

Opinion: The importance and future development of perturbed parameter ensembles in climate and atmospheric science

Ken S. Carslaw¹, Leighton Regayre², Ulrike Proske³, Andrew Gettelman⁴, David M. H. Sexton⁵, Yun Qian⁴, Lauren R. Marshall⁶, Oliver Wild⁷, Marcus van Lier-Walqui⁸, Annika Oertel⁹, Saloua Peatier¹⁰,
5 Ben Yang¹¹, Jill S Johnson¹², Sihan Li¹³, Daniel T. McCoy¹⁴, Benjamin M. Sanderson¹⁵, Christina J. Williamson¹⁶, Gregory S. Elsaesser¹⁷, Kuniko Yamazaki⁵, Ben B. Booth⁵.

1. School of Earth and Environment, University of Leeds, UK.
2. School of Earth and Environment, University of Leeds, UK and Met Office, Exeter, UK.
- 10 3. Hydrology and Environmental Hydraulics Group, Wageningen University, Wageningen, The Netherlands.
4. Pacific Northwest National Laboratory, Richland, WA, USA.
5. Met Office, Exeter, UK.
6. School of Earth and Environmental Sciences, University of St Andrews, St Andrews, UK.
7. Lancaster Environment Centre, Lancaster University, Lancaster, UK.
- 15 8. NASA Goddard Institute for Space Studies, New York, NY, USA and Center for Climate Systems Research (CCSR), The Earth Institute, Columbia University, NY, New York, USA.
9. Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research Troposphere Research, Karlsruhe, Germany.
10. Department of Meteorology, University of Reading, UK.
- 20 11. School of Atmospheric Sciences, Nanjing University, Nanjing, China.
12. School of Mathematical and Physical Sciences, University of Sheffield, Sheffield, UK.
13. School of Geography and Planning, University of Sheffield, Sheffield, UK.
14. Department of Atmospheric Science, University of Wyoming, Laramie, WY, USA.
15. CICERO Center for International Climate Research, Oslo, Norway.
- 25 16. Finnish Meteorological Institute, Helsinki, Finland and Institute for Atmospheric and Earth System Research/Physics, University of Helsinki, Finland.
17. Department of Applied Physics and Mathematics, Columbia University, New York, NY, USA and NASA Goddard Institute for Space Studies, New York, NY, USA.

30 **Correspondence to:** Ken S. Carslaw (k.s.carslaw@leeds.ac.uk)

Abstract. A grand challenge in climate science is to translate advances in our fundamental understanding into reduced uncertainty in climate projections. Model uncertainty, characterized for example by the spread of simulations of future climate projections, has changed little over the past few decades despite major advances in model complexity, resolution, and the growing number of intercomparison projects and observational datasets. Here we argue that the use of perturbed parameter ensembles (PPEs) would accelerate our understanding of uncertainty in its broadest sense and help identify strategies for
35 reducing it. We make eleven recommendations for future research priorities, drawing on existing studies that have used PPEs to guide model development and simplification, understand inter-model differences, more fully characterize the plausible

spread in climate projections, , define observational requirements, and to enhance our understanding of complex, nonlinear processes. These studies extend across climate, weather, atmospheric chemistry, clouds, aerosols and renewable energy using process-based high-resolution models through to global-scale models. Although increases in model complexity, resolution and intercomparison projects consume most computing resources today, we argue that, in synergy with these efforts, PPEs are essential for fully characterizing model uncertainty and improving model reliability.

1. The challenges of climate modelling

The future evolution of atmospheric, climate and Earth system models involves several well-motivated yet partly competing priorities, each placing increasing demands on computing resources. Most model development is motivated by the goal of improving model fidelity – the extent to which model simulations reproduce the observed state and behaviour of the climate and Earth system. To that end, the community has focused primarily on increasing process-level detail (complexity) and spatial resolution, as well as on running initial-condition ensembles to characterize decadal and regional climate variability by sampling internal variability. This effort has produced impressive model simulations that look and behave in many ways like the real world.

The other vital aspect of modelling alongside fidelity is reliability – the extent to which models produce consistent and trustworthy results across multiple scenarios. The primary approach used to assess model reliability is the model intercomparison project (MIP), where the spread of simulations serves as a rough, yet incomplete, proxy for reliability (see Table 1 for a glossary of uncertainty terms used in this article). The number of climate MIPs has grown from two standardized experimental protocols in the 1990s to 322 in phase 6 of the Coupled Model Intercomparison Project (CMIP) in 2017 ([Durack et al., 2025](#); [Stevens, 2024](#)). MIPs have revealed systematic model biases, improved our understanding of the climate system, and guided international assessments and policy. However, MIPs represent an incomplete and unquantified mixture of model structural differences, parametric uncertainties and internal variability (Knutti et al., 2010).

The incompleteness of MIPs has two important consequences. Firstly, MIPs represent the main way that we communicate plausible uncertainties to wider science and impact communities – for example through a subset of CMIP simulations that inform IPCC’s Impacts, Adaptation and Vulnerability assessments (WG2). An incomplete representation of uncertainties may therefore have conveyed an unjustifiably high level of confidence, inconsistent with the underlying physical modelling. The second consequence relates to how we move forward. Without a full understanding of the magnitude and causes of model spread, we lack the information required to improve model reliability, assuming this is possible.

Although the community emphasizes model complexity, greater complexity does not equate to greater reliability (Baartman et al., 2020; Knüsel and Baumberger, 2020; Proske et al., 2024; Puy et al., 2022). This is essentially the over-fitting problem – see Fig. 1. Over more than 30 years and six phases of CMIP, increasingly detailed components of the climate system have been evaluated and compared across models (Durack et al., 2025). Yet, for key emergent climate metrics—such as climate sensitivity, cloud feedback, and aerosol radiative forcing—the spread among models has remained substantial. The famous quote from George Box “All models are wrong, but some are useful” was followed by perhaps the more relevant statement

that “The scientist cannot obtain a ‘correct’ [model] by excessive elaboration” (Box, 1976; Carslaw et al., 2018). Climate model spread might persist or even grow as we obtain new knowledge (Knutti and Sedláček, 2013), and we may at present even be underestimating the spread. But it is also the case that efforts to enhance model complexity have not been matched by efforts to understand and reduce the uncertainty that it introduces. In short, our understanding of model uncertainty is incomplete, and we are certainly not on a path to reducing it.

Increases in model resolution will further enhance model fidelity (Slingo et al., 2022). However, some critical processes will always be parameterized even in km-scale models (Morrison et al., 2020), which will continue to affect model reliability. Furthermore, any high-resolution MIP to assess reliability would still represent an unquantified mixture of structural differences, parametric uncertainties and internal variability, making assessment of reliability perhaps even more challenging.

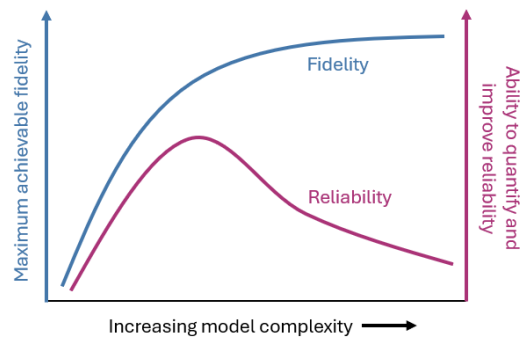


Figure 1. Schematic of the model complexity-reliability challenge. Models with greater complexity can in principle achieve higher fidelity. However, higher-complexity models have a larger number of uncertain processes in them, which become increasingly difficult to verify against observations. This is often referred to as over-fitting. Complex models are also slower to run, which limits how the tools of uncertainty quantification and reliability assessment, such as PPEs, can be applied.

What is the role of perturbed parameter ensembles? The prevailing view is that the uncertainty caused by adjustable quantities inside a model (the parameters – see Table 1) is just a small part of the overall uncertainty challenge – mainly just a component of model tuning. While tuning is essential and universal (Hourdin et al., 2017; Mauritsen et al., 2012), expansion into full perturbed parameter ensembles (PPEs) in which combinations of parameters are systematically perturbed, is given much lower priority than structural changes to models. When a model is structurally deficient—due to low resolution or missing processes—addressing these issues is prioritized, while PPEs are seen as merely “over-polishing” a structurally deficient model. This narrow perspective is one we challenge here.

In this article, we highlight how PPE science has expanded since the 2015 Workshop on Uncertainty Quantification in Climate Modeling and Projection (Qian et al., 2016) to address the *overall* challenge of model uncertainty and to provide an increasingly deep, process-level understanding of model behaviour. In particular, we argue that PPEs can help to more fully characterize current model uncertainty and provide a path toward reducing it. We start with a brief introduction to PPEs before summarizing

the range of recent PPE studies in atmospheric and climate science, how they are contributing to wider climate science challenges, and how these efforts could be further developed.

100

Table 1. Glossary of uncertainty terms referred to in this article. Terms in *italic* are defined elsewhere in the table.

Term	Definition
Model uncertainty	The range of possible outcomes that results from incomplete or imperfect representation of the climate system in a model, including processes, inputs and assumptions.
Parameter	Adjustable quantity inside a model, most often within the model parameterization equations.
Parametric uncertainty	Uncertainty in model output variables caused by uncertainty in the values of <i>parameters</i> of the model, when the structure of the model (equations and processes) is fixed. In <i>perturbed parameter ensembles</i> , parameters have been expanded to include perturbation of environmental boundary conditions and other inputs like topography.
Structural uncertainty	Uncertainty arising from the way a model is built – its structure, equations, resolution, assumptions, and simplifications rather than the specific values of <i>parameters</i> or inputs.
Epistemic uncertainty	Uncertainty in a model’s predictions stemming from what we don’t know about the system being modelled, including uncertainty in structural choices, process representations and <i>parameter</i> values.
Internal variability	Uncertainty in model simulations arising from natural fluctuations in the climate system, even when external forcings, the model structure and <i>parameters</i> are fixed. Its size can be estimated by initial condition ensembles (Maher et al., 2021).
Perturbed parameter ensemble (PPE)	A set of model simulations created by systematically varying uncertain <i>parameters</i> within a model to explore the effect on model behaviour. PPEs are ‘designed’ ensembles in that the variations in output are due to defined differences between the ensemble members.
Multi-model ensemble (MME)	Simulations of a set of models usually following a prescribed ‘experimental protocol’, although not ‘designed’ in the same way as a PPE – see <i>perturbed parameter ensemble</i> .
Fidelity	The extent to which model simulations reproduce the observed state and behaviour of the climate system, in particular how well the model captures these in sufficient detail.
Model spread (sometimes called diversity)	The range of outcomes produced by a set of climate models in response to the same scenario or inputs. Model spread in an <i>MME</i> does not fully represent the model <i>reliability</i> because it is an unquantified mixture of model <i>structural uncertainty</i> , <i>parametric uncertainty</i> and <i>internal variability</i> unless these are standardized or suppressed in the experimental design.
Reliability	The extent to which models produce consistent and trustworthy results across multiple scenarios or applications contexts. A proxy for reliability is the <i>model spread</i> in a <i>multi-model ensemble</i> , but it is incomplete because the models do not represent the true range of <i>epistemic</i>

uncertainty. Reliability can manifest differently when observations are available to test *model performance*, as opposed to when making projections, where models may diverge despite good *model performance* due to *model equifinality*.

Robustness	Stability of model results across varying assumptions or uncertainties in the model.
Model performance	The goodness of fit or agreement of a model with observations.
Structural deficiency	A systematic limitation or flaw in the way a model represents the underlying processes (e.g., incomplete, wrong or over-simplified) leading to reduced realism.
Equifinality	The concept that different models, or different combinations of model parameters and structures within a model, can produce equally good <i>model performance</i> . Different models may diverge when making predictions, introducing previously hidden uncertainty and unreliability.
Constrain	To reduce uncertainty in model simulations. Constraint does not mean improvement of <i>model performance</i> , but a narrowing of the quantified uncertainty range, for example using methods of <i>history-matching</i> .
History-matching	A systematic, statistical process of finding combinations of <i>parameter settings (model variants)</i> that match observations within their uncertainty range. Unlike <i>tuning</i> , history matching starts with a dense sample of model <i>parametric uncertainty</i> and aims to rule out observationally implausible model variants rather than to find a single “best” model.
Model variant	In PPEs, a model variant is a specific version of a model with different parameter choices. More widely, the term may encompass different choices of specific model process representations within the same overall model structure or family.
Calibration	A systematic, statistical process of adjusting model <i>parameters</i> to find the model variant(s) that best match observations within their uncertainty range. Unlike <i>history-matching</i> , calibration aims to match observations as closely as possible to improve <i>model performance</i> . Unlike <i>tuning</i> , calibration usually starts with a sampling of model uncertainty. It may result in several model variants with similar <i>model performance</i> because of <i>equifinality</i> .
Tuning	In contrast to <i>calibration</i> and <i>history matching</i> , tuning is often a more informal or heuristic process of adjusting model <i>parameters</i> to improve <i>model performance</i> without first sampling model uncertainty.
Representativeness error or uncertainty	The error introduced when comparing simulated and observed quantities that represent different spatial or temporal domains—e.g., comparing fine-scale observations with coarse-scale model grid-box values.

2. A PPE primer

105 A perturbed parameter ensemble (PPE) is a set (ensemble) of model simulations in which each simulation has a different combination of selected parameters. In a traditional PPE, the parameters are most-often quantities in the model’s defining equations (parameterizations) and the purpose is to determine how combinations of parameter perturbations affect the model outputs. A PPE therefore explores the joint effects of changes in several parameters across a multi-dimensional ‘parameter space’ that cannot be learned by adjusting one parameter at a time. The increasingly diverse application of PPEs in climate

and atmospheric science originated in pioneering work on idealised experiments that doubled CO₂ concentrations (Murphy et al., 2004; Stainforth et al., 2005).

The PPE approach has been extended to include a wider range of quantities that can control the model outputs, such as gas and aerosol emission factors and, for some types of model set-up, boundary conditions (or ‘forcings’) like sea surface temperature or humidity. One benefit of this hybrid approach is that model behaviour and uncertainty can be explored in a consistent way over a range of environmental conditions (Wellmann et al., 2018).

PPEs are often combined with statistical (Lee et al., 2011; Qian et al., 2018) or machine learning (Elsaesser et al., 2025; Gettelman et al., 2024) emulation to create a much larger sample of input/output combinations for statistical analysis (Fig. 2). This is often necessary due to the high computational cost of the model, which limits simulations to a very sparse coverage of the multi-dimensional parameter space. Emulators are trained to learn the relationship between parameter settings and model outputs and, once trained, are fast to compute. Emulators enable model outputs to be estimated, ideally with associated emulator uncertainty, for any combination of parameter settings within the parameter space, which then provides a way to scale up from, say, 100 PPE members to millions of evaluations for statistical analysis. Typically, an ensemble of around 5-10 simulations per parameter produces a useful emulator, making PPEs a highly efficient computational method for exploring model behaviour. New ML/AI methods are enabling emulators to be built efficiently, reducing the cost of sampling parameter space for PPEs (Elsaesser et al., 2025; Gettelman et al., 2024).

Although PPEs are distinct from initial-condition ensembles (Maher et al., 2021), a PPE may sample internal variability, and methods have been developed to account for it in building emulators (Rostron et al., 2020; Sansom et al., 2024).

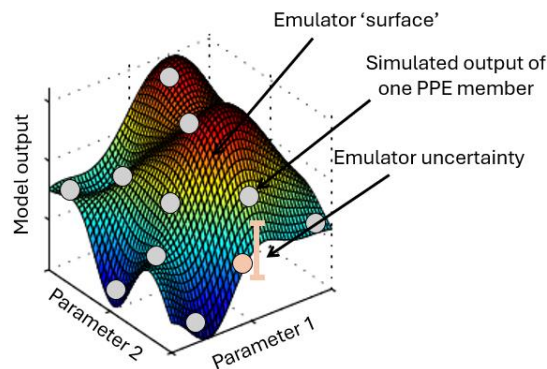


Figure 2. A perturbed parameter ensemble and an emulator trained on the data to describe a mapping between the input parameter values and the model outputs across a multi-dimensional parameter space. The emulator can then be used to generate millions of ‘model variants’ for any combination of parameter values for statistical analysis.

- Define ‘variants’ of a model that span a range of behaviours, such as the “calibrated physics ensemble” of GISS ModelE for submission to CMIP6 (Elsaesser et al., 2025), climate model variants that span a range of climate sensitivities (Hourdin et al., 2023; Peatier et al., 2022), and samples of extreme climates to support climate adaptation (Leach et al., 2022). PPEs show that the parametric uncertainty of a single model can be similar to that of multi-model ensembles (Duffy et al., 2023).
- Define and categorise structural errors within a climate model by identifying trade-offs between errors in different locations and fields and biases that are orthogonal to those resulting from parameter perturbations (Peatier et al., 2024).

In *aerosol*, global-scale PPEs have been used to:

- Understand and quantify model parameters controlling aerosol properties (Carslaw et al., 2013; Fanourgakis et al., 2019; Hamilton et al., 2014; Lee et al., 2011, 2012, 2013) and radiative forcing (Bhatti et al., 2026; Carslaw et al., 2013; Regayre et al., 2015, 2014), and to inform wider climate model tuning efforts (Sexton et al., 2021).
- Observationally constrain uncertainty in aerosol radiative forcing (Carzon et al., 2023; Johnson et al., 2018, 2020; Lee et al., 2016; McCoy et al., 2020; Regayre et al., 2018, 2020; Watson-Parris et al., 2020) and aerosol–cloud adjustments (Mikkelsen et al., 2025a; Song et al., 2024). In the stratosphere, a PPE explored how combinations of eruption properties affect volcanic radiative forcing (Marshall et al., 2019). As is the case for the whole climate system, parametric uncertainty in aerosol radiative forcing in one model can be comparable to the multi-model spread (Regayre et al., 2018; Yoshioka et al., 2019).
- Quantify the effect of aerosol on climate change, including climate threshold exceedances (Peace et al., 2020) and the climatic effect of historical volcanic eruptions using ice core sulfate measurements (Marshall et al., 2021).
- Expose model structural deficiencies in the representation of aerosols and clouds (Prévost et al., 2025; Regayre et al., 2023).
- Simulate the aerosol effects on East Asian climate and their parametric uncertainties associated with emissions and cloud microphysics (Yan et al., 2015).

In *chemistry*, PPEs have been used to:

- Quantify the sensitivity of ozone, hydroxyl radicals and methane lifetime to atmospheric conditions, uncertain processes and emissions across three global models, identifying the cause of differing model responses (Wild et al., 2020).
- Robustly calibrate model parameters through observational constraint to measured gas concentrations (Ryan and Wild, 2021). In a regional chemistry model, PPEs were used to define the important parameters of a complex organic chemistry scheme (Reyes-Villegas et al., 2023) and on the global scale to expose potential structural deficiencies in a secondary organic aerosol scheme (Sengupta et al., 2021).

In *cloud physics*, PPEs have been applied from large-eddy scale to global scale.

- 190 • For shallow clouds, PPEs at the large-eddy scale with model domains of tens of km and resolutions of tens of metres have exposed how combinations of cloud-controlling factors (environmental conditions such as above-cloud humidity and potential temperature) affect cloud evolution and adjustments to changes in aerosol (Glassmeier et al., 2019; Sansom et al., 2024, 2025). They have provided considerable insight into how stratocumulus cloud-field properties vary over a wide range of states and to understand the timescales of cloud response to shipping emissions (Glassmeier et al., 2021). For mixed-phase clouds, they have exposed how microphysical processes have a non-
195 linear interacting effect on cloud properties (Huang et al., 2026).
- For deep convective clouds, the key parameters controlling liquid and ice hydrometeors and precipitation (Johnson et al., 2015) and anvil cirrus (Hawker et al., 2021) in high-resolution (60 m to 250 m) models have been identified, as well as the relative importance of model parameters and environmental conditions for forecasting deep convection and hail (Wellmann et al., 2018, 2020). On the global scale, the response of circulation and cloud
200 responses to surface warming has been analysed to understand the results from multi-model ensembles (Schiro et al., 2019).
- On regional climate scales, Qian et al. (2024) used a set of short PPE simulations to understand how cloud forcing depends on different sets of parameters over different regions.

In *meteorology*, PPEs have been used to:

- 205 • Understand how environmental conditions and uncertainties in cloud microphysics affect the evolution of weather-scale phenomena. Studies with 500 m grid spacing examined how 11 environmental conditions like boundary layer height and soil moisture interact to affect the characteristics of sea breezes under dry (Igel et al., 2018) and moist (Park et al., 2020) convective conditions. At 3.3 km grid spacing a PPE has been used to quantify the relative importance of uncertainties in low-level temperature and moisture compared to uncertainty in cloud microphysical processes in numerical weather predictions of warm conveyor belts (Oertel et al., 2025). This PPE was also used to
210 quantify sensitivities of cirrus cloud properties and water vapor transport into the upper troposphere/lower stratosphere region to cloud microphysical parameters (Schwenk et al., 2025).
- Guide model development for improved tracking of wind and solar energy, as well as the design of major field studies (Berg et al., 2021; Yang et al., 2017, 2019). These studies examined the sensitivity of turbine-height wind
215 speeds to parameters in planetary boundary layer and surface-layer schemes, quantifying both parametric and structural sensitivities. The parametric sensitivity of solar irradiance to model parameters related to cloud and aerosol processes has also been assessed in WRF-Solar, allowing these parameters to be optimized to enhance irradiance forecasting skill by up to 33% (Liu et al., 2022b, c).

4. How can PPEs contribute to the future of climate modelling?

220 In this ‘golden age of climate modeling’ (Betancourt, 2022) we argue that PPEs ought to be included alongside the increases
in complexity, resolution, initial-condition ensembles and model intercomparison projects. Effort in this direction would adjust
the balance towards greater consideration of model reliability (consistent and trustworthy results) and plausible model spread
alongside existing efforts to improve model fidelity (models resembling the real world). We make eleven recommendations in
six areas:

225

1. Use PPEs to understand model structural differences and deficiencies, and to define priorities for model development or simplification.

PPEs provide a very effective way to disentangle structural and parametric causes of model–observation
biases, which would provide valuable information for prioritising model developments. For example, such separation would
help to prevent over-tuning of parameters, a practice that is likely to obscure structural deficiencies and the need for targeted
230 model developments. It would also help to prevent the addition of process-level details to a model where a comprehensive
sampling of parameter space might remove biases – i.e., where the need for structural changes is not supported by observational
evidence. PPEs offer a rigorous framework for identifying where model developments would be most effective, reducing
reliance on the interests or parameterization-specific expertise of development teams.

PPEs have exposed structural errors when a comprehensive sampling of the model’s parameter space fails to capture observed
235 system behaviour (Peatier et al., 2024) and when there are inconsistencies among observational constraints (Regayre et al.,
2023). Successful efforts to understand structural differences and to detect deficiencies have so far involved global climate
models (Furtado et al., 2023; Rostron et al., 2020; Sanderson, 2011; Shiogama et al., 2014; Tsushima et al., 2020; Yokohata
et al., 2010), land surface models (Hawkins et al., 2019; McNeall et al., 2016), ocean models (Williamson et al., 2015), and
regional and global chemistry and aerosol models (Regayre et al., 2023; Reyes-Villegas et al., 2023; Sengupta et al., 2021).

240 Alongside the detection of model deficiencies, PPEs can also help to identify opportunities for model simplification (Proske
et al., 2022, 2023). Developing parsimonious models that maintain simplicity without sacrificing fidelity increases model
robustness, computational efficiency, and interpretability.

The effort needed to operationalize this activity—integrating PPEs into the model development cycle—should be seen in the
context of the current, suboptimal approach to model development. A well-designed PPE can be exploited by an entire
245 development team to consistently tackle biases. There have been several barriers to wider adoption of the PPE approach, such
as the challenge of selecting appropriate parameters and perturbation ranges (which often requires input from multiple
developers), designing the simulations, implementing the perturbations in models, defining efficient workflows and simulation
submission scripts, and the large data volumes that are produced. However, we estimate that PPEs have now been developed
across a wide range of model types by over twenty research groups in small research teams and large modelling centres, which
250 has reduced some of the knowledge barriers. Barriers to wider operationalisation have been successfully tackled at several
modelling centres with streamlined PPE workflows (Elsaesser et al., 2025; Yarger et al., 2024). We argue that

operationalization could enable a step-change in model evaluation and development that takes us beyond the current incremental advancements between MIPs.

255 *We recommend developing a deeper understanding of how PPEs in combination with observations can be used to detect structural deficiencies in models. This should include new methods, e.g., using machine learning and structurally diverse models, to relate model-observation biases to deficiencies in the way that processes are represented in models.*

260 *We also recommend building PPEs into the model development cycle rather than being a separate activity after the tuned model version has been released. Such integration would enable model developers to understand the strengths and limitations of their parameterisation schemes and where model development is most needed and would be most effective.*

2. Include PPEs in MIPs to link known uncertainties through to adaptation and mitigation decision making, to understand inter-model differences, and to lay the foundation for rigorous multi-model calibration.

265 Multi-Model Ensembles (MMEs) are the main way that we communicate plausible uncertainties to wider science and impact communities. Adaptation, for example, is informed through a subset of CMIP simulations that are considered in IPCC's Impacts, Adaptation and Vulnerability assessments (WG2). However, we lack a mechanism to capture the wider uncertainties associated with PPEs in current MMEs (Fig. 4a), even when modelling centres are aware of them. For example, Golaz et al. (2013) showed that there were multiple parameter configurations of their GFDL climate model consistent with observed present-day climate, but which simulated very different historical temperature evolutions. PPE ensembles subsequent to IPCC's 4th Assessment MIP (CMIP3) (Shiogama et al., 2014; Watanabe et al., 2012) showed that the CMIP3 MIROC3 configuration was right on the lower end of the 4-10 K PPE range of climate sensitivities consistent with MIROC3's formulation. Absence of this PPE context led to the impression of a more confident CMIP3 uncertainty range than supported by the known model uncertainties. There is likely to be similar tension for modelling centres for the upcoming CMIP7. For example, PPEs show how the development of the UK climate model has led to improvements in radiative fluxes compared to observations (Rostron et al., 2025), but with typically larger climate feedbacks than the previous model configuration (that was deemed in CMIP6 to already be warmer than the IPCC's likely range). In the absence of a mechanism to capture PPE configurations in MMEs, insights from potential alternative configurations will not be captured in CMIP7 and will therefore not be visible to the wider science and impact communities.

280 Systematic inclusion of PPEs within MMEs opens up the potential for fuller, more rigorous multi-model calibration. A multi-model PPE (MMPPE) combined with emulators to generate dense samples of model data (Fig. 2) would offer new opportunities to exploit statistical history matching (Craig et al., 1997; Lee et al., 2016; Williamson et al., 2013) or other calibration methods to observationally constrain climate metrics – Fig. 4b. In contrast, the unquantified mix of dozens of model structural, parametric and initial-condition uncertainties represented by an extremely small number of MME members makes it statistically inappropriate to constrain the spread by down-weighting single models (Knutti et al., 2010) – Fig. 4c. History

matching has been used successfully to constrain individual PPEs and would be readily extendable to multi-model PPEs regardless of whether they followed a standardized protocol. Submission of a few PPE members to MIPs that span a range of model behaviours or climate metrics would be an efficient initial approach (Elsaesser et al., 2025; Hourdin et al., 2023; Peatier et al., 2022).

290

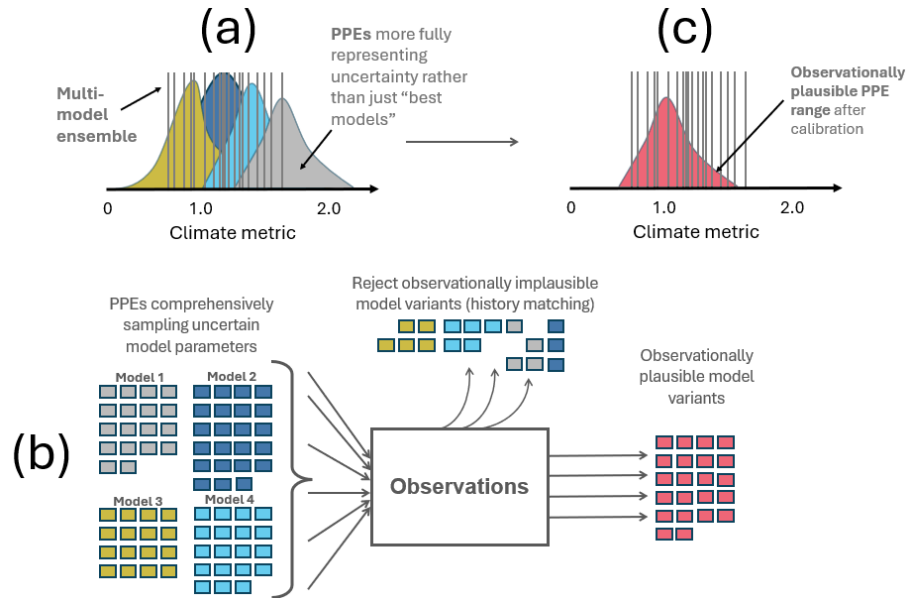


Figure 4. Schematic illustrating two aspects of PPEs. (a) PPEs more fully represent uncertainties than a multi-model ensemble (MME) of well-configured best estimates from several modelling centres. The MME (grey vertical lines) represents an unquantified mixture of structural differences, parametric uncertainties and internal variability, making it difficult to understand and ultimately reduce the spread in predictions or projections. PPEs (shown by pdfs) provide a way of cleanly separating structural and parametric uncertainty. (b) Statistically rigorous model calibration can be used to constrain the multi-model PPE range –shown in (c) – by defining a set of observationally plausible model variants, which is not possible with an MME alone. Variants are typically ruled out using a scoring approach like the implausibility metric (Craig et al., 1997; Williamson et al., 2013).

300

We recommend submission of selected PPE simulations to multi-model intercomparison projects to (i) remove the emphasis on over-tuned “best” models so that plausible uncertainties can be more fully communicated to wider science and impacts communities, (ii) more-comprehensively define model reliability, (iii) gain a deeper insight into the causes of multi-model spread, and (iv) enable statistically rigorous calibration of MMEs against observations. Multi-model PPEs (MMPPEs) will require an improved level of parameterization documentation, although efforts to design MMPPEs may naturally bring this about.

305

310 *We also recommend efforts to separate internal variability and parametric uncertainty so that PPEs of models exhibiting dynamic variability can be more easily compared. This could be achieved by running an initial condition ensemble in combination with a PPE.*

3. Adopt a common set of constraining observations and associated uncertainties. While observations are extensively used in MIPs (Waliser et al., 2020) to document and compare model skill, with MMPPEs they could also be used for model calibration (Fig. 4). This would require systematic quantification of observational uncertainties and model-observation
315 representation errors (Johnson et al., 2020; Schutgens et al., 2017, 2016). The challenge is particularly large because inspection of individual uncertainties is infeasible when comparing millions of model variants (estimates from emulators) against thousands of observations in an automated statistical manner (Johnson et al., 2020). Reliable information on observation uncertainty is also required in the detection of potential structural deficiencies (point 1) to avoid misattributing the causes of model-observations biases.

320

We recommend defining sets of observations for model constraint, including consideration of observation uncertainties and their representation errors in a form that would enable automated calibration of very large sets of model output.

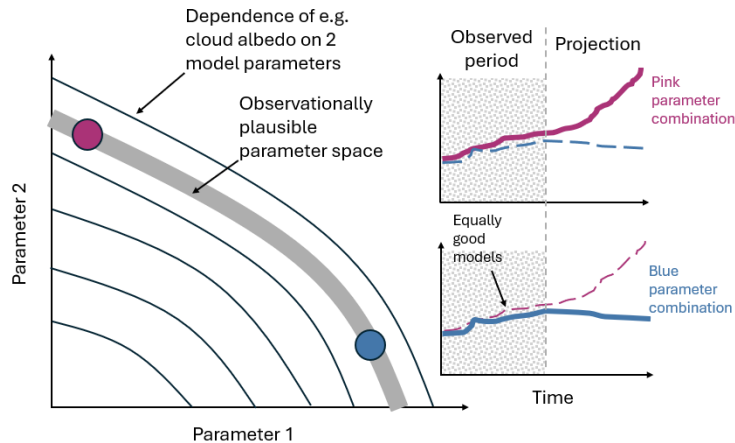
325 *We also recommend closer collaboration of the modelling and observational communities to understand the requirements of model uncertainty quantification. This could include (i) efforts towards systematic measurement strategies in field campaigns that consider the representativeness of the measurements (Kahn et al., 2023) and how they would be ingested into automated model calibration; (ii) consideration of structural uncertainty in retrieval products through, for example, retrieval bundles (Chiu et al., 2024; Elsaesser et al., 2025).*

330 **4. Account for and address model equifinality as a major cause of model unreliability.** Equifinality is where different combinations of model parameters or structures produce simulations that cannot be distinguished within the uncertainty of observations (Beven, 2006), but may diverge and cause large uncertainty in predictions or retrodictions (e.g., projections or historical climate simulations – see Fig. 5). In one example related to aerosols (Lee et al., 2016), equifinality was so strong that aerosol observations with arbitrarily small uncertainty had a very weak effect on the uncertainty in aerosol radiative forcing
335 despite both quantities having common uncertain parameters. Equifinality has profound implications for our interpretation of model performance (how well models match observations), which does not guarantee reliable historical or future simulations (Golaz et al., 2013). Without a PPE, it is impossible to assess the scale of this problem.

340 *Research is needed to identify the model processes responsible for compensating effects so that strategies can be developed to reduce their effect on model constraint. There are ways forward given that equifinality often occurs because observations of state variables (such as aerosol concentration) do not constrain the compensating processes, which we highlight in point 5.*

We recommend research to understand the causes of error compensation (equifinality) and ways to minimize the effects.

345 We also recommend submission of equifinal model variants developed using PPE methods to future MIPs to better characterize model reliability and to understand which projection metrics are most affected by equifinality.



350 **Figure 5. Schematic of how equifinality affects model reliability. During the period with observations (stippled), the pink and blue parameter combinations have similar model performance and appear to be reliable, but they can diverge and become unreliable in projections. Poor model projection reliability can therefore be masked when simulating only one of the parameter combinations, but will become visible with a PPE.**

355 **5. Use PPEs to identify new observations and new ways of using observations to constrain models.** PPEs are a valuable tool to not only leverage existing observations better, but also to plan future observations targeted at model constraint. This objective was highlighted by community consensus in the US-CLIVAR sponsored workshop Micro2Macro: Origins of Climate Change Uncertainty (McCoy et al., 2025). There are opportunities, for example, to understand where observations are effective or ineffective at reducing model uncertainty, the processes that cause the remaining uncertainty, and which observations would help to reduce it (Regayre et al., 2025). There are also opportunities to use observations in more creative ways, such as observational constraint of model processes rather than just simulated states, which would help to reduce equifinality.

360 Ultimately, we observe states, not processes, and we must stitch these states together with models to infer causation (Feingold et al., 2025; Hume, 1751). Outside of a few cases (Christensen et al., 2022; Malavelle et al., 2017; McCoy et al., 2018), it is very hard to infer causality from observations of state because of equifinality -- many combinations of processes can yield the same state (Gryspeerd et al., 2019; Stevens and Feingold, 2009; Wood et al., 2012). PPEs provide an opportunity to infer causation from observations (Mikkelsen et al., 2025b).

We recommend using PPEs to develop ways of identifying new observations or new combinations of existing observations with the greatest potential to constrain model uncertainty.

We also recommend exploiting PPEs to infer causation and process-level understanding from observations of atmospheric state.

6. Exploit PPEs of hierarchies of models to understand system behaviour and develop parameterizations.

Parameterization development typically involves a hierarchy of models and several steps, including understanding key process interactions across diverse environmental conditions, conducting sensitivity tests, and performing optimization and simplification (Gettelman, 2023). PPEs are ideal for tackling all these steps. PPEs of high-resolution process-based models have been performed in recent years and, as described above, they have helped to understand the joint effects of co-varying environmental conditions (Park et al., 2020; Sansom et al., 2024, 2025; Wellmann et al., 2020). Early research on connecting several models is very promising (Couvreur et al., 2021; Hourdin et al., 2021) and there is considerable scope to test or develop regime-aware parameterizations in a more comprehensive way than is possible with single perturbation studies or intercomparisons of process-based models (Blossey et al., 2016; Zhang et al., 2013).

We recommend applying PPEs across hierarchies of multi-scale models with varying levels of process sophistication to enhance our understanding of complex, nonlinear processes and to develop parameterisations for large-scale models.

5. Conclusions

In summary, we argue that PPEs:

1. Improve our understanding and quantification of model uncertainty in the widest sense by cleanly separating structural and parametric causes of uncertainty.
2. Provide deep insight into the causes of multi-model spread.
3. More fully characterize the plausible spread in climate projections, which is vital to better communicate current knowledge to downstream science, impacts and decisions.
4. Generate rich datasets that can be used for statistically rigorous observational calibration of model outputs to better characterize and reduce multi-model spread.
5. Guide the judicious selection of model structural developments or simplifications.
6. Guide observational strategies and ways of using existing observations to achieve the greatest possible constraint of models.
7. Provide a way to disentangle the interacting environmental drivers of system behaviour and to infer causation from observations of atmospheric state.

8. Contribute to the development of parameterizations by linking processes and sensitivities across a hierarchy of models.

We started this opinion piece by pointing out that there are several essentially competing efforts in climate modelling –
400 complexity, resolution and initial condition ensembles. To this list we add perturbed parameter ensembles, which we argue are
making a substantial contribution to assessing and improving model reliability and understanding model behaviour, which are
vital given the high societal cost of model uncertainty (Hope, 2015). Increased model complexity and resolution will
undoubtedly provide additional detailed and actionable information, for example about some types of extreme or local events,
and there are proposals to consolidate such efforts (Stevens et al., 2024). However, the development of climate models over
405 several decades has shown that, while our understanding of fundamental processes (knowledge uncertainty) has improved, we
have not translated this improved knowledge into commensurate improvements in our understanding of model reliability. This
lack of translation is what Box meant by “The scientist cannot obtain a ‘correct’ [model] by excessive elaboration” (Box,
1976), but we argue that PPEs provide the best opportunity to obtain models that are as ‘correct’ as possible. Critically,
applying PPEs as outlined here would enable more robust estimates of model uncertainty, empowering stakeholders to
410 realistically evaluate risk and make informed decisions based on information from the Earth system modelling community,
rather than relying on a single high-fidelity model variant. Treading a new path, with increased emphasis on the use of PPEs
in model evaluation and development will help to address the model uncertainty challenge.

Acknowledgements

The authors acknowledge funding from NERC (NE/X013901/1 and NE/Y001028/1); EU (821205); Research Council of
415 Finland (359166); Department of Energy (DOE) (DE-SC0024161, DE-SC0021270, DE-SC0022323, and DE-SC0023151);
the Transregional Collaborative Research Center SFB/TRR 165 “Waves to Weather”, the Italia—Deutschland science-4-
services network in weather and climate (4823IDEAP6) funded by the German Federal Ministry of Digital and Transport;
NASA Model Analysis and Prediction program (MAP) (#80NSSC21K1498), the NSF STC Learning the Earth with Artificial
Intelligence and Physics (LEAP) (2019625 and AGS-2203001); NASA (MAP NNH16ZDA001); Swiss National Science
420 Foundation Post Doc Mobility Grant (217899). LR, DS and KY were supported by the Met Office Hadley Centre Climate
Programme funded by DSIT. The Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by
the Battelle Memorial Institute under contract DE-AC05-76RL01830.

Author contributions

All authors contributed to the conceptualization and writing.

425 **Competing interests**

KC, AO and YQ are members of the ACP editorial board.

References

430 Baartman, J. E. M., Melsen, L. A., Moore, D., and van der Ploeg, M. J.: On the complexity of model complexity: Viewpoints across the geosciences, *CATENA*, 186, 104261, <https://doi.org/10.1016/j.catena.2019.104261>, 2020.

Bhatti, Y. A., Watson-Parris, D., Regayre, L. A., Jia, H., Neubauer, D., Im, U., Svenhag, C., Schutgens, N., Tsikerdekis, A., Nenes, A., Irfan, M., van Dierenhoven, B., Arifi, A., Fu, G., and Hasekamp, O. P.: Uncertainty in aerosol effective radiative forcing from anthropogenic and natural aerosol parameters in ECHAM6.3-HAM2.3, *Atmospheric Chemistry and Physics*, 26, 435 269–293, <https://doi.org/10.5194/acp-26-269-2026>, 2026.

Bellprat, O., Kotlarski, S., Lüthi, D., and Schär, C.: Exploring Perturbed Physics Ensembles in a Regional Climate Model, 25, 4582–4599, <https://doi.org/10.1175/JCLI-D-11-00275.1>, 2012.

440 Berg, L. K., Liu, Y., Yang, B., Qian, Y., Krishnamurthy, R., Sheridan, L., and Olson, J.: Time Evolution and Diurnal Variability of the Parametric Sensitivity of Turbine-Height Winds in the MYNN-EDMF Parameterization, *Journal of Geophysical Research: Atmospheres*, 126, e2020JD034000, <https://doi.org/10.1029/2020JD034000>, 2021.

Betancourt, M.: Are We Entering The Golden Age Of Climate Modeling?, *Eos*, <https://doi.org/10.1029/2022EO220538>, 2022.
445

Beven, K.: A manifesto for the equifinality thesis, *Journal of Hydrology*, 320, 18–36, <https://doi.org/10.1016/j.jhydrol.2005.07.007>, 2006.

Blossey, P. N., Bretherton, C. S., Cheng, A., Endo, S., Heus, T., Lock, A. P., and van der Dussen, J. J.: CGILS Phase 2 LES 450 intercomparison of response of subtropical marine low cloud regimes to CO₂ quadrupling and a CMIP3 composite forcing change, *Journal of Advances in Modeling Earth Systems*, 8, 1714–1726, <https://doi.org/10.1002/2016MS000765>, 2016.

Box, G. E. P.: Science and Statistics, *Journal of the American Statistical Association*, 71, 791–799, <https://doi.org/10.1080/01621459.1976.10480949>, 1976.
455

Brown, J. K., Dorheim, K., Mu, D., Snyder, A., Tebaldi, C., and Bond-Lamberty, B.: The Effect of Different Climate Sensitivity Priors on Projected Climate: A Probabilistic Analysis, *Geophysical Research Letters*, 52, e2024GL113505, <https://doi.org/10.1029/2024GL113505>, 2025.

460 Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V., Woodhouse, M. T., Regayre, L. A., and Pierce, J. R.: Large contribution of natural aerosols to uncertainty in indirect forcing, *Nature*, 503, 67–71, <https://doi.org/10.1038/nature12674>, 2013.

Carslaw, K. S., Lee, L. A., Regayre, L. A., and Johnson, J. S.: Climate Models Are Uncertain, but We Can Do Something
465 About It, *Eos*, 99, <https://doi.org/10.1029/2018EO093757>, 2018.

Carzon, J., Abreu, B., Regayre, L., Carslaw, K., Deaconu, L., Stier, P., Gordon, H., and Kuusela, M.: Statistical constraints on climate model parameters using a scalable cloud-based inference framework, *Environmental Data Science*, 2, e24, <https://doi.org/10.1017/eds.2023.12>, 2023.

470

Chiu, C., Ma, P.-L., Silber, I., Theisen, A., Williams, C., Ghate, V., O'Brien, J., Zhang, D., Chen, H., Comstock, J., Elsaesser, G., Feng, Y.-C., Gupta, S., Gustafson, W., Muradyan, P., Sockol, A., Yuan, T., Zhang, Y., Zheng, X., and Zhu, Z.: ARM Cloud and Precipitation Measurements and Science Group (CPMSG) 2024 Workshop Report, U.S. Department of Energy, Atmospheric Radiation Measurement user facility, Richland, Washington, 2024.

475

Christensen, M. W., Gettelman, A., Cermak, J., Dagan, G., Diamond, M., Douglas, A., Feingold, G., Glassmeier, F., Goren, T., Grosvenor, D. P., Gryspeerdt, E., Kahn, R., Li, Z., Ma, P.-L., Malavelle, F., McCoy, I. L., McCoy, D. T., McFarquhar, G., Mülmenstädt, J., Pal, S., Possner, A., Povey, A., Quaas, J., Rosenfeld, D., Schmidt, A., Schrödner, R., Sorooshian, A., Stier, P., Toll, V., Watson-Parris, D., Wood, R., Yang, M., and Yuan, T.: Opportunistic experiments to constrain aerosol effective
480 radiative forcing, *Atmos. Chem. Phys.*, 22, 641–674, <https://doi.org/10.5194/acp-22-641-2022>, 2022.

Couvreur, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., Rio, C., Audouin, O., Salter, J., Bazile, E., Brient, F., Favot, F., Honnert, R., Lefebvre, M.-P., Madeleine, J.-B., Rodier, Q., and Xu, W.: Process-Based Climate Model Development Harnessing Machine Learning: I. A Calibration Tool for Parameterization Improvement, *Journal of Advances in Modeling Earth Systems*, 13, e2020MS002217, <https://doi.org/10.1029/2020MS002217>, 2021.

485

Craig, P. S., Goldstein, M., Scheult, A. H., and Smith, J. A.: Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments, in: *Case Studies in Bayesian Statistics*, New York, NY, 37–93, https://doi.org/10.1007/978-1-4612-2290-3_2, 1997.

- 490 Duffy, M. L., Medeiros, B., Gettelman, A., and Eidhammer, T.: Perturbing Parameters to Understand Cloud Contributions to Climate Change,
- Durack, P. J., Taylor, K. E., Gleckler, P. J., Meehl, G. A., Lawrence, B. N., Covey, C., Stouffer, R. J., Levavasseur, G., Ben-Nasser, A., Denvil, S., Stockhause, M., Gregory, J. M., Juckes, M., Ames, S. K., Antonio, F., Bader, D. C., Dunne, J. P., Ellis,
495 D., Eyring, V., Fiore, S. L., Joussaume, S., Kershaw, P., Lamarque, J.-F., Lautenschlager, M., Lee, J., Mauzey, C. F., Mizielinski, M., Nassisi, P., Nuzzo, A., O'Rourke, E., Painter, J., Potter, G. L., Rodriguez, S., and Williams, D. N.: The Coupled Model Intercomparison Project (CMIP): Reviewing project history, evolution, infrastructure and implementation, *EGUsphere*, 1–74, <https://doi.org/10.5194/egusphere-2024-3729>, 2025.
- 500 Eidhammer, T., Gettelman, A., Thayer-Calder, K., Watson-Parris, D., Elsaesser, G., Morrison, H., van Lier-Walqui, M., Song, C., and McCoy, D.: An extensible perturbed parameter ensemble for the Community Atmosphere Model version 6, *Geoscientific Model Development*, 17, 7835–7853, <https://doi.org/10.5194/gmd-17-7835-2024>, 2024.
- Elsaesser, G. S., van Lier-Walqui, M., Yang, Q., Kelley, M., Ackerman, A. S., Fridlind, A. M., Cesana, G. V., Schmidt, G. A.,
505 Wu, J., Behrangi, A., Camargo, S. J., De, B., Inoue, K., Leitmann-Niimi, N. M., and Strong, J. D. O.: Using Machine Learning to Generate a GISS ModelE Calibrated Physics Ensemble (CPE), *Journal of Advances in Modeling Earth Systems*, 17, e2024MS004713, <https://doi.org/10.1029/2024MS004713>, 2025.
- Fanourgakis, G. S., Kanakidou, M., Nenes, A., Bauer, S. E., Bergman, T., Carslaw, K. S., Grini, A., Hamilton, D. S., Johnson,
510 J. S., Karydis, V. A., Kirkevåg, A., Kodros, J. K., Lohmann, U., Luo, G., Makkonen, R., Matsui, H., Neubauer, D., Pierce, J. R., Schmale, J., Stier, P., Tsigaridis, K., van Noije, T., Wang, H., Watson-Parris, D., Westervelt, D. M., Yang, Y., Yoshioka, M., Daskalakis, N., Decesari, S., Gysel-Beer, M., Kalivitis, N., Liu, X., Mahowald, N. M., Myriokefalitakis, S., Schrödner, R., Sfakianaki, M., Tsimpidi, A. P., Wu, M., and Yu, F.: Evaluation of global simulations of aerosol particle and cloud condensation nuclei number, with implications for cloud droplet formation, *Atmospheric Chemistry and Physics*, 19, 8591–
515 8617, <https://doi.org/10.5194/acp-19-8591-2019>, 2019.
- Feingold, G., Glassmeier, F., Zhang, J., and Hoffmann, F.: Opinion: Inferring Process from Snapshots of Cloud Systems, *EGUsphere*, 2025, 1–28, <https://doi.org/10.5194/egusphere-2025-1869>, 2025.
- 520 Furtado, K., Tsushima, Y., Field, P. R., Rostron, J., and Sexton, D.: The Relationship Between the Present-Day Seasonal Cycles of Clouds in the Mid-Latitudes and Cloud-Radiative Feedback, *Geophysical Research Letters*, 50, e2023GL103902, <https://doi.org/10.1029/2023GL103902>, 2023.

- 525 Gettelman, A.: Rainbows and climate change: a tutorial on climate model diagnostics and parameterization, *Geoscientific Model Development*, 16, 4937–4956, <https://doi.org/10.5194/gmd-16-4937-2023>, 2023.
- 530 Gettelman, A., Eidhammer, T., Duffy, M. L., McCoy, D. T., Song, C., and Watson-Parris, D.: The Interaction Between Climate Forcing and Feedbacks, *Journal of Geophysical Research: Atmospheres*, 129, e2024JD040857, <https://doi.org/10.1029/2024JD040857>, 2024.
- Glassmeier, F., Hoffmann, F., Johnson, J., Yamaguchi, T., Carslaw, K., and Feingold, G.: An emulator approach to stratocumulus susceptibility, *ATMOSPHERIC CHEMISTRY AND PHYSICS*, 19, <https://doi.org/10.5194/acp-19-10191-2019>, 2019.
- 535 Glassmeier, F., Hoffmann, F., Johnson, J., Yamaguchi, T., Carslaw, K., and Feingold, G.: Aerosol-cloud-climate cooling overestimated by ship-track data, *SCIENCE*, 371, 485–+, <https://doi.org/10.1126/science.abd3980>, 2021.
- Golaz, J.-C., Horowitz, L. W., and Levy II, H.: Cloud tuning in a coupled climate model: Impact on 20th century warming, *Geophysical Research Letters*, 40, 2246–2251, <https://doi.org/10.1002/grl.50232>, 2013.
- 540 Gryspeerdt, E., Goren, T., Sourdeval, O., Quaas, J., Mülmenstädt, J., Dipu, S., Unglaub, C., Gettelman, A., and Christensen, M.: Constraining the aerosol influence on cloud liquid water path, *Atmos. Chem. Phys.*, 19, 5331–5347, <https://doi.org/10.5194/acp-19-5331-2019>, 2019.
- 545 Hamilton, D. S., Lee, L. A., Pringle, K. J., Reddington, C. L., Spracklen, D. V., and Carslaw, K. S.: Occurrence of pristine aerosol environments on a polluted planet, *Proceedings of the National Academy of Sciences*, 111, 18466–18471, <https://doi.org/10.1073/pnas.1415440111>, 2014.
- 550 Hawker, R. E., Miltenberger, A. K., Johnson, J. S., Wilkinson, J. M., Hill, A. A., Shipway, B. J., Field, P. R., Murray, B. J., and Carslaw, K. S.: Model emulation to understand the joint effects of ice-nucleating particles and secondary ice production on deep convective anvil cirrus, *Atmospheric Chemistry and Physics*, 21, 17315–17343, <https://doi.org/10.5194/acp-21-17315-2021>, 2021.
- 555 Hawkins, L. R., Rupp, D. E., McNeall, D. J., Li, S., Betts, R. A., Mote, P. W., Sparrow, S. N., and Wallom, D. C. H.: Parametric Sensitivity of Vegetation Dynamics in the TRIFFID Model and the Associated Uncertainty in Projected Climate Change Impacts on Western U.S. Forests, *Journal of Advances in Modeling Earth Systems*, 11, 2787–2813, <https://doi.org/10.1029/2018MS001577>, 2019.

560 Hope, C.: The \$10 trillion value of better information about the transient climate response, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 373, 20140429, <https://doi.org/10.1098/rsta.2014.0429>, 2015.

565 Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., and Williamson, D.: The Art and Science of Climate Model Tuning,

Hourdin, F., Williamson, D., Rio, C., Couvreur, F., Roehrig, R., Villefranque, N., Musat, I., Fairhead, L., Diallo, F. B., and Volodina, V.: Process-Based Climate Model Development Harnessing Machine Learning: II. Model Calibration From Single Column to Global, *Journal of Advances in Modeling Earth Systems*, 13, e2020MS002225, <https://doi.org/10.1029/2020MS002225>, 2021.

570 Hourdin, F., Ferster, B., Deshayes, J., Mignot, J., Musat, I., and Williamson, D.: Toward machine-assisted tuning avoiding the underestimation of uncertainty in climate change projections, *Science Advances*, 9, eadf2758, <https://doi.org/10.1126/sciadv.adf2758>, 2023.

575 Huang, X., Field, P. R., Herbert, R. J., Murray, B. J., van den Heuvel, F., Grosvenor, D. P., Sansom, R. W. N., and Carslaw, K. S.: Interacting effects of droplet number and ice formation processes on mixed-phase cold-air outbreak clouds, *EGUsphere*, 1–49, <https://doi.org/10.5194/egusphere-2026-311>, 2026.

580 Hume, D.: *Philosophical essays concerning human understanding*, Second edition, with additions and corrections., Printed for M. Cooper, London, 1751.

Igel, A. L., Heever, S. C. van den, and Johnson, J. S.: Meteorological and Land Surface Properties Impacting Sea Breeze Extent and Aerosol Distribution in a Dry Environment, *Journal of Geophysical Research: Atmospheres*, 123, 22–37, <https://doi.org/10.1002/2017JD027339>, 2018.

585 Jiang, Y., Chen, L., Li, H., and Zhu, Y.: Parametric sensitivity analysis of East Asian summer-mean precipitation simulations by perturbed parameter ensemble experiments in CAM6, *Atmospheric and Oceanic Science Letters*, 100667, <https://doi.org/10.1016/j.aosl.2025.100667>, 2025.

- 590 Johnson, J. S., Cui, Z., Lee, L. A., Gosling, J. P., Blyth, A. M., and Carslaw, K. S.: Evaluating uncertainty in convective cloud microphysics using statistical emulation, *Journal of Advances in Modeling Earth Systems*, 7, 162–187, <https://doi.org/10.1002/2014MS000383>, 2015.
- Johnson, J. S., Regayre, L. A., Yoshioka, M., Pringle, K. J., Lee, L. A., Sexton, D. M. H., Rostron, J. W., Booth, B. B. B., and
595 Carslaw, K. S.: The importance of comprehensive parameter sampling and multiple observations for robust constraint of aerosol radiative forcing, *Atmospheric Chemistry and Physics*, 18, 13031–13053, <https://doi.org/10.5194/acp-18-13031-2018>, 2018.
- Johnson, J. S., Regayre, L. A., Yoshioka, M., Pringle, K. J., Turnock, S. T., Browse, J., Sexton, D. M. H., Rostron, J. W.,
600 Schutgens, N. A. J., Partridge, D. G., Liu, D., Allan, J. D., Coe, H., Ding, A., Cohen, D. D., Atanacio, A., Vakkari, V., Asmi, E., and Carslaw, K. S.: Robust observational constraint of uncertain aerosol processes and emissions in a climate model and the effect on aerosol radiative forcing, *Atmospheric Chemistry and Physics*, 20, 9491–9524, <https://doi.org/10.5194/acp-20-9491-2020>, 2020.
- 605 Kahn, R. A., Andrews, E., Brock, C. A., Chin, M., Feingold, G., Gettelman, A., Levy, R. C., Murphy, D. M., Nenes, A., Pierce, J. R., Popp, T., Redemann, J., Sayer, A. M., da Silva, A. M., Sogacheva, L., and Stier, P.: Reducing Aerosol Forcing Uncertainty by Combining Models With Satellite and Within-The-Atmosphere Observations: A Three-Way Street, *Reviews of Geophysics*, 61, e2022RG000796, <https://doi.org/10.1029/2022RG000796>, 2023.
- 610 Knüsel, B. and Baumberger, C.: Understanding climate phenomena with data-driven models, *Studies in History and Philosophy of Science Part A*, 84, 46–56, <https://doi.org/10.1016/j.shpsa.2020.08.003>, 2020.
- Knutti, R. and Sedláček, J.: Robustness and uncertainties in the new CMIP5 climate model projections, *Nature Clim Change*, 3, 369–373, <https://doi.org/10.1038/nclimate1716>, 2013.
- 615 Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., and Meehl, G. , <https://doi.org/10.1175/2009JCLI3361.1>, 2010.
- Leach, N. J., Watson, P. A. G., Sparrow, S. N., Wallom, D. C. H., and Sexton, D. M. H.: Generating samples of extreme winters to support climate adaptation, *Weather and Climate Extremes*, 36, 100419, <https://doi.org/10.1016/j.wace.2022.100419>, 2022.
- 620 Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., and Spracklen, D. V.: Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, *Atmospheric Chemistry and Physics*, 11, 12253–12273, <https://doi.org/10.5194/acp-11-12253-2011>, 2011.

- 625 Lee, L. A., Carslaw, K. S., Pringle, K. J., and Mann, G. W.: Mapping the uncertainty in global CCN using emulation, *Atmospheric Chemistry and Physics*, 12, 9739–9751, <https://doi.org/10.5194/acp-12-9739-2012>, 2012.
- Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., Pierce, J. R., and Carslaw, K. S.: The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei, *Atmospheric Chemistry and Physics*, 13, 8879–8914, <https://doi.org/10.5194/acp-13-8879-2013>, 2013.
- 630 Lee, L. A., Reddington, C. L., and Carslaw, K. S.: On the relationship between aerosol model uncertainty and radiative forcing uncertainty, *PNAS*, 113, 5820–5827, <https://doi.org/10.1073/pnas.1507050113>, 2016.
- Li, S., Rupp, D. E., Hawkins, L., Mote, P. W., McNeall, D., Sparrow, S. N., Wallom, D. C. H., Betts, R. A., and Wettstein, J. J.: Reducing climate model biases by exploring parameter space with large ensembles of climate model simulations and statistical emulation, *Geoscientific Model Development*, 12, 3017–3043, <https://doi.org/10.5194/gmd-12-3017-2019>, 2019.
- 635 Liu, S., Yang, B., Guo, Z., Wang, M., Qian, Y., Huang, A., and Zhang, Y.: Quantifying the local and remote impacts of sub-grid physical processes on the Southeast Pacific sea surface fluxes in the Community Atmosphere Model version 5 by a limited-area parameter perturbation approach, *International Journal of Climatology*, 42, 1369–1387, <https://doi.org/10.1002/joc.7308>, 2022a.
- Liu, Y., Qian, Y., Feng, S., Berg, L. K., Juliano, T. W., Jiménez, P. A., Gritmit, E., and Liu, Y.: Calibration of cloud and aerosol related parameters for solar irradiance forecasts in WRF-solar, *Solar Energy*, 241, 1–12, <https://doi.org/10.1016/j.solener.2022.05.064>, 2022b.
- 645 Liu, Y., Qian, Y., Feng, S., Berg, L. K., Juliano, T. W., Jiménez, P. A., and Liu, Y.: Sensitivity of solar irradiance to model parameters in cloud and aerosol treatments of WRF-solar, *Solar Energy*, 233, 446–460, <https://doi.org/10.1016/j.solener.2022.01.061>, 2022c.
- 650 Maher, N., Milinski, S., and Ludwig, R.: Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble, *Earth System Dynamics*, 12, 401–418, <https://doi.org/10.5194/esd-12-401-2021>, 2021.
- 655 Malavelle, F. F., Haywood, J. M., Jones, A., Gettelman, A., Clarisse, L., Bauduin, S., Allan, R. P., Karset, I. H. H., Kristjánsson, J. E., Oreopoulos, L., Cho, N., Lee, D., Bellouin, N., Boucher, O., Grosvenor, D. P., Carslaw, K. S., Dhomse, S., Mann, G. W., Schmidt, A., Coe, H., Hartley, M. E., Dalvi, M., Hill, A. A., Johnson, B. T., Johnson, C. E., Knight, J. R., O’Connor, F.

M., Partridge, D. G., Stier, P., Myhre, G., Platnick, S., Stephens, G. L., Takahashi, H., and Thordarson, T.: Strong constraints on aerosol–cloud interactions from volcanic eruptions, *Nature*, 546, 485–491, <https://doi.org/10.1038/nature22974>, 2019.

660

Marshall, L., Johnson, J., Mann, G., Lee, L., Dhomse, S., Regayre, L., Yoshioka, M., Carslaw, K., and Schmidt, A.: Exploring How Eruption Source Parameters Affect Volcanic Radiative Forcing Using Statistical Emulation, *JOURNAL OF GEOPHYSICAL RESEARCH-ATMOSPHERES*, 124, 964–985, <https://doi.org/10.1029/2018JD028675>, 2019.

665 Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., Haak, H., Jungelaus, J., Klocke, D., Matei, D., Mikolajewicz, U., Notz, D., Pincus, R., Schmidt, H., and Tomassini, L.: Tuning the climate of a global model, *Journal of Advances in Modeling Earth Systems*, 4, <https://doi.org/10.1029/2012MS000154>, 2012.

670 Marshall, L., Schmidt, A., Johnson, J., Mann, G., Lee, L., Rigby, R., and Carslaw, K.: Unknown Eruption Source Parameters Cause Large Uncertainty in Historical Volcanic Radiative Forcing Reconstructions, *JOURNAL OF GEOPHYSICAL RESEARCH-ATMOSPHERES*, 126, <https://doi.org/10.1029/2020JD033578>, 2021.

675 McCoy, A. D., Wood, A. R., Burrows, A. S. M., Fridlind, A. A., Igel, A. A., Jen, A. C., Regayre, A. L., Saito, A. M., Watson-Parris, A. D., Cannistraci, E. A., and Patterson, E. M.: Micro2Macro: Origins of Climate Change Uncertainty: A US CLIVAR Workshop Report, U.S. CLIVAR Project Office, 2025.

McCoy, D. T., Bender, F. A. M., Grosvenor, D. P., Mohrmann, J. K., Hartmann, D. L., Wood, R., and Field, P. R.: Predicting decadal trends in cloud droplet number concentration using reanalysis and satellite data, *Atmos. Chem. Phys.*, 18, 2035–2047, <https://doi.org/10.5194/acp-18-2035-2018>, 2018.

680

McCoy, I. L., McCoy, D. T., Wood, R., Regayre, L., Watson-Parris, D., Grosvenor, D. P., Mulcahy, J. P., Hu, Y., Bender, F. A.-M., Field, P. R., Carslaw, K. S., and Gordon, H.: The hemispheric contrast in cloud microphysical properties constrains aerosol forcing, *Proceedings of the National Academy of Sciences*, 117, 18998–19006, <https://doi.org/10.1073/pnas.1922502117>, 2020.

685

McNeill, D., Williams, J., Booth, B., Betts, R., Challenor, P., Wiltshire, A., and Sexton, D.: The impact of structural error on parameter constraint in a climate model, *Earth System Dynamics*, 7, 917–935, <https://doi.org/10.5194/esd-7-917-2016>, 2016.

690 Mikkelsen, A., McCoy, D. T., Eidhammer, T., Gettelman, A., Song, C., Gordon, H., and McCoy, I. L.: Constraining aerosol–cloud adjustments by uniting surface observations with a perturbed parameter ensemble, *Atmospheric Chemistry and Physics*, 25, 4547–4570, <https://doi.org/10.5194/acp-25-4547-2025>, 2025a.

Mikkelsen, A., McCoy, D. T., Eidhammer, T., Gettelman, A., Song, C., Gordon, H., and McCoy, I. L.: Constraining aerosol–cloud adjustments by uniting surface observations with a perturbed parameter ensemble, *Atmos. Chem. Phys.*, 25, 4547–4570, 695 <https://doi.org/10.5194/acp-25-4547-2025>, 2025b.

Morrison, H., van Lier-Walqui, M., Fridlind, A. M., Grabowski, W. W., Harrington, J. Y., Hoose, C., Korolev, A., Kumjian, M. R., Milbrandt, J. A., Pawlowska, H., Posselt, D. J., Prat, O. P., Reimel, K. J., Shima, S.-I., van Dierenhoven, B., and Xue, L.: Confronting the Challenge of Modeling Cloud and Precipitation Microphysics, *Journal of Advances in Modeling Earth* 700 *Systems*, 12, e2019MS001689, <https://doi.org/10.1029/2019MS001689>, 2020.

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J., Collins, M., and Stainforth, D. A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations, *Nature*, 430, 768–772, 705 <https://doi.org/10.1038/nature02771>, 2004.

Oertel, A., Miltenberger, A. K., Grams, C. M., and Hoose, C.: Sensitivities of warm conveyor belt ascent, associated precipitation characteristics and large-scale flow pattern: Insights from a perturbed parameter ensemble, *Quarterly Journal of the Royal Meteorological Society*, n/a, e4986, <https://doi.org/10.1002/qj.4986>, 2025.

710 Park, J. M., Heever, S. C. van den, Igel, A. L., Grant, L. D., Johnson, J. S., Saleeby, S. M., Miller, S. D., and Reid, J. S.: Environmental Controls on Tropical Sea Breeze Convection and Resulting Aerosol Redistribution, *Journal of Geophysical Research: Atmospheres*, 125, e2019JD031699, <https://doi.org/10.1029/2019JD031699>, 2020.

Peace, A. H., Carslaw, K. S., Lee, L. A., Regayre, L. A., Booth, B. B. B., Johnson, J. S., and Bernie, D.: Effect of aerosol 715 radiative forcing uncertainty on projected exceedance year of a 1.5 °C global temperature rise, *Environ. Res. Lett.*, 15, 0940a6, <https://doi.org/10.1088/1748-9326/aba20c>, 2020.

Peace, A. H., Booth, B. B. B., Regayre, L. A., Carslaw, K. S., Sexton, D. M. H., Bonfils, C. J. W., and Rostron, J. W.: Evaluating uncertainty in aerosol forcing of tropical precipitation shifts, *Earth System Dynamics*, 13, 1215–1232, 720 <https://doi.org/10.5194/esd-13-1215-2022>, 2022.

Peatier, S., Sanderson, B. M., Terray, L., and Roehrig, R.: Investigating Parametric Dependence of Climate Feedbacks in the Atmospheric Component of CNRM-CM6-1, *Geophysical Research Letters*, 49, e2021GL095084, 725 <https://doi.org/10.1029/2021GL095084>, 2022.

- Peatier, S., Sanderson, B. M., and Terray, L.: Exploration of diverse solutions for the calibration of imperfect climate models, *Earth System Dynamics*, 15, 987–1014, <https://doi.org/10.5194/esd-15-987-2024>, 2024.
- 730 Prévost, L. M. C., Regayre, L. A., Johnson, J. S., McNeill, D., Milton, S., and Carslaw, K. S.: Detection of structural deficiencies in a global aerosol model to explain limits in parametric uncertainty reduction, <https://doi.org/10.5194/egusphere-2025-4795>, 7 October 2025.
- Proske, U., Ferrachat, S., Neubauer, D., Staab, M., and Lohmann, U.: Assessing the potential for simplification in global climate model cloud microphysics, *Atmospheric Chemistry and Physics*, 22, 4737–4762, <https://doi.org/10.5194/acp-22-4737-2022>, 2022.
- 735 Proske, U., Ferrachat, S., Klampt, S., Abeling, M., and Lohmann, U.: Addressing Complexity in Global Aerosol Climate Model Cloud Microphysics, *Journal of Advances in Modeling Earth Systems*, 15, e2022MS003571, <https://doi.org/10.1029/2022MS003571>, 2023.
- 740 Proske, U., Ferrachat, S., and Lohmann, U.: Developing a climatological simplification of aerosols to enter the cloud microphysics of a global climate model, *Atmospheric Chemistry and Physics*, 24, 5907–5933, <https://doi.org/10.5194/acp-24-5907-2024>, 2024.
- 745 Puy, A., Beneventano, P., Levin, S. A., Lo Piano, S., Portaluri, T., and Saltelli, A.: Models with higher effective dimensions tend to produce more uncertain estimates, *Science Advances*, 8, eabn9450, <https://doi.org/10.1126/sciadv.abn9450>, 2022.
- 750 Qian, Y., Yan, H., Hou, Z., Johannesson, G., Klein, S., Lucas, D., Neale, R., Rasch, P., Swiler, L., Tannahill, J., Wang, H., Wang, M., and Zhao, C.: Parametric sensitivity analysis of precipitation at global and local scales in the Community Atmosphere Model CAM5, *Journal of Advances in Modeling Earth Systems*, 7, 382–411, <https://doi.org/10.1002/2014MS000354>, 2015.
- Qian, Y., Jackson, C., Giorgi, F., Booth, B., Duan, Q., Forest, C., Higdon, D., Hou, Z. , <https://doi.org/10.1175/BAMS-D-15-00297.1>, 2016.
- 755 Qian, Y., Wan, H., Yang, B., Golaz, J.-C., Harrop, B., Hou, Z., Larson, V. E., Leung, L. R., Lin, G., Lin, W., Ma, P.-L., Ma, H.-Y., Rasch, P., Singh, B., Wang, H., Xie, S., and Zhang, K.: Parametric Sensitivity and Uncertainty Quantification in the Version 1 of E3SM Atmosphere Model Based on Short Perturbed Parameter Ensemble Simulations, *Journal of Geophysical Research: Atmospheres*, 123, 13,046–13,073, <https://doi.org/10.1029/2018JD028927>, 2018.

- 760 Regayre, L., Pringle, K., Lee, L., Rap, A., Browse, J., Mann, G., Reddington, C., Carslaw, K., Booth, B., and Woodhouse, M.: The Climatic Importance of Uncertainties in Regional Aerosol-Cloud Radiative Forcings over Recent Decades, *JOURNAL OF CLIMATE*, 28, 6589–6607, <https://doi.org/10.1175/JCLI-D-15-0127.1>, 2015.
- Regayre, L. A., Pringle, K. J., Booth, B. B. B., Lee, L. A., Mann, G. W., Browse, J., Woodhouse, M. T., Rap, A., Reddington,
765 C. L., and Carslaw, K. S.: Uncertainty in the magnitude of aerosol-cloud radiative forcing over recent decades, *Geophysical Research Letters*, 41, 9040–9049, <https://doi.org/10.1002/2014GL062029>, 2014.
- Regayre, L. A., Johnson, J. S., Yoshioka, M., Pringle, K. J., Sexton, D. M. H., Booth, B. B. B., Lee, L. A., Bellouin, N., and
770 Carslaw, K. S.: Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF, *Atmospheric Chemistry and Physics*, 18, 9975–10006, <https://doi.org/10.5194/acp-18-9975-2018>, 2018.
- Regayre, L. A., Schmale, J., Johnson, J. S., Tatzelt, C., Baccarini, A., Henning, S., Yoshioka, M., Stratmann, F., Gysel-Beer,
M., Grosvenor, D. P., and Carslaw, K. S.: The value of remote marine aerosol measurements for constraining radiative forcing
775 uncertainty, *Atmospheric Chemistry and Physics*, 20, 10063–10072, <https://doi.org/10.5194/acp-20-10063-2020>, 2020.
- Regayre, L. A., Deaconu, L., Grosvenor, D. P., Sexton, D. M. H., Symonds, C., Langton, T., Watson-Paris, D., Mulcahy, J.
P., Pringle, K. J., Richardson, M., Johnson, J. S., Rostron, J. W., Gordon, H., Lister, G., Stier, P., and Carslaw, K. S.: Identifying
climate model structural inconsistencies allows for tight constraint of aerosol radiative forcing, *Atmospheric Chemistry and
780 Physics*, 23, 8749–8768, <https://doi.org/10.5194/acp-23-8749-2023>, 2023.
- Regayre, L. A., Prévost, L. M. C., Ghosh, K., Johnson, J. S., Oakley, J. E., Owen, J., Webb, I., and Carslaw, K. S.: Remaining
aerosol forcing uncertainty after observational constraint and the processes that cause it, *Atmos. Chem. Phys.*, Submitted,
2025.
- 785 Reyes-Villegas, E., Lowe, D., Johnson, J., Carslaw, K., Darbyshire, E., Flynn, M., Allan, J., Coe, H., Chen, Y., Wild, O.,
Archer-Nicholls, S., Archibald, A., Singh, S., Shrivastava, M., Zaveri, R., Singh, V., Beig, G., Sokhi, R., and McFiggans, G.:
Simulating organic aerosol in Delhi with WRF-Chem using the volatility-basis-set approach: exploring model uncertainty with
a Gaussian process emulator, *ATMOSPHERIC CHEMISTRY AND PHYSICS*, 23, 5763–5782, [https://doi.org/10.5194/acp-
23-5763-2023](https://doi.org/10.5194/acp-23-5763-2023), 2023.
- 790 Rostron, J. W., Sexton, D. M. H., McSweeney, C. F., Yamazaki, K., Andrews, T., Furtado, K., Ringer, M. A., and Tsushima,
Y.: The impact of performance filtering on climate feedbacks in a perturbed parameter ensemble, *Clim Dyn*, 55, 521–551,
<https://doi.org/10.1007/s00382-020-05281-8>, 2020.

795 Rostron, J. W., Sexton, D. M. H., Furtado, K., and Tsushima, Y.: A clearer view of systematic errors in model development:
two practical approaches using perturbed parameter ensembles. | Research Square, <https://doi.org/10.21203/rs.3.rs-5025285/v1>, 19 March 2025.

Ryan, E. and Wild, O.: Calibrating a global atmospheric chemistry transport model using Gaussian process emulation and
800 ground-level concentrations of ozone and carbon monoxide, *Geoscientific Model Development*, 14, 5373–5391,
<https://doi.org/10.5194/gmd-14-5373-2021>, 2021.

Sanderson, B. M.: A Multimodel Study of Parametric Uncertainty in Predictions of Climate Response to Rising Greenhouse
Gas Concentrations,
805

Sansom, R. W. N., Carslaw, K. S., Johnson, J. S., and Lee, L.: An Emulator of Stratocumulus Cloud Response to Two Cloud-
Controlling Factors Accounting for Internal Variability, *Journal of Advances in Modeling Earth Systems*, 16,
e2023MS004179, <https://doi.org/10.1029/2023MS004179>, 2024.

810 Sansom, R. W. N., Johnson, J. S., Regayre, L. A., Lee, L. A., and Carslaw, K. S.: Strong control of the stratocumulus-to-
cumulus transition time by aerosol: analysis of the joint roles of several cloud-controlling factors using Gaussian process
emulation, *EGUsphere*, 1–30, <https://doi.org/10.5194/egusphere-2025-3104>, 2025.

Schiro, K. A., Su, H., Wang, Y., Langenbrunner, B., Jiang, J. H., and Neelin, J. D.: Relationships Between Tropical Ascent
815 and High Cloud Fraction Changes With Warming Revealed by Perturbation Physics Experiments in CAM5, *Geophysical
Research Letters*, 46, 10112–10121, <https://doi.org/10.1029/2019GL083026>, 2019.

Schutgens, N., Tsyro, S., Gryspeerdt, E., Goto, D., Weigum, N., Schulz, M., and Stier, P.: On the spatio-temporal
representativeness of observations, *Atmospheric Chemistry and Physics*, 17, 9761–9780, [https://doi.org/10.5194/acp-17-9761-](https://doi.org/10.5194/acp-17-9761-2017)
820 2017, 2017.

Schutgens, N. A. J., Gryspeerdt, E., Weigum, N., Tsyro, S., Goto, D., Schulz, M., and Stier, P.: Will a perfect model agree
with perfect observations? The impact of spatial sampling, *Atmospheric Chemistry and Physics*, 16, 6335–6353,
<https://doi.org/10.5194/acp-16-6335-2016>, 2016.

825 Schwenk, C., Miltenberger, A., and Oertel, A.: Microphysical Parameter Choices Modulate Ice Content and Relative Humidity
in the Outflow of a Warm Conveyor Belt, *EGUsphere*, 1–43, <https://doi.org/10.5194/egusphere-2025-1816>, 2025.

830 Sengupta, K., Pringle, K., Johnson, J., Reddington, C., Browse, J., Scott, C., and Carslaw, K.: A global model perturbed
parameter ensemble study of secondary organic aerosol formation, *ATMOSPHERIC CHEMISTRY AND PHYSICS*, 21,
2693–2723, <https://doi.org/10.5194/acp-21-2693-2021>, 2021.

835 Sexton, D. M. H., McSweeney, C. F., Rostron, J. W., Yamazaki, K., Booth, B. B. B., Murphy, J. M., Regayre, L., Johnson, J.
S., and Karmalkar, A. V.: A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 1: selecting
the parameter combinations, *Clim Dyn*, 56, 3395–3436, <https://doi.org/10.1007/s00382-021-05709-9>, 2021.

840 Shiogama, H., Watanabe, M., Ogura, T., Yokohata, T., and Kimoto, M.: Multi-parameter multi-physics ensemble (MPMPE):
a new approach exploring the uncertainties of climate sensitivity, *Atmospheric Science Letters*, 15, 97–102,
<https://doi.org/10.1002/asl2.472>, 2014.

845 Slingo, J., Bates, P., Bauer, P., Belcher, S., Palmer, T., Stephens, G., Stevens, B., Stocker, T., and Teutsch, G.: Ambitious
partnership needed for reliable climate prediction, *Nat. Clim. Chang.*, 12, 499–503, <https://doi.org/10.1038/s41558-022-01384-8>, 2022.

845 Song, C., McCoy, D. T., Eidhammer, T., Gettelman, A., McCoy, I. L., Watson-Parris, D., Wall, C. J., Elsaesser, G., and Wood,
R.: Buffering of Aerosol-Cloud Adjustments by Coupling Between Radiative Susceptibility and Precipitation Efficiency,
Geophysical Research Letters, 51, e2024GL108663, <https://doi.org/10.1029/2024GL108663>, 2024.

850 Stainforth, D. A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D. J., Kettleborough, J. A., Knight, S., Martin, A.,
Murphy, J. M., Piani, C., Sexton, D., Smith, L. A., Spicer, R. A., Thorpe, A. J., and Allen, M. R.: Uncertainty in predictions
of the climate response to rising levels of greenhouse gases, *Nature*, 433, 403–406, <https://doi.org/10.1038/nature03301>, 2005.

855 Stevens, B. and Feingold, G.: Untangling aerosol effects on clouds and precipitation in a buffered system, *Nature*, 461, 607–
613, 2009.

855 Stevens, B.: A Perspective on the Future of CMIP, *AGU Advances*, 5, e2023AV001086,
<https://doi.org/10.1029/2023AV001086>, 2024.

860 Stevens, B., Adami, S., Ali, T., Anzt, H., Aslan, Z., Attinger, S., Bäck, J., Baehr, J., Bauer, P., Bernier, N., Bishop, B.,
Bockelmann, H., Bony, S., Brasseur, G., Bresch, D. N., Breyer, S., Brunet, G., Buttigieg, P. L., Cao, J., Castet, C., Cheng, Y.,
Dey Choudhury, A., Coen, D., Crewell, S., Dabholkar, A., Dai, Q., Doblas-Reyes, F., Durran, D., El Gaidi, A., Ewen, C.,

- Exarchou, E., Eyring, V., Falkenhoff, F., Farrell, D., Forster, P. M., Frassoni, A., Frauen, C., Fuhrer, O., Gani, S., Gerber, E., Goldfarb, D., Grieger, J., Gruber, N., Hazeleger, W., Herken, R., Hewitt, C., Hoefler, T., Hsu, H.-H., Jacob, D., Jahn, A., Jakob, C., Jung, T., Kadow, C., Kang, I.-S., Kang, S., Kashinath, K., Kleinen-von Königslöw, K., Klocke, D., Kloenne, U., Klöwer, M., Kodama, C., Kollet, S., Kölling, T., Kontkanen, J., Kopp, S., Koran, M., Kulmala, M., Lappalainen, H., Latifi, F., Lawrence, B., Lee, J. Y., Lejeun, Q., Lessig, C., Li, C., Lippert, T., Luterbacher, J., Manninen, P., Marotzke, J., Matsouoka, S., Merchant, C., Messmer, P., Michel, G., Michielsen, K., Miyakawa, T., Müller, J., Munir, R., Narayanasetti, S., Ndiaye, O., Nobre, C., Oberg, A., Oki, R., Özkan-Haller, T., Palmer, T., Posey, S., Prein, A., Primus, O., Pritchard, M., Pullen, J., Putrasahan, D., et al.: Earth Virtualization Engines (EVE), *Earth System Science Data*, 16, 2113–2122, <https://doi.org/10.5194/essd-16-2113-2024>, 2024.
- 865
- Tsushima, Y., Ringer, M. A., Martin, G. M., Rostron, J. W., and Sexton, D. M. H.: Investigating physical constraints on climate feedbacks using a perturbed parameter ensemble, *Clim Dyn*, 55, 1159–1185, <https://doi.org/10.1007/s00382-020-05318-y>, 2020.
- 875
- Waliser, D., Gleckler, P. J., Ferraro, R., Taylor, K. E., Ames, S., Biard, J., Bosilovich, M. G., Brown, O., Chepfer, H., Cinquini, L., Durack, P. J., Eyring, V., Mathieu, P.-P., Lee, T., Pinnock, S., Potter, G. L., Rixen, M., Saunders, R., Schulz, J., Thépaut, J.-N., and Tuma, M.: Observations for Model Intercomparison Project (Obs4MIPs): status for CMIP6, *Geoscientific Model Development*, 13, 2945–2958, <https://doi.org/10.5194/gmd-13-2945-2020>, 2020.
- 880
- Watanabe, M., Shiogama, H., Yokohata, T., Kamae, Y., Yoshimori, M., Ogura, T., Annan, J. D., Hargreaves, J. C., Emori, S., and Kimoto, M.: Using a Multiphysics Ensemble for Exploring Diversity in Cloud–Shortwave Feedback in GCMs, *Watson-Parris, D.: Integrating Top-Down Energetic Constraints With Bottom-Up Process-Based Constraints for More Accurate Projections of Future Warming, Geophysical Research Letters*, 52, e2024GL114269, <https://doi.org/10.1029/2024GL114269>, 2025.
- 885
- Watson-Parris, D., Bellouin, N., Deaconu, L., Schutgens, N., Yoshioka, M., Regayre, L., Pringle, K., Johnson, J., Smith, C., Carslaw, K., and Stier, P.: Constraining Uncertainty in Aerosol Direct Forcing, *GEOPHYSICAL RESEARCH LETTERS*, 47, <https://doi.org/10.1029/2020GL087141>, 2020.
- 890
- Wellmann, C., Barrett, A. I., Johnson, J. S., Kunz, M., Vogel, B., Carslaw, K. S., and Hoose, C.: Using Emulators to Understand the Sensitivity of Deep Convective Clouds and Hail to Environmental Conditions, *Journal of Advances in Modeling Earth Systems*, 10, 3103–3122, <https://doi.org/10.1029/2018MS001465>, 2018.

- 895 Wellmann, C., Barrett, A. I., Johnson, J. S., Kunz, M., Vogel, B., Carslaw, K. S., and Hoose, C.: Comparing the impact of environmental conditions and microphysics on the forecast uncertainty of deep convective clouds and hail, *Atmospheric Chemistry and Physics*, 20, 2201–2219, <https://doi.org/10.5194/acp-20-2201-2020>, 2020.
- Wild, O., Voulgarakis, A., O'Connor, F., Lamarque, J.-F., Ryan, E. M., and Lee, L.: Global sensitivity analysis of chemistry–
900 climate model budgets of tropospheric ozone and OH: exploring model diversity, *Atmospheric Chemistry and Physics*, 20, 4047–4058, <https://doi.org/10.5194/acp-20-4047-2020>, 2020.
- Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., and Yamazaki, K.: History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble, *Clim Dyn*,
905 41, 1703–1729, <https://doi.org/10.1007/s00382-013-1896-4>, 2013.
- Williamson, D., Blaker, A. T., Hampton, C., and Salter, J.: Identifying and removing structural biases in climate models with history matching, *Clim Dyn*, 45, 1299–1324, <https://doi.org/10.1007/s00382-014-2378-z>, 2015.
- Wood, R., Leon, D., Lebsack, M., Snider, J., and Clarke, A. D.: Precipitation driving of droplet concentration variability in
910 marine low clouds, *J Geophys Res-Atmos*, 117, n/a-n/a, <https://doi.org/10.1029/2012jd018305>, 2012.
- Yamazaki, K., Sexton, D. M. H., Rostron, J. W., McSweeney, C. F., Murphy, J. M., and Harris, G. R.: A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 2: global performance and future changes, *Clim Dyn*, 56, 3437–3471, <https://doi.org/10.1007/s00382-020-05608-5>, 2021.
- 915 Yamazaki, K., Jackson, L. C., and Sexton, D. M. H.: Prediction of slowdown of the Atlantic Meridional Overturning Circulation in coupled model simulations, *Clim Dyn*, 62, 5197–5217, <https://doi.org/10.1007/s00382-024-07159-5>, 2024.
- Yan, H., Qian, Y., Zhao, C., Wang, H., Wang, M., Yang, B., Liu, X., and Fu, Q.: A new approach to modeling aerosol effects on East Asian climate: Parametric uncertainties associated with emissions, cloud microphysics, and their interactions, *Journal of Geophysical Research: Atmospheres*, 120, 8905–8924, <https://doi.org/10.1002/2015JD023442>, 2015.
- 920 Yang, B., Qian, Y., Berg, L. K., Ma, P.-L., Wharton, S., Bulaevskaya, V., Yan, H., Hou, Z., and Shaw, W. J.: Sensitivity of Turbine-Height Wind Speeds to Parameters in Planetary Boundary-Layer and Surface-Layer Schemes in the Weather Research and Forecasting Model, *Boundary-Layer Meteorol*, 162, 117–142, <https://doi.org/10.1007/s10546-016-0185-2>, 2017.
- 925 Yang, B., Berg, L. K., Qian, Y., Wang, C., Hou, Z., Liu, Y., Shin, H. H., Hong, S., and Pekour, M.: Parametric and Structural Sensitivities of Turbine-Height Wind Speeds in the Boundary Layer Parameterizations in the Weather Research and

Forecasting Model, *Journal of Geophysical Research: Atmospheres*, 124, 5951–5969, <https://doi.org/10.1029/2018JD029691>, 2019.

930

Yang, B., Guo, Z., Song, F., Zhang, Y., Zhou, T., and Qian, Y.: Fast and Slow Responses of Atmospheric Energy Budgets to Perturbed Cloud and Convection Processes in an Atmospheric Global Climate Model, *Geophysical Research Letters*, 50, e2023GL104305, <https://doi.org/10.1029/2023GL104305>, 2023.

935 Yarger, D., Wagman, B. M., Chowdhary, K., and Shand, L.: Autocalibration of the E3SM Version 2 Atmosphere Model Using a PCA-Based Surrogate for Spatial Fields, *Journal of Advances in Modeling Earth Systems*, 16, e2023MS003961, <https://doi.org/10.1029/2023MS003961>, 2024.

940 Yokohata, T., Webb, M. J., Collins, M., Williams, K. D., Yoshimori, M., Hargreaves, J. C., and Annan, J. D.: Structural Similarities and Differences in Climate Responses to CO₂ Increase between Two Perturbed Physics Ensembles,

Yoshioka, M., Regayre, L. A., Pringle, K. J., Johnson, J. S., Mann, G. W., Partridge, D. G., Sexton, D. M. H., Lister, G. M. S., Schutgens, N., Stier, P., Kipling, Z., Bellouin, N., Browse, J., Booth, B. B. B., Johnson, C. E., Johnson, B., Mollard, J. D. P., Lee, L., and Carslaw, K. S.: Ensembles of Global Climate Model Variants Designed for the Quantification and Constraint of
945 Uncertainty in Aerosols and Their Radiative Forcing, *Journal of Advances in Modeling Earth Systems*, 11, 3728–3754, <https://doi.org/10.1029/2019MS001628>, 2019.

Zhang, H., Wang, M., Guo, Z., Zhou, C., Zhou, T., Qian, Y., Larson, V. E., Ghan, S., Ovchinnikov, M., Bogenschutz, P. A., and Gettelman, A.: Low-Cloud Feedback in CAM5-CLUBB: Physical Mechanisms and Parameter Sensitivity Analysis,
950 *Journal of Advances in Modeling Earth Systems*, 10, 2844–2864, <https://doi.org/10.1029/2018MS001423>, 2018.

Zhang, M., Bretherton, C. S., Blossey, P. N., Austin, P. H., Bacmeister, J. T., Bony, S., Brient, F., Cheedela, S. K., Cheng, A., Del Genio, A. D., De Roode, S. R., Endo, S., Franklin, C. N., Golaz, J.-C., Hannay, C., Heus, T., Isotta, F. A., Dufresne, J.-L., Kang, I.-S., Kawai, H., Köhler, M., Larson, V. E., Liu, Y., Lock, A. P., Lohmann, U., Khairoutdinov, M. F., Molod, A. M.,
955 Neggers, R. A. J., Rasch, P., Sandu, I., Senkbeil, R., Siebesma, A. P., Siegenthaler-Le Drian, C., Stevens, B., Suarez, M. J., Xu, K.-M., von Salzen, K., Webb, M. J., Wolf, A., and Zhao, M.: CGILS: Results from the first phase of an international project to understand the physical mechanisms of low cloud feedbacks in single column models, *Journal of Advances in Modeling Earth Systems*, 5, 826–842, <https://doi.org/10.1002/2013MS000246>, 2013.

960 Zhang, X., He, B., Guo, Z., Sexton, D. M. H., Rostron, J. W., and Furtado, K.: Sensitivities of the Asian Summer Monsoon Simulations to Physical Parameters for the Perturbed Parameter Ensemble of HadGEM3-GC3.05, *Geophysical Research Letters*, 50, e2022GL101826, <https://doi.org/10.1029/2022GL101826>, 2023.