Micro and macro parametric uncertainty in climate change prediction: a large ensemble perspective

Francisco de Melo Viríssimo[a] and David A. Stainforth[a]

[a] *Grantham Research Institute on Climate Change and the Environment, London School of Economics and Political Science, London, WC2A 2AE, United Kingdom*

Corresponding author: Francisco de Melo Viríssimo, f.de-melo-virissimo@lse.ac.uk

Early Online Release: This preliminary version has been accepted for publication in *Bulletin of the American Meteorological Society*, may be fully cited, and has been assigned DOI 10.1175/BAMS-D-24-0064.1. The final typeset copyedited article will replace the EOR at the above DOI when it is published.

ABSTRACT: Earth system models (ESMs) are widely used to make projections of the future behaviour of the earth's climate in the context of anthropogenic climate change. Setting aside uncertainties stemming from the design and implementation of the model, there nevertheless remain substantial uncertainties with such projections. Two important ones arise from uncertainties in (a) the initial conditions and (b) the values of parameters within the model. Here we systematically investigate the latter: the consequences of parametric uncertainty as might be explored by perturbed parameter ensembles. Utilising a low-dimensional system with key characteristics of a climate model, we examine two types of parametric uncertainty through a large ensemble approach. The first, micro-parametric uncertainty, is akin to micro-initial condition uncertainty and explores a situation where one knows the relevant parameter values well but not perfectly. The second, macro-parametric uncertainty, explores the situation where there may be substantial uncertainty in parameter values. We also investigate how they interact with each other and with micro-initial condition uncertainty. In general, we find that micro-parametric uncertainty can lead to a much broader range of states than in initial condition ensembles, with the resulting standard deviations being over 2.5-3.5 times higher for slow- and fast-mixing variables alike. Additionally, we show that the scale of the effect may be even larger with macro-parametric uncertainty. Finally, we discuss the implications for ensemble design and interpretation, and particularly how these results indicate the need for more complex ensemble designs when making projections of climate change within ESMs.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

SIGNIFICANCE STATEMENT: This study presents a systematic discussion of the sources and consequences of parametric uncertainty in climate models. Two types of parametric uncertainty are found depending on the source and magnitude of uncertainty: a micro and macro uncertainty, with the latter being subdivided into two types. While parametric ensembles are generally found to quantify a broader range of plausible states and behaviour than initial condition ensembles, the results are varied and dependent on multiple factors, such as the number of parameters perturbed and the output variable of interest. Together, these results shed some light on how to better design informative climate model ensembles, particularly given the computational power demands of running such models.

1. Introduction

Earth System Models (ESMs) are highly complex and nonlinear mathematical representations of the Earth's spheres, including the atmosphere, ocean, land and cryosphere, and their interactions (Dijkstra 2013). They are widely used to make projections of future climate under scenarios for future anthropogenic greenhouse gas emissions within projects such as the Coupled Model Intercomparison Project (CMIP; Eyring et al. (2016)). These projections provide key inputs to a range of high profile publications, such as the United Nation's Intergovernmental Panel on Climate Change (IPCC) assessment reports (IPCC 2023).

Being highly complex and nonlinear, ESMs are too complicated to be tractable analytically or qualitatively. Indeed, they are inherently computer-based models, discretised in space and time and with an extraordinarily large number of degrees of freedom. For instance, a typical 6th generation CMIP model can have something like $10^7 - 10^9$ degrees of freedom, making them very expensive to study even with high-performance computers.

ESMs also include a large number of parameters. While some of the parameters arise from physical laws and have well defined meanings, such as the Coriolis parameter, many others are within "parameterisations": simplified, closed-form mathematical representations of processes that cannot be resolved explicitly due to either limited resolution or lack of scientific understanding. Examples include the "1/3" power-law ratio in cloud micro-physics (Liu and Hallett 1997), sub-shelf melting parameterisations in ice sheet models (Favier et al. 2019), the Gent-McWilliams eddy parameterisation in ocean circulation dynamics (Gent and McWilliams 1990; Gent 2011), and the

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Martin curve for particulate organic carbon attenuation in marine biogeochemistry (Martin et al. 1987; de Melo Viríssimo et al. 2022), to mention just a few. Overall, this results in ESMs having hundreds of free parameters that must be specified by either the model developer or user.

In practice, determining these parameters is difficult. Even when they have a well-defined meaning within a physical law, their values are still inherently uncertain due to measurement limitations (Youden 1961), effectively meaning that many values are plausible within a measured uncertainty range. There is also a question over whether the value of such parameters in reality are necessarily suited to reproducing the behaviour of the system as a whole in a multi-component, complex model where some processes are missing and others are different from those occurring in reality (Martin et al. 2024). Other parameters may not be directly measurable but their values may still be guided by observations of the processes they represent, or they might instead be chosen to maximise the ability of the model to reproduce a collection of observations of the system more broadly. The latter is the widely acknowledged process of model calibration or "tuning" (Hourdin et al. 2017), which may be performed manually (so-called "expert" tuning) or objectively via optimisation algorithms (Bellprat et al. 2012b; Tett et al. 2017). There is, however, still a question of whether the ability to reproduce such observations (and pass a "climate Turing test" (Palmer 2016)) is an adequate target measure given missing processes, compensation of errors, and the extrapolatory nature of the model-based projections. Relatedly, some parameters may be expected to be space- or time-dependent but are represented in the model by a single value.

Together, these obstacles to determining parameter values result in ESMs having a large number of poorly constrained parameters, which one might expect to be an important source of uncertainty in climate change projections. It is natural, therefore, to talk about the consequences of parametric uncertainty when considering such climate models: how might different parameter values affect their results?

a. Quantifying the consequences of parametric uncertainty in models

One way of quantifying this is to run what are called *perturbed parameter ensembles* (PPEs[1]; Parker (2013); Carslaw et al. (2018)). These are collections of model simulations where each one is generated by perturbing uncertain parameters (individually or collectively) to explore

[1]Sometimes the acronym "PPE" is used to refer to perturbed physics ensembles, which can go beyond parameter uncertainty and consider uncertainties in the model formulation as well.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

the variability that this uncertainty generates in the model forecast (Hargreaves 2010). Note that this is very different to multi-model ensembles (or MMEs), which are ensembles-of-opportunity and the approach taken by projects such as CMIP (Masson and Knutti 2011; Knutti et al. 2013; Eyring et al. 2016). PPEs are also different to the so-called stochastic parameterisations, a common feature in modern weather and atmosphere models, which attempts to emulate the uncertainty in unresolved physical processes in general circulation models by multiplying the sub-grid parameterisations (or parameters) by a time-dependent stochastic random variable (Buizza et al. 1999; Palmer 2001; Berner et al. 2017), resulting in a time-dependent parameterisation.

PPEs are not new. Indeed, there is a broad literature on the subject, ranging from the very large PPEs produced by climateprediction.net (Stainforth et al. 2005; Sanderson et al. 2008), to discussion around the implications of ensemble design and interpretation (Stainforth et al. 2007a,b), to large ad hoc PPEs in regional (Bellprat et al. 2012a) and global (Murphy et al. 2004; Collins et al. 2011) climate models, to considerations on robustness of bifurcation points (Knopf et al. 2005) and effective number of degrees of freedom (Peatier et al. 2024) in multi-parameter spaces, to cite a few. However, a systematic analysis of the consequences of parametric uncertainty in the context of climate modelling remains largely absent. This is an important gap, because quantifying parametric uncertainty is a key element in the design of ensembles that aim to be informative, and it is unclear to what extent available approaches such as ad hoc sampling are adequate. This paper aims to fill this gap through a broad investigation of parametric uncertainty in a low-dimensional nonlinear system that encapsulates key characteristics of a climate model. We approach the problem by considering the system dynamics in an ensemble distribution sense rather than studying single trajectories.

We explore two aspects of parameter uncertainty. The first, which we call micro parametric uncertainty (micro PU) by comparison with micro initial condition uncertainty (micro ICU; Stainforth et al. (2007a); de Melo Viríssimo et al. (2024)), is used to quantify the response uncertainty in a situation where we believe we know the parameter values well, but not perfectly. In this case, we explore small, residual uncertainties in the parameter values, where the order of magnitude of the uncertainty is much smaller than the parameter value itself. This exploration is done with a *micro perturbed parameter ensemble* (micro PPE): an ensemble of simulations with the same initial condition but different parameter values within the micro uncertainty range. This is defined

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

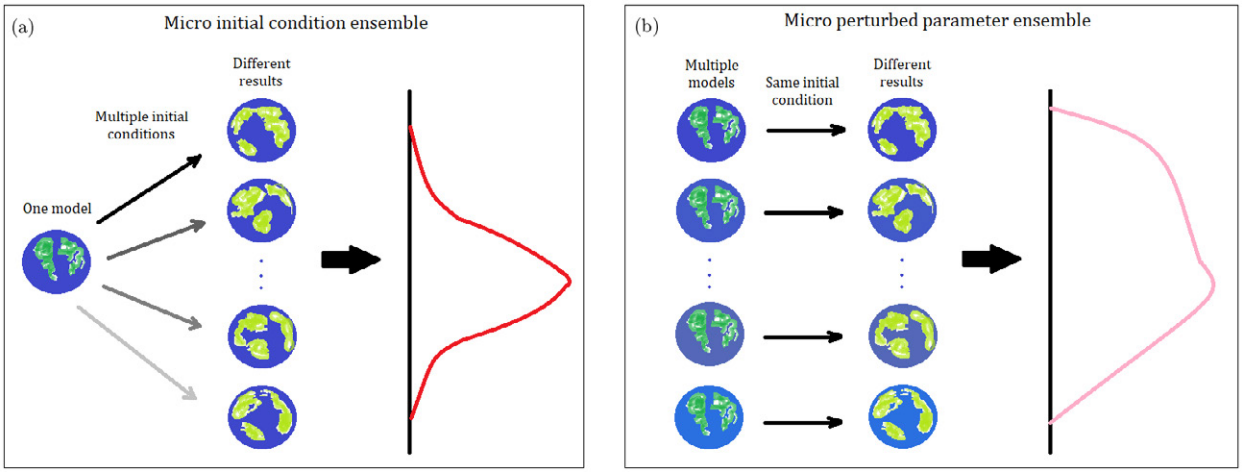

FIG. 1. The conceptual representation of micro ensembles as "parallel realisations". (a) micro initial condition ensemble: each arrow represents the evolution of the same planet, obeying the same physical laws, but starting from slightly different initial conditions. As the uncertainty grows exponentially, these lead to different future states of the planet, which can then be summarised as a distribution; (b) micro perturbed parameter ensemble: each globe representing a slightly different version of the same planet. Each of them is evolving from the same initial state, but end up at different states, which can also be summarised as a distribution.

by analogy to a *micro initial condition ensemble* (micro ICE), in which residual uncertainties in the *initial state* of the system is explored. However, the concepts behind micro ICE and micro PPE are rather different, as illustrated in Figure 1. Conceptually speaking, a micro ICE represents the evolution of equal copies of the same universe, obeying the same physical laws but starting from slightly different initial conditions (Herein et al. 2017; Tél et al. 2020), as shown in Figure 1(a). By contrast, a micro PPE represents the evolution of very similar but different universes, obeying slightly perturbed versions of the same physical laws, but starting from the same initial conditions, as shown in Figure 1(b).

The second aspect of parameter uncertainty that we explore we call macro parametric uncertainty (macro PU), again by comparison with macro ICU (de Melo Viríssimo et al. 2024). Here we explore the consequences large uncertainties in parameter values, a situation which would occur if, for instance, tuning leads to inappropriate optimisation due to compensation of errors in a model which does not contain certain processes that exist in reality. In this case, the order of magnitude of the uncertainty may be comparable with the parameter value itself.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-24-0064.1.

In what follows, we quantify and separate the consequences of micro ICU, micro PU and macro PU, and shed some light on how to design PPEs to better explore the future range of possibilities as constrained by today's models of the Earth system.

2. A low-dimensional dynamical systems approach

This work uses a low-dimensional coupled model with key characteristics of an atmosphere-ocean model. It consists of the Stommel 61 (S61) ocean model (Stommel 1961) coupled to the Lorenz 84 (L84) atmospheric model (Lorenz 1984; Provenzale and Balmforth 1999). Together, we refer to this as Lorenz 84-Stommel 61 (L84-S61) model (Roebber 1995; Veen et al. 2001). The model consists of the following equations:

$$\frac{d}{dt}X = -Y^2 - Z^2 - aX + a(F_0(t) + F_1 T) \tag{1}$$

$$\frac{d}{dt}Y = XY - bXZ - Y + G_0 + G_1(T_{av} - T) \tag{2}$$

$$\frac{d}{dt}Z = bXY + XZ - Z \tag{3}$$

$$\frac{d}{dt}T = k_a(\gamma X - T) - |f(T,S)|T - k_w T \tag{4}$$

$$\frac{d}{dt}S = \delta_0 + \delta_1(Y^2 + Z^2) - |f(T,S)|S - k_w S \tag{5}$$

where

$$f(T,S) = \omega T - \epsilon S \tag{6}$$

$$F_0(t) = F_m + M\cos(((2\pi t/73) - (\pi/12)) + F_{\mathrm{CC}}(t) \tag{7}$$

and

$$F_{\mathrm{CC}}(t) = \begin{cases} 0 & \text{if } t < t_{\mathrm{start}} \\ (H/73)(t - t_{\mathrm{start}}) & \text{if } t_{\mathrm{start}} \le t \le t_{\mathrm{end}} \\ (H/73)(t_{\mathrm{end}} - t_{\mathrm{start}}) & \text{if } t_{\mathrm{end}} < t. \end{cases} \tag{8}$$

Equations (1) to (3) describe the evolution of the high-frequency, atmospheric variables X, Y, Z, where X represents the intensity of the westerly wind, and Y and Z are the Fourier amplitudes

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

characterising a chain of large-scale eddies, which transport heat towards the pole at a rate proportional to their amplitude. Equations (4) and (5) describe the evolution of the slow ocean variables T, S, where T and S denote the pole-equator temperature and salinity differences, respectively. An important component of the ocean model is the thermohaline circulation (THC) f, which is given by Equation (6). The system dynamics is also modulated by an external forcing $F_0(t)$, given by Equation (7), which represents the cross-latitude (equator-to-pole) heating contrast. This consists of a seasonal cycle of amplitude M over a mean forcing F_m, together with a time-dependent linear forcing increase F_{CC} (given by Equation (8)) between times $t = t_{\mathrm{start}}$ and $t = t_{\mathrm{end}}$. This linear increase allows us to study exogenously driven changes to the state of the system; something that can be considered "climate change" within this system. It provides a parallel with the changing radiative forcing applied in ESMs in multi-decadal simulations of anthropogenic climate change. All quantities in the model are non-dimensional, with one unit of time (hereafter Lorenz time unit, or LTU) being equivalent to 5 days, implying 73 LTUs per year (hence the factor of 73 in Equations (7) and (8)). A detailed description of the unforced L84-S61 model can be found in Veen et al. (2001). The forcing $F_0(t)$ was introduced in Daron (2012), where further details can be found.

Despite its reduced number of equations, the model shares some key mathematical properties with state-of-the-art ESMs (de Melo Viríssimo and Stainforth 2023): it is complex, nonlinear, multi-component, multi-scale and, importantly, it is chaotic. In particular, the model represents distinct timescales of motion: a fast "atmosphere" and a slow "ocean", which are allowed to dynamically interact and exchange information with each other. More importantly, its low number of degrees of freedom allows for extensive computational studies using very large ensembles but at modest computational cost - something infeasible with spatially-resolving ESMs. For instance, a 1,000 member ensemble run for 200 years with a 1.2 hours time step and daily outputs would take somewhere between 2 and 5 hours on a personal laptop. This is in sharp contrast with EMSs, where even low resolution models can take 10-14 days to run a single 1,000-year simulation (Holden et al. 2016), not to mention CMIP-like models such as UKESM, which can take 1-2 months to run a single simulation for 80-100 years on a multi-core high-performance computer (Kuhlbrodt et al. 2015; Stringer 2017).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Low-dimensional, Lorenz-like models such as L84-S61 have been widely used, with great success, to study conceptual and practical questions in weather and climate sciences over the past 40 years. These include the irregularity of trajectories (Lorenz 1984), range of predictability (Lorenz 1996), initial condition ensemble design (Daron and Stainforth 2013), climate tipping points (Ashwin and Newman 2021), metastability (Mehling et al. 2024), initial condition uncertainty (de Melo Viríssimo et al. 2024), deep learning-based data assimilation (Bocquet et al. 2024), hysteresis in extreme events (Bódai and Tél 2012), to mention a few. Readers interested in a comprehensive discussion on the value of such models in geosciences should consult Ghil (2019) and Dijkstra (2024).

The L84-S61 model contains a total of 16 parameters, with 15 of these potentially open to physical constraint[2]. Hence, the dimension of the parameter space is much higher than the dimension of the state space. This is a distinct feature of the L84-S61 model by comparison to its spatially-resolving counterparts: in general, ESMs have a much higher dimensional state space than parameter space - primarily due to their discretisation in three spatial dimensions. All parameters for the L84-S61 model and their reference values are described in Table 1. The parameters in the forcing function $F_0(t)$ are at the bottom of Table 1.

In this work, we consider perturbations to 14 of the 16 parameters, leaving aside the average forcing F_m and the rate of change H. These parameters are, of course, important, but the latter parallels the choice of future human actions and the former the solar energy input, in ESMs. Here we wish to constrain our focus to the internal dynamics of the system so in this study we keep them constant.

a. Experimental design

The L84-S61 model consists of a non-autonomous system of ODEs (Sell 1967). These are solved using a 4th-order Runge-Kutta method (Iserles 2008), with time step 0.01 LTUs (1.2 hours). This time step is shorter than the 4-hour time step originally used by Lorenz (1984) and allows us to better capture the impact of hourly variability in the daily outputs, while offering a good balance between accuracy and computational cost. The output frequency is 0.2 LTUs (1 day), with all the results presented in this paper as 1-year averages. The ensembles run in this work are designed as follows.

[2]The only exception is the rate of forcing change H, which parallels the choice of future radiative forcing scenario in ESMs.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Parameter	Value	Description
F_1	0.02	Coupling parameter for the equator-pole temperature difference
G_0	1	Reference value for the land-sea temperature difference
G_1	0.01	Coupling parameter for the land-sea temperature difference
a	0.25	Damping coefficient of the westerly winds
b	4	Displacement of the waves due to interaction with the westerly wind
T_{av}	30	Standard temperature contrast between the polar and the equatorial box
γ	30	Proportionality constant between the westerly wind and non-homogeneous forcing by solar heating
k_w	$1.8 \cdot 10^{-5}$	Coefficient of internal diffusion in the ocean
k_a	$1.8 \cdot 10^{-4}$	Coefficient of heat exchange between the ocean and atmosphere
ω	$1.3 \cdot 10^{-4}$	Coefficient derived from the linearised equation of state
ϵ	$1.1 \cdot 10^{-3}$	Coefficient derived from the linearised equation of state
δ_0	$7.8 \cdot 10^{-7}$	Coefficient for the atmospheric water transport
δ_1	$9.6 \cdot 10^{-8}$	Coupling parameter for the wind dependent atmospheric water transport
M	1	Magnitude of the seasonal cycle
F_m	7	1-year mean value of the seasonal variation function $F_0(t)$ when $H = 0$
H	0.01	Externally forced rate of change (per year) of the 1-year mean of $F_0(t)$

TABLE 1. Description of the parameters and their reference values used in the L84-S61 model. The three parameters in the bottom controls the forcing function $F_0(t)$. All parameter values presented here were taken from Daron and Stainforth (2013).

Given an initial condition $\mathbf{X}_0 = (X_{0,1}, \ldots, X_{0,N_X}) \equiv (X_0, Y_0, Z_0, T_0, S_0)$ in the state space and the set of model parameters $\mathbf{P} = (p_1, \ldots, p_{N_P}) \equiv (F_1, G_0, G_1, \ldots, M)$:

- **Micro initial condition ensembles:** We randomly sample another 1,000 initial conditions such that, for each dependent variable, the sample is normally distributed around each $X_{0,i}$ with standard deviation given by σ_{X_i}. For the purpose of this paper, micro initial condition ensembles correspond to each σ_{X_i} being two orders of magnitude lower than $X_{0,i}$. For each initial condition in the sample, we run a simulation starting from this initial condition with the same set of parameters $\mathbf{P}$.

- **Micro perturbed parameter ensembles:** We randomly sample another 1,000 parameter values such that, for each parameter included, the sample is normally distributed around P_j with standard deviation given by σ_{P_j}. For the purpose of this paper, micro perturbed parameter ensembles correspond to each σ_{P_j} being two orders of magnitude lower than P_j. For each parameter value in the sample, we run a simulation starting at the same initial condition $\mathbf{X}_0$.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

- **Macro perturbed physics ensembles:** These consist of either multiple micro ICEs or multiple micro PPEs, with one parameter P_l being replaced by $P_l + \tilde{P}_l$ in each ensemble, where $\tilde{P}_l$ has the same order of magnitude as P_l.

Hence, each ensemble has 1,001 members. All simulations are run for 200 years (after spinup). The first 100 years are under "climate change" at a rate $H = 0.01$ per year (one unit per 100 years), thereby taking the mean value of the forcing from $F_m = 7$ to $F_m + 1 = 8$. This initial run is then followed by 100 years of simulation without further driven change. The initial condition used in these experiments is taken after an initial spin up of 3,000 years with $H = 0$. A detailed description of each experiment run in this study can be found in the Supplementary Materials.

3. Micro initial condition uncertainty

In a perfect model scenario (Smith 2002), where the mathematical formulation of natural processes and their parameter values are assumed to be known and certain, there still remains uncertainty related to the initial state of the system. This is true even if a complete set of observations for this state is available, because such observations can never be perfect, and therefore a micro uncertainty is always present. In the context of climate models (or any chaotic model) this uncertainty, however small, is enough to lead to different future states of the system. This can be significant on both weather- (Lorenz 1963) and climate-relevant timescales (Deser et al. 2012; Hawkins et al. 2016; de Melo Viríssimo et al. 2024).

The impact of this micro uncertainty can be quantified with a large *initial condition ensemble* (ICE). The results of one such large micro ICE are shown (as a heatmap) in Figure 2(a) - for a slowly mixing variable, the ocean temperature difference T - and Figure 3(a) - for a fast mixing variable, the atmospheric variable X. Here, the property of being large shows its value: large ensembles can capture aspects which are not identifiable with either single trajectories or small ensembles. This includes important qualitative features of the ensemble dynamics. For instance, while the central initial condition trajectory (shown in red) shows an increase in T in the first few years (Figure 2(a)), the vast majority of trajectories actually decrease in the first 10 years or so, telling us the most likely outcome for this system under these conditions.

The value of large ensembles is also evident in the atmospheric variables. For instance, with a large ICE, it is possible to see that an initial condition ensemble quickly breaks into three nearly

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

distinct branches at $X \approx 0.4$, $X \approx 0.8$ and $X \approx 1.1$, representing a trimodal distribution of likely states. It also shows that under the forcing timeseries prescribed by H, this trimodality can persist for as long as 60-80 years before merging into a central, broader branch. Even though the central initial condition trajectory (shown in red) alternates between these branches, it does not provide much detail regarding the shape and intensity of the different branches in a given year, nor how they change over time.

4. Micro parametric uncertainty

Although micro ICEs are important to assess the model's internal variability (Mankin et al. 2020; Deser et al. 2020) and the time-dependent distributions under driven climate change (Daron and Stainforth 2013; de Melo Viríssimo et al. 2024), they are not sufficient to quantify the range of possible future climatic conditions. Even if one assumes that the model **formulation** is perfect, there still remains uncertainty around parameter values, as discussed in Section 1, which reflects the fact that alternative, but very similar models may be just as plausible. One way to quantify this uncertainty is via micro perturbed-parameter ensembles (PPEs), in which the initial condition $\mathbf{X}_0$ is fixed but the parameter values $\mathbf{P}$ are slightly changed.

Generally, PPE studies in ESMs tend to focus on a single or small set of parameters, a constraint imposed by the computational cost of running such models. The choice of parameters is usually ad hoc, based on "expert" judgement, and consists of those where uncertainty is thought (or known), a priori, to affect the answer to a particular question, such as the representation of clouds or convection in a global circulation model (Stainforth et al. 2005; Sanderson et al. 2008). To illustrate this scenario, we run large micro PPEs for three different sets of parameters chosen ad hoc: the coupling parameters F_1, G_1, which affect the strength of the coupling of the ocean temperature T with the atmospheric component of the model; the atmospheric parameters a, b, which control the damping and the interaction of Y and Z with the westerly wind X; and the seasonal cycle amplitude M. These parameters were chosen ad hoc, but focused on probing three different parts of the model: the coupling between its ocean and atmospheric components, the internal dynamics and feedbacks within the atmospheric component, and the variability in the equator-to-pole seasonality.

Figures 2(b) and 3(b) show the resulting distributions of T and X, respectively, from one of these ensembles: a large micro PPE of the coupling parameters F_1, G_1. In both cases, despite all

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

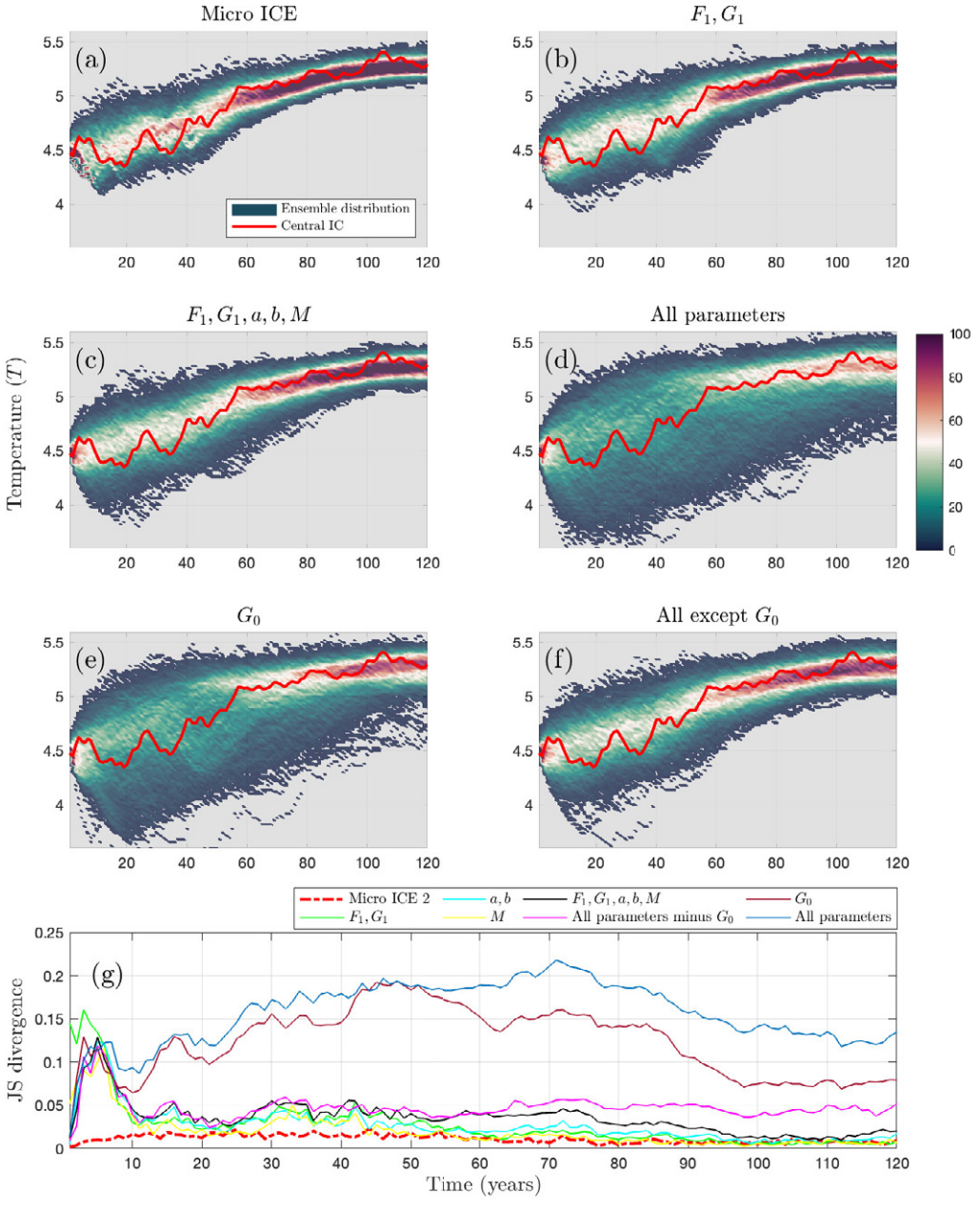

FIG. 2. Micro ICE distribution and micro PPE distributions for different parameter values. The figure shows 120 years of ocean temperature T distribution, consisting of 100 years of climate change followed by 20 years of stationary climate. (a) Micro IC; (b) F_1, G_1; (c) F_1, G_1, a, b, M; (d) All 14 parameters; (e) G_0; (f) All 14 parameters except G_0. Panel (g) shows the Jensen-Shannon divergence comparing the micro ICE with a second micro ICE (red dash-dot line) and several micro PPEs (solid lines). The red solid line in panels (a-f) shows the trajectory evolving from the central initial condition.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

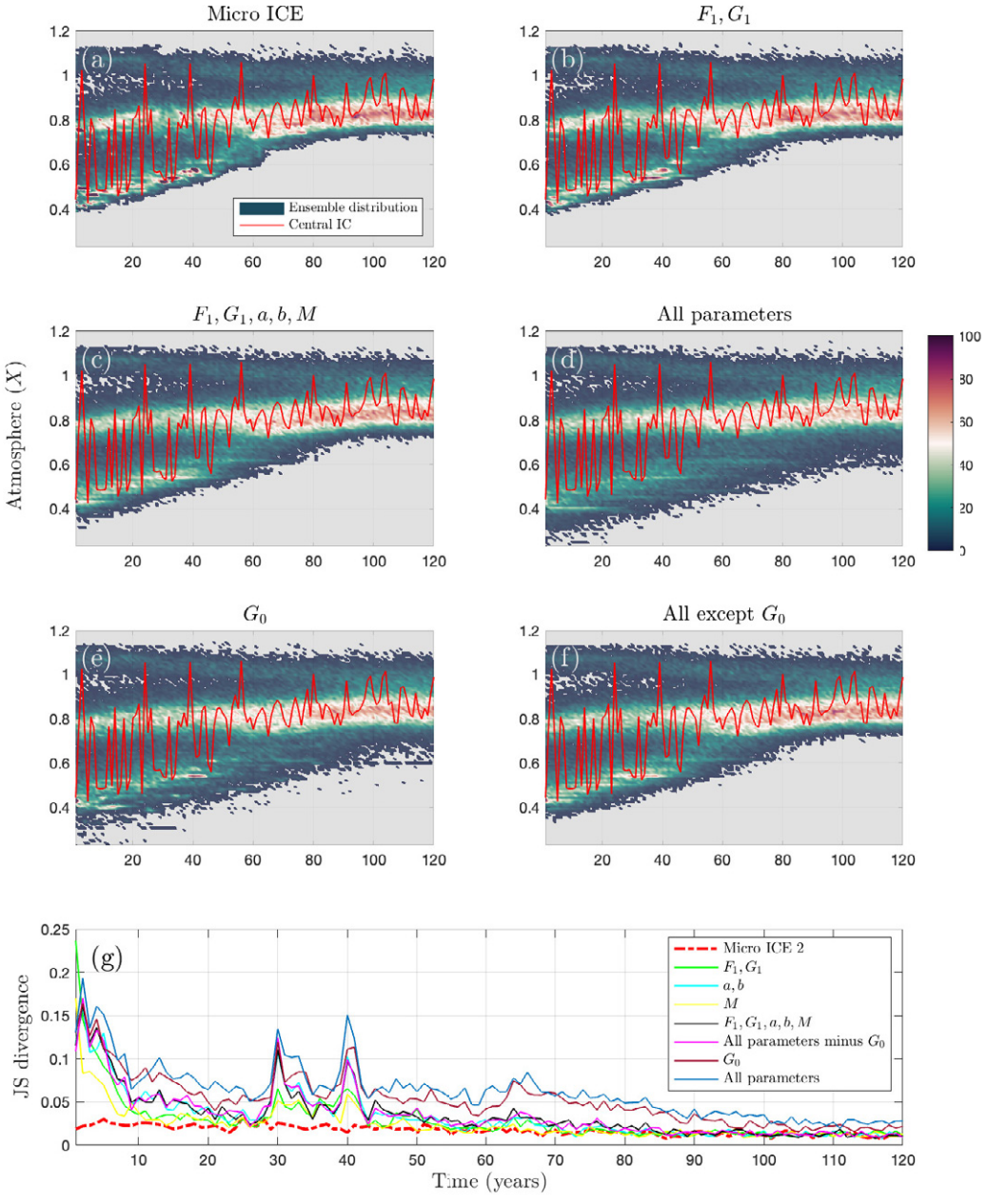

FIG. 3. Micro ICE distribution and micro PPE distributions for different parameter values. The figure shows 120 years of atmosphere variable X distribution, consisting of 100 years of climate change followed by 20 years of stationary climate. (a) Micro IC; (b) F_1, G_1; (c) F_1, G_1, a, b, M; (d) All 14 parameters; (e) G_0; (f) All 14 parameters except G_0. Panel (g) shows the Jensen-Shannon divergence comparing the micro ICE with a second micro ICE (red dash-dot line) and several micro PPEs (solid lines). The red solid line in panels (a-f) shows the trajectory evolving from the central initial condition.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-24-0064.1.

ensemble members having the same initial condition, the micro uncertainty in the parameters lead to the ensemble trajectories quickly diverging from each other, so that they densely populate a large range of values. The resulting distributions are largely similar to those obtained from a micro ICE: they covers a similar range of values, and the frequency of outcomes is also similar - for both T and X.

Even though this micro PPE and the micro ICE distributions look broadly similar, there are nevertheless, visible qualitative distinctions for the slowly-mixing variable T. For instance, the micro PPE does not produce the intricate structure of high probabilities in years 30-40. This might perhaps be expected given the fundamental conceptual differences between these ensembles: while a micro ICE samples the system's attractor, a micro PPE is a union of single trajectories from different attractors, resulting in a distribution that looks more homogeneous and symmetric than the micro ICE distribution. Similar results are obtained when perturbing the parameters a, b, and the seasonal cycle amplitude M (see Supplementary Materials).

a. Many parameter perturbation

As discussed in the previous section, the results of an ad hoc perturbation of individual or pair of parameters are largely similar to that of a micro ICE, and are also similar to each other. But what about perturbing all five ad hoc parameters at once? How does that compare to the individual and pairwise micro PPEs? The results in Figures 2(c) and 3(c), in principle suggest that the distributions look very similar to the others - although an attentive eye might notice some further signs of difference, such as the distributions covering a slightly larger range of values, particularly for the slow ocean variable T.

One could further ask what would be the result of a *full* micro PPE - i.e. a micro PPE applied to all 14 parameters. Would that be similar to a micro ICE or the individual micro PPEs? The answer seems to be a clear no. For the slowly-mixing ocean variable T, Figure 2(d) shows a distribution that is much broader when the micro uncertainty is applied to all parameters, and also flatter in most of the first 60 years of evolution. This is an important difference because it indicates that any of this broad range of distinctive states could be similarly plausible in this multi-decadal period - in contrast with the message from the other micro ensembles in Figures 2(a-c). Regarding the fast-mixing atmospheric variable X, Figure 3(d) shows similar results to those of Figures 3(a-c),

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

except for the lower branch whose tails persist for a significantly longer time. These results indicate that focusing on a small number of parameters can substantially underestimate uncertainty.

b. How different are the micro ICEs and the micro PPEs from each other?

Figures 2(a-d) and 3(a-d) suggest that the distributions from these micro ICEs and micro PPEs can be very similar in some cases. But to what extent do they actually carry similar information?

To answer this question, we statistically compare the micro ICE distribution with each of the micro PPE distributions, for each of the five dependent variables in our conceptual climate model. To do so, we compute the dissimilarity of each micro PPE with respect to the micro ICE using the Jensen-Shannon (JS) divergence (Menéndez et al. 1997; Lin 1991). The JS divergence is an information-theoretic divergence measure based on the Kullback-Leibler divergence (Shlens 2014), which quantifies how close two probability distributions, say D_1, D_2, are. It returns a value between 0 and $log(2) \approx 0.69$, where the value of zero means that the distributions contain the same information, and the value of $log(2)$ means that they are completely different[3]. The JS divergence is a popular statistical measure (Lin 1991), and is widely used in areas such as signal processing, pattern recognition and several branches of artificial intelligence such as generative learning (Sutter et al. 2020).

The results from this computation are shown in Figures 2(g) and 3(g) for the variables T and X respectively (see Supplementary Material for other variables). The red dash-dot line compares the reference micro ICE with a second randomly generated micro ICE, showing the level of divergence that should be expected between two similar ensembles of the same size and type. As expected it indicates a very low divergence between the two micro ICEs. All the micro PPEs, on the other hand, are more divergent from the reference micro ICE, with the all-parameter PPE diverging substantially more than the others (blue line). On further investigation, it turned out that most of this difference arises from a single, originally overlooked, parameter, G_0, whose single parameter JS divergence is given by the red bordeaux line. The resulting distributions, shown in Figures 2(e) and 3(e), confirm that it is indeed much closer to an all-parameters PPE than those in panels (a-d). Nevertheless, the PPE for all-parameters-excluding-G_0, presented in the solid magenta line in Figures 2(g), also shows substantial dissimilarity from the single and pairwise PPEs, for variable T

[3]The relationship between JS, the mean and standard deviation of the distributions is usually nonlinear. As an example, given two normal distributions $(0, 1)$ and $(0, \sigma)$, a JS of 0.1, 0.2 and 0.3 can be achieved by increasing $sigma$ from 1 to 2,3 and 5 respectively (Supplementary Materials).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

and particularly over the long-run beyond about 60 years. This is also confirmed by the distribution shown in the panel (f) of Figures 2. The results are less clear for the fast-mixing variable X, as shown in panels (f,g) of Figure 3(g)).

To better see the impact of the different micro PPEs on the fast-mixing variables, we look at the 30-year climatologies of each distribution, shown in panels (a-c) of Figure 4 for T and panels (d-f) for X. There, the first 29 years correspond to the same yearly distributions presented in Figures 2 and 3, while the remaining 91 years are presented as 30-year rolling averages.

Looking at the smooth trajectories, we can see that the trajectories are different, even for slow variables. Additionally, the impact of perturbing multiple parameters becomes much clearer once the inter-annual variability is smoothed out, indicating the the darker areas of the distributions in panels (b,e) of Figure 4 correspond to genuinely different climatological trajectories which are solely due to the a micro level of uncertainty in the model parameters. Additional, panels (c,f) of Figure 4 show clearly the difference between ICE, single and multiple PPEs, particularly for X, where the JS divergence for the all-parameters PPE is much larger than that observed in Figure 3(g) for example. The clearer difference between the annual and 30-year rolling mean distributions for X is due to the role of the slow dynamics of T, which integrates X over time (Lucarini and Chekroun 2023). As a fast variable, perturbations to X tend to grow very quickly and so they are not so noticeable in interannual timescales. However, in climatological timescales, the effect of perturbations becomes very clear and mirrors that of T.

Looking at the changes in standard deviations, again measured with respect to the micro ICE standard deviation, provides a more familiar illustration of the differences in these distributions. They are shown in panels (a,b) of Figure 5 for the annual distributions, and in panels (c,d) for the 30-year rolling mean distributions. The panels show that perturbing multiple parameters lead to a substantial increase in standard deviation with respect to the micro ICE. For the slow-mixing ocean variable, the magnitude of the increase can reach over 350% when all parameters are perturbed. For the fast-mixing atmosphere variable, this increase exceed 250% when looking at rolling 30-year mean for example.

This demonstrates the value of perturbing multiple parameters at once and the dangers of relying on ad hoc parameter selection. If the purpose of the ensemble is to project possible future states, perturbing single parameters, even if that parameter (say, G_0) accounts for much of the uncertainty,

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

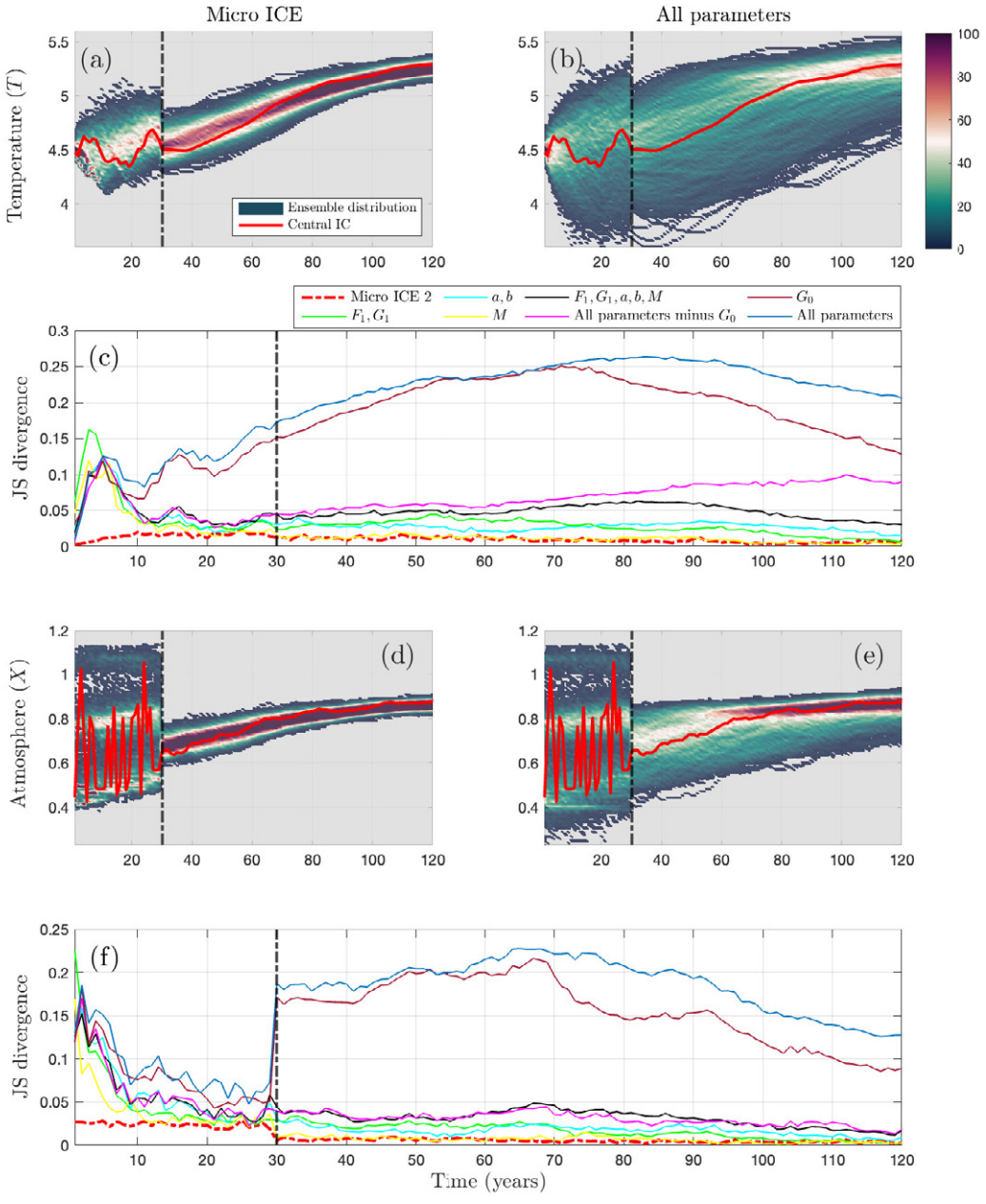

FIG. 4. 30-Year climatologies for micro ICE distribution (panels (a,d)) and micro PPE distribution for all 14 parameters (panels (b,e)). The figure shows 120 years of temperature T (panels (a-b)) and atmosphere variable X (panels (d-e)) distributions, consisting of 100 years of climate change followed by 20 years of stationary climate. The first 29 years shows the yearly time series (the same as in Figures 2 and 3), and the remaining 90 years show the rolling 30-year mean computed year by year. Similarly, panels (c,f) show the Jensen-Shannon divergence comparing the micro ICE with a second micro ICE (red dash-dot line) and several micro PPEs (solid lines) for T and X respectively. In all panels, the vertical black dash-dot line indicates the start of the 30-year rolling mean.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-24-0064.1.

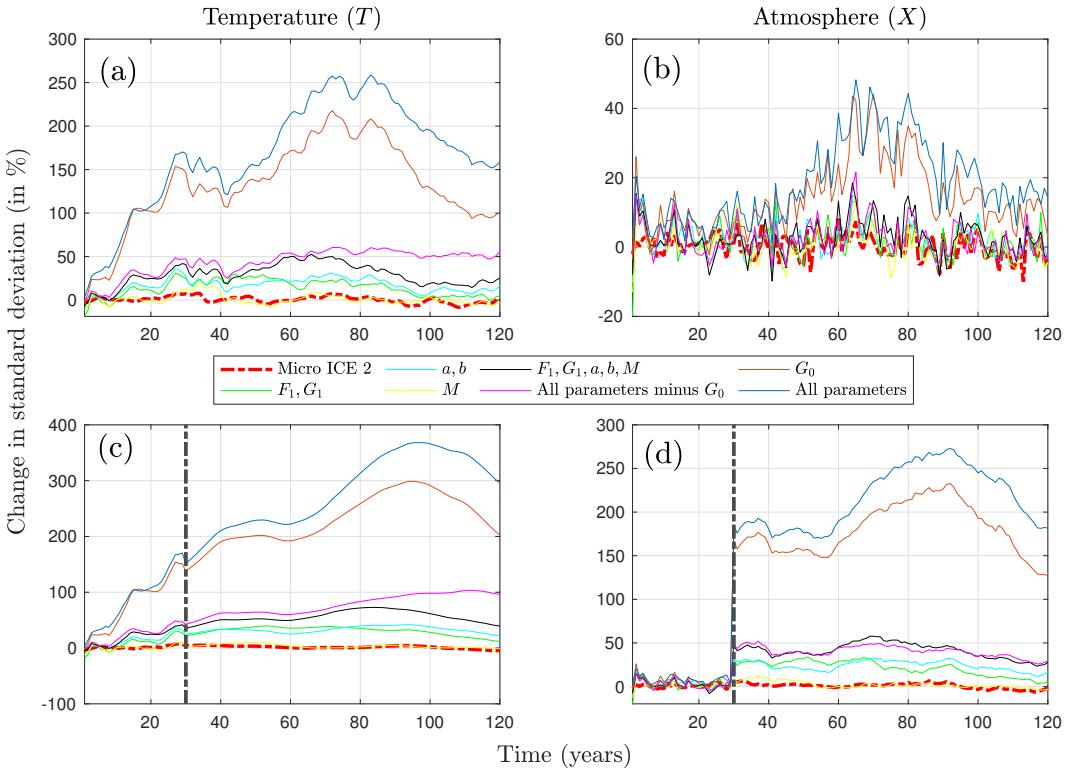

FIG. 5. Change in standard deviation (in percentage) for a second micro ICE (red dash-dot line) and several micro PPEs (solid lines) with respect to the standard deviation from the micro ICE, for the temperature T (left column) and atmosphere variable X (right column), consisting of 100 years of climate change followed by 20 years of stationary climate. Panels (a,b) show the change in standard deviation for the annual distributions (as in Figures 2 and 3)), while panels (c,d) shows the change with for the 30-year rolling mean (as in Figure 4)). In panels (c,d) the first 29 years coincide with those in panels (a,b), with the vertical black dash-dot line indicating the start of the 30-year rolling mean.

might not give the complete picture and might leave out some (potentially important) information. This has direct implications for the design of ensembles of ESMs for projecting future real-world climate, as current ensembles generally perturb only a small subset of parameters.

But what type of information could an ad hoc ensemble be missing?

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-24-0064.1.

c. Micro PPEs and extreme outcomes

One aspect of the differences between the micro PPEs that is responsible for their JS divergences is the fraction of ensemble members projecting an extreme outcome (in a particular year) with respect to the micro ICE. Here we take an extreme outcome in a given year to be a value more than three standard deviations from the micro ICE mean for that year[4]. How the fraction of ensemble members with extreme outcomes varies over time is shown, for each micro ensemble, in Figures 6(a) and 7(a) for the variables T and X respectively. The cumulative rates are presented in Figures 6(b-c) and 7(b-c). Results for the 30-year rolling mean are presented in panels (d-f) of Figures 6 and 7.

As previously suggested by Figures 2 to 5, the micro ICE and individual and pairwise micro PPEs could severely underestimate the likelihood of extreme outcomes by comparison with the all-parameter PPE - whether looking at inter-annual of panels (a-c) or rolling 30-year mean in panels (d-f) in Figures 6 and 7. In the case of the variable T, they are not even additive - with the long term projection at year 120 for the all-parameter PPE resulting in about 20% of ensemble members being extreme, about a third higher than the 15% of the G_0 PPE and the all-parameter-minus-G_0 PPE combined (Figure 6(a)). Similarly, the cumulative numbers suggest about 22% of the all-parameter micro PPE being extreme, versus 20% of the all-parameter-minus-G_0 micro PPE and G_0 micro PPE combined (Figure 6(c)). This suggests that perturbing all parameters at once may be necessary to best capture this information (whether on a yearly basis or cumulatively). Such information from PPEs could be important beyond climate science - for instance in planning and adaptation to climate change. For example, suppose that a decision maker has to make a decision that depends on the information provided by a T-like variable from an operational ESM - say, hypothetically, the projected threat posed to their country's seafood industry by changes in sea temperature. Were such information to look like Figures 6(b,e), it is very unlikely that the decision made if provided with the bottom six lines of Figure 6(a,d) would be the same as if provided with one of the top three lines. Such model-based information could affect plans to adapt to the potential impacts projected, and also influence actions to mitigate them if an alternative scenario could reduce their likelihood.

[4]This is an extreme in an ensemble sense and is different to traditional concepts of extreme events in a climate sense, which usually refer to extremes within a particular timeseries and measured against a 30-year climatology (Arguez and Vose 2011).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

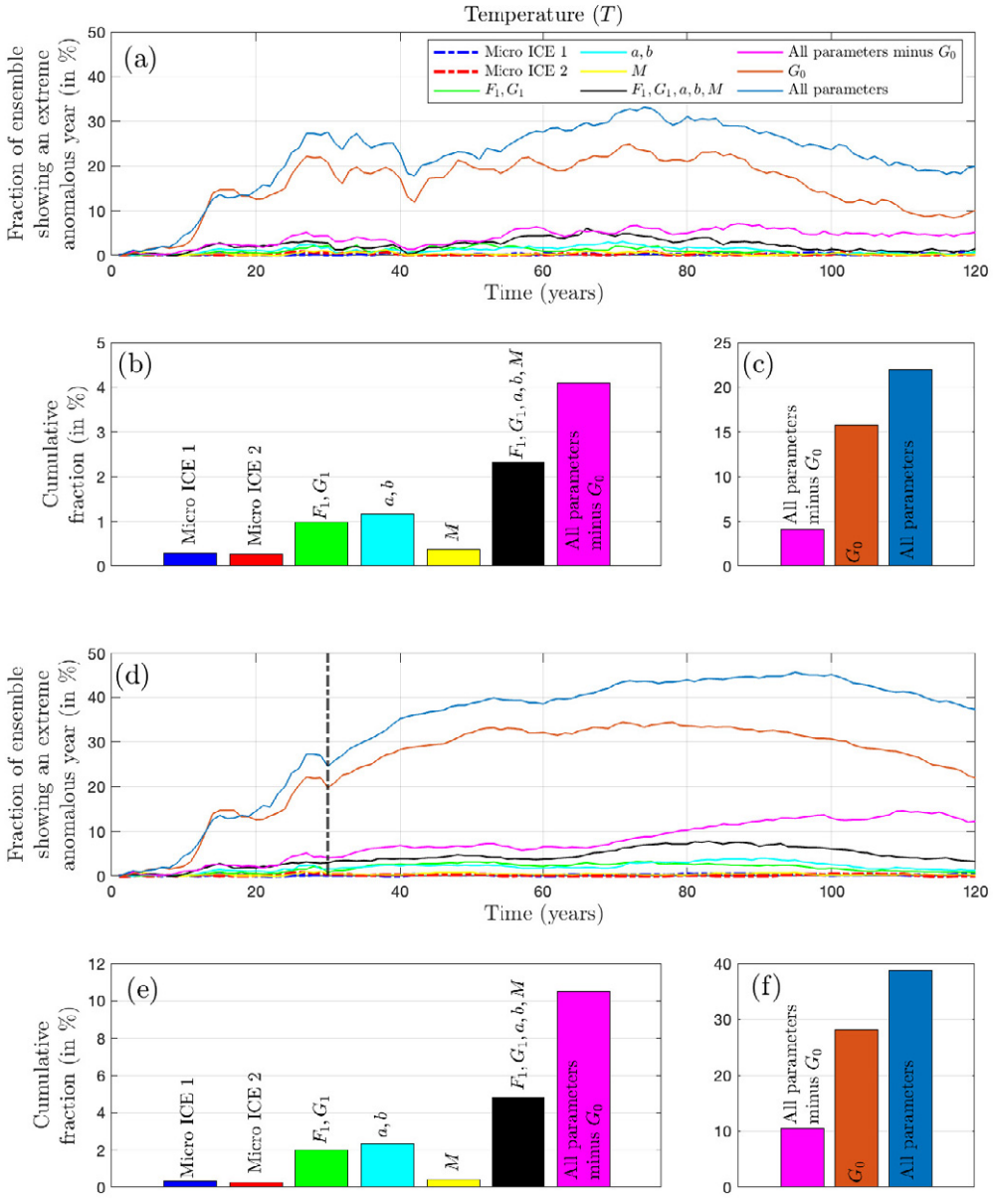

FIG. 6. Comparing a micro ICE with micro PPEs for the ocean temperature T for the annual (panels (a-c)) and 30-year rolling mean (panels (d-f)) distributions. Panels (a,d) show the fraction of ensemble members projecting an extreme outcome (here considered as states over 3 standard deviations from the micro ICE mean) over time. Panels panels (b-c) and (e-f) show the cumulative fraction from (a) and (d), respectively, for the entire 120-year period simulated. In panel (d), the vertical black dash-dot line denotes the start of the 30-year rolling mean. Note the change in scale from panel (b) to (c) and (e) to (f).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

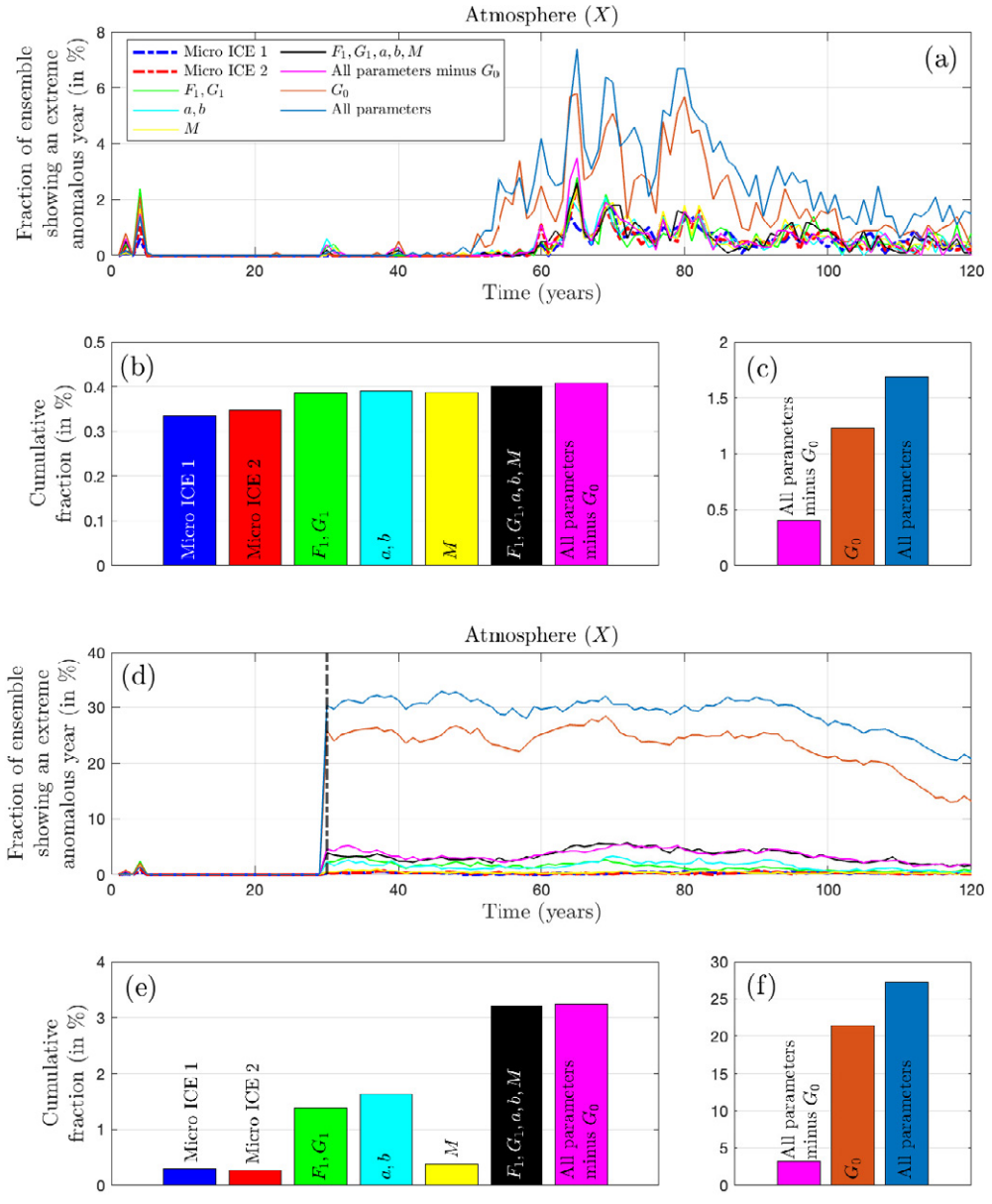

FIG. 7. Comparing a micro ICE with micro PPEs for the atmosphere variable X for the annual (panels (a-c)) and 30-year rolling mean (panels (d-f)) distributions. Panels (a,d) show the fraction of ensemble members projecting an extreme outcome (here considered as states over 3 standard deviations from the micro ICE mean) over time. Panels panels (b-c) and (e-f) show the cumulative fraction from (a) and (d), respectively, for the entire 120-year period simulated. In panel (d), the vertical black dash-dot line denotes the start of the 30-year rolling mean. Note the change in scale from panel (b) to (c) and (e) to (f).

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-24-0064.1.

5. Macro parametric perturbation

In practice it is often the case that the range of plausible parameter values is quite large, such that the uncertainty might have a magnitude comparable with the parameter value itself. In this case, one might choose to explore this uncertainty with a handful of different parameter values but otherwise assume a perfect model scenario within which uncertainty is quantified by a micro ICE alone. Here we call this a *macro parametric ensemble of type I*. This is illustrated in the left columns of Figures 8 and 9 for the ocean temperature difference T and atmospheric variable X respectively.

The results from these macro PPE (of type I) are notably different from those of the micro PPEs shown in Figures 2 and 3. Large changes to single parameters can lead to very different micro ICEs. Note that, without any perturbation to the parameters, a "macro PPE" is simply the micro ICE shown in Figures 2(a) and 3(a); this is reproduced in blue in the left column of Figures 8 and 9 for reference. As one could expect, the impact depends on the variable, and on the parameters considered. For instance, macro perturbing the parameter F_1 results in distributions that are not substantially different from the reference ensemble. Whether one doubles F_1 (representing a stronger $X-T$ coupling) or halves it (representing a weaker coupling), the distributions for T and X are only shifted up and down by a small amount.

For other parameters, however, the impact can be much more dramatic. This is the case for the parameter b, where reducing it by 1/8 dramatically changes the behaviour of the T distribution, shown in red in Figure 8(c). Increasing it by 1/8 leads to something even more distinct, as shown in green in Figure 8(c). This different behaviour is due to the parameter change being sufficient to change the system attracting set from a strange attractor to a limit cycle, towards which the ensemble is initially drawn (see Supplement). Were such a change seen in an ESM, it might be considered unrealistic and the model version (or parameter value) might be ruled out of consideration. However, such behaviour can be hard to identify with single trajectories or small ensembles, which again demonstrates why carefully designed ensembles should be a high priority in the field of climate modelling. Furthermore, including such simulations in a PPE can skew the overall distribution and its statistics, highlighting the importance of explicitly defining what behaviour is deemed acceptable in a model. Ideally this should be done in the design phase to avoid the risks of in-sample bias (Stainforth 2023).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Beyond what we have discussed, there is also the issue of how to approach macro parametric sampling. For example, perturbing the parameter M by 10% (Figures 8(e) and 9(e)) has little effect on the long-term distributions (beyond 80 years) but reveals distinctive behaviour, for $M = 1.1$, in the near term (less than 40 years). By contrast increasing the perturbation to 30% (Figures 8(g) and 9(g)) removes the near term distinctive behaviour but leads to greater changes in the long-term distributions; in this case the long term ensemble mean is inversely affected by the strength of the perturbation. This diversity of behaviour, which is seen in both slow and fast variables, illustrates why a careful sampling design incorporating multiple macro parametric perturbations is needed to better capture the range of outcomes in a macro uncertainty scenario.

a. A second type of macro PPE: combining micro and macro parametric perturbations

If on top of the macro uncertainty in one (or a set of) specific parameters we consider the "micro" uncertainty in all parameters (including the macro-perturbed ones), we then have a *macro parametric ensemble of type II*. This is illustrated in the right columns of Figures 8 and 9. As is the case for the micro PPEs shown in Figures 2 to 7, a macro PPE of type II leads to a larger spread of ensemble trajectories, resulting in distributions that are substantially broader than that given by a macro PPE of type I. For example, in the case of the variable T, the transient behaviour seen in the macro PPEs of type I for $M = 1.1$ and $b = 3.5$ is largely absent in type II distributions. Replacing a micro ICE by a micro PPE substantially increases the standard deviation and can change the shape of the distributions (Figure 8(h)), again highlighting the importance of micro PPEs in capturing the range of plausible values. Micro ICEs alone seem to be substantially more constrained.

The difference between type I and type II macro PPEs are also seen in the atmospheric variable X. For example, the type I macro PPEs for b in Figure 9(c) show highly constrained behaviour for $b = 4.5$ between years 50 and 70 which is absent in the type II macro PPEs in in Figure 9(d). Similar contrasting behaviour between type I and type II can be seen for the other macro PPEs, making clear the qualitative differences between micro ICEs and micro PPEs.

6. Consequences for ensemble design and uncertainty quantification in ESMs

Quantifying the downstream effects of parametric uncertainty in ESMs remains a primary challenge in climate science. The relevance and impact of parametric uncertainty can, however, be seen

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

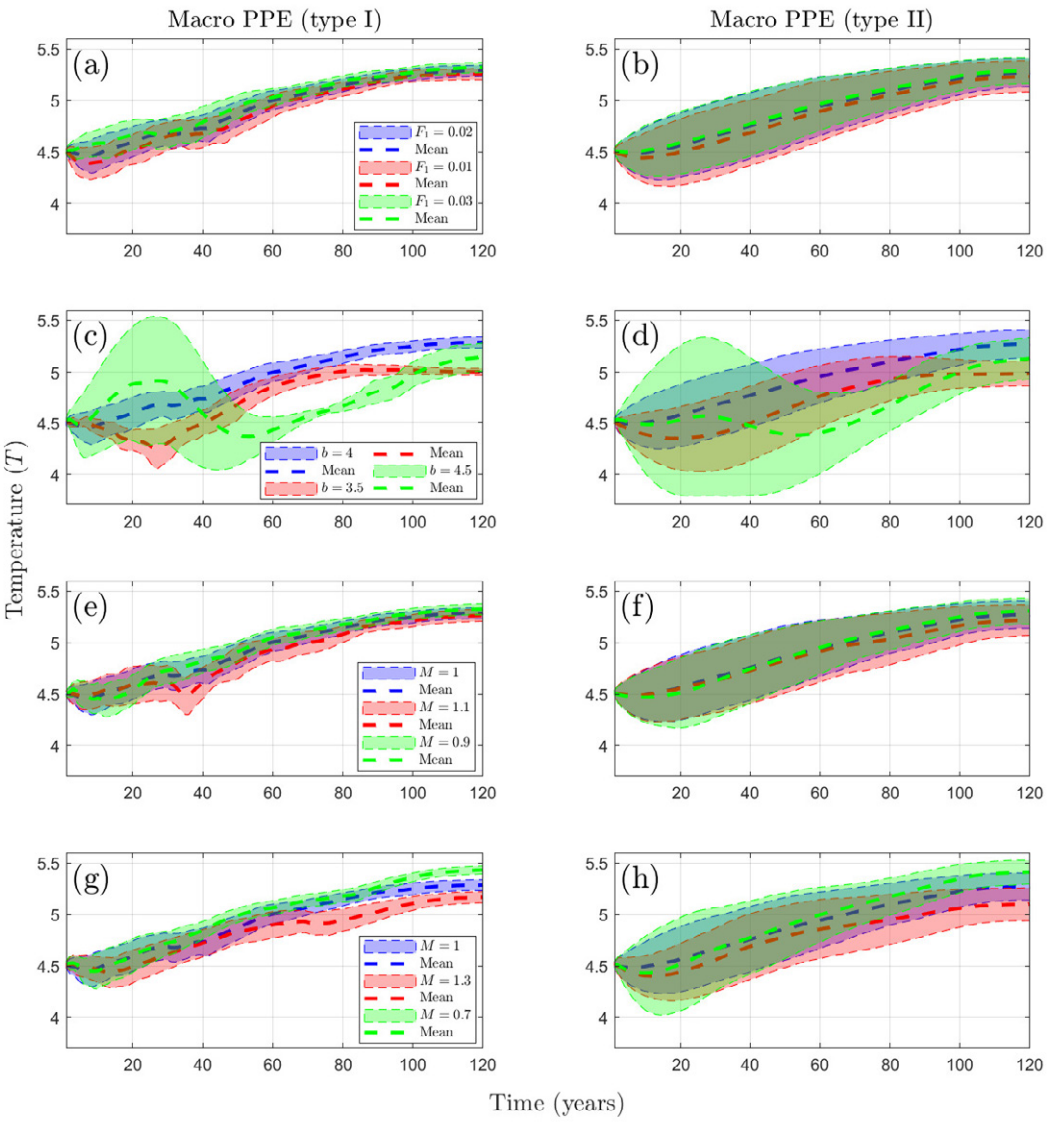

FIG. 8. Macro PPE of types I (left column) and II (right column), for different parameters. The figure shows 120 years of ocean temperature T distribution, consisting of 100 years of climate change followed by 20 years of stationary climate. The distributions are shown as the mean value (thick dash line) plus/minus one standard deviation from the mean (shade bordered by the thin dash line). The distribution in blue always correspond to the micro ICE in the left column, and to the all-parameter micro PPE in the right column. Each panel shows the macro PPE for 3 different macro perturbations: panels (a,b) F_1; panels (c,d) b; panels (e-h) M.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

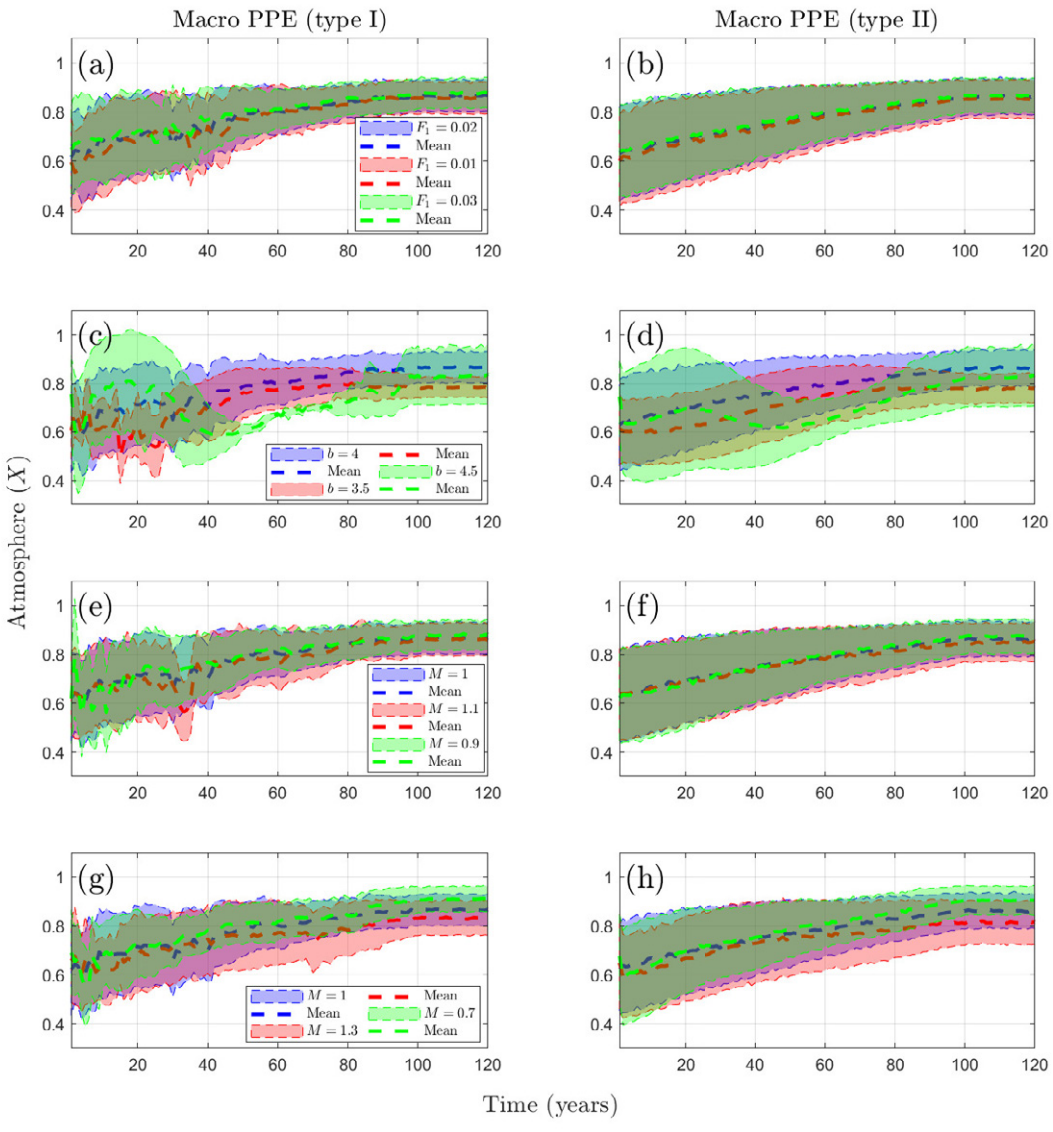

FIG. 9. Macro PPE of types I (left column) and II (right column), for different parameters. The figure shows 120 years of atmosphere variable X distribution, consisting of 100 years of climate change followed by 20 years of stationary climate. The distributions are shown as the mean value (thick dash line) plus/minus one standard deviation from the mean (shade bordered by the thin dash line). The distribution in blue always correspond to the micro ICE in the left column, and to the all-parameter micro PPE in the right column. Each panel shows the macro PPE for 3 different macro perturbations: panels (a,b) F_1; panels (c,d) b; panels (e-h) M.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-24-0064.1.

and studied in simple conceptual models such as L84-S61, and this can help us design efficient assessments of uncertainty in ESMs. As we have seen, when compared to a micro ICE, perturbing multiple parameters lead to an increase of over 300% in the standard deviation, and there is no reason in principle to expect the impact to be different in state-of-the-art ESMs. Furthermore, it is not clear a priori which parameters are the most sensitive controls of the systems behaviour and response to "climate change". While there are several sensitivity analysis techniques (e.g. Morris (1991)) to aid such assessments, in real ESMs with hundreds to thousands of parameters and millions of degrees of freedom, they usually rely on the selection of few observational targets and a small (order $10^1 - 10^2$) subset of parameters (Strobach et al. 2022; Shi et al. 2019; Raj Deepak S. N. and Monahan 2024) - leaving most of the parameters outside the analysis. Crucially, the own concept of sensitivity analysis is unclear, particularly when looking at global quantities, posing a challenge to the objective application of such techniques in ESMs (Razavi and Gupta 2015; Gupta and Razavi 2018).

We note that, as might be expected, the results are dependent on the magnitude of σ_{X_i} and σ_{P_j}. Reducing σ_{X_i} and σ_{P_j} by one order of magnitude might reduce the uncertainty and make the distributions more similar; while increasing them by one order of magnitude could make the micro PPEs much closer to a macro PPE (see Supplementary Materials). Likewise, some of the results such as the shape of the distributions also depend on the location of the initial condition in the attractor (Hawkins et al. 2016; de Melo Viríssimo et al. 2024) (Supplementary Materials). But these do not change the conclusions of the manuscript. In fact, they do support the case for more complex ensembles to be considered, by adding new elements that, although beyond of the scope of this manuscript, should be further explored.

The results presented in this paper relied on the ability to run many very large ensembles - something enabled by our decision to use a low-dimensional conceptual model. Studying parametric uncertainty by running such large PPEs in state-of-the-art ESMs is currently a distant possibility. In the meantime, conceptual studies like this can help us approach ensemble design and interpretation questions from a theoretical perspective before implementing them in computationally expensive models. Parallel to that, the development of data-driven machine learning techniques shows promise in providing guidance in the efficient exploration of parameter space (Cleary et al. 2021; Dunbar et al. 2021), but as illustrated herein the information on the response behaviour

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

may well not exist in previous simulations and this inevitably limits how much information such techniques can provide - suggesting the need for caution when utilising them.

The key question, then, is how to best design parametric ensembles of both low and high dimensional models of real world systems, given the limited resources currently available? While this remains essentially an open question, the results from this paper suggests a few issues worthy of consideration when designing and interpreting such ensembles:

- Micro PPEs that sample the entire parameter space - rather than a single or a few parameters - may well give a better representation of the variability of response to exogenous forcing even if the ensembles are not large. Sampling single or a small combination of parameters leaves large regions of the parameter space unexplored.
- Micro ICEs alone are not enough to capture the range of plausible extreme responses, however large the micro ICE might be. Micro PPEs seem likely to do a better job, and should be considered when this information is sought.
- For complex, multi-component systems, parameters can often be substantially uncertain, partly because compensation of errors can lead to misleading tuning. In this context, macro perturbations are important to explore the wide range of plausible behaviour and responses to driven, extrapolatory changes such as anthropogenic climate change. However, interpreting them requires that they embed micro ICEs and (or) micro PPEs as sub-ensembles.
- Micro ICEs may be better for investigating and understanding particular dynamical changes. Micro PPEs result in more homogeneously spread distributions but they explore the domain of potential behaviour, and therefore potential extrapolatory responses, better.

Beyond the design of ensembles, another challenge remains in transforming the results from parametric ensembles into probabilities (Webster and Sokolov 2000; Murphy et al. 2007). This work demonstrates how difficult this will be. In fact, even were we to have expert priors for parameter values, the arbitrary nature of parameter space (Stainforth 2023) and the lurking potential presence of structural instability in climate models[5] represent conceptual barriers that have not yet begun to be addressed. Probabilities might be welcomed by decision makers across all sectors

[5]This phenomenon is sometimes referred as the *Hawkmoth Effect*, in analogy to Ed Lorenz's Butterfly Effect for initial condition (Thompson 2013; Frigg et al. 2014; Thompson and Smith 2019; Lam 2021).

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

of society, but it is crucial for modellers to be clear when decision-relevant probabilities are not available. Nevertheless well designed PPEs and ensembles in general could help provide them with better information.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Acknowledgments. This work was funded by a UK Natural Environment Research Council grant (ODESSS, agreement number NE/V011790/1) in support of F.d.M.V. and D.A.S. For the purpose of open access, the authors have applied a Creative Commons attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Data availability statement. All data used in this work consist on model output generated by the authors, and is freely available on Zenodo (de Melo Viríssimo and Stainforth 2024). All figures in this work were generated by the authors.

References

Arguez, A., and R. S. Vose, 2011: The definition of the standard WMO climate normal: The key to deriving alternative climate normals. *Bulletin of the American Meteorological Society*, **92 (6)**, 699 – 704, https://doi.org/https://doi.org/10.1175/2010BAMS2955.1.

Ashwin, P., and J. Newman, 2021: Physical invariant measures and tipping probabilities for chaotic attractors of asymptotically autonomous systems. *The European Physical Journal Special Topics*, **230**, 3235–3248, https://doi.org/epjs/s11734-021-00114-z.

Bellprat, O., S. Kotlarski, D. Lüthi, and C. Schär, 2012a: Exploring perturbed physics ensembles in a regional climate model. *Journal of Climate*, **25 (13)**, 4582 – 4599, https://doi.org/https://doi.org/10.1175/JCLI-D-11-00275.1.

Bellprat, O., S. Kotlarski, D. Lüthi, and C. Schär, 2012b: Objective calibration of regional climate models. *Journal of Geophysical Research: Atmospheres*, **117 (D23)**, https://doi.org/https://doi.org/10.1029/2012JD018262.

Berner, J., and Coauthors, 2017: Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, **98 (3)**, 565 – 588, https://doi.org/10.1175/BAMS-D-15-00268.1.

Bocquet, M., A. Farchi, T. S. Finn, C. Durand, S. Cheng, Y. Chen, I. Pasmans, and A. Carrassi, 2024: Accurate deep learning-based filtering for chaotic dynamics by identifying instabilities without an ensemble. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **34 (9)**, 091 104, https://doi.org/10.1063/5.0230837.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Buizza, R., M. Milleer, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ecmwf ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society*, **125 (560)**, 2887–2908, https://doi.org/https://doi.org/10.1002/qj.49712556006, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49712556006.

Bódai, T., and T. Tél, 2012: Annual variability in a conceptual climate model: Snapshot attractors, hysteresis in extreme events, and climate sensitivity. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **22 (2)**, 023 110, https://doi.org/10.1063/1.3697984.

Carslaw, K. S., L. A. Lee, L. A. Regayre, and J. S. Johnson, 2018: Climate models are uncertain, but we can do something about it. *Eos*, **99**, https://doi.org/10.1029/2018EO093757.

Cleary, E., A. Garbuno-Inigo, S. Lan, T. Schneider, and A. M. Stuart, 2021: Calibrate, emulate, sample. *Journal of Computational Physics*, **424**, 109 716, https://doi.org/https://doi.org/10.1016/j.jcp.2020.109716.

Collins, M., B. B. B. Booth, B. Bhaskaran, G. R. Harris, J. M. Murphy, D. M. H. Sexton, and M. J. Webb, 2011: Climate model errors, feedbacks and forcings: a comparison of perturbed physics and multi-model ensembles. *Climate Dynamics*, **36**, 1737–1766, https://doi.org/10.1007/s00382-010-0808-0.

Daron, J. D., 2012: Examining the decision-relevance of climate model information for the insurance industry. Ph.D. thesis, London School of Economics and Political Science.

Daron, J. D., and D. A. Stainforth, 2013: On predicting climate under climate change. *Environmental Research Letters*, **8 (3)**, 034 021, https://doi.org/10.1088/1748-9326/8/3/034021.

de Melo Viríssimo, F., A. P. Martin, and S. A. Henson, 2022: Influence of seasonal variability in flux attenuation on global organic carbon fluxes and nutrient distributions. *Global Biogeochemical Cycles*, **36 (2)**, e2021GB007 101, https://doi.org/10.1029/2021GB007101.

de Melo Viríssimo, F., and D. Stainforth, 2023: A low-dimensional dynamical systems approach to climate ensemble design and interpretation. URL https://doi.org/10.5194/egusphere-egu23-14755, abstract EGU23-14755.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

de Melo Viríssimo, F., and D. A. Stainforth, 2024: Model output used in the manuscript "Micro and macro parametric uncertainty in climate change prediction: a largeensemble perspective". *Zenodo*, https://doi.org/10.5281/zenodo.13200874.

de Melo Viríssimo, F., D. A. Stainforth, and J. Bröcker, 2024: The evolution of a non-autonomous chaotic system under non-periodic forcing: A climate change example. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **34 (1)**, 013 136, https://doi.org/10.1063/5.0180870, https://pubs.aip.org/aip/cha/article-pdf/doi/10.1063/5.0180870/19329638/013136_1_5.0180870.pdf.

Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012: Communication of the role of natural variability in future North American climate. *Nature Climate Change*, **2**, 775–779, https://doi.org/10.1038/nclimate1562.

Deser, C., and Coauthors, 2020: Insights from earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, **10**, 277–286, https://doi.org/10.1038/s41558-020-0731-2.

Dijkstra, H. A., 2013: *Nonlinear Climate Dynamics*. Cambridge University Press, https://doi.org/10.1017/CBO9781139034135.

Dijkstra, H. A., 2024: The role of conceptual models in climate research. *Physica D: Nonlinear Phenomena*, **457**, 133 984, https://doi.org/10.1016/j.physd.2023.133984.

Dunbar, O. R. A., A. Garbuno-Inigo, T. Schneider, and A. M. Stuart, 2021: Calibration and uncertainty quantification of convective parameters in an idealized gcm. *Journal of Advances in Modeling Earth Systems*, **13 (9)**, e2020MS002 454, https://doi.org/https://doi.org/10.1029/2020MS002454, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002454.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, **9 (5)**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Favier, L., N. C. Jourdain, A. Jenkins, N. Merino, G. Durand, O. Gagliardini, F. Gillet-Chaulet, and P. Mathiot, 2019: Assessment of sub-shelf melting parameterisations using the ocean–ice-

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

sheet coupled model nemo(v3.6)–elmer/ice(v8.3). *Geoscientific Model Development*, **12 (6)**, 2255–2283, https://doi.org/10.5194/gmd-12-2255-2019.

Frigg, R., S. Bradley, H. Du, and L. A. Smith, 2014: Laplace's demon and the adventures of his apprentices. *Philosophy of Science*, **81 (1)**, 31–59, https://doi.org/10.1086/674416.

Gent, P. R., 2011: The Gent–McWilliams parameterization: 20/20 hindsight. *Ocean Modelling*, **39 (1)**, 2–9, https://doi.org/https://doi.org/10.1016/j.ocemod.2010.08.002.

Gent, P. R., and J. C. McWilliams, 1990: Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, **20 (1)**, 150 – 155, https://doi.org/https://doi.org/10.1175/1520-0485(1990)020⟨0150:IMIOCM⟩2.0.CO;2.

Ghil, M., 2019: A century of nonlinearity in the geosciences. *Earth and Space Science*, **6 (7)**, 1007–1042, https://doi.org/https://doi.org/10.1029/2019EA000599, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019EA000599.

Gupta, H. V., and S. Razavi, 2018: Revisiting the basis of sensitivity analysis for dynamical earth system models. *Water Resources Research*, **54 (11)**, 8692–8717, https://doi.org/10.1029/2018WR022668.

Hargreaves, J. C., 2010: Skill and uncertainty in climate models. *WIREs Climate Change*, **1 (4)**, 556–564, https://doi.org/https://doi.org/10.1002/wcc.58, https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.58.

Hawkins, E., R. S. Smith, J. M. Gregory, and D. A. Stainforth, 2016: Irreducible uncertainty in near-term climate projections. *Climate Dynamics*, **46**, 3807–3819, https://doi.org/10.1007/s00382-015-2806-8.

Herein, M., G. Drótos, T. Haszpra, J. Márfy, and T. Tél, 2017: The theory of parallel climate realizations as a new framework for teleconnection analysis. *Scientific Reports*, **7**, 44 529, https://doi.org/10.1038/srep44529.

Holden, P. B., N. R. Edwards, K. Fraedrich, E. Kirk, F. Lunkeit, and X. Zhu, 2016: Plasim–genie v1.0: a new intermediate complexity aogcm. *Geoscientific Model Development*, **9 (9)**, 3347–3361, https://doi.org/10.5194/gmd-9-3347-2016.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Hourdin, F., and Coauthors, 2017: The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, **98 (3)**, 589 – 602, https://doi.org/10.1175/BAMS-D-15-00135.1.

IPCC, 2023: *Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, https://doi.org/10.59327/IPCC/AR6-9789291691647, URL https://doi.org/10.59327/IPCC/AR6-9789291691647, [Core Writing Team, H. Lee and J. Romero (eds.)].

Iserles, A., 2008: *A First Course in the Numerical Analysis of Differential Equations*. 2nd ed., Cambridge Texts in Applied Mathematics, Cambridge University Press.

Knopf, B., H. Held, and H. J. Schellnhuber, 2005: Forced versus coupled dynamics in earth system modelling and prediction. *Nonlinear Processes in Geophysics*, **12 (2)**, 311–320, https://doi.org/10.5194/npg-12-311-2005.

Knutti, R., D. Masson, and A. Gettelman, 2013: Climate model genealogy: Generation cmip5 and how we got there. *Geophysical Research Letters*, **40 (6)**, 1194–1199, https://doi.org/https://doi.org/10.1002/grl.50256, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/grl.50256.

Kuhlbrodt, T., C. Jones, A. Sellar, and R. S. Smith, 2015: Recommending a resolution for UKESM-LO. A UKESM project report. Tech. rep., https://www.met.reading.ac.uk/~till/lit/recommending-resolution-ukesm.pdf, Met Office and National Environmental Research Council (NERC). Last checked on 16.05.2024.

Lam, V., 2021: Climate modelling and structural stability. *Euro Jnl Phil Sci*, **11**, 98, https://doi.org/10.1007/s13194-021-00414-0.

Lin, J., 1991: Divergence measures based on the shannon entropy. *IEEE Transactions on Information Theory*, **37 (1)**, 145–151, https://doi.org/10.1109/18.61115.

Liu, Y., and J. Hallett, 1997: The '1/3' power law between effective radius and liquid-water content. *Quarterly Journal of the Royal Meteorological Society*, **123 (542)**, 1789–1795, https://doi.org/https://doi.org/10.1002/qj.49712354220, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49712354220.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *Journal of Atmospheric Sciences*, **20 (2)**, 130 – 141, https://doi.org/10.1175/1520-0469(1963)020⟨0130:DNF⟩2.0.CO;2.

Accepted for publication in *Bulletin of the American Meteorological Society*. DOI 10.1175/BAMS-D-24-0064.1.

Lorenz, E. N., 1984: Irregularity: a fundamental property of the atmosphere. *Tellus A: Dynamic Meteorology and Oceanography*, **36 (2)**, 98–110, https://doi.org/10.3402/tellusa.v36i2.11473, https://doi.org/10.3402/tellusa.v36i2.11473.

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Reading, Vol. 1, 1–18.

Lucarini, V., and M. D. Chekroun, 2023: Theoretical tools for understanding the climate crisis from hasselmann's programme and beyond. *Nature Reviews Physics*, **5**, 744–765, https://doi.org/10.1038/s42254-023-00650-8.

Mankin, J. S., F. Lehner, S. Coats, and K. A. McKinnon, 2020: The value of initial condition large ensembles to robust adaptation decision-making. *Earth's Future*, **8 (10)**, e2012EF001 610, https://doi.org/https://doi.org/10.1029/2020EF001610, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020EF001610.

Martin, A. P., and Coauthors, 2024: When to add a new process to a model – and when not: A marine biogeochemical perspective. *Ecological Modelling*, **498**, 110 870, https://doi.org/j.ecolmodel.2024.110870.

Martin, J. H., G. A. Knauer, D. M. Karl, and W. W. Broenkow, 1987: Vertex: carbon cycling in the northeast pacific. *Deep Sea Research Part A. Oceanographic Research Papers*, **34 (2)**, 267–285, https://doi.org/https://doi.org/10.1016/0198-0149(87)90086-0.

Masson, D., and R. Knutti, 2011: Climate model genealogy. *Geophysical Research Letters*, **38 (8)**, https://doi.org/https://doi.org/10.1029/2011GL046864, https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011GL046864.

Mehling, O., R. Börner, and V. Lucarini, 2024: Limits to predictability of the asymptotic state of the atlantic meridional overturning circulation in a conceptual climate model. *Physica D: Nonlinear Phenomena*, **459**, 134 043, https://doi.org/https://doi.org/10.1016/j.physd.2023.134043.

Menéndez, M., J. Pardo, L. Pardo, and M. Pardo, 1997: The jensen-shannon divergence. *Journal of the Franklin Institute*, **334 (2)**, 307–318, https://doi.org/https://doi.org/10.1016/S0016-0032(96)00063-4.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Morris, M. D., 1991: Factorial sampling plans for preliminary computational experiments. *Technometrics*, **33 (2)**, 161–174, https://doi.org/10.1080/00401706.1991.10484804.

Murphy, J., B. Booth, M. Collins, G. Harris, D. Sexton, and M. Webb, 2007: A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **365 (1857)**, 1993–2028, https://doi.org/10.1098/rsta.2007.2077, https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2007.2077.

Murphy, J. M., D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth, 2004: Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, **430**, 768–772, https://doi.org/10.1038/nature02771.

Palmer, T. N., 2001: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society*, **127 (572)**, 279–304, https://doi.org/https://doi.org/10.1002/qj.49712757202, https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49712757202.

Palmer, T. N., 2016: A personal perspective on modelling the climate system. *Proc. R. Soc. A.*, **472**, 20150 772, https://doi.org/10.1098/rspa.2015.0772.

Parker, W. S., 2013: Ensemble modeling, uncertainty and robust predictions. *WIREs Climate Change*, **4 (3)**, 213–223, https://doi.org/https://doi.org/10.1002/wcc.220, https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcc.220.

Peatier, S., B. M. Sanderson, and L. Terray, 2024: Exploration of diverse solutions for the calibration of imperfect climate models. *Earth System Dynamics*, **15 (4)**, 987–1014, https://doi.org/10.5194/esd-15-987-2024.

Provenzale, A., and N. J. Balmforth, 1999: Chaos and structures in geophysics and astrophysics. *Proceedings of GFD (Geophysical Fluid Dynamics of Woods Hole Oceanographic Institution*.

Raj Deepak S. N., C. S., and A. H. Monahan, 2024: A global sensitivity analysis of parameter uncertainty in the classic model. *Atmosphere-Ocean*, **62 (5)**, 347–359, https://doi.org/10.1080/07055900.2024.2396426.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Razavi, S., and H. V. Gupta, 2015: What do we mean by sensitivity analysis? the need for comprehensive characterization of "global" sensitivity in earth and environmental systems models. *Water Resources Research*, **51 (5)**, 3070–3092, https://doi.org/10.1002/2014WR016527.

Roebber, P. J., 1995: Climate variability in a low-order coupled atmosphere-ocean model. *Tellus A: Dynamic Meteorology and Oceanography*, **47 (4)**, 473–494, https://doi.org/10.3402/tellusa.v47i4.11534, https://doi.org/10.3402/tellusa.v47i4.11534.

Sanderson, B. M., C. Piani, W. J. Ingram, D. A. Stone, and M. R. Allen, 2008: Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. *Climate Dynamics*, **30**, 175–190, https://doi.org/10.1007/s00382-007-0280-7.

Sell, G. R., 1967: Nonautonomous differential equations and topological dynamics i. the basic theory. *Transactions of the American Mathematical Society*, **127 (2)**, 241–262.

Shi, Y., W. Gong, Q. Duan, J. Charles, C. Xiao, and H. Wang, 2019: How parameter specification of an earth system model of intermediate complexity influences its climate simulations. *Progress in Earth and Planetary Science*, **6**, 46, https://doi.org/10.1186/s40645-019-0294-x.

Shlens, J., 2014: Notes on kullback-leibler divergence and likelihood. *CoRR*, **abs/1404.2000**, 1404.2000.

Smith, L. A., 2002: What might we learn from climate forecasts? *Proceedings of the National Academy of Sciences*, **99 (suppl_1)**, 2487–2492, https://doi.org/10.1073/pnas.012580599, https://www.pnas.org/doi/pdf/10.1073/pnas.012580599.

Stainforth, D., M. Allen, E. Tredger, and L. Smith, 2007a: Confidence, uncertainty and decision-support relevance in climate predictions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **365 (1857)**, 2145–2161, https://doi.org/10.1098/rsta.2007.2074, https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2007.2074.

Stainforth, D., and Coauthors, 2005: Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, **433**, 403–406, https://doi.org/10.1038/nature03301.

Stainforth, D. A., 2023: *Predicting Our Climate Future: What We Know, What We Don't Know, And What We Can't Know*. 1st ed., Oxford University Press.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Stainforth, D. A., T. E. Downing, R. Washington, A. Lopez, and M. New, 2007b: Issues in the interpretation of climate model ensembles to inform decisions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **365 (1857)**, 2163–2177, https://doi.org/10.1098/rsta.2007.2073, https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2007.2073.

Stommel, H., 1961: Thermohaline convection with two stable regimes of flow. *Tellus*, **13 (2)**, 224–230, https://doi.org/10.3402/tellusa.v13i2.9491, https://doi.org/10.3402/tellusa.v13i2.9491.

Stringer, M., 2017: Timings for all UKESM configurations. Last checked on 16.05.2024, https://www.met.reading.ac.uk/~marc/metOffice/myDocs/profiling/timingsCray2/ukesm_all/.

Strobach, E., A. Molod, D. Barahona, A. Trayanov, D. Menemenlis, and G. Forget, 2022: Earth system model parameter adjustment using a green's functions approach. *Geoscientific Model Development*, **15 (5)**, 2309–2324, https://doi.org/10.5194/gmd-15-2309-2022.

Sutter, T., I. Daunhawer, and J. Vogt, 2020: Multimodal generative learning utilizing jensen-shannon-divergence. *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., Curran Associates, Inc., Vol. 33, 6100–6110, URL https://proceedings.neurips.cc/paper_files/paper/2020/file/43bb733c1b62a5e374c63cb22fa457b4-Paper.pdf.

Tett, S. F. B., K. Yamazaki, M. J. Mineter, C. Cartis, and N. Eizenberg, 2017: Calibrating climate models using inverse methods: case studies with hadam3, hadam3p and hadcm3. *Geoscientific Model Development*, **10 (9)**, 3567–3589, https://doi.org/10.5194/gmd-10-3567-2017.

Thompson, E. L., 2013: Modelling north atlantic storms in a changing climate. Ph.D. thesis, Imperial College London.

Thompson, E. L., and L. A. Smith, 2019: Escape from model-land. *Economics*, **13 (1)**, 20190 040, https://doi.org/10.5018/economics-ejournal.ja.2019-40.

Tél, T., T. Bódai, G. Drótos, T. Haszpra, M. Herein, B. Kaszás, and M. Vincze, 2020: The theory of parallel climate realizations. *J Stat Phys*, **179**, 1496–1530, https://doi.org/10.1007/s10955-019-02445-7.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.

Veen, L. V., T. Opsteegh, and F. Verhulst, 2001: Active and passive ocean regimes in a low-order climate model. *Tellus A: Dynamic Meteorology and Oceanography*, **53 (5)**, 599–615, https://doi.org/10.3402/tellusa.v53i5.12229, https://doi.org/10.3402/tellusa.v53i5.12229.

Webster, M. D., and A. P. Sokolov, 2000: A methodology for quantifying uncertainty in climate projections. *Climatic Change*, **46**, 417–446, https://doi.org/10.1023/A:1005685317358.

Youden, W. J., 1961: Systematic errors in physical constants. *Physics Today*, **14 (9)**, 32–43, https://doi.org/10.1063/1.3057731, https://pubs.aip.org/physicstoday/article-pdf/14/9/32/8257152/32_1_online.pdf.

Accepted for publication in *Bulletin of the American Meteorological Society.* DOI 10.1175/BAMS-D-24-0064.1.