



## ABSTRACT

5  
6 We describe a new approach for emulating the output of a fully coupled climate model under  
7 arbitrary forcing scenarios that is based on a small set of precomputed runs from the model.  
8 We express temperature and precipitation as simple functions of the past trajectory of at-  
9 mospheric CO<sub>2</sub> concentrations and fit a statistical model using a limited set of training runs.  
10 We demonstrate that the approach is a useful and computationally efficient alternative to  
11 pattern scaling that captures the nonlinear evolution of spatial patterns of climate anomalies  
12 inherent in transient climates. The approach does as well as pattern scaling in all circum-  
13 stances and substantially better in many; it is not computationally demanding; and, once  
14 the statistical model is fit, produces emulated climate output effectively instantaneously. It  
15 may therefore find wide application in climate impacts assessments and other policy analyses  
16 requiring rapid climate projections.

# 1. Introduction

The wide consensus among the scientific community that climate is changing and will almost certainly produce detrimental impacts for humanity (IPCC AR4; Meehl et al. (2007)) means that attention is increasingly turning to evaluating the magnitude of those impacts and possible policies to reduce them. Atmosphere-Ocean General Circulation Models (AOGCMs) are state-of-art tools for producing climate predictions based on our best understanding of the radiative effects of CO<sub>2</sub> and other anthropogenic forcing agents and the complex dynamical feedbacks of the Earth's climate system. However, the computational demands of AOGCMs preclude or limit their use in the context of integrated assessment models (IAMs) used to estimate climate damages and the cost-benefit trade-offs of potential mitigation actions. Analyses that involve optimal policy determination or uncertainty quantification require repeated iterations of climate projections in response to forcing trajectories over the decadal or centennial timescale, which is computationally prohibitive with AOGCMs. For IAMs whose only climate input is global mean temperature (GMT), climate projections can be provided instead by simple energy-balance models tuned to the climate sensitivity of AOGCMs. Climate changes and impacts will not be uniform across the Earth, however, and more advanced IAMs may require regional climate predictions. There is increasing need for techniques that can capture the regional information provided by AOGCMs and produce tools useful for the impacts assessment community.

The most common approach for producing such regional projections has been to use “pattern scaling” to downscale the projections of simple global energy-balance models. Pattern scaling relies on the assumption that regional climate responses are a linear function of global climate response, so that regional climate evolution can be captured by scaling a single pattern to the global mean temperature. The technique was introduced by Santer et al. (1990) as a means of comparing spatial patterns of climate response from different GCMs and has been widely used in subsequent years (e.g. Hulme and Raper (1995); Hulme and Brown (1998); Cabre et al. (2010); Dessai et al. (2005); Fowler et al. (2007); Harris

44 et al. (2006); Murphy et al. (2007)). Different possible techniques for obtaining patterns are  
45 reviewed in Mitchell (2003).

46 The linearity assumption has been shown to be reasonable for centennial-scale projec-  
47 tions (e.g. Mitchell et al. (1999); Giorgi (2008)), but on some timescales the technique will  
48 be inappropriate, since different parts of the Earth warm at different rates. Furthermore,  
49 if the regional pattern of climate response were a function of the magnitude of warming,  
50 a single pattern would also not accurately capture the climate response to arbitrary CO<sub>2</sub>  
51 scenarios even in equilibrium cases. Using the HadCM2 model, Mitchell (2003) showed that  
52 both the rate and the magnitude of forcing changes influence patterns of regional climate,  
53 and suggested approaches to pattern construction to minimize errors.

54 We propose to overcome some of the limitations of pattern scaling through an alternative  
55 emulation approach based on a collection of precomputed climate model runs that allows  
56 us to capture rate dependencies in regional climate evolution. This collection of runs, or  
57 training set, is used to obtain estimates of the parameters in simple statistical models that  
58 describe temperature and precipitation as a function of past trajectories of radiative forcing  
59 due to CO<sub>2</sub>. The resulting tool allows us to reproduce (emulate) the output of an AOGCM  
60 under a large range of forcing scenarios. Once the emulator is constructed, emulation of a cli-  
61 mate scenario is effectively instantaneous, as it would be under pattern scaling. In contrast,  
62 climate projection from a state-of-the-art model can still take days to weeks even on the most  
63 powerful platforms. Since our training set is used only to estimate statistical parameters,  
64 the emulator is determined by a set of regional parameter values and requires negligible data  
65 storage. The simplicity and robustness of statistical emulation based on a modest training  
66 set makes it a promising tool for impacts assessment. Similar ideas have been previously  
67 proposed by Mitchell (2003), though execution was precluded because of lack of suitable  
68 collection of model runs, and recently explored by Holden and Edwards (2010) (See Section  
69 5 for comparison of approaches).

70 In the remainder of this paper, Section 2 describes the collection of climate runs on which

71 our emulator is based; Section 3 introduces the statistical models for annual temperature  
72 and precipitation at a regional level and shows an example of emulation; and Section 4  
73 develops emulation diagnostics and uses them both to assess the influence of training set  
74 size on emulation quality and to compare our emulation to pattern scaling. Finally, Section  
75 5 discusses our approach in comparison to other techniques for computer model emulation.  
76 We describe the particular requirements and characteristics of climate emulation over forcing  
77 scenarios, for which both the inputs and outputs are time series, and provide suggestions to  
78 guide future emulation approaches.

## 79 **2. Precomputed climate runs**

80 To explore the problem of emulating climate under arbitrary forcing scenarios, we built  
81 a collection of climate model runs to be used for training and prediction. These runs are  
82 driven by different trajectories of future CO<sub>2</sub> concentration and have different initial condi-  
83 tions, but all are performed with the same model and same representation of model physics.  
84 Simulations were performed with the Community Climate System Model Version 3 (CCSM3)  
85 (Yeager et al. (2006), Collins et al. (2006)) at a relatively modest T31 atmospheric resolution  
86 ( $\approx 3.75^\circ \times 3.75^\circ$ ) and nominally 3 degree ocean resolution, a configuration that allows us to  
87 run multiple realizations of a wide range of multi-century scenarios. Since we are interested  
88 in capturing the effects of changing CO<sub>2</sub> on climate, in all runs all other greenhouse gases  
89 and aerosols are held fixed at their pre-industrial values.

90 The AOGCM runs used in the work described here consist of five scenarios, three with  
91 gradual rise and then stabilization of CO<sub>2</sub> and two with abrupt changes (Figure 1). All  
92 scenarios follow estimated historical CO<sub>2</sub> concentrations from 1870 to 2010 and then branch  
93 off into different future trajectories of evolving CO<sub>2</sub> over the subsequent 189-439 years (end  
94 years range from 2199 to 2449). We denote the five scenarios as *fast*, *moderate*, *slow*, *jump*  
95 and *drop*. To enhance our ability to distinguish changes in mean climate from internal

96 variability, we simulated five realizations of each scenario with different initial conditions:  
97 specifically, we used restart files from years 410, 420, 430, 440 and 450 of the NCAR b30.048  
98 pre-industrial control run (Collins et al. 2006). In total, our collection of runs consists of  
99 more than 10,000 model years, though individual emulators used in this paper are trained  
100 using subsets of the runs.

101 Multiple realizations of each scenario are useful both in producing emulators and in eval-  
102 uating emulator performance. We treat the five realizations of each scenario as statistically  
103 independent because they were generated with decadal spaced restart files. The chaotic  
104 nature of the climate system means that changes in any initial conditions other than those  
105 of the deepest ocean are expected to produce essentially independent results after approxi-  
106 mately a decade (e.g. Branstator and Teng (2010), Collins (2002), Collins and Allen (2002))  
107 so we believe this assumption of independence reasonable. For similar reasons, runs under  
108 different scenarios but the same restart year should be very nearly independent within a few  
109 years after the scenarios diverge, but since all scenarios are identical before 2010, the results  
110 for runs with the same restart year are also identical until 2010. We avoid this problem by  
111 using runs with different restart years in our training sets.

112 The choice of scenarios for the precomputed runs was not based on any formal design  
113 criteria and is not meant to be optimal in any sense. We deliberately chose some scenarios  
114 that were somewhat realistic and others with large changes in CO<sub>2</sub> in order to be able to  
115 distinguish short and long term effects, but in general we sought simply to reproduce the  
116 kind of runs that would typically be available in pre-existing archives of climate model out-  
117 put. Impacts assessments often require emulation of multiple AOGCMs, but it would be  
118 prohibitively difficult for an individual research group to run multiple climate models to gen-  
119 erate optimal libraries for emulation. It is therefore useful to develop emulation techniques  
120 that are not critically sensitive to the characteristics of their training sets and that can make  
121 use of existing community multi-model resources such as the CMIP5 archive (Taylor et al.  
122 2003).

### 3. Statistical models for temperature and precipitation

In this work, we emulate annual mean temperature and precipitation in climate projections with simple statistical models that involve a mean function that varies in time plus a stochastic term. For the mean function, we chose simple functional forms relating temperature  $T$  and precipitation  $P$  to past trajectories of  $\text{CO}_2$  that capture physically justified relationships. We train emulators based on various subsets of our precomputed climate model runs, fitting the parameters of the statistical models using standard statistical methods. (See Supplementary Material for more details.) The resulting emulators can then predict annual temperature and precipitation for arbitrary climate forcing scenarios. In the emulations shown here, we fit the statistical models not at native climate model spatial resolution ( $48 \times 96$  gridpoints for T31 resolution) but aggregated at subcontinental scale in 47 regions. The regions are modifications of those defined by Ruosteenoja et al. (2003), subdivided over the oceans to ensure that we separately emulate regions of qualitatively different precipitation response (Figure 4, or see Figure S1 for regional codes). Without regional aggregation, obtaining a stable fit of the statistical models parameters for  $T$  and  $P$  would require a significantly larger training set. Emulation can be extended to the grid scale through regional pattern scaling (see Section 4).

#### *a. Temperature*

A long body of research suggests that within the range of  $\text{CO}_2$  concentrations likely to be produced by anthropogenic activity, equilibrium global mean temperature change is proportional to  $\log[\text{CO}_{2r}]$ , where  $[\text{CO}_{2r}]$  is the ratio between current and preindustrial  $\text{CO}_2$  concentrations (Manabe and Wetherald 1967; Forster et al. 2007). For policy analysis purposes, however, emulating equilibrium climate is less relevant than understanding the spatio-temporal climate changes that populations will face over the next century. We seek here to emulate the transient climate response when climate is a function not only of the present

148 value of  $[\text{CO}_{2r}]$  but also of its past history. As mentioned before, even if pattern scaling were  
 149 sufficient to reproduce equilibrium climate, i.e. if the equilibrium spatial distribution of tem-  
 150 perature were linear with  $\log[\text{CO}_{2r}]$ , it would not be sufficient in transient climates. Because  
 151 different regions of the Earth warm at different rates, the spatial distribution of temperature  
 152 anomalies in a given year during warming will not be a multiple of the equilibrium pattern.

153 For emulation of temperature, we propose a representation that captures this dependence  
 154 on past trajectories of  $\text{CO}_2$  via an infinite distributed lag model (Judge et al. (1980) ch.10)  
 155 in which current temperature is dependent on a weighted sum of past  $\log[\text{CO}_{2r}](t)$ :

$$T(t) = \beta_0 + \beta_1 \frac{1}{2} (\log[\text{CO}_{2r}](t) + \log[\text{CO}_{2r}](t - 1)) + \beta_2 \sum_{i=2}^{+\infty} w_i \log[\text{CO}_{2r}](t - i) + \varepsilon(t), \quad (1)$$

156 where  $T(t)$  is the temperature at year  $t$ . Because temperature may show some auto-  
 157 correlation, we assume the stochastic term  $\varepsilon(t)$  is an autoregressive model of order 1:  
 158  $\varepsilon(t) = \phi\varepsilon(t - 1) + \nu(t)$ , where  $\nu$  is a Gaussian white noise with unknown variance  $\sigma^2$ .  
 159 This model is able to capture the modest dependence in temperatures across years.

160 The  $\beta$  coefficients in Equation (1) are physically interpretable:  $\beta_0$  is preindustrial temper-  
 161 ature;  $\beta_1$  is the near-term response to changes in  $\text{CO}_2$ , and  $\beta_2$  the slower response dependent  
 162 on  $\text{CO}_2$  levels in prior years. This form gives us the flexibility to represent a temperature  
 163 response characterized by multiple adjustment timescales, and is especially important when  
 164 emulating scenarios with abrupt  $\text{CO}_2$  changes. Using the average  $\log[\text{CO}_{2r}]$  over years  $t$  and  
 165  $t - 1$  for the short-term effect is somewhat arbitrary, but we have experimented with other  
 166 forms for this term and not found anything clearly superior. Because we expect the influence  
 167 of past radiative forcing to decrease as we go back in time, the weights  $w_i$  in the long-term  
 168 component should be chosen to decrease with the trajectory year  $i$ . We choose here a simple  
 169 exponential decay of the weighting of past years:  $w_i = \rho^{-2}(1 - \rho)\rho^i$  with  $0 < \rho < 1$  so  
 170  $\sum_{i=2}^{\infty} w_i = 1$ . (Note that we could also have taken the infinite sum in (1) to start at 0 rather  
 171 than 2. The resulting fitted models would be negligibly different.) The model parameters are  
 172 then the three  $\beta_j$ s,  $\rho$ ,  $\phi$  and  $\sigma^2$ . The first four parameters capture the mean evolution of the  
 173 climate system averaged over initial conditions, a deterministic function of  $\text{CO}_2$  trajectory,

174 and the final two parameters describe the stochastic variability in the climate state about  
175 this mean, which differs between realizations (initial conditions). We discuss emulation of  
176 the stochastic behavior of both temperature and precipitation in Section 3c.

177 It is important to point out several assumptions implicit in the choice of our functional  
178 form for temperature. First, the model assumes that, on average, equilibrium spatial tem-  
179 perature patterns are linear with  $\log[\text{CO}_{2r}]$ , since when sufficient time has passed after  
180 stabilization of  $\text{CO}_2$  concentration, emulated mean temperature approaches

$$\beta_0 + (\beta_1 + \beta_2) \log[\text{CO}_{2r}]_{stab},$$

181 where the subscript *stab* indicates the  $\text{CO}_{2r}$  level after stabilization. This assumption would  
182 likely break down in cases of extreme  $\text{CO}_2$  changes. Second, our functional form is ap-  
183 propriate only for centennial-scale or shorter emulation scenarios. Although in principle  
184 our approach allows us to emulate climate in any year for arbitrary  $\text{CO}_2$  scenarios, Equa-  
185 tion (1) should not be used for emulating considerably beyond the several-century time  
186 span of the training runs. This constraint arises not only because statistical models can-  
187 not be expected to capture processes not represented in the training set, but also because  
188 the simple exponential weights used here do not capture well the combined behavior of the  
189 decadal/centennial-scale warming of the upper ocean and the long-tail warming of the deep  
190 ocean over thousands of years (see Supplementary Figure S4).

191 To construct an emulator, we derive parameter estimates from one or more training runs.  
192 (By “run” we mean a climate projection driven by a given scenario and begun from given  
193 initial conditions.) Throughout this manuscript, we focus on an emulator generated with a  
194 training set consisting of two runs: one realization each of the *fast* and *jump* scenarios with  
195 different restart years. The resulting emulator appears to track accurately the overall trend of  
196 out-of-training-set climate scenarios. Figure 2a-b shows emulations of the mean temperature  
197 trajectory for the *slow* and *drop* scenarios, superimposed with all five realizations of actual  
198 CCSM3 output for these scenarios. Emulation of the *drop* scenario does show slight misfit  
199 immediately following the sudden drop in  $\text{CO}_2$ . This misfit can be reduced by using a more

200 complex functional form, but introducing additional terms can lead to instability of the  
201 fit, and we consider the emulation of this physically extreme scenario to be reasonably good  
202 under the circumstances. (See Section 4 for a more extensive evaluation of emulation fidelity,  
203 and see Supplementary Materials for a table of parameter estimates and their standard errors  
204 for all regions.)

205 *b. Precipitation*

206 Precipitation in transient climates has been frequently described as a combination of a  
207 “fast” response that is a function of the changed forcing agent and a “slow” linear response to  
208 evolving temperature. The fast response is negative in the case of CO<sub>2</sub>, so that in scenarios  
209 with rising CO<sub>2</sub>, precipitation at a given temperature is lower than it would be at equilibrium  
210 for that temperature (Andrews and Forster 2010). The transient precipitation response was  
211 first discussed in detail by Allen and Ingram (2002), and the fast/slow framework became  
212 commonly accepted in later works (e.g. Bala et al. (2010); Cao et al. (2011)). These findings  
213 motivate the following regression model for precipitation (though see McInerney and Moyer  
214 (2012) for further discussion of underlying physics):

$$P(t) = \gamma_0 + \gamma_1 \hat{T}(t) + \gamma_2 \log[\text{CO}_{2r}](t) + \eta(t), \quad (2)$$

215 with  $\gamma_1 \hat{T}(t)$  and  $\gamma_2 \log[\text{CO}_{2r}](t)$  the “slow” and “fast” terms, respectively, and  $\hat{T}(t)$  the mean  
216 emulated temperature from (1). We use  $\hat{T}(t)$  rather than  $T(t)$ , the actual temperature in  
217 year  $t$ , because the physical processes underlying the model are likely distinct from those  
218 driving stochastic interannual variability. Since we found no clear evidence for dependence  
219 in the stochastic terms for precipitation in this model, the stochastic term  $\eta(t)$  is simply  
220 assumed to be Gaussian white noise with unknown variance  $\tau^2$ . Once  $\hat{T}(t)$  is obtained from  
221 fitting (1), the parameters in (2) are estimated using linear regression. Joint emulation of  
222 temperature and precipitation including their stochastic components would require modeling  
223 the corresponding stochastic terms  $\varepsilon(t)$  for temperature and  $\eta(t)$  for precipitation jointly,

224 which we do not attempt here.

225 The resulting emulated mean precipitation again matches well the overall trend in the  
226 CCSM3 output, although variability in precipitation is much larger than in temperature and  
227 trend prediction is therefore less informative (Figure 2c-d). We chose to show the equatorial  
228 West Pacific in Figure 2 because this region demonstrates one feature of our emulation that  
229 stands out in scenarios of abrupt CO<sub>2</sub> change: a sharp spike in precipitation coincident  
230 with a drop in CO<sub>2</sub> (Figure 2d), such that precipitation momentarily increases even while  
231 temperature is decreasing. This effect has a well-founded physical interpretation and has  
232 been shown clearly above variability in AOGCM output in more extreme scenarios in several  
233 recent works (e.g. Wu et al. (2010), McInerney and Moyer (2012)). Linear pattern scaling  
234 with global mean temperature change cannot capture this effect.

235 *c. Stochastic temperature and precipitation components*

236 While the mean emulations shown in Figure 2 capture the dependence of temperature  
237 and precipitation on CO<sub>2</sub> trajectories, impacts assessments may require emulation that fully  
238 reproduces an actual climate simulation, including short-term variability. Many applications  
239 would therefore require addition of stochastic components to the mean emulator. A simple  
240 initial approach is to simulate this variability from our stochastic models and estimated  
241 parameters. This method implicitly assumes that the statistical characteristics of the error  
242 terms are invariant over time for any scenario and are the same for all scenarios. That  
243 assumption is unlikely to be exactly true, but appears to provide a satisfactory approximation  
244 for most regions in the scenarios tested here. That is, the simple stochastic model appears to  
245 capture the variability in the actual realizations of the CCSM3 temperature and precipitation  
246 (Figure 3a-b, which show emulated full simulations including stochastic components for  
247 the cases of Figure 2a and c, along with corresponding actual CCSM3 realizations). More  
248 quantitatively, CCSM3 output can be compared with the 95% prediction bands based on  
249 the emulators (Figure 3c-d). For the cases shown, the empirical coverage of the prediction

250 intervals are 0.9531 and 0.9545 for temperature and precipitation, respectively, very close to  
251 the nominal coverage of 0.95. Figure S3 shows empirical coverages for temperature for all  
252 regions and both the *slow* and *drop* scenarios; the results are close to 95% in all regions other  
253 than the Southern Ocean. The fact that our model does not provide an accurate substitute  
254 for CCSM3 output in the Southern Ocean is not unexpected because upwelling from the deep  
255 ocean complicates temperature evolution there. Misfit for the Southern Ocean is evident in  
256 multiple diagnostics of emulation performance; see Section 4.

## 257 **4. Diagnostics, training set size, and comparison with** 258 **pattern scaling**

### 259 *a. Evaluating the fit*

260 The appropriate evaluation of emulator performance depends on the purpose for which  
261 the emulator is used. For impacts assessments that have previously relied on global pattern  
262 scaling, one possible performance criterion is exceeding the emulation fidelity provided by  
263 pattern scaling. Other criteria could be that emulation error is small relative to differences  
264 in climate projections between AOGCMs, or small relative to initial conditions uncertainty  
265 in the emulated AOGCM. We discuss here various approaches to evaluating emulator per-  
266 formance. Evaluations are aided by having multiple realizations for each prediction scenario,  
267 allowing us to distinguish the mean climate trajectories from the stochastic component with-  
268 out assuming our mean model is correct. The test of empirical coverage of 95% prediction  
269 intervals discussed in Section 3 is one type of emulator evaluation, but not the most relevant  
270 for the main focus of this work, emulation of change in mean climate. We therefore seek  
271 additional diagnostics.

272 Even if our emulation model (1) were strictly correct for all scenarios, the mean emulator  
273 generated from it would retain some uncertainty due to the limited size of the training

274 set used to estimate the model parameters. Confidence bands for the estimated regression  
 275 function provide a natural way to quantify this uncertainty. Figure 3e-f shows the pointwise  
 276 95% confidence bands along with the average of the five available CCSM3 realizations. (See  
 277 Supplementary Materials for details.) The width of these bands are small relative to internal  
 278 variability and agree well by eye with the average of the five CCSM3 realizations.

279 These confidence bands assume that the underlying statistical model is correct. We  
 280 consider two additional indices whose validity does not depend on knowing the form of  
 281 the mean function. The index  $I_1$  measures emulation performance relative to the optimal  
 282 emulation possible given initial condition uncertainty and  $I_2$  the trend in the data relative  
 283 to initial condition uncertainty (i.e. how much of the variation in a climate time series could  
 284 be explained by an emulator).

285 The first index is related to what statisticians call the lack-of-fit statistic (see e.g. Mont-  
 286 gomery (2012)). Let  $T_r(t)$  denote temperature for year  $t = 1, \dots, n$  (here,  $t = 1$  corresponds  
 287 to the year 2010, the year the scenarios diverge) and realization  $r = 1, \dots, R$  (here,  $R = 5$ ).  
 288 We compare the sum of squared deviations of the actual realizations from the emulated mean  
 289 temperatures  $\hat{T}(t)$  to the sum of squared deviations of realizations from the average across  
 290 realizations  $\bar{T}(t) = \frac{1}{R} \sum_{r=1}^R T_r(t)$ :

$$I_1 = \frac{\sum_{r=1}^R \sum_{t=1}^n (T_r(t) - \hat{T}(t))^2}{\frac{R}{R-1} \sum_{r=1}^R \sum_{t=1}^n (T_r(t) - \bar{T}(t))^2} = \frac{N_1}{O_1}. \quad (3)$$

291 The numerator  $N_1$  measures the actual performance of the emulator. The denominator  
 292  $O_1$  makes use of the multiple realizations we have under each scenario to give an unbiased  
 293 estimate of the sum of squared errors for a hypothetical “perfect” emulator that, for each  
 294 year  $t$ , reproduces the average temperature over an infinite number of realizations. The  
 295 factor of  $\frac{R}{R-1}$  in  $O_1$  takes account of the fact that we do not know this perfect emulator but  
 296 use  $\bar{T}(t)$  as an estimate of it. A value of 1 for  $I_1$  is therefore the best possible performance  
 297 from an emulator. (Occasional values less than 1 may however arise due to random variation  
 298 in  $N_1$  and  $O_1$ .)

299 A value  $I_1$  close to 1 has different implications depending on the noise in the model output

300 being emulated. In particular, if the noise is large compared to the trend in the data, then  
 301  $I_1$  will likely be close to 1 even if the emulation poorly captures the small underlying trend.  
 302 To quantify the degree of variation in the data attributable to the trend, we construct an  
 303 index whose denominator is that of  $I_1$  but whose numerator now describes the trend itself:

$$I_2 = \frac{\frac{n}{n-1} \sum_{r=1}^R \sum_{t=1}^n (T_r(t) - \bar{T}_r)^2}{\frac{R}{R-1} \sum_{r=1}^R \sum_{t=1}^n (T_r(t) - \bar{T}(t))^2} \quad (4)$$

304 where  $\bar{T}_r$  is the mean across time of each realization:  $\bar{T}_r = \frac{1}{n} \sum_{t=1}^n T_r(t)$ . Note that this index  
 305 depends only on the AOGCM data and is completely independent of the emulation. If the  
 306 mean AOGCM data shows no trend, then the numerator and the denominator are unbiased  
 307 estimates of the same quantity and  $I_2$  should be close to 1. The conditions  $I_2 \gg 1$  and  
 308  $I_1 \approx 1$  would mean that there is a trend to emulate and that the emulator captures it well.  
 309 If  $I_1$  is comparable to  $I_2$ , then the emulator would not be useful for tracking the evolution of  
 310 the mean. As interannual variability in precipitation is larger relative to trend than it is in  
 311 temperature (e.g. Figure 2, see also Deser et al. (2012)),  $I_2$  values tend to be much smaller  
 312 for precipitation than for temperature (compare Figures 4 and 9).

313 These indices suggest that the temperature emulator described previously in Section 3  
 314 (trained by one realization each of the *fast* and *jump* scenarios) produces near-optimal mean  
 315 emulation of nearly all regions in the physically reasonable *slow* stabilization scenario and  
 316 only modestly degraded quality in the extreme *drop* scenario. (Figure 4 shows  $I_1$  and  $I_2$   
 317 values for all regions.) For the *slow* scenario, the emulated mean functions are essentially  
 318 optimal ( $I_1$  very nearly 1) throughout the northern hemisphere and equatorial region, and  
 319 close to optimal ( $I_1 \leq 1.13$ ) everywhere except in part of the Southern Ocean. For the *drop*  
 320 scenario, unsurprisingly, the emulator predictions perform substantially worse in all regions,  
 321 but even here, we believe this lack of fit may be small compared to other possible sources  
 322 of error in forecasting climate, such as differences between AOGCMs or differences between  
 323 AOGCMs and reality, and so would still serve as a useful emulator. The largest discrepancies  
 324 arise for both scenarios in a single portion of the Southern Ocean. Values of  $I_1$  substantially  
 325 larger than 1 are not necessarily associated with a poor skill of the emulator relative to other

326 techniques, but do indicate that the statistical model for the region could be improved.

327 In the end, whether an emulator of an AOGCM is adequate will depend on the specific  
328 application. Because we make no effort to capture spatial dependence in the stochastic terms  
329 between regions, the emulator would be less appropriate for studies that involve large-scale  
330 spatial correlations in weather, e.g. global droughts or jet stream shifts. (See Castruccio and  
331 Stein (2013) for one approach to emulating the stochastic component of annual tempera-  
332 tures in climate model output that captures both spatial and temporal dependence.) We  
333 also do not capture any dependence between the stochastic components of temperature and  
334 precipitation within a region. However, for an impacts assessment requiring annual tempera-  
335 tures in a given region, any differences between the emulated temperature and the AOGCM  
336 temperature showed in, for example, Figure 3a would most likely be inconsequential.

337 *b. Training set size: how many scenarios/realizations?*

338 One of the advantages of our approach is that it permits emulation with a relatively small  
339 training set of precomputed runs. To determine the trade-off between size of the training set  
340 and goodness of fit, we examined the performance of the emulator with a varying number  
341 of scenarios and realizations. Investigating the impact of the number of realizations on  
342 emulation quality is the more straightforward test, involving computing  $I_1$  for temperature  
343 emulation over a range of number of realizations used. Figure 5b shows results from an  
344 experiment in which the *moderate* scenario was emulated with from 1 to 5 realizations of  
345 the *fast* scenario as the training set. Increasing the number of realizations of each training  
346 scenario produces more accurate emulations, but the difference between the use of even 1  
347 and 2 realizations is small, and there is diminishing return gained from further increasing  
348 the number of realizations in the training set. Increasing the number of realizations further  
349 also does not reduce the misfit of the outlier regions with highest  $I_1$  values, all of which lie  
350 in the Southern Ocean.

351 Testing the value added by additional scenarios is a less well-defined problem, since

352 different choices of scenarios will affect the emulation differently. Nevertheless, we attempt  
353 a test by conducting emulations with increasing numbers of scenarios. Again we emulate  
354 temperature in the *moderate* scenario beginning with a training set consisting of a single  
355 realization of *slow* and successively adding to the training set *fast*, *jump*, and *drop* (Figure  
356 5a), which is a rough attempt to order the training scenarios from most to least similar  
357 to the prediction scenario. The results show that addition of scenarios first improves and  
358 then degrades the emulation. We interpret this result as implying that our simple statistical  
359 model cannot perfectly represent all scenarios, i.e. that best values of  $(\beta_0, \beta_1, \beta_2)$  and  $\rho$  in  
360 Equation (1) vary somewhat with scenario. Including scenarios in the training set very  
361 different from the one emulated can then result in worse performance. Figure 5 shows that  
362 even a single *slow* or a single *fast* realization yields a fairly good emulator of the *moderate*  
363 scenario. However, we would be cautious about building emulators when AOGCM output  
364 is available for only one scenario since that would leave no opportunity to check for stability  
365 of the regression parameters across scenarios.

366 Our tests suggest that the choice of training set is not especially crucial if prediction and  
367 training scenarios are similar, but more care would be needed for emulating extreme sce-  
368 narios. One approach might be to choose different training sets according to the prediction  
369 scenario. In this case one algorithm might be to (1) order the available forcing scenarios  
370 in the training set by their similarity to the prediction scenario; (2) fit the emulator using  
371 first only the nearest training scenario, then the two nearest, and so on; and (3) choose the  
372 emulator with the smallest training set that offers stable parameter estimations as measured  
373 by the width of the 95% confidence bands for the mean emulator (e.g. Figure 3e-f). Further  
374 research would be needed to actually apply this approach in the context of integrated assess-  
375 ments over many possible scenarios, both to define the notion of similarity and to automate  
376 implementation. In this work we have focused simply on demonstrating that in some circum-  
377 stances, emulation requires only a limited training set of a few scenarios and realizations.  
378 This finding supports the utility of statistical emulation based on modest training sets for

379 uses such as policy analysis or model intercomparison.

380 *c. Comparison with pattern scaling*

381 One of the motivations for our approach to statistical emulation is to offer an improve-  
382 ment on pattern scaling by capturing the dependencies on rate of forcing change that make  
383 transient climates different from equilibrium ones. We therefore test the fidelity of our  
384 mean emulation against pattern scaling to global mean temperature. To provide a direct  
385 comparison, we first evaluate performance of the regional climate projections generated by  
386 our statistical mean emulator to regional projections generated by pattern-scaling to global  
387 mean temperature (GMT). Second, we evaluate an extension of our approach that allows  
388 us to emulate climate at native model spatial resolution, again comparing to GMT pattern  
389 scaling. The latter test may be more relevant for policy analysis purposes, since impacts  
390 assessments often require fine-scale climate projections. We perform grid-scale emulation by  
391 a hybrid approach, first statistically emulating regional temperature and precipitation and  
392 then downscaling by pattern scaling to the regional mean temperatures.

393 For the comparison of regional emulation, we construct patterns of temperature and pre-  
394 cipitation for our 47 regions from all realizations in our training set (*fast* and *jump*). Pattern  
395 scaling assumes that all regional temperature anomalies  $T_i(t) - T_{i,PI}$  are linear with global  
396 mean temperature anomaly  $T_{GM}(t) - T_{GM,PI}$ . (The subscripts PI and GM denote preindus-  
397 trial values and global mean, respectively.) We derive the pattern by linear regression on all  
398 data in the training set assuming

$$T_i(t) - T_{i,PI} = \alpha_i (T_{GM}(t) - T_{GM,PI}) + \varepsilon_i(t) \quad (5)$$

399 and estimating  $\alpha_i$  by least squares. Patterns for temperature and precipitation are shown  
400 in Figures 6 and 7, with the fitted relationship between the regional climate variable and  
401 GMT shown in red. These figures provide a visual check on the linearity assumption behind  
402 pattern scaling and on the variability in regional temperature and precipitation.

403 GMT in a typical pattern scaling emulation would usually be obtained by running an  
 404 energy-balance model tuned to match the climate sensitivity of the AOGCM to be emulated.  
 405 Here we forgo the use of an additional external model and instead simply use the GMT from  
 406 our statistical emulator. This simplification gives pattern scaling a slight artificial advantage  
 407 over a more realistic comparison. Nevertheless, when comparing to emulation of temperature  
 408 in the same scenarios shown previously (*slow* and *drop*), statistical emulation matches or  
 409 outperforms pattern scaling in most regions (Figure 8). Comparing Figures 4 and 8, we  
 410 see that for the *slow* scenario, which has the smallest transient response and emulation  
 411 is easiest, the regional differences in performance for our emulator and pattern scaling as  
 412 measured by  $I_1$  are small; these differences are much larger for the more challenging *drop*  
 413 scenario. For precipitation,  $I_2$  values are much smaller than for temperature (see Figure 9),  
 414 so the differences in  $I_1$  values for the two emulators are unsurprisingly smaller. Nevertheless,  
 415 in both prediction scenarios used here, statistical emulation conveys an advantage in most  
 416 regions outside the Southern Ocean (which is problematic for both methods).

417 For a grid-scale comparison, we use a hybrid approach, emulating regional temperature  
 418 and precipitation and then downscaling by applying pattern scaling at the regional level.  
 419 This approach consists of four steps:

- 420 **1.** For each region  $i$ , use the training set to fit parameters for regional  $T_i$  and  $P_i$
  - 421 **2.** With those parameters, statistically emulate regional  $T_i$  and  $P_i$  for the prediction scenario
  - 422 **3.** For each region  $i$ , use the training set to obtain regional patterns of grid-scale  $T$  and  $P$
  - 423 **4.** Predict grid-scale  $T$  and  $P$  by multiplying the regional patterns by emulated regional  $T_i$
- 424 This approach retains the benefits of statistical emulation in capturing nonlinearities in  
 425 regional climate evolution but allows projections at small spatial scale.

426 Step 3, estimating for each region  $i$  a grid-resolution pattern that scales with respect  
 427 to regional temperature, is mathematically similar to the global pattern scaler described  
 428 previously, where we obtained a regional-resolution pattern that scales with respect to global  
 429 mean temperature. For  $T$  emulation, we use all data in the training set to fit the parameters

430 in

$$T((L, \ell), t) - T_{\text{PI}}(L, \ell) = \alpha_{(L, \ell)}(T_i(t) - T_{i, \text{PI}}) + \varepsilon_{(L, \ell)}(t), \quad (6)$$

431 where  $T((L, \ell), t)$  is temperature at a model gridpoint at latitude  $L$  and longitude  $\ell$ , and  
432 the  $i$  subscript again refers to subcontinental regions. The grid-level parameters  $\alpha_{(L, \ell)}$  are  
433 estimated by least squares. We compare this hybrid pattern scaling-emulator with the simple  
434 global pattern scaling described previously: the pattern is at grid level and the scaler is GMT,  
435 which we obtain from our statistical emulation. In the case of temperature emulation, the  
436 very simple hybrid approach outperforms pattern scaling for most gridpoints outside the  
437 polar regions, particularly for the continental areas of greatest interest for impacts assessment  
438 (Figure 10).

## 439 5. Alternative emulation strategies

440 In the previous section we compared our climate model emulation approach to pattern  
441 scaling, the most commonly-used approach for emulation of climate model output in the im-  
442 pacts assessment community (see e.g. Santer et al. (1990); Hulme and Raper (1995); Hulme  
443 and Brown (1998); Cabre et al. (2010); Dessai et al. (2005); Fowler et al. (2007); Harris  
444 et al. (2006); Murphy et al. (2007)). However, interest is growing in alternative approaches,  
445 and it is therefore useful to compare our technique with more complex emulation strategies  
446 proposed in the recent literature (Rougier et al. (2009); Holden and Edwards (2010); Wilks  
447 (2012); Vecchi et al. (2011); Murphy et al. (2007)). These strategies include the empirical  
448 orthogonal function (EOF) regression of Holden and Edwards (2010) and Gaussian process  
449 modeling (henceforth “GP modeling”), a standard method for emulating the output of de-  
450 terministic computer models (Sacks et al. 1989; Santner et al. 2003; Kennedy and O’Hagan  
451 2001; Oakley and O’Hagan 2002; Rougier et al. 2009; O’Hagan 2006). For climate models,  
452 Gaussian processes have mainly been used to emulate over physical parameters, although  
453 (Holden and Edwards 2010) raise the prospect of using Gaussian processes for forcing sce-

454 nario emulation. A number of authors have built emulators over physical parameters in  
455 order to calibrate a climate model (Sanso et al. 2008; Sanso and Forest 2009; Bhat et al.  
456 2012; Drignei et al. 2008).

457 The GP approach to computer model emulation assumes that the output of interest  
458 is a Gaussian process in some set of inputs that vary across model runs. Among others,  
459 Challenor et al. (2010) and Rougier (2008) have discussed extensions to the GP approach to  
460 multivariate climate output, and several authors have proposed approaches for multivariate,  
461 time-dependent output: projection on a lower dimensional space via principal component  
462 analysis (Wilkinson 2010; Higdon et al. 2008) or wavelet decomposition (Bayarri et al. 2007),  
463 choice of a single representative output (Challenor et al. 2006) or a spatial aggregated average  
464 of it (Hankin 2005), kernel mixing and matrix identities (Bhat et al. 2012), and dynamically  
465 autoregressive models (Fei and West 2009).

466 Using Gaussian processes to emulate computer models is attractive in many circum-  
467 stances because it does not require the prior assumption of any particular parametric form  
468 for the relationship between inputs and outputs and provides an internally consistent ap-  
469 proach to estimating the uncertainties of the emulator based on the GP model (Sacks et al.  
470 1989; Oakley and O’Hagan 2002). This flexibility comes at some cost, since it is intrinsically  
471 difficult to estimate an arbitrary function nonparametrically in high dimensions. Neverthe-  
472 less, to give a specific example, Challenor et al. (2006) fit a GP emulator to climate model  
473 output with 17 input parameters and only 100 model runs. This fitting is aided by the fact  
474 that most of the input parameters appear to have little impact on the output of interest.  
475 Emulation over physical parameters that are globally constant has been done with very few  
476 model runs by exploiting the information available in a spatially resolved climate model that  
477 provides many informative outputs about these parameters from each run (Sanso et al. 2008;  
478 Sanso and Forest 2009; Bhat et al. 2012). In contrast, for forcing scenario emulation, we  
479 should not assume that any of the statistical parameters in our emulators (1) and (2) are  
480 constant across all regions, since accounting for regional differences in patterns of climate

481 change is the whole point of our approach. We instead exploit the multiple observations in  
482 *time* rather than in space to build an emulator with few runs.

483 In our view, emulating a long time-series of spatially resolved climate variables over a  
484 wide range of forcing scenarios is a highly specialized problem, and general techniques for  
485 multivariate computer model emulation are not the most appropriate tools to approach it.  
486 Choosing an appropriate emulation strategy requires recognition of three key issues: 1) the  
487 desired output variables are a function of the previous history of CO<sub>2</sub> or other forcings and  
488 so the emulator inputs should be functions of past trajectories; 2) because climate response  
489 is dependent only on these past trajectories, the statistical model that relates model inputs  
490 to outputs is the same for any given year (i.e., the  $\beta_j$ 's and  $\rho$  in (1) do not depend on  $t$ );  
491 and 3) the appropriate means of reducing the dimensionality of the problem is not to limit  
492 the inputs, which would reduce the types of forcing trajectories that can be emulated, but  
493 instead to reduce the number of parameters that need to be fit by using a structured model  
494 of the functional form describing climate response.

495 Reducing climate emulation to a tractable problem necessarily involves some compro-  
496 mises. The trade-offs of different choices are illustrated by comparing our approach to that  
497 of Holden and Edwards (2010), whose goal is the most similar to ours among published  
498 works on climate model emulation of which we are aware. Holden and Edwards (2010)  
499 share our motivation of using a collection of climate runs and relatively simple statistical  
500 techniques to produce computationally efficient climate predictions for the purposes of in-  
501 tegrated assessment modeling, although they include both forcing scenarios and 19 climate  
502 model parameters as inputs, whereas we only consider forcing scenarios. Both their and our  
503 approach limit the number of parameters that need to be estimated in the statistical model,  
504 although with some noticeable differences.

505 Holden and Edwards (2010) emulate decadal average temperature at a single time period  
506 (2100) based on annual CO<sub>2</sub> levels between 2005 and 2105. If one were to directly regress  
507 each output for this problem (temperature changes for each pixel of the model) on the

508 100 inputs (CO<sub>2</sub> in each year from 2005 to 2105), the resulting parameter estimates would  
509 likely be unstable and yield problematic predictions under some CO<sub>2</sub> trajectories. To obtain  
510 outputs with a higher signal to noise ratio, Holden and Edwards (2010) consider just the five  
511 principal EOFs rather than results for each individual grid point as the outputs. To reduce  
512 the number of regression parameters that need to be estimated for each output, they consider  
513 only CO<sub>2</sub> trajectories following a specific functional form (a cubic polynomial), so that the  
514 regression is made on the three polynomial parameters (the polynomial is constrained to  
515 equal a fixed value in 2005) rather than on each of the 100 years of the CO<sub>2</sub> time series. The  
516 emulation problem thereby simplifies to a regression of five outputs on three parameters of a  
517 CO<sub>2</sub> trajectory. This simplicity permits Holden and Edwards (2010) to extend their analysis  
518 to include emulation over physical parameters.

519 These choices make emulation possible, but with several limitations. Reducing spatial  
520 dimensionality of grid level output by using EOFs rather than our use of sub-continental  
521 regions is a reasonable choice, though we believe the regional approach makes interpretation  
522 of results somewhat easier. However, restricting CO<sub>2</sub> trajectories to some simple functional  
523 form described by a small number of parameters (such as cubic polynomials) forgoes the  
524 flexibility needed for integrated assessment problems in which CO<sub>2</sub> emissions must be allowed  
525 to vary with economic activity, whose own growth may be complex. The restriction to cubic  
526 polynomials also precludes modeling scenarios with abrupt changes in CO<sub>2</sub> levels.

527 A more fundamental set of limitations results from formulating the output as a function  
528 of the CO<sub>2</sub> concentrations for a fixed set of years (which we term a “fixed timeframe trajec-  
529 tory”) rather than as a past trajectory of CO<sub>2</sub> concentrations. Specifically, when using fixed  
530 timeframe trajectories, the only model output that can be used for emulation are results for  
531 those years over which the prediction is sought. By contrast, using past trajectories permits  
532 use of any model runs covering any years to build a single emulator that allows predictions  
533 for all years. The limitation is less apparent in Holden and Edwards (2010) because they  
534 make only a single prediction in time (a change in decadal averages). If, however, their

535 collection of climate model runs were used to predict temperature in an earlier period such  
536 as 2021–2030, then the fixed timeframe approach would require excluding all available model  
537 output after 2030. Furthermore, and perhaps more importantly, with a fixed timeframe tra-  
538 jectory, one would have to build and fit a new statistical model for each time point at which  
539 one wants to predict, whereas past trajectories can be used to generate a single emulator for  
540 predictions at all time points. Because the past trajectory approach uses all information in  
541 the training runs to build a single emulator, we can produce a stable emulator with much  
542 less training data. In some circumstances, we were able to build an effective emulator based  
543 on a single run (see Figure 5) and can predict a whole series of annual average temperatures,  
544 whereas Holden and Edwards (2010) use 245 runs and predict only a single temperature  
545 (itself a decadal average). As we have noted, Holden and Edwards (2010) also include vari-  
546 ation in climate model parameters, but even with a fixed climate model parametrization,  
547 they would need at least three runs to estimate the three parameters related to their cubic  
548 polynomial representation of the forcing scenario. The requirement for a large training set in  
549 turn led Holden and Edwards (2010) to use a climate model of only intermediate complexity,  
550 GENIE-2 (Lenton et al. 2007).

551 While the functional form we chose in (1) is somewhat arbitrary, no further increase  
552 in complexity seemed warranted. With the runs available to us, explorations with several  
553 more complex functional forms did not yield substantially better emulation performance  
554 (lower  $I_1$ ) for centennial-scale predictions. On the other hand, models with fewer parameters  
555 than (1) that we have considered resulted in noticeable degradation of prediction skills for  
556 some scenarios. Our finding that temperature emulations in the somewhat realistic *slow*  
557 scenario yield  $I_1$  values very near 1 in nearly all regions (e.g. Figure 4a) implies that even  
558 the simple approach we describe leaves little room to further improve emulation of the mean  
559 temperature evolution over timescales typical of impacts assessments.

560 Although our emulators of mean trajectories worked very well in some circumstances,  
561 there is still room for improvement in several categories: for precipitation (where trend is

562 small relative to variability), for scenarios with extremely rapid CO<sub>2</sub> changes, and for longer-  
563 time-scale scenarios. In all cases, a larger collection of climate model runs would be necessary  
564 to explore these issues. Multiple millennial-scale training runs would allow adding a second  
565 lag term in the statistical models to account for the qualitatively different climate response  
566 at long timescales. Runs with substantial jumps in consecutive years could address the misfit  
567 after rapid CO<sub>2</sub> changes by allowing separate contributions from each of the two most recent  
568 years rather than taking their average. Finally, a larger collection of scenarios might make it  
569 feasible to allow the regression parameters to vary smoothly in some way with the prediction  
570 scenario, or, more in keeping with the approach here, the past trajectory. That is, we could  
571 construct a model that views these parameters as a function of the past trajectory, possibly  
572 as a multivariate GP after some dimension reduction on the past trajectory.

## 573 **6. Conclusions**

574 Statistical emulation of climate model output from computationally demanding AOGCMs  
575 has the potential to make climate projections capturing the full temporal dynamics of tran-  
576 sient climates readily available for impacts assessment, policy analysis, and other applica-  
577 tions. Developing methods that can function reasonably well with very small training sets  
578 is essential, however, to permit emulation to be a widely useful tool. The simple statistical  
579 approach we have outlined here permits us to credibly emulate climate model output with a  
580 very small training set, even in some cases of severe scenario extrapolations. Small training  
581 set size is permitted by two key aspects of our approach: treating emulation inputs (CO<sub>2</sub>  
582 concentrations here) as past trajectories rather than fixed timeframe trajectories and using  
583 simple, physically-based statistical models that capture the relationships between CO<sub>2</sub> and  
584 temperature or precipitation. The consequence is that a small training set produces rich  
585 results.

586 While the collection of runs used here was based on a fairly coarse spatial resolution

587 climate model, the proven efficiency of our emulator should permit its use for emulating  
588 more state-of-the-art models based on quite small training sets. This approach performs  
589 at least as well as pattern scaling in all circumstances we have examined and substantially  
590 better in many. It therefore can be seen as a natural alternative for fast climate impacts  
591 assessments, saving orders of magnitude in computational time over running a full AOGCM.

592 *Acknowledgments.*

593 The authors thank Ray Pierrehumbert and participants in the University of Chicago's  
594 2008 Workshop on Modeling Uncertainty in Integrated Climate Assessment Models for valu-  
595 able discussion that helped initiate this project, and Matt Huber, Lan Zhao, and Wonjun  
596 Lee for computational assistance. We also thank the reviewers for many helpful recommen-  
597 dations leading to improvements in the substance and presentation in this paper. This work  
598 is part of the Center for Robust Decision-making on Climate and Energy Policy (RDCEP);  
599 funding was provided by grants from the University of Chicago (UC) Energy Initiative; from  
600 the UC and Argonne National Lab (ANL) (under contract DE-AC02-06CH11357 from the  
601 U.S. Department of Energy); from STATMOS, an NSF-funded Network (NSF-DMS awards  
602 1106862, 1106974 and 1107046); and from the NSF Decision Making Under Uncertainty  
603 program (NSF grant SES- 0951576). Simulations were performed on "Fusion", a 320-node  
604 computing cluster operated by the Laboratory Computing Resource Center at Argonne Na-  
605 tional Laboratory and on TeraGrid resources operated by Purdue University. Gary Strand  
606 (NCAR) provided CCSM3 restart files. Data storage was provided by PADS (NSF grant  
607 OCI-0821678) at the Computation Institute, a joint initiative between the UC and ANL.

## REFERENCES

- 610 Allen, M. and W. J. Ingram, 2002: Constraints on future changes in climate and the hydro-  
611 logic cycle. *Nature*, **419**, 224–232.
- 612 Andrews, T. and P. M. Forster, 2010: The transient response of global-mean precipitation  
613 to increasing carbon dioxide levels. *Environmental Research Letters*, **5**, 025212.
- 614 Bala, G., K. Caldeira, and R. Nemani, 2010: Fast versus slow response in climate change:  
615 implications for the global hydrological cycle. *Climate Dynamics*, **35**, 423–434.
- 616 Bayarri, M., et al., 2007: Computer model validation with functional output. *Annals of*  
617 *Statistics*, **35**, 1874–1906.
- 618 Bhat, K., M. Haran, R. Olson, and K. Keller, 2012: Inferring likelihoods and climate system  
619 characteristics from climate models and multiple tracers. *Environmetrics*, **23** (4), 345–362.
- 620 Branstator, G. and H. Teng, 2010: Two limits of initial-value decadal predictability in a  
621 CGCM. *Journal of Climate*, **23**, 6292–6311.
- 622 Cabre, M., S. Solman, and M. Nunez, 2010: Creating regional climate change scenarios over  
623 southern South America for the 2020’s and 2050’s using the pattern scaling technique:  
624 validity and limitations. *Climatic Change*, **98** (3-4), 449–469.
- 625 Cao, L., G. Bala, and K. Caldeira, 2011: Why is there a short-term increase in global  
626 precipitation in response to diminished CO<sub>2</sub> forcing? *Geophysical Research Letters*, **38**,  
627 L06703.
- 628 Castruccio, S. and M. Stein, 2013: Global space-time models for climate ensembles. *Annals*  
629 *of Applied Statistics*, in press.

- 630 Challenor, P., R. Hankin, and M. R., 2006: Towards the probability of rapid climate  
631 change. *Avoiding Dangerous Climate Change*, H. Schellnhuber, W. Cramer, N. Nakicen-  
632 ovic, T. Wigley, and G. Yohe, Eds., Cambridge University Press, Cambridge, 55–63.
- 633 Challenor, P., D. McNeall, and J. Gattiker, 2010: Assessing the probability of rare climate  
634 events. *The Oxford Handbook of Applied Bayesian Analysis*, A. O’Hagan and M. West,  
635 Eds., Oxford University Press, Oxford, 403–430.
- 636 Collins, M., 2002: Climate predictability on interannual to decadal time scales: the initial  
637 value problem. *Climate Dynamics*, **19**, 671–692.
- 638 Collins, M. and M. R. Allen, 2002: Assessing the relative roles of initial and boundary  
639 conditions in interannual to decadal climate predictability. *Journal of Climate*, **15**, 3104–  
640 3109.
- 641 Collins, W., et al., 2006: The Community Climate System Model: CCSM3. *Journal of*  
642 *Climate*, **19**, 2122–2143, doi:10.1175/JCLI3761.1.
- 643 Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change  
644 projections: the role of internal variability. *Climate Dynamics*, **38**, 527–546.
- 645 Dessai, S., X. Lu, and M. Hulme, 2005: Limited sensitivity analysis of regional climate  
646 change probabilities for the 21st century. *Journal of Geophysical Research: Atmospheres*,  
647 **110 (D19)**.
- 648 Drignei, D., C. E. Forest, and D. Nychka, 2008: Parameter estimation for computationally  
649 intensive nonlinear regression with an application to climate modeling. *Annals of Applied*  
650 *Statistics*, **2 (4)**, 1217–1230.
- 651 Fei, L. and M. West, 2009: A dynamic modelling strategy for Bayesian computer emulation.  
652 *Bayesian Analysis*, **4 (2)**, 393–412.

- 653 Forster, P., et al., 2007: Changes in atmospheric constituents and in radiative forcing. *Cli-*  
654 *mate Change 2007: The Physical Science Basis. Contribution of Working Group I to the*  
655 *Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon,  
656 D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller, Eds.,  
657 Cambridge University Press, Cambridge.
- 658 Fowler, H. J., S. Blenkinsop, and C. Tebaldi, 2007: Linking climate change modelling to  
659 impacts studies: recent advances in downscaling techniques for hydrological modelling.  
660 *International Journal of Climatology*, **27** (12), 1547–1578.
- 661 Giorgi, F., 2008: A simple equation for regional climate change and associated uncertainties.  
662 *Journal of Climate*, **21**, 1589–1604.
- 663 Hankin, R., 2005: Introducing BACCO, an R bundle for Bayesian analysis of computer code  
664 output. *Journal of Statistical Software*, **14**, 1–21.
- 665 Harris, G., D. Sexton, B. Booth, M. Collins, J. Murphy, and M. Webb, 2006: Frequency  
666 distributions of transient regional climate change from perturbed physics ensembles of  
667 general circulation model simulations. *Climate Dynamics*, **27**, 357–375.
- 668 Higdon, D., J. Gattiker, B. Williams, and M. Rightley, 2008: Computer model calibration  
669 using high dimensional output. *Journal of the American Statistical Association*, **103**, 570–  
670 583.
- 671 Holden, P. B. and N. R. Edwards, 2010: Dimensionally reduced emulation of an AOGCM for  
672 application to integrated assessment modelling. *Geophysical Research Letters*, **37**, L21707.
- 673 Hulme, M. and O. Brown, 1998: Portraying climate scenario uncertainties in relation to  
674 tolerable regional climate change. *Climate Research*, **10**, 1–14.
- 675 Hulme, M. and S. Raper, 1995: An integrated framework to address climate change (ES-

676 CAPE) and further developments of the global and regional climate modules (MAGICC).  
677 *Energy Policy*, **23**, 347–355.

678 Judge, G. et al., 1980: *The Theory and Practice of Econometrics*. Wiley, New York.

679 Kennedy, M. C. and A. O’Hagan, 2001: Bayesian calibration of computer models. *Journal*  
680 *of the Royal Statistical Society: Series B (Statistical Methodology)*, **63 (3)**, 425–464.

681 Lenton, T. et al., 2007: Effects of atmospheric dynamics and ocean resolution on bi-stability  
682 of the thermohaline circulation examined using the Grid Enabled Integrated Earth system  
683 modelling (GENIE) framework. *Climate Dynamics*, **29**, 591–613.

684 Manabe, S. and R. T. Wetherald, 1967: Thermal equilibrium of the atmosphere with a given  
685 distribution of relative humidity. *Journal of The Atmospheric Sciences*, **24**, 241–259.

686 McInerney, D. and E. Moyer, 2012: Direct and disequilibrium effects on precipitation in tran-  
687 sient climates. *Atmospheric Chemistry and Physics Discussions*, **12 (8)**, 19 649–19 681.

688 Meehl, G. A., et al., 2007: Global climate projections. *Climate Change 2007: The Physical*  
689 *Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the*  
690 *Intergovernmental Panel on Climate Change*, S. Solomon, D. Qin, M. Manning, Z. Chen,  
691 M. Marquis, K. B. Averyt, M. Tignor, and H. L. Miller, Eds., Cambridge University Press,  
692 Cambridge.

693 Mitchell, J., T. C. Johns, M. Eagles, W. J. Ingram, and R. A. Davis, 1999: Towards the  
694 construction of climate change scenarios. *Climatic Change*, **41**, 547–581, doi:10.1023/A:  
695 1005466909820.

696 Mitchell, T., 2003: Pattern scaling: An examination of the accuracy of the technique for  
697 describing future climates. *Climatic Change*, **60**, 217–242, doi:10.1023/A:1026035305597.

698 Montgomery, D., 2012: *Design and Analysis of Experiments*. eighth ed., Wiley, New York.

- 699 Murphy, J., B. Booth, M. Collins, G. Harris, D. Sexton, and M. Webb, 2007: A methodology  
700 for probabilistic predictions of regional climate change from perturbed physics ensembles.  
701 *Philosophical Transactions of the Royal Society A*, **365**, 1993–2028.
- 702 Oakley, O. and A. O’Hagan, 2002: Bayesian inference for the uncertainty distribution of  
703 computer model outputs. *Biometrika*, **89**, 769–784.
- 704 O’Hagan, A., 2006: Bayesian analysis of computer code output: A tutorial. *Reliability En-*  
705 *gineering and System Safety*, **91**, 1290–1300.
- 706 Rougier, J., 2008: Efficient emulators for multivariate deterministic functions. *Journal of*  
707 *Computational and Graphical Statistics*, **17** (4), 827–843.
- 708 Rougier, J., D. Sexton, J. Murphy, and D. Stainforth, 2009: Analyzing the climate sensitivity  
709 of the HadSM3 climate model using ensembles from different but related experiments.  
710 *Journal of Climate*, **22**, 3540–3557.
- 711 Ruosteenoja, K., T. Carter, K. Jylha, and H. Tuomenvirta, 2003: Future climate in world  
712 regions: an intercomparison of model-based projections for the new IPCC emissions sce-  
713 narios. Tech. rep., Finnish Environment Institute, Helsinki, 83 pp.
- 714 Sacks, J., W. Welch, T. Mitchell, and H. Wynn, 1989: Design and analysis of computer  
715 experiments. *Statistical Science*, **4** (4), 409–423.
- 716 Sanso, B. and C. Forest, 2009: Statistical calibration of climate system properties. *Journal*  
717 *of the Royal Statistical Society: Series C (Applied Statistics)*, **58** (4), 485–503.
- 718 Sanso, B., C. Forest, and D. Zantedeschi, 2008: Inferring climate system properties using a  
719 computer model. *Bayesian Analysis*, **3** (1), 1–37.
- 720 Santer, B. D., T. M. L. Wigley, M. E. Schlesinger, and J. F. B. Mitchell, 1990: Developing  
721 climate scenarios from equilibrium GCM results. Tech. Rep. 47, Max-Planck-Institut-für-  
722 Meteorologie, Hamburg.

- 723 Santner, T. J., W. B., and N. W., 2003: *The Design and Analysis of Computer Experiments*.  
724 Springer-Verlag, New York.
- 725 Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2003: An overview of CMIP5 and the  
726 experiment design. *Bulletin of the American Meteorological Society*, **93**, 485–498.
- 727 Vecchi, G. A., M. Zhao, H. Wang, G. Villarini, A. Rosati, A. Kumar, I. Held, and R. Gudgel,  
728 2011: Statistical-dynamical predictions of seasonal North Atlantic hurricane activity.  
729 *Monthly Weather Review*, **139**, 1070–1082.
- 730 Wilkinson, R. D., 2010: Bayesian calibration of expensive multivariate computer experi-  
731 ments. *Large-Scale Inverse Problems and Quantification of Uncertainty*, John Wiley &  
732 Sons, Ltd, Chichester, UK, 195–215, doi:10.1002/9780470685853.ch10.
- 733 Wilks, D. S., 2012: “Superparameterization” and statistical emulation in the Lorenz ’96  
734 system. *Quarterly Journal of the Royal Meteorological Society*, **138 (666)**, 1379–1387.
- 735 Wu, P., R. Wood, J. Ridley, and J. Lowe, 2010: Temporary acceleration of the hydrological  
736 cycle in response to a CO2 rampdown. *Geophysical Research Letters*, **37**, L12705, doi:  
737 10.1029/2010GL043730.
- 738 Yeager, S., C. Shields, W. Large, and J. Hack, 2006: The low-resolution CCSM3. *Journal of*  
739 *Climate*, **19**, 2545–2566, doi:10.1175/JCLI3744.1.

## 740 List of Figures

- 741 1 CO<sub>2</sub> scenarios used for building the collection of runs. We refer to these  
742 throughout the paper as 1- *slow*, 2- *moderate*, 3- *fast*, 4- *jump* and 5- *drop*  
743 scenario. All scenarios start at year 1870. Some scenarios extend beyond the  
744 range shown here: *slow*, *moderate* and *fast* end at year 2449, while *jump* ends  
745 at 2199 and *drop* at 2399. 34
- 746 2 An example of temperature emulation for the North Pacific West (NPW)  
747 region (a-b), chosen as representative of a region with significant change, and  
748 of precipitation emulation for the Equatorial Pacific West (EPW) region (c-d),  
749 chosen to highlight interesting transient precipitation behavior. Panels a and c  
750 show the emulated *slow* scenario and b and d the *drop* scenario. The emulator  
751 was trained by one realization each of the *fast* and *jump* scenarios. The  
752 solid red line represents the emulated mean function and the gray lines shows  
753 the five CCSM3 realizations for the scenarios. Emulation captures expected  
754 transient precipitation behavior in which precipitation anomaly is a function  
755 of the rate of change in radiative forcing. Note that the trend in temperature  
756 is larger relative to stochastic variability than it is for precipitation. We  
757 define diagnostics of emulation goodness-of-fit  $I_1$  and trend-vs-variability  $I_2$   
758 in Section 4a. Values of  $(I_1, I_2)$  for the emulations shown here in panels a-d  
759 are  $(1.01, 11.23)$ ,  $(1.94, 35.82)$ ,  $(1.02, 1.18)$  and  $(1.09, 1.41)$ , respectively;  $I_2$  is  
760 much larger for temperature, as expected. 35

761 3 An example of uncertainty quantification for temperature in the North Pacific  
762 West (NPW) region (a,c,e), and for precipitation emulation for the Equato-  
763 rial Pacific West (EPW) region (b,d,f). All panels show the emulated *slow*  
764 scenario. The emulator was trained by one realization each of the *fast* and  
765 *jump* scenarios. The first row (a-b) shows an example of emulated realiza-  
766 tions. The gray lines represent the five CCSM3 realizations and the red lines  
767 the five emulated realizations (with an offset of 1 degree for temperature and  
768 1000 mm/yr for precipitation). The actual runs and those simulated via the  
769 emulator appear to be qualitatively similar. The second row (c-d) shows the  
770 five superimposed CCSM3 realizations in gray, and the dashed red lines the  
771 95% prediction bands from the emulator. Empirical coverage is 0.9531 for  
772 (c) and 0.9545 for (d), very close to the nominal 95% level. The bottom row  
773 (e-f) plots the mean across the five CCSM3 realizations of the *slow* scenario in  
774 gray, and the dashed red lines represent the pointwise 95% confidence bands  
775 based on the emulator. The bands are very narrow, especially for tempera-  
776 ture, highlighting the ability of the emulator to capture the mean trend with  
777 very high precision. 36

778 4 Emulation indices for all regions for the regional temperature emulation de-  
779 scribed in the text and shown in Figure 2. Large font is the “emulation opti-  
780 mality” index  $I_1$  ( $\times 100$ ) and small font the “trend” index  $I_2$ . Low  $I_2$  means  
781 there is little trend relative to noise and the  $I_1$  index is not informative, even  
782 if close to 100 (optimal emulation). Top panel shows the *slow* and bottom  
783 panel the *drop* scenarios. Emulation is worse for the physically extreme drop  
784 scenario, as expected, but is generally close to the optimal value of 1 in most  
785 inhabited regions. All indices have been computed between year 2010 and the  
786 farthest time point (2449 for *slow*, 2399 for *drop*). 37

787	5	Boxplots of the fit index ( $I_1$ ) for various numbers of realizations (b) and sce-	
788		narios (a) in the training set for mean emulation of the <i>moderate</i> stabilization	
789		scenario. The training sets for the realization test (b) are made up of 1-5 re-	
790		realizations of the <i>fast</i> scenario; those for the scenario test (a) are made up of a	
791		single realization of <i>slow</i> and then adding, successively, one realization of <i>fast</i> ,	
792		<i>jump</i> and <i>drop</i> . Panel (b) shows adding realizations of a single scenario offers	
793		a modest benefit and (a) shows that adding scenarios too dissimilar from the	
794		test case can actually degrade emulator performance. Box-and-whisker plots	
795		exclude severe outliers, which are shown with their regional codes. Four of	
796		the five outliers lie in the polar regions; see supplementary material Figure S1	
797		for locations.	38
798	6	Construction of regional pattern scaling for temperature: linear regressions of	
799		regional temperature anomalies on GMT. Data used are 60 years from 2010 to	
800		2070 (we picture a subset of the data in this figure for visualization purposes)	
801		for all 47 regions in the standard training set consisting of the <i>fast</i> and <i>jump</i>	
802		scenarios. The two scenarios are shown in different colors. Regional subfigures	
803		are arranged to approximate their geographic distribution (north at top) to	
804		give an idea of spatial patterns. Subfigures share a consistent y-axis scale, so	
805		that differences in warming rate and variability may be seen by eye.	39
806	7	Construction of regional pattern scaling for precipitation: linear regressions	
807		of regional precipitation anomalies on GMT, as in Figure 6 for temperature.	
808		Because precipitation anomalies differ widely between regions, y-axis scales	
809		(in % anomaly) are shown separately for each subfigure.	40

810 8 Comparison between statistical emulation and pattern scaling for regional  
811 temperature. Training set, predicted scenarios, and time range for calculating  
812 indices are as in Figure 4. The top number shown in each region is the log  
813 ratio of the temperature fit indices  $I_1$  for the statistical model (numerator)  
814 and pattern scaling (denominator), multiplied by 100 for clarity. Negative  
815 numbers mean that statistical emulation outperforms pattern scaling. The  
816 small type gives the trend index  $I_2$ , which does not depend on the emulator.  
817 For the *slow* scenario (top), the median log ratio across all the regions times  
818 100 is  $-1.35$  (with 10% and 90% quantiles are of  $-2.94$  and  $0.93$ ), indicating  
819 a modest advantage from statistical emulation. Statistical emulation provides  
820 stronger benefits for the *drop* scenario: the median log ratio is  $-7.42$  (with  
821 10% and 90% quantiles of  $-30.08$  and  $10.68$ ). 41

822 9 Comparison between statistical emulation and pattern scaling for regional  
823 precipitation. Quantities shown are as in Figure 8. The high variability in  
824 precipitation leads to smaller  $I_2$  values and reduces the distinction between  
825 emulation methods. For the *slow* scenario, the median log ratio of  $I_1$  across  
826 all regions (times 100) is  $-0.40$  (with 10% and 90% quantiles of  $-1.74$  and  
827  $0.23$ ); for the *drop* scenario it is  $-0.93$  (with 10% and 90% quantiles of  $-5.70$   
828 and  $5.91$ ). 42

829 10 Emulating temperature at grid resolution and comparison with pattern scal-  
830 ing. On the left, log ratio of the fit index ( $I_1$ ) for statistical emulation of  
831 the *drop* scenario over pattern scaling. This is the grid-scaled case of Figure  
832 8 bottom panel. Negative values (blue) indicate that statistical emulation  
833 outperforms pattern scaling. On the right, the average log ratio for different  
834 latitude bands. Statistical emulation generally outperforms pattern scaling  
835 outside the polar regions. 43

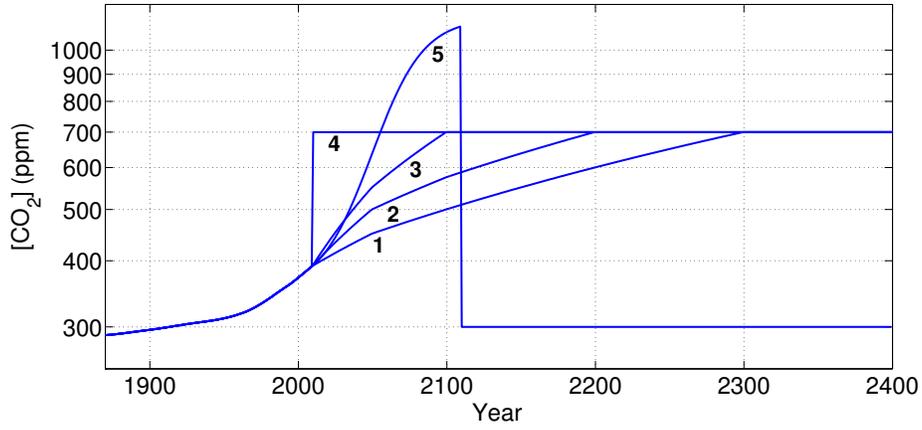


FIG. 1. CO<sub>2</sub> scenarios used for building the collection of runs. We refer to these throughout the paper as 1- *slow*, 2- *moderate*, 3- *fast*, 4- *jump* and 5- *drop* scenario. All scenarios start at year 1870. Some scenarios extend beyond the range shown here: *slow*, *moderate* and *fast* end at year 2449, while *jump* ends at 2199 and *drop* at 2399.

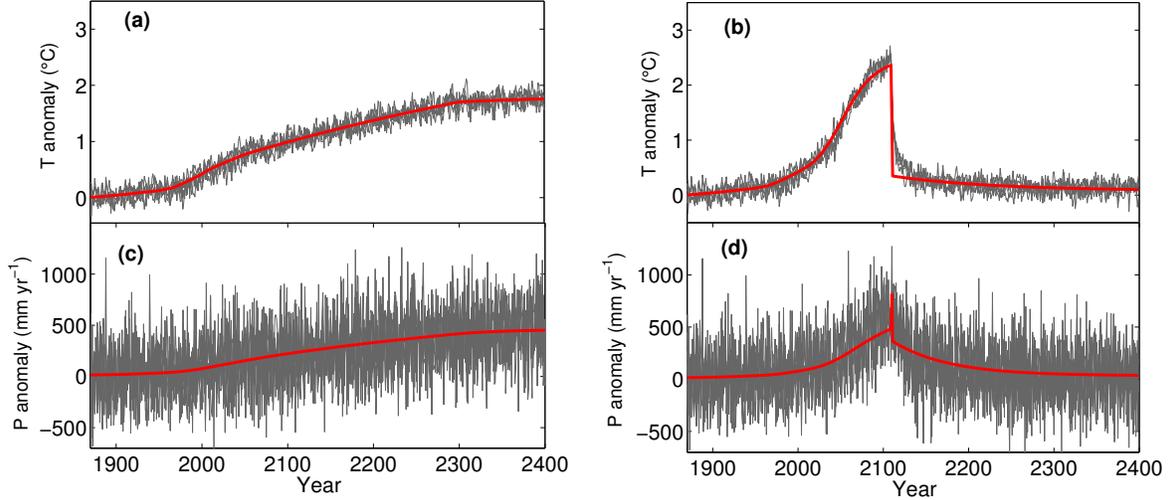


FIG. 2. An example of temperature emulation for the North Pacific West (NPW) region (a-b), chosen as representative of a region with significant change, and of precipitation emulation for the Equatorial Pacific West (EPW) region (c-d), chosen to highlight interesting transient precipitation behavior. Panels a and c show the emulated *slow* scenario and b and d the *drop* scenario. The emulator was trained by one realization each of the *fast* and *jump* scenarios. The solid red line represents the emulated mean function and the gray lines shows the five CCSM3 realizations for the scenarios. Emulation captures expected transient precipitation behavior in which precipitation anomaly is a function of the rate of change in radiative forcing. Note that the trend in temperature is larger relative to stochastic variability than it is for precipitation. We define diagnostics of emulation goodness-of-fit  $I_1$  and trend-vs-variability  $I_2$  in Section 4a. Values of  $(I_1, I_2)$  for the emulations shown here in panels a-d are  $(1.01, 11.23)$ ,  $(1.94, 35.82)$ ,  $(1.02, 1.18)$  and  $(1.09, 1.41)$ , respectively;  $I_2$  is much larger for temperature, as expected.

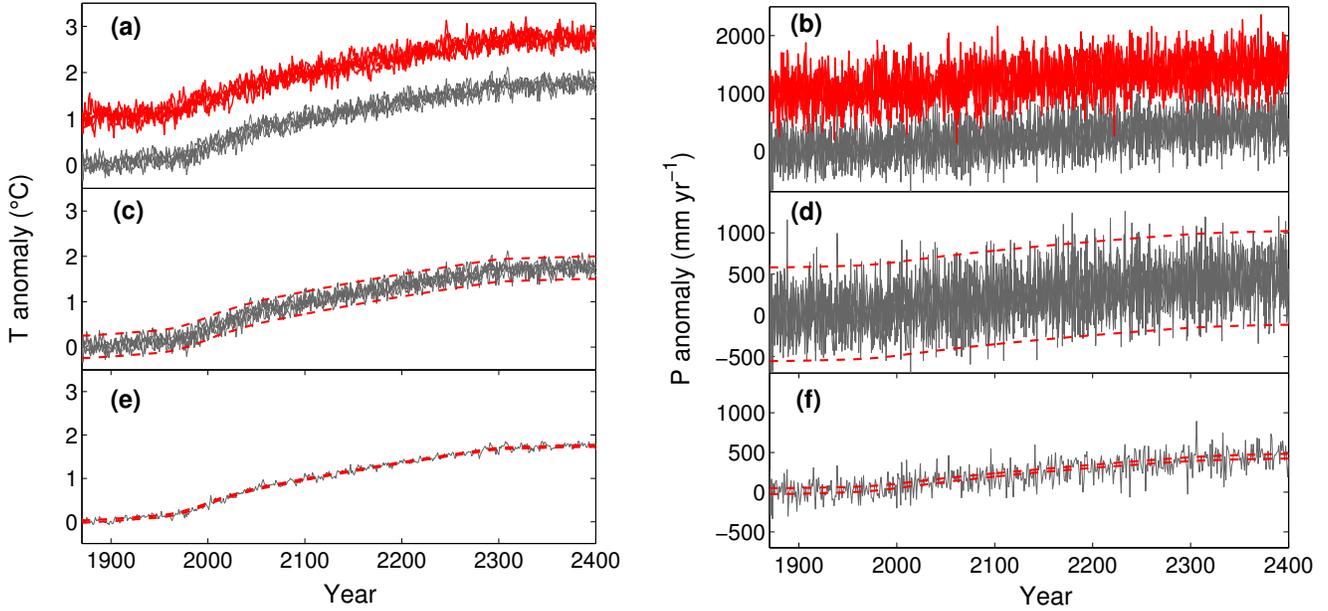


FIG. 3. An example of uncertainty quantification for temperature in the North Pacific West (NPW) region (a,c,e), and for precipitation emulation for the Equatorial Pacific West (EPW) region (b,d,f). All panels show the emulated *slow* scenario. The emulator was trained by one realization each of the *fast* and *jump* scenarios. The first row (a-b) shows an example of emulated realizations. The gray lines represent the five CCSM3 realizations and the red lines the five emulated realizations (with an offset of 1 degree for temperature and 1000 mm/yr for precipitation). The actual runs and those simulated via the emulator appear to be qualitatively similar. The second row (c-d) shows the five superimposed CCSM3 realizations in gray, and the dashed red lines the 95% prediction bands from the emulator. Empirical coverage is 0.9531 for (c) and 0.9545 for (d), very close to the nominal 95% level. The bottom row (e-f) plots the mean across the five CCSM3 realizations of the *slow* scenario in gray, and the dashed red lines represent the pointwise 95% confidence bands based on the emulator. The bands are very narrow, especially for temperature, highlighting the ability of the emulator to capture the mean trend with very high precision.

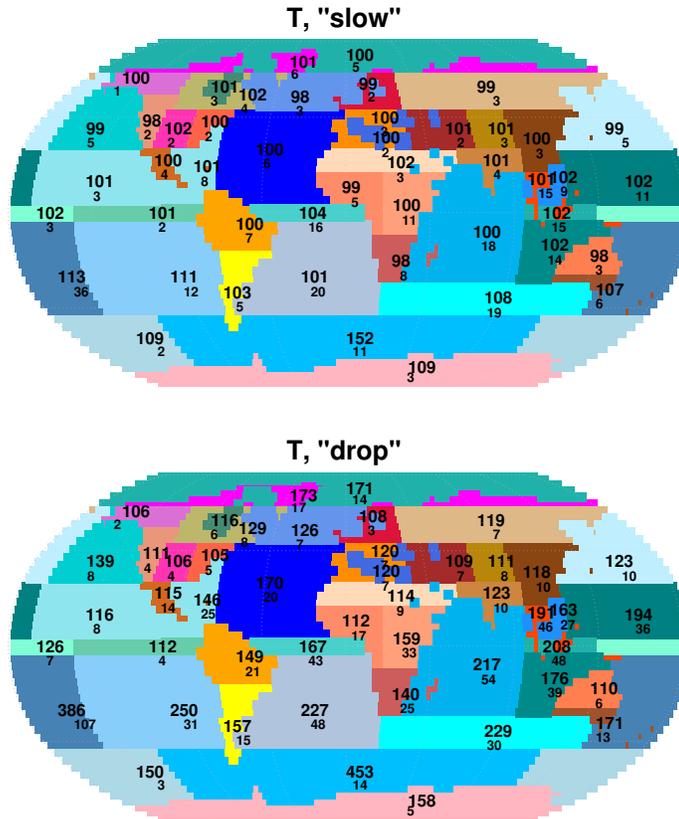


FIG. 4. Emulation indices for all regions for the regional temperature emulation described in the text and shown in Figure 2. Large font is the “emulation optimality” index  $I_1$  ( $\times 100$ ) and small font the “trend” index  $I_2$ . Low  $I_2$  means there is little trend relative to noise and the  $I_1$  index is not informative, even if close to 100 (optimal emulation). Top panel shows the *slow* and bottom panel the *drop* scenarios. Emulation is worse for the physically extreme drop scenario, as expected, but is generally close to the optimal value of 1 in most inhabited regions. All indices have been computed between year 2010 and the farthest time point (2449 for *slow*, 2399 for *drop*).

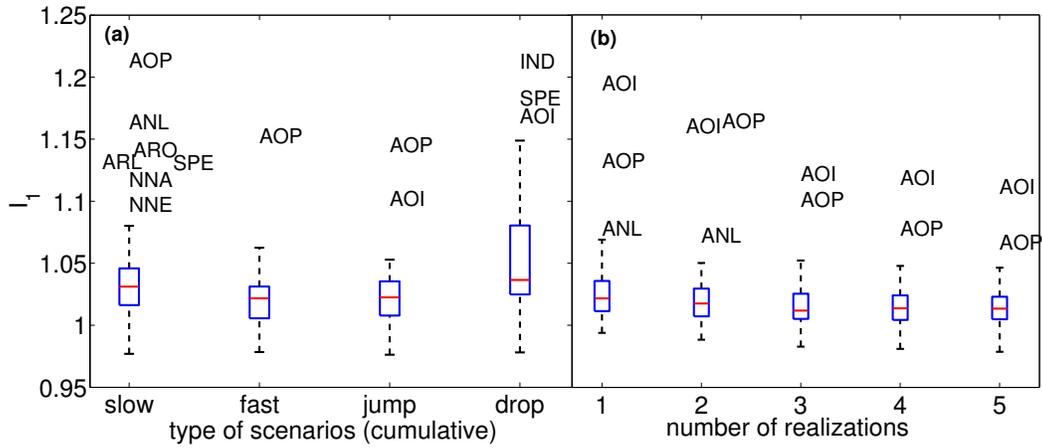


FIG. 5. Boxplots of the fit index ( $I_1$ ) for various numbers of realizations (b) and scenarios (a) in the training set for mean emulation of the *moderate* stabilization scenario. The training sets for the realization test (b) are made up of 1-5 realizations of the *fast* scenario; those for the scenario test (a) are made up of a single realization of *slow* and then adding, successively, one realization of *fast*, *jump* and *drop*. Panel (b) shows adding realizations of a single scenario offers a modest benefit and (a) shows that adding scenarios too dissimilar from the test case can actually degrade emulator performance. Box-and-whisker plots exclude severe outliers, which are shown with their regional codes. Four of the five outliers lie in the polar regions; see supplementary material Figure S1 for locations.

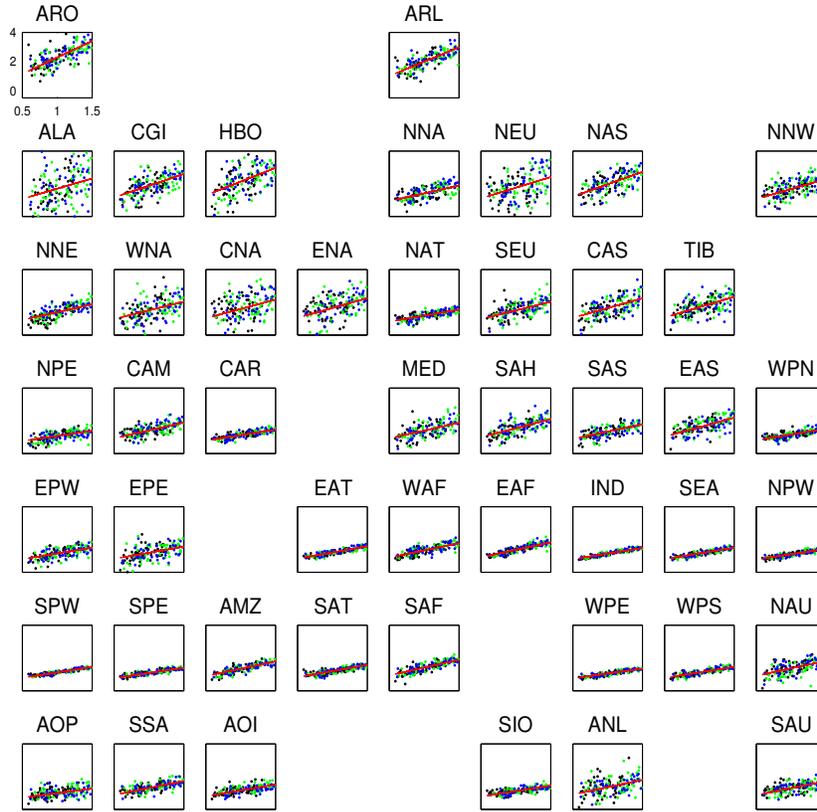


FIG. 6. Construction of regional pattern scaling for temperature: linear regressions of regional temperature anomalies on GMT. Data used are 60 years from 2010 to 2070 (we picture a subset of the data in this figure for visualization purposes) for all 47 regions in the standard training set consisting of the *fast* and *jump* scenarios. The two scenarios are shown in different colors. Regional subfigures are arranged to approximate their geographic distribution (north at top) to give an idea of spatial patterns. Subfigures share a consistent y-axis scale, so that differences in warming rate and variability may be seen by eye.

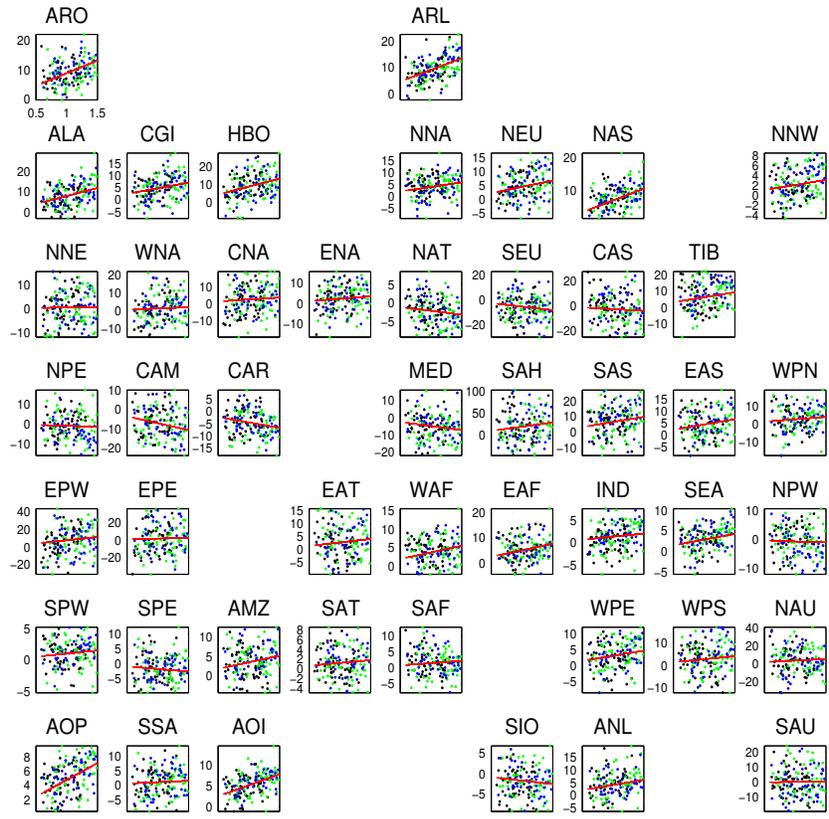


FIG. 7. Construction of regional pattern scaling for precipitation: linear regressions of regional precipitation anomalies on GMT, as in Figure 6 for temperature. Because precipitation anomalies differ widely between regions, y-axis scales (in % anomaly) are shown separately for each subfigure.

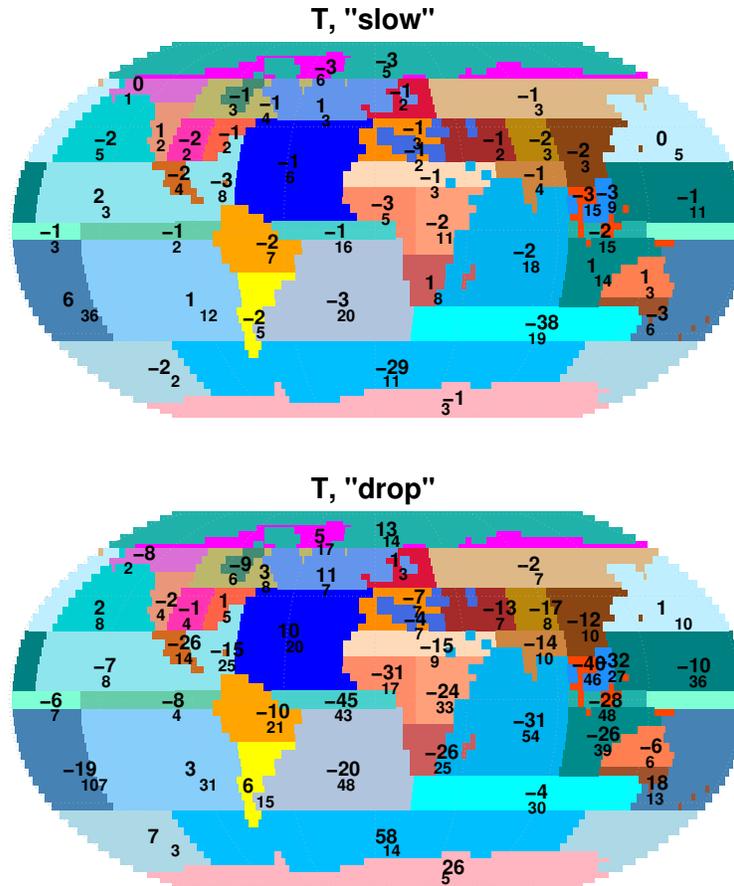


FIG. 8. Comparison between statistical emulation and pattern scaling for regional temperature. Training set, predicted scenarios, and time range for calculating indices are as in Figure 4. The top number shown in each region is the log ratio of the temperature fit indices  $I_1$  for the statistical model (numerator) and pattern scaling (denominator), multiplied by 100 for clarity. Negative numbers mean that statistical emulation outperforms pattern scaling. The small type gives the trend index  $I_2$ , which does not depend on the emulator. For the *slow* scenario (top), the median log ratio across all the regions times 100 is  $-1.35$  (with 10% and 90% quantiles are of  $-2.94$  and  $0.93$ ), indicating a modest advantage from statistical emulation. Statistical emulation provides stronger benefits for the *drop* scenario: the median log ratio is  $-7.42$  (with 10% and 90% quantiles of  $-30.08$  and  $10.68$ ).

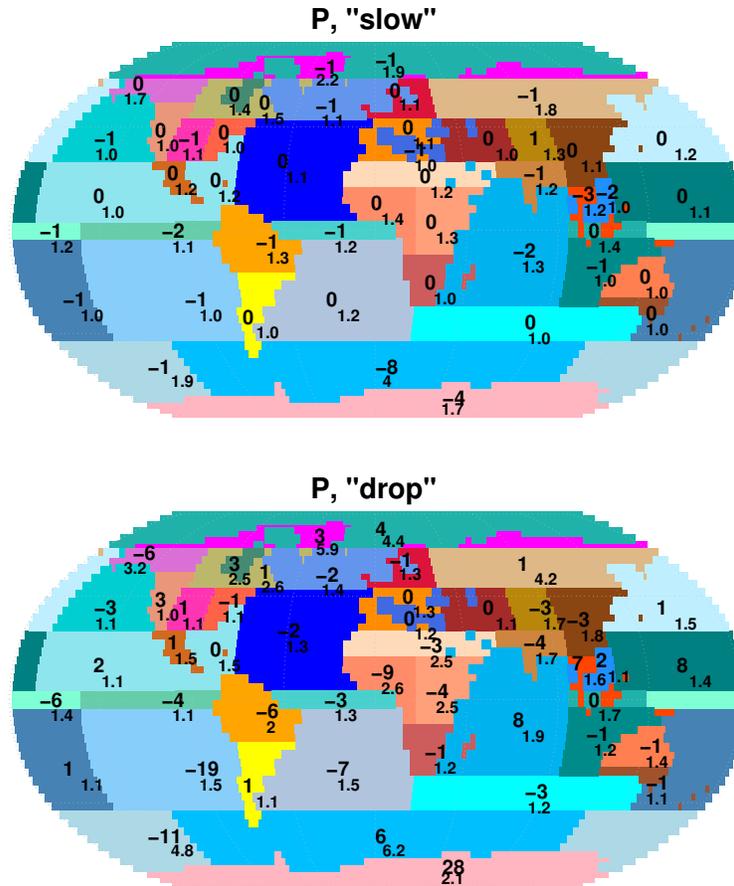


FIG. 9. Comparison between statistical emulation and pattern scaling for regional precipitation. Quantities shown are as in Figure 8. The high variability in precipitation leads to smaller  $I_2$  values and reduces the distinction between emulation methods. For the *slow* scenario, the median log ratio of  $I_1$  across all regions (times 100) is  $-0.40$  (with 10% and 90% quantiles of  $-1.74$  and  $0.23$ ); for the *drop* scenario it is  $-0.93$  (with 10% and 90% quantiles of  $-5.70$  and  $5.91$ ).

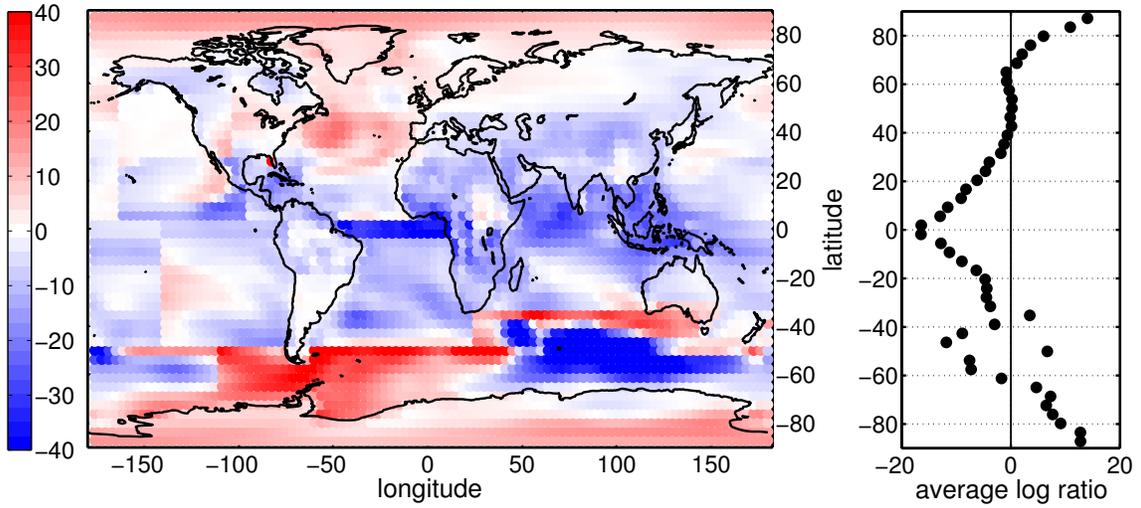


FIG. 10. Emulating temperature at grid resolution and comparison with pattern scaling. On the left, log ratio of the fit index ( $I_1$ ) for statistical emulation of the *drop* scenario over pattern scaling. This is the grid-scaled case of Figure 8 bottom panel. Negative values (blue) indicate that statistical emulation outperforms pattern scaling. On the right, the average log ratio for different latitude bands. Statistical emulation generally outperforms pattern scaling outside the polar regions.