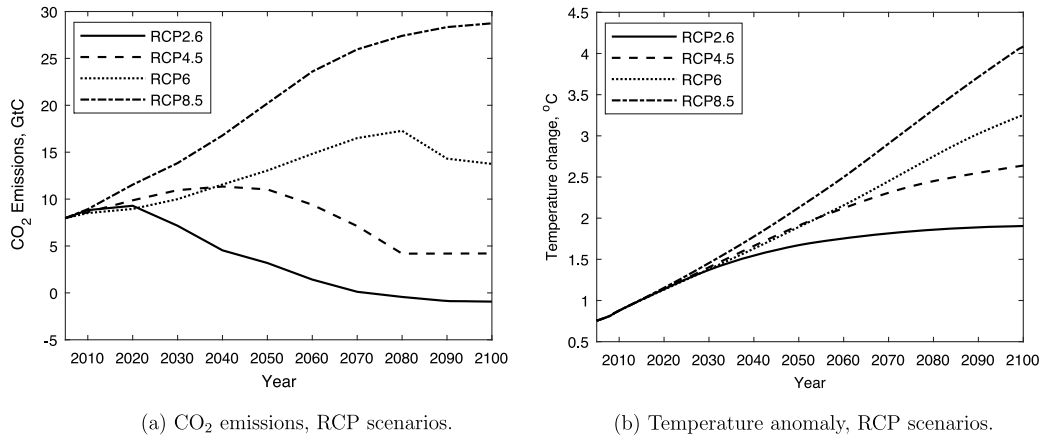
(a) CO_2 emissions, RCP scenarios.

(b) Temperature anomaly, RCP scenarios.

Fig. 1. RCP emissions scenarios (a) and corresponding predictions of temperature (b).
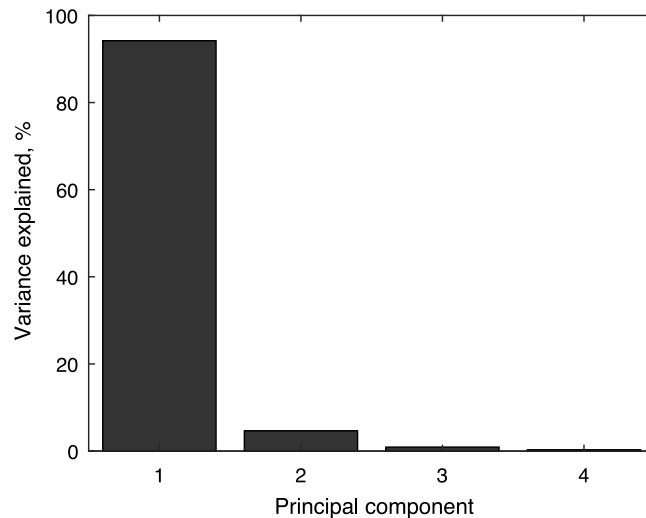
Fig. 2. Variance decomposition from principal component analysis of the RCP scenarios.

The RCPs are an example of a scenario set designed to provide reference pathways to scientists, integrated assessment modelers, and policy makers (Moss et al., 2010). They were selected to represent scientific agreement with regard to the probable trajectories of future climate and socioeconomic conditions. It is not the ultimate goal of such universal scenarios to serve as an input for approximations. We show below that common scenarios such as the RCPs are not suitable for estimating model parameters, and that the shape of scenarios should be determined by the purpose of their use.

Principal component analysis indicates that the RCPs are not likely to be a good set when it comes to estimating an emulator. Fig. 2 displays the variance decomposition implied by PCA, and shows that the first principal component carries more than 94% of the total variance in the set; the first principal component, which is a linear combination of the four RCPs, provides nearly as much information about emissions as the four scenarios do collectively.

The statistical approach to emulator construction uses the emissions–temperature pairs to estimate a single time series model where temperature is the dependent variable and lagged dependent and independent variables are on the right-hand side. The structure is similar to that of a pooled cross-section problem. Even though we use terms like "statistical approach", it needs to be emphasized that there is no underlying stochastic structure to the problem. The problem is really one of approximation, where we want to find a simple dynamic model relating temperature to emissions with small prediction errors.

2.3. Orthogonal emissions scenarios

To motivate our choice of emissions scenarios for the construction of an emulator, it is helpful to take a closer look at the underlying problem. Suppose some scientists or policy makers would like to consider an additional scenario of CO_2

emissions, $E(t)$, for years $t = t_0, t_1, \ldots, t_L$. Unfortunately, they cannot quickly access a complex dynamic climate model to forecast the resulting temperature anomalies, and therefore must resort to an emulator. Such an emulator relies on inputs from certain previous emissions scenarios, $E_k(t)$, $k = 0, 1, \ldots, m$, for which temperature predictions were gathered from that aforementioned complex model. But in which cases will an emulator trained on some existing scenarios, $E_k(t)$, provide a good approximation for the new scenario, $E(t)$?

We argue that the most likely or most popular emissions scenarios, such as the four afore-mentioned RCPs, are not necessarily a good set of input scenarios for training an emulator, because such scenarios are likely to be strongly correlated. (The four RCPs are a case in point.) Instead, a desirable condition of a good training set of emissions scenarios is that they span as large a set of reasonable scenarios as possible. Put differently, for any reasonable future scenario we would like there to exist weights, a_k, $k = 0, 1, \ldots, m$, such that

$$E(t) \approx \sum_{k=0}^{m} a_k E_k(t) \tag{1}$$

for all t.

The first of several questions that naturally arise is whether such an approximation of a new scenario by using previous scenarios is even possible. An answer in the affirmative will require that the set of scenarios employed for the approximation offers sufficient flexibility and scope to approximate any new (reasonable) scenario. RCP scenarios such as those depicted in Fig. 1(a) are (discrete) time series of CO_2 emissions with fixed start and end dates. Therefore, a natural method of approximating them is a least squares regression approach using a suitable set of basis vectors.

Trefethen (2013) provides an excellent introduction to the approximation of one-dimensional functions in theory and practice. He strongly advocates interpolation methods relying on Chebyshev polynomials for the practical approximation of functions on intervals, and literally refers to such methods as "unbeatable". Moreover, this particular family of orthogonal polynomials appears also to be the most popular choice for approximation problems in which we want to use more points than the maximum order of the Chebyshev polynomials. When the number of conditions exceeds the number of basis elements we need to resort to a regression approach.

In this paper we use Chebyshev polynomials to create basis vectors of CO_2 emissions. These newly created scenario vectors, E_k, are then used as explanatory variables in a regression with a new emissions scenario vector, E, as the dependent variable. If the scenario vectors E_k are a good basis for the approximation of an arbitrary scenario E, they can also be a good input set for creating an emulator that ultimately needs to provide forecasts for that new scenario.

2.3.1. Chebyshev polynomials and regression

For an excellent treatment of Chebyshev polynomials we refer interested readers to Trefethen (2013). Here, we only provide a brief introduction to Chebyshev polynomials, which suffices for our purposes.

Without loss of generality, consider the interval $[-1, 1]$. Consider the following recurrence relation:

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_k(x) = 2xP_{k-1}(x) - P_{k-2}(x) \text{ for } k = 2, 3, \ldots .$$

The polynomial[6] P_k is called the kth *Chebyshev polynomial*. The Chebyshev polynomials P_0 through P_7 are as follows:

$$\begin{aligned}
&P_0(x) = 1 &&P_4(x) = 8x^4 - 8x^2 + 1 \\
&P_1(x) = x &&P_5(x) = 16x^5 - 20x^3 + 5x \\
&P_2(x) = 2x^2 - 1 &&P_6(x) = 32x^6 - 48x^4 + 18x^2 - 1 \\
&P_3(x) = 4x^3 - 3x &&P_7(x) = 64x^7 - 112x^5 + 56x^3 - 7x.
\end{aligned}$$

For each $k \geq 1$, P_k is a polynomial of degree k with a leading coefficient 2^{k-1}; it satisfies $-1 \leq P_k(x) \leq 1$ for $x \in [-1, 1]$ and it has the k zeros

$$x_n = \cos\left(\frac{(2n+1)\pi}{2k}\right) \quad \text{for } n = 0, 1, \ldots, k - 1,$$

which all lie in the interval $[-1, 1]$. These zeros are also called *Chebyshev nodes*.

For our study, an important property of Chebyshev polynomials is their discrete orthogonality property. If $\{x_n\}_{n=0}^{k-1}$ are the k zeros of P_k, then for all $i, j \leq k$

$$\sum_{n=0}^{k-1} P_i(x_n)P_j(x_n) = \begin{cases} 0 & : i \neq j, \\ k & : i = j = 0, \\ k/2 & : i = j \neq 0. \end{cases}$$

So, for $i \neq j$, the vectors of the values of P_i and of P_j at the Chebyshev nodes $\{x_n\}_{n=0}^{k-1}$ are orthogonal and thus uncorrelated.

---

[6] We deviate from the standard notation and denote Chebyshev polynomials by P instead of by T or C. We need the letters T and C for the time series of temperatures and cumulative CO_2 emissions, respectively.
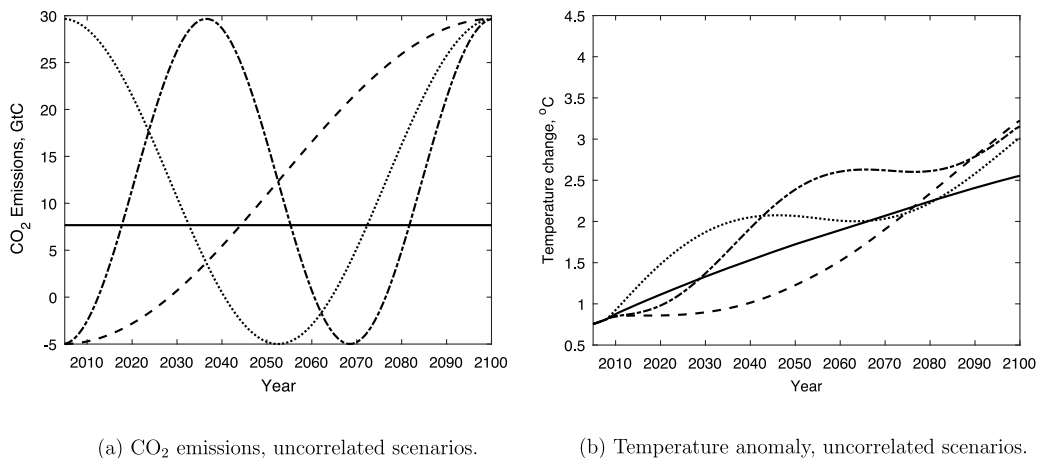
(a) CO_2 emissions, uncorrelated scenarios.

(b) Temperature anomaly, uncorrelated scenarios.

Fig. 3. Uncorrelated emissions scenarios (a) and corresponding predictions of temperature (b).

All Chebyshev zeros lie in the interval $[-1, 1]$. For an approximation on a general interval, $[a, b]$, we use the standard linear transformation,

$$t = \left(\frac{b - a}{2}\right)x + \frac{a + b}{2}, \tag{2}$$

for $t \in [a, b]$ and $x \in [-1, 1]$.

To apply the standard Chebyshev regression approach, we could now proceed as follows: First, choose a number of n Chebyshev nodes – that is, all zeros of P_n – and evaluate the Chebyshev polynomials of degrees 0 to m at these n points in order to obtain $m + 1$ basis vectors of length n. Second, determine the values of the to-be-approximated new scenario at the n nodes. Since the n nodes would not coincide with integer values – the years 2005, 2006, . . . , 2100 – we would need to use interpolated values from the new scenario. Third, determine the regression coefficient by minimizing some criterion—for example, the sum of squared residuals.

We do not pursue this standard approach because it encounters the following problem: Both the input scenarios and the temperature output of a typical climate change model are annual figures—that is, they are given for the years 2005, 2006, . . . , 2100. Therefore, we must determine the values of the Chebyshev polynomials at these integer points, instead of at the Chebyshev nodes, when we want to enter the corresponding base scenario into a model such as MAGICC. But then the adjusted values of the base vectors at the integer points are no longer uncorrelated. While some small correlation among the base scenario vectors may do only little harm to the regression, we would like to avoid it if possible. Therefore, we generate the data for the regression approach in a slightly different way.

2.3.2. Design of uncorrelated scenarios

The first step of our regression approach is identical to that of the general approach described above. We choose n to be the number of years in our simulated data; here $n = 96$, since we consider scenarios for the time period 2005–2100. For the n zeros of P_n, we evaluate the Chebyshev polynomials of degrees 0 to m at these n points in order to obtain $m + 1$ basis vectors of length n. By construction, these $m + 1$ basis vectors are pairwise uncorrelated. Now we treat the ith element of each basis n-vector as the value for the year $2004 + i$. As a result, in the second step, we no longer need to interpolate values for the new scenario but can just take the given scenario values. Using this different vector as the independent variable we can determine regression coefficients and check whether the new scenario can be approximated well.

When we determine and evaluate the linear approximation, $\sum_{k=0}^{m} a_k E_k(t)$, for a new scenario, $E(t)$, the transformation of the Chebyshev polynomials to the domain [2005, 2100] suffices and we do not need to adjust the range value. However, before we can feed the corresponding emissions scenarios into a model such as MAGICC, we also need to linearly transform the range of the polynomials in order to obtain CO_2 emissions scenarios of reasonable sizes. Since the four RCPs in Fig. 1(a) span a range of 29.6708 ℃, we use (2) to transform the m scenarios based on the Chebyshev polynomials of degrees 1 to m to the range $[-5, 29.6708]$. The zero-degree polynomial is an exception: it corresponds to a scenario of constant annual emissions, or a steady state of the economy, and is therefore set to the last historical value of CO_2 emissions.

The resulting values represent CO_2 emissions levels for the $n = 96$ years, 2005–2100. For the baseline case for our analysis, we set $m = 3$ so that using four uncorrelated scenarios can be compared to using the four RCPs. Fig. 3(a) depicts the resulting CO_2 emissions scenarios. When we enter them into MAGICC we must also specify scenarios for some other gases. For simplicity, in this study we set the annual emissions of all other gases in the years 2005–2100 to their average
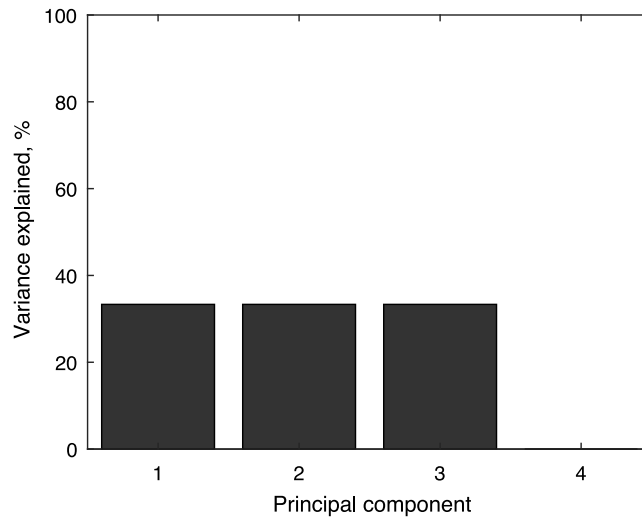
Fig. 4. Variance decomposition from principal component analysis of the four uncorrelated scenarios.

(a) Approximation of RCP scenarios.
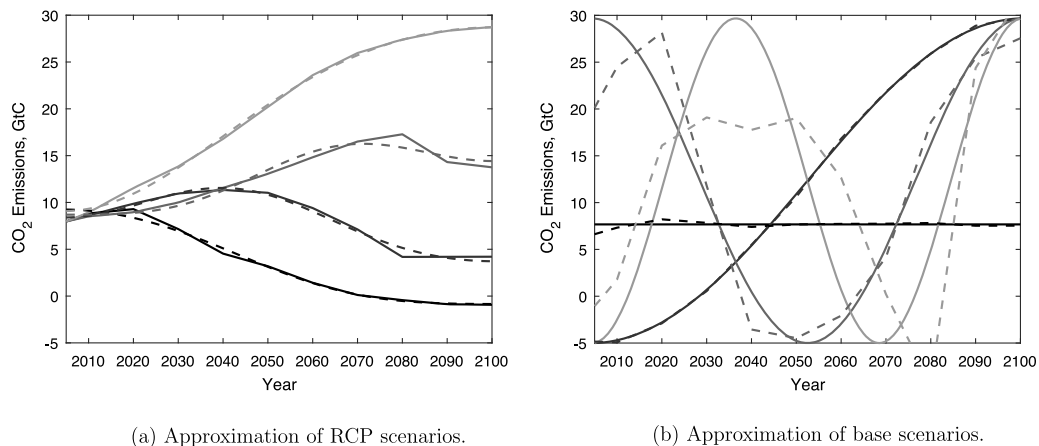

(b) Approximation of base scenarios.

Fig. 5. Approximation of the RCP scenarios with uncorrelated base scenarios (a) and approximation of base scenarios with the RCP scenarios (b).

levels across the four RCP scenarios. The emissions of all gases for all years prior to 2005 are kept at their historical values. Fig. 3(b) shows the resulting temperature anomalies forecasted by MAGICC for the four uncorrelated scenarios.

Fig. 4 shows the variance decomposition from a principal component analysis of the four base scenarios. By construction, each of the first three components explains a third of the variance. The degree-zero polynomial (a constant) obviously does not carry any variance but is nevertheless included in the set of the base scenarios. Recall that a good set of input scenarios for emulation should provide a base for the linear approximation (1). For a successful linear approximation, the set of basis vectors has to include a constant.

The set of base scenarios, besides carrying the zero correlation property, should trigger a strong enough response in the approximated system. The advantage of the scenarios based on the Chebyshev polynomials is that they include the chosen extreme values of emissions. Each Chebyshev polynomial of degree $k \geq 1$ has $k + 1$ extrema, with their values at the endpoints of the range $[-1, 1]$. Each corresponding emissions scenario scaled to a range $[a, b]$ reaches the minimum and maximum emissions levels a and b at those extreme points. This property ensures that we observe the simulated response of the climate system when the CO_2 emissions levels reach their extreme values, at a slower or faster rate.

Clearly, future emissions paths will not look anything like the base scenarios based on the degree two or degree three Chebyshev polynomials. But the task here is not to discuss what the likely emissions paths are. The task is to find a collection of emissions scenarios that enable us to extract as much information as possible from a mathematical model of the climate such as MAGICC. The key property of the uncorrelated base scenarios is that they produce good approximations of the standard RCP scenarios, which are of particular interest in the literature. Fig. 5(a) demonstrates that the designed uncorrelated scenarios, while unrealistic, provide an excellent basis for close approximations of the RCPs; in fact, the root-mean-squared error (RMSE) of the approximation is only 0.36 gigatons of carbon per year ($GtCyr^{-1}$). The RCPs, designed

to represent likely paths, do not form a suitable collection for emulator construction. Fig. 5(b) shows that they do not approximate the uncorrelated base scenarios well, and produce an RMSE of 2.55 GtCyr^{-1}. This result is consistent with PCA, which shows that the four standard RCPs together contain little more information than the best one on its own.

2.4. Construction of an emulator

With the uncorrelated emissions scenarios at hand, we can proceed to building an emulator of a climate model of our choice. To generate the data for emulation, we run the four designed emissions scenarios in MAGICC and collect its temperature forecasts for each of them. The particular structure of an emulator is defined by its purpose and potential application. In this paper we provide an example of an emulator that can be first of all used in integrated assessment modeling, where dynamic models need a simple way of producing forecasts for temperature rise if they are to assess its feedback effect on the economy.

Our emulator takes the following general form of a dynamic linear model:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \sum_{j=2}^{J} \beta_j X_{j,t-1} + \varepsilon_t, \tag{3}$$

where Y_t is the predicted variable in year t, its first lag is the first covariate, $X_{j,t}$ is the further jth covariate in year t, and J is the number of covariates in the model. To capture the residual autocorrelation, we assume that model errors follow an ARMA(1,1) process and that its residuals are normally distributed,

$$\varepsilon_t = a\varepsilon_{t-1} + bu_{t-1} + u_t, \quad u_t \overset{i.i.d}{\sim} N(0, \sigma^2). \tag{4}$$

The temperature anomaly forecasts produced by MAGICC are the benchmark values of the predicted variable Y. The candidates for covariates come from the input scenarios and depend on the ultimate goal of the emulation. An emulator that predicts temperature response to CO_2 emissions can serve as a substitute for the entire climate module in an integrated assessment model. Alternatively, an IAM can preserve its carbon cycle representation as a distinct feature—in this case, emulation covers the transition from concentrations to temperature rise. Consequently, we consider CO_2 emissions, cumulative CO_2 emissions, and CO_2 concentrations as the available set of variables to include in X.

For each potential set of covariates, we pool the simulated data from the four runs of the climate model together and find the best values for the parameters using a maximum likelihood estimator. To test our model, we take the CO_2 emissions scenarios given by the four RCPs—they constitute our out-of-sample testing set. We use the average RMSE of predictions for this out-of-sample set of scenarios to assess the quality of their predictions.

3. Results

Using different combinations of the available covariates, we choose an exact form for the emulator (3). This section presents the results for the best two- and three-dimensional models, according to the quality of their out-of-sample predictions.

3.1. Best-performing low-dimensional model

In all our specifications, the dependent variable is the temperature anomaly T with respect to the preindustrial temperature level—that is, the increase in temperature since 1765. We find that the following model specification with cumulative CO_2 emissions, C_t, as a single exogenous variable produces the best predictions:

$$T_t = \beta_0 + \beta_1 T_{t-1} + \beta_2 C_{t-1} + \varepsilon_t,$$
$$\varepsilon_t = a\varepsilon_{t-1} + bu_{t-1} + u_t, \quad u_t \overset{i.i.d}{\sim} N(0, \sigma^2). \tag{5}$$

Here, cumulative CO_2 emissions, C_t, are accumulated from the year 1765 to t and measured in GtC · 10^3. The scenarios themselves provide values for the lagged cumulative emissions starting in 2005. For each uncorrelated emissions scenario, we obtain temperature anomalies from MAGICC for $t \in \{2005, 2006, \ldots, 2100\}$. And so, the number of data points for the estimation of this model is 96 times the number of scenarios, m. The first row of Table 1 reports the estimated values for the parameters of model (5) and the model's average errors of out-of-sample prediction.[7] The average RMSE across all four testing scenarios is only about 0.03 ℃; the prediction errors are on average 1.62% of the level of temperature rise.

---

[7] The estimated high values of the coefficient a bring the error process (4) close to unit root, suggesting a transformation into a first-difference model (Hendry and Juselius, 2000). For our purpose, however, it is essential to preserve the model in the form of a law of motion for temperature, so that it can be implemented in a typical dynamic integrated assessment model. By the design of our problem, we are approximating deterministic, gradually evolving time series paths and – however small the estimated errors are – we indeed expect the autocorrelation to be very high. To adequately model the dynamics of the process for IAMs, we keep the current autoregressive distributed lag form of model (5) and let the errors capture the autocorrelation, with the most parsimonious specification. The specific ARMA(1,1) is the shortest form that absorbs this autocorrelation to the extent that the Phillips–Perron test rejects the null hypothesis of unit root. Model (5) is thus not suitable for analyzing the intrinsic interaction between the involved time series processes —it serves merely for the task of approximating the dynamics of the system with maximum simplicity.

Table 1
Approximation results for different model specifications. The last two columns report the average errors of prediction for the testing set of the RCP scenarios. RMSE is the root-mean-squared error; MAPE is the mean absolute percentage error. The estimated standard errors of the coefficients are given in parentheses.

Model	β_0	β_1	β_2	β_3	a	b	σ	RMSE	MAPE
(5)	0.2500 (0.0344)	0.7650 (0.0227)	0.3632 (0.0345)		0.9805 (0.0031)	0.2128 (0.0636)	0.0022 (0.0001)	0.0338	1.62%
(6)	0.1188 (0.0401)	0.6874 (0.0193)	0.1503 (0.0089)		0.9878 (0.0012)	0.1854 (0.0558)	0.0018 (0.0001)	0.0426	1.75%
(7)	0.1230 (0.0396)	0.6820 (0.0209)	0.0286 (0.0399)	0.1445 (0.0121)	0.9873 (0.0013)	0.1832 (0.0572)	0.0018 (0.0001)	0.0411	1.68%
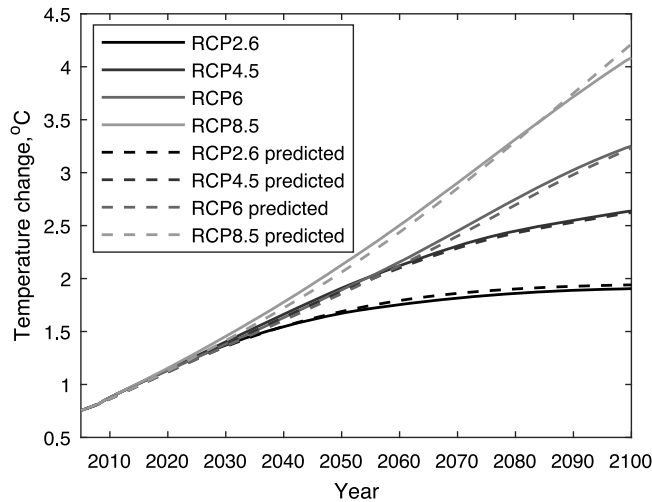
Fig. 6. Out-of-sample temperature predictions of the best-fitting model (5).

Fig. 6 demonstrates that the model produces very accurate predictions. With cumulative emissions being the best candidate for an exogenous variable, our results are consistent with recent studies that suggest a linear–proportional relationship between global warming and the level of cumulative CO_2 emissions (Allen et al., 2009; Zickfeld et al., 2009; Matthews et al., 2009; IPCC, 2013). In other words, even when only CO_2 emissions data are available, the temperature anomaly can be approximated instantly, leaving out the emulation of the carbon cycle and any other intermediary steps.

The required number of scenarios becomes a very important question when complex, computationally expensive models, such as AOGCMs, are emulated. Our approach suggests that emulation does not necessarily benefit from running the complex model as many times as possible. Fig. 7 demonstrates that only up to three runs contribute significantly to the precision of the approximation, after which the prediction errors level off. The key to successful emulation in such a case is that the input data set is efficiently designed before the emulated climate models are run. Furthermore, our recommended prediction model can be embedded within a dynamic system of equations, such as is often employed in integrated assessment models. Many IAMs – in particular those focusing on intrinsic uncertainty in social decision-making – suffer from poor representations of the climate system due to computational constraints. Given the accuracy of our predictions, integrated assessment modelers may now include this simple yet accurate low-dimensional mapping of emissions to temperature levels in their models.

Given the structure of the two dynamic equations from above, $T_t = f(T_{t-1}, C_{t-1}, \varepsilon_t)$ and $\varepsilon_t = g(\varepsilon_{t-1}, u_{t-1}, u_t)$, an economic IAM would also require adding a dynamic equation for cumulative emissions, $C_t = h(C_{t-1}, E_t)$, where E_t is some emissions scenario resulting from the model's economic framework.

3.2. Alternative specifications

Here, we present alternative functional forms of two-dimensional and three-dimensional representations of the climate system.

The alternative two-dimensional model includes CO_2 concentrations, S_t, measured in $ppm \cdot 10^2$ in year t as an exogenous covariate,

$$T_t = \beta_0 + \beta_1 T_{t-1} + \beta_2 S_{t-1} + \varepsilon_t,$$
$$\varepsilon_t = a\varepsilon_{t-1} + bu_{t-1} + u_t, \quad u_t \overset{i.i.d}{\sim} N(0, \sigma^2). \tag{6}$$

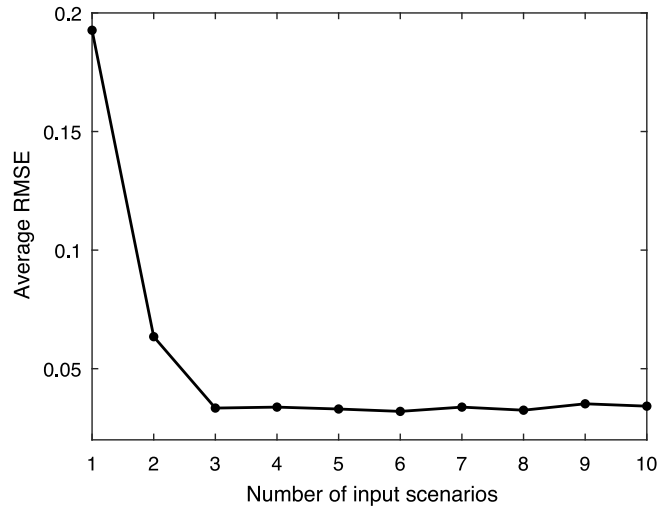**Fig. 7.** Average out-of-sample prediction error as a function of the number of uncorrelated scenarios used as input.

(a) Two-dimensional model (6).

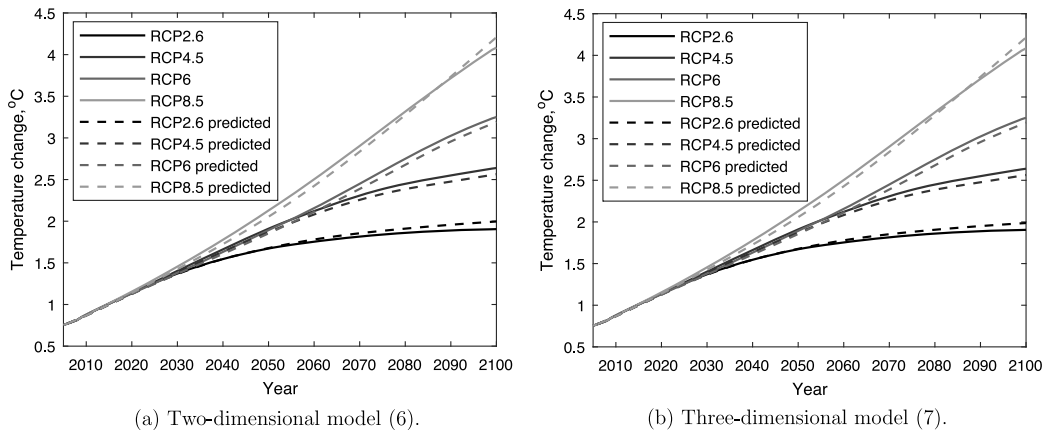(b) Three-dimensional model (7).

Fig. 8. Out-of-sample temperature predictions of the alternative models.

The results reported in Table 1 show that cumulative CO_2 emissions are a better predictor; however, if only CO_2 concentrations are available, the resulting emulator would also generate good predictions.

The suggested three-dimensional model includes both CO_2 concentrations and cumulative emissions,

$$T_t = \beta_0 + \beta_1 T_{t-1} + \beta_2 S_{t-1} + \beta_3 C_{t-1} + \varepsilon_t,$$
$$\varepsilon_t = a\varepsilon_{t-1} + bu_{t-1} + u_t, \quad u_t \overset{i.i.d}{\sim} N(0, \sigma^2). \tag{7}$$

The results in the last row of Table 1 indicate that the extended model does not outperform the two lower-dimensional ones.

Fig. 8 shows the out-of-sample predictions of temperature levels produced by the two alternative models. Visually, the approximations delivered by the alternative models look similar to those of the best model (see (5)).

3.3. Performance verification

Advances in climate and economic research bring with them new knowledge about likely paths of socioeconomic development and estimated climate impacts. An ongoing process of scenario creation addresses the research community's need for new scenarios that are consistent with the current understanding of possible global developments and that reflect the associated uncertainty (Moss et al., 2010). The most recent example of such is the concept of Shared Socioeconomic Pathways, which complement existing scenarios by including various paths of socioeconomic development (Riahi et al., 2016; O'Neill et al., 2014).
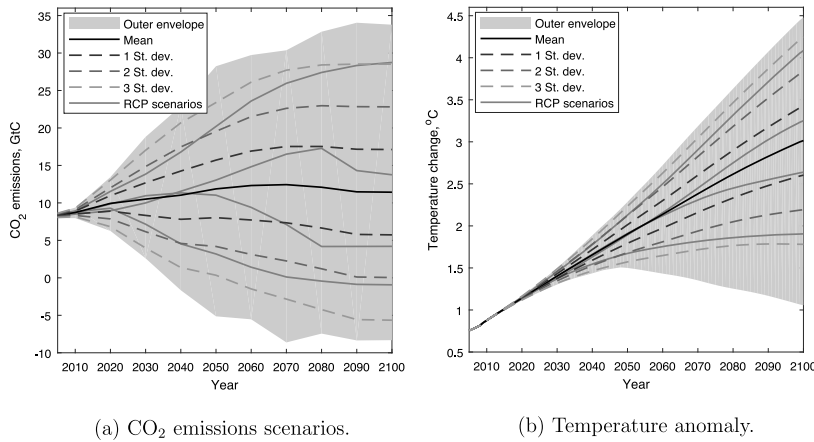
(a) CO$_2$ emissions scenarios.

(b) Temperature anomaly.

Fig. 9. Statistical distributions of the simulated scenarios (a) and of the corresponding temperature predictions of MAGICC (b).

As new generations of scenarios are incorporated into integrated assessment modeling, they become a new common base for scientific research in this field. However, models trained on traditionally used scenarios, such as the RCPs, might perform poorly as these new scenarios come into play.

It is therefore important for us to ensure that the proposed emulator works on the entire range of scenarios considered plausible in the literature. Since the RCPs were created to represent the wide range of scenarios present in the prior literature, we would like to assess the performance of our model on this range of scenarios. In particular, we construct a stochastic process that allows us to generate a large testing set of CO$_2$ emissions paths similar in their nature to those present in the RCPs. We then verify whether the temperature levels predicted by our emulator for the simulated scenarios are close to those produced by MAGICC.

In principle, the following simple stochastic process can generate the decadal CO$_2$ emissions paths of the four RCP scenarios[8]:

$$E_\tau = E_{\tau-1} + \varepsilon_\tau, \ \ \varepsilon_\tau \sim N(\mu_\tau, \sigma_\tau), \tag{8}$$

where E_τ is the annual CO$_2$ emissions level in decade τ. The parameters μ_τ and σ_τ can be estimated from the four RCP scenarios for each τ. Using the method of moments,

$$\hat{\mu}_\tau = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_{i,\tau}, \quad \hat{\sigma}_\tau = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\varepsilon}_{i,\tau} - \hat{\mu}_\tau\right)^2}, \ \text{for } \tau = 2010, 2020, \ldots, 2100,$$

where $\hat{\varepsilon}_{i,\tau} = E_{i,\tau} - E_{i,\tau-1}$ and $n = 4$ for each decade τ.

We generate a testing set of 10,000 realizations of the stochastic process (8) (Fig. 9) and run these realizations in MAGICC to obtain the benchmark predictions of temperature levels. We then compare the predictions of our emulator with the benchmark values in the cases of training the emulator on the uncorrelated scenarios and on the RCP scenarios. In the case of using the uncorrelated scenarios as a training set and the best-fitting model (5) as an emulator, we obtain an average RMSE of only 0.02 ℃. The corresponding error in the case of using the RCP scenarios is almost twice as big. A statistical comparison of the performance of the uncorrelated scenarios with that of the RCP scenarios as input sets for models (5) and (6) (Fig. 10) confirms that the average prediction error decreases significantly when the designed input paths are uncorrelated.

3.4. Performance on alternative model settings

The predictions of complex climate models are known to diverge greatly from one another—even when calibrated to the same observational data and run with the same forcing scenarios they span a wide range of possible system forecasts. These discrepancies might stem from differences in assumptions, modeled components, and the structure of those components (Tebaldi and Knutti, 2007), and can have a significant impact on the accuracy of emulating models (Meinshausen et al., 2011a).

So far, we have used in our analysis only one (default) combination of the 20 AOGCMs and 10 carbon cycle models emulated in MAGICC (Meinshausen et al., 2011b). However, as documented by numerous intercomparison projects, there

---

[8] Because the RCP scenarios are specified on a decadal scale, we first generate decadal scenarios and then linearly interpolate them to use the resulting annual paths.
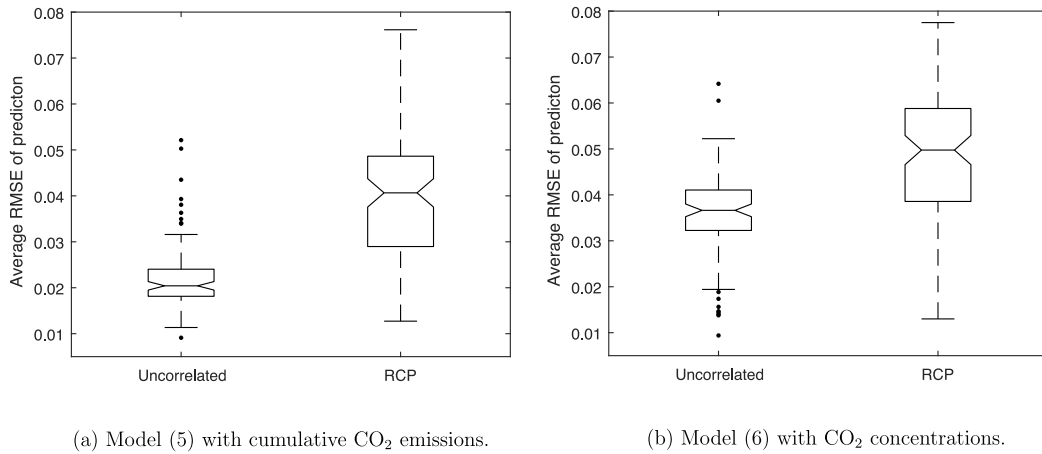
(a) Model (5) with cumulative CO_2 emissions.   (b) Model (6) with CO_2 concentrations.

Fig. 10. Out-of-sample prediction errors of models (5) (a) and (6) (b) for simulated scenarios in the cases of using covariates from uncorrelated scenarios and RCP scenarios for training the models.
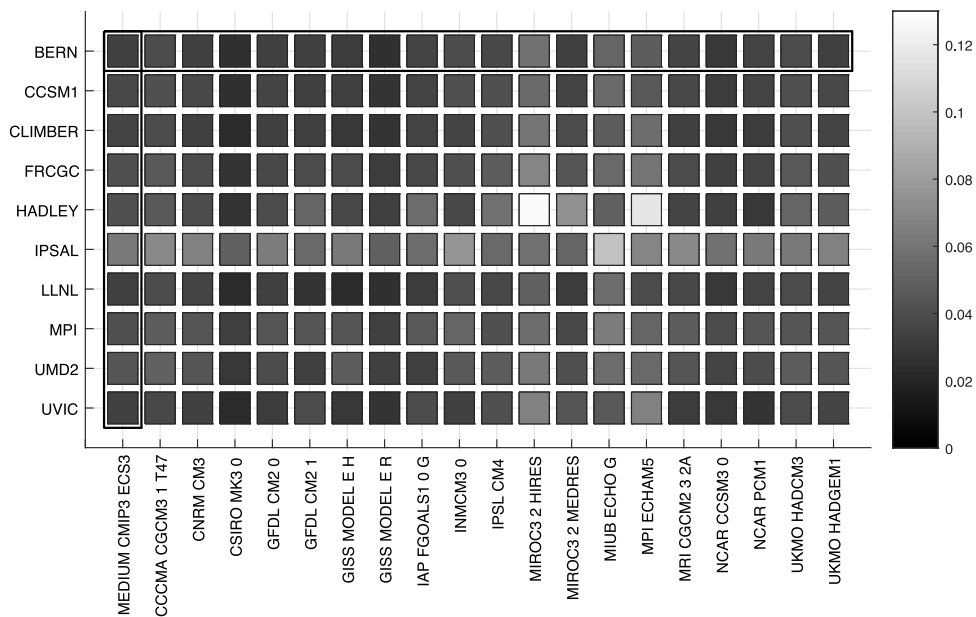

Fig. 11. RMSEs for combinations of 20 AOGCM settings (horizontal axis) and 10 carbon cycle settings (vertical axis). The settings used in MAGICC as default values are marked with black rectangles.

is great variability among complex models in terms of their predictions of climate response to emissions scenarios (Taylor et al., 2012; Palmer et al., 2005).

We would like, therefore, to ensure that our approach is not restricted to a single climate model, but may be applied to any of the existing ones. To do so, we additionally check whether the method performs equally well for the 200 different sets of parameters obtained by combining each of the 20 AOGCM settings with each of the 10 carbon cycle settings emulated with MAGICC. Fig. 11 shows the heat map of the magnitude of RMSE in each of these 200 cases.

We find that the average RMSE across all settings is only about 0.04 ℃, and that in most of the 200 model combinations in MAGICC the average RMSE does not exceed 0.07 ℃. In general, we find that all combinations produce low approximation errors.[9] The insights from Fig. 11 could be useful for improving emulation exercises in the future, and for estimating the certainty levels of models' predictions. Overall, we conclude that our emulation technique and the recommended low-dimensional model perform very well for the plausible settings of the underlying model parameters.

---

[9] There are some notable differences among the individual climate and carbon cycle models, such as "Hadley" and "IPSAL", which are both known for strong carbon cycle feedbacks.

4. Conclusion

New emissions scenarios and socioeconomic pathways are constructed on a regular basis. As the work on the next IPCC Assessment Report has commenced, emulation of large and complex climate models will certainly be on the research agenda. The known resource limitations of running large climate models call for efficient emulation techniques. Our study complements the existing emulation literature by addressing the task of designing efficient input scenarios for emulation. We recommend the use of uncorrelated emissions scenarios for an efficient yet accurate approximation of climate models. These uncorrelated scenarios, based on Chebyshev polynomials, display quite unrealistic emissions paths. However, the purpose of using such scenarios is purely technical—namely, to extract as much information as possible from the complex model.

Using the global temperature anomaly as a predicted response variable, we produce an econometric model—a low-dimensional system of mapping emissions to temperature levels for the twenty-first century. Our simulations confirm that the model performs well on conventional scenarios: the precision of approximation stays high under various settings of climate and carbon cycle parameters. The designed system of equations can be directly implemented in the dynamic stochastic general equilibrium models often used in macroeconomics, allowing one to study optimal policies for dealing with global warming under conditions of uncertainty in terms of social decision-making.

Appendix

List of acronyms and abbreviations

AOGCM	Atmosphere–Ocean General Circulation Model
IAM	Integrated assessment model
IPCC	Intergovernmental Panel on Climate Change
MAGICC	Model for the Assessment of Greenhouse-gas Induced Climate Change
PCA	Principal component analysis
RCP	Representative Concentration Pathway

References

Allen, M.R., Frame, D.J., Huntingford, C., Jones, C.D., Lowe, J.A., Meinshausen, M., Meinshausen, N., 2009. Warming caused by cumulative carbon emissions towards the trillionth tonne. Nature 458 (7242), 1163–1166. http://dx.doi.org/10.1038/nature08019.

Brumm, J., Scheidegger, S., 2017. Using adaptive sparse grids to solve high-dimensional dynamic models. Econometrica 85 (5), 1575–1612. http://dx.doi.org/10.3982/ECTA12216.

Cai, Y., Judd, K.L., Lontzek, T.S., 2015. The Social Cost of Carbon with Economic and Climate Risks, arXiv:1504.06909, https://arxiv.org/abs/1504.06909.

Castelletti, A., Galelli, S., Ratto, M., Soncini-Sessa, R., Young, P., 2012. A general framework for dynamic emulation modelling in environmental problems. Environ. Model. Softw. 34, 5–18. http://dx.doi.org/10.1016/j.envsoft.2012.01.002.

Castruccio, S., Genton, M.G., 2016. Compressing an ensemble with statistical models: An algorithm for global 3D spatio-temporal temperature. Technometrics 58 (3), 319–328. http://dx.doi.org/10.1080/00401706.2015.1027068.

Castruccio, S., McInerney, D.J., Stein, M.L., Crouch, F.L., Jacob, R.L., Moyer, E.J., 2014. Statistical emulation of climate model projections based on precomputed GCM runs. J. Clim. 27 (5), 1829–1844. http://dx.doi.org/10.1175/JCLI-D-13-00099.1.

Collins, W.D., Bitz, C.M., Blackmon, M.L., Bonan, G.B., Bretherton, C.S., Carton, J.A., Chang, P., Doney, S.C., Hack, J.J., Henderson, T.B., Kiehl, J.T., Large, W.G., McKenna, D.S., Santer, B.D., Smith, R.D., 2006. The community climate system model version 3 (CCSM3). J. Clim. 19 (11), 2122–2143. http://dx.doi.org/10.1175/JCLI3761.1.

Crost, B., Traeger, C.P., 2013. Optimal climate policy: Uncertainty versus Monte Carlo. Econom. Lett. 120 (3), 552–558. http://dx.doi.org/10.1016/j.econlet.2013.05.019.

Dringnei, D., Forest, C., Nychka, D., 2008. Parameter stimation for computationally intensive nonlinear regression with an application to climate modeling. Ann. Appl. Stat. 2 (4), 1217–1230. http://dx.doi.org/10.1214/08-AOAS210.

EPA, 2010. Peer Review of ADAGE and IGEM, Environmental Protection Agency, U.S. Washington, D.C..

Guinness, J., Hammerling, D., 2018. Compression and conditional emulation of climate model output. J. Amer. Statist. Assoc. 113 (521), 56–67. http://dx.doi.org/10.1080/01621459.2017.1395339.

Hendry, D.F., Juselius, K., 2000. Explaining cointegration analysis: Part 1. Energy J. 21 (1), 1–42. https://www.jstor.org/stable/41322853.

Holden, P., Edwards, N., 2010. Dimensionally reduced emulation of an AOGCM for application to integrated assessment modelling. Geophys. Res. Lett. 37 (21), L21707. http://dx.doi.org/10.1029/2010GL045137.

Hope, C., 2013. Critical issues for the calculation of the social cost of CO_2: why the estimates from PAGE09 are higher than those from PAGE2002. Clim. Change 117 (3), 531–543. http://dx.doi.org/10.1007/s10584-012-0633-z.

IPCC, 2013. Summary for policymakers. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, http://dx.doi.org/10.1017/CBO9781107415324.004.

IPCC, 2014. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)], IPCC, Geneva, Switzerland, 151 pp. in IPCC AR5 Synthesis Report website, http://ar5-syr.ipcc.ch/.

IPCC-DDC, 1998. IPCC Data Distribution Centre, www.ipcc-data.org.

IPCC-TGICA, 2007. General Guidelines on the Use of Scenario Data for Climate Impact and Adaptation Assessment, Version 2. Prepared by T.R. Carter on behalf of the Intergovernmental Panel on Climate Change, Task Group on Data and Scenario Support for Impact and Climate Assessment, 66 pp., http://www.ipcc-data.org/guidelines/index.html#ftnote3.

Jensen, S., Traeger, C.P., 2014. Optimal climate change mitigation under long-term growth uncertainty: stochastic integrated assessment and analytic findings. Eur. Econ. Rev. 69, 104–125. http://dx.doi.org/10.1016/j.euroecorev.2014.01.008.

Judd, K.L., Maliar, L., Maliar, S., 2011. Numerically stable and accurate stochastic simulation approaches for solving dynamic economic models. Quantit. Econom. 2, 173–210. http://dx.doi.org/10.3982/QE14.

Kaufmann, R.K., Kauppi, H., Mann, M.L., Stock, J.H., 2013. Does temperature contain a stochastic trend: linking statistical results to physical mechanisms. Climatic Change 118 (3–4), 729–743. http://dx.doi.org/10.1007/s10584-012-0683-2.

Lontzek, T.S., Cai, Y., Judd, K.L., Lenton, T.M., 2015. Stochastic integrated assessment of climate tipping points indicates the need for strict climate policy. Nature Clim. Change 5, 441–444. http://dx.doi.org/10.1038/nclimate2570.

Maliar, L., Maliar, S., 2015. Merging simulation and projection approaches to solve high-dimensional problems with an application to a new Keynesian model. Quantit. Econom. 6 (1), 1–47. http://dx.doi.org/10.3982/QE364.

Mastrandrea, M.D., Schneider, S.H., 2004. Probabilistic integrated assessment of "dangerous" climate change. Science 304 (5670), 571–575. http://dx.doi.org/10.1126/science.1094147.

Matthews, H.D., Gillett, N.P., Stott, P.A., Zickfeld, K., 2009. The proportionality of global warming to cumulative carbon emissions. Nature 459 (7248), 829–832. http://dx.doi.org/10.1038/nature08047.

Meinshausen, M., Raper, S., Wigley, T., 2011a. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6–Part 1: Model description and calibration. Atmos. Chem. Phys. 11 (4), 1417–1456. http://dx.doi.org/10.5194/acp-11-1417-2011.

Meinshausen, M., Wigley, T., Raper, S., 2011b. Emulating atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6–Part 2: Applications. Atmos. Chem. Phys. 11 (4), 1457–1471. http://dx.doi.org/10.5194/acp-11-1457-2011.

Moss, R.H., Edmonds, J.A., Hibbard, K.A., Manning, M.R., Rose, S.K., van Vuuren, D.P., Carter, T.R., Emori, S., Kainuma, M., Kram, T., Meehl, G.A., Mitchell, J.F.B., Nakicenovic, N., Riahi, K., Smith, S.J., Stouffer, R.J., Thomson, A.M., Weyant, J.P., Wilbanks, T.J., 2010. The next generation of scenarios for climate change research and assessment. Nature 463, 747–756. http://dx.doi.org/10.1038/nature08823.

Newbold, S.C., Griffiths, C., Moore, C., Wolverton, A., Kopits, E., 2013. A rapid assessment model for understanding the social cost of carbon. Clim. Change Econom. 04 (01), 1350001. http://dx.doi.org/10.1142/S2010007813500012.

O'Neill, B.C., Kriegler, E., Riahi, K., Ebi, K.L., Hallegatte, S., Carter, T.R., Mathur, R., van Vuuren, D.P., 2014. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. Clim. Change 122 (3), 387–400. http://dx.doi.org/10.1007/s10584-013-0905-2.

Palmer, T.N., Shutts, G.J., Hagedorn, R., Doblas-Reyes, F.J., Jung, T., Leutbecher, M., 2005. Representing model uncertainty in weather and climate prediction. Ann. Rev. Earth Planet. Sci. 33, 163–193. http://dx.doi.org/10.1146/annurev.earth.33.092203.122552.

Pretis, F., 2019. Econometric Modelling of Climate Systems: The Equivalence of Energy Balance Models and Cointegrated Vector Autoregressions, J. Econom., forthcoming.

Riahi, K., van Vuuren, D.P., Kriegler, E., Edmonds, J., O'Neill, B.C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J.C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Silva, L.A.D., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J.C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., Tavoni, M., 2016. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global Environ. Change http://dx.doi.org/10.1016/j.gloenvcha.2016.05.009.

Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the Experiment Design. Bull. Amer. Meteorol. Soc. 93 (4), 485–498. http://dx.doi.org/10.1175/BAMS-D-11-00094.1.

Tebaldi, C., Knutti, R., 2007. The use of the multi-model ensemble in probabilistic climate projections. Phil. Trans. R. Soc. A 365 (1857), 2053–2075. http://dx.doi.org/10.1098/rsta.2007.2076.

Trefethen, L.N., 2013. Approximation Theory and Approximation Practice. SIAM.

van Vuuren, D.P., Edmonds, J., Kainuma, M., Riahi, K., Thomson, A., Hibbard, K., Hurtt, G.C., Kram, T., Krey, V., Lamarque, J.-F., Masui, T., Meinshausen, M., Nakicenovic, N., Smith, S.K., 2011. The representative concentration pathways: an overview. Clim. Change 109, 5–31. http://dx.doi.org/10.1007/s10584-011-0148-z.

Webster, M., Santen, N., Parpas, P., 2012. An approximate dynamic programming framework for modeling global climate policy under decision-dependent uncertainty. Computat. Manag. Sci. 9 (3), 339–362. http://dx.doi.org/10.1007/s10287-012-0147-1.

Yeager, S.G., Shields, C.A., Large, W.G., Hack, J.J., 2006. The low-resolution CCSM3. J. Clim. 19 (11), 2545–2566. http://dx.doi.org/10.1175/JCLI3744.1.

Young, P.C., Ratto, M., 2011. Statistical emulation of large linear dynamic models. Technometrics 53 (1), 29–43. http://dx.doi.org/10.1198/TECH.2010.07151.

Zickfeld, K., Eby, M., Matthews, H.D., Weaver, A.J., 2009. Setting cumulative emissions targets to reduce the risk of dangerous climate change. Proc. Natl. Acad. Sci. 106 (38), 16129–16134. http://dx.doi.org/10.1073/pnas.0805800106.