Using a Gaussian Process Emulator to approximate short-term climate response patterns to greenhouse gas and aerosol forcings.

Laura A. Mansfield[1,2,3*], Peer J. Nowack[4,5,1,], Edmund M. Ryan[6], Oliver Wild[7], and Apostolos Voulgarakis[1,8,9]

[1]Department of Physics, Imperial College London, United Kingdom.
[2]School of Mathematics and Statistics, University of Reading, United Kingdom.
[3]Atmospheric, Oceanic and Planetary Physics, University of Oxford, United Kingdom.
[4]Climatic Research Unit, School of Environmental Sciences, University of East Anglia, United Kingdom.
[5]Institute of Theoretical Informatics and Institute of Meteorology and Climate Research (IMKASF), Karlsruhe Institute of Technology, Germany.
[6]AWE Aldermaston, Reading, Berkshire, UK
[7]Lancaster Environment Centre, Lancaster University, Lancaster, United Kingdom.
[8]Leverhulme Centre for Wildfires, Environment and Society, Imperial College London, United Kingdom.
[9]School of Chemical and Environmental Engineering, Technical University of Crete, Greece

Correspondence to: Laura A. Mansfield (laura.mansfield@physics.ox.ac.uk)

Abstract. We present a Gaussian process emulator for estimating fast surface temperature response patterns to a range of different climate forcing agents, including both long-lived greenhouse gases and short-lived pollutants such as aerosols. This emulator is trained on simulations driven by perturbations to emissions (for short-lived pollutants) and concentrations (for long-lived greenhouse gases) using a full-complexity global climate model and predicts the response in the first five years after the forcing, at a small fraction of the computational cost. We outline the emulator design, including the choice of pollutant perturbations and the input space covered by the training data. We show that the emulator performs well in most regions of the chosen input space, except under very large aerosol perturbations. A global sensitivity analysis is carried out to characterize and understand emission-response relationships for each pollutant. We find similar large-scale patterns of sensitivity to aerosol pollutants released in different regions. Our work demonstrates the potential for rapid climate change projections subject to various climate forcing agents, focusing on the short-term response. This paves the way for the future development of more policy-relevant emulators, that can target decadal and multi-decadal timescales.

1 Introduction

Climate change projections are essential to inform society and policymakers about possible scenario-dependent impacts. To estimate the state of the future climate, we rely on Global Climate Models (GCMs) to run numerical simulations of the Earth system, including atmospheric, oceanic, land and sea-ice components and their interactions. These GCMs project the climate at a time in the future, given the current climate state and an emissions scenario that describes how the emissions/concentrations of greenhouse gases, aerosols and other atmospheric constituents are expected to evolve. Since future emissions are highly uncertain, GCMs must be run for a range of different possible scenarios. The output from a GCM can add to our understanding of the individual and net effects of different anthropogenic forcings on the climate system, which has long been of interest to climate scientists. Furthermore, understanding the climate response to emissions can inform important policy decisions.

The high complexity and computational cost of climate and atmospheric modelling puts a limit on the number and range of climate projection studies that can be carried out. This has led to a recent increase in application of surrogate models that

reproduce climate relevant variables given known inputs, at a much lower cost (Bolton & Zanna, 2019; Castruccio et al., 2014; Kaltenborn et al., 2023; Mansfield et al., 2020; Nowack et al., 2018; Rasp et al., 2018; Ryan et al., 2018; Watson-Parris, 2021; Wild et al., 2020). Here, we use a type of surrogate model called a *Gaussian process (GP) emulator*, which have become popular for aiding atmospheric and climate model development, through sensitivity analyses, uncertainty quantification and calibration (Carslaw et al., 2013; Lee et al., 2012; Ryan et al., 2018; Salter & Williamson, 2016; Williamson et al., 2015).

Climate model emulators are beginning to emerge with the primary aim of climate change projection. One approach currently gaining traction involves learning an *autoregressive model of the full climate state*, i.e., by predicting all climate variables at one timestep, given all climate variables at the previous timestep (e.g., Boussard et al., 2023; Cachay et al., 2024; Guan et al., 2024; Watt-Meyer et al., 2023, 2025). These require large-scale neural networks to capture all climate model variables and often suffer from instability when implemented over long timescales. These studies are also limited to only exploring a few forcings that have been included during training, such as through sea surface temperature forcing or CO_2 forcing (Watt-Meyer et al., 2025), and are yet to explore a wide range of climate change responses to different scenarios. In contrast, an alternative approach is to directly *learn the climate model response to a change in forcing*, often referred to as 'reduced complexity models' (Nicholls et al., 2020). This approach leverages simpler machine learning models to emulate the change in the climate state given a change in forcing, typically focusing on the response to different CO_2 scenarios (Bao et al., 2016; Castruccio et al., 2014; Foley et al., 2016). Most past emulation studies estimate the *global mean* response to forcings over a timeseries (Castruccio et al., 2014; Meinshausen et al., 2011; Smith, Forster, et al., 2018), with some extending this to include spatial patterns (Bao et al., 2016; Beusch et al., 2022; Dewey et al., 2025; Nath et al., 2022). The ability to predict climate response to changing forcings with quick, simple models is particularly valuable for policy-makers as they can be embedded within Integrated Assessment Models (IAMs, e.g., Foley et al., 2016) or can inform climate change scenario pathways (e.g., Durack et al., 2025; Forster et al., 2021).

In this paper, we present an emulator based on the HadGEM3 GCM, that predicts the short-term global surface temperature response to a wide range of forcing types, including regional aerosol emissions and global greenhouse gas concentration changes. This is the fast adjustment of the surface temperature in the first 5-years following the perturbation. This fast adjustment emulator predicts the response to an abrupt change in emissions, such as the simulations run in idealized forcing-response studies (e.g., Myhre et al., 2017; Richardson et al., 2019; Stohl et al., 2015; Samset et al., 2018; Smith et al., 2018). The emulator is designed to predict the entire spatial pattern of the global climate change response, rather than just the global mean, which could prove useful in future policy and impact studies. Our approach lays the foundations for the development of long-term climate response emulators which would be valuable in informing policy. Furthermore, we propose that this type of emulator could be coupled to other emulators designed to predict across different timescales. For instance, it could be combined with the emulator developed in Mansfield et al. (2020), which predicts the long-term temperature response given the short-term temperature response. Additionally, while such a fast-response emulator is not directly applicable for long-term

policy design, it can have real-world use for predicting regional and global temperature responses to short-term forcings that are triggered by episodic pollution sources such as wildfires or sudden anthropogenic pollution decreases (e.g. as during the COVID-19 pandemic).

In Section 2, we describe the emulator design including the choices of input variables and perturbation ranges. Section 3 considers the performance of the trained emulator on a set of test scenarios, on both regional and spatially resolved scales. In Section 4, the emulator is used to carry out a global sensitivity analysis to explore the main effects of each perturbation and the sensitivity of each region to local and remote emission perturbations. Section 5 makes the concluding remarks on the study.

2 Emulator Design

2.1 Global Climate Model Configuration

The GP emulator is trained on simulations from the HadGEM3 model, in the third Global Coupled configuration 3.1 (GC3.1) (K. D. Williams et al., 2018) which consists of the global atmosphere-land (GA7/GL7.1) (Walters et al., 2019), global ocean (GO6) (Storkey et al., 2018) and sea-ice (GSI8.1) (Ridley et al., 2018). This configuration uses the modal version of the Global Model of Aerosol Processes (GLOMAP-mode), two-moment aerosol microphysics scheme (Mann et al., 2010). The model resolution is 1.25° latitude by 1.875° longitude, giving a grid cell of roughly 208 km by 139 km at the Equator and a total of $144 \times 192 = 27{,}648$ grid cells. The same model configuration was employed for historical simulations in (Andrews et al., 2020), which formed part of the UK contribution to the sixth Coupled Model Intercomparison Project (CMIP6). `Present-day' conditions are taken from the end of these simulations, and correspond to the year 2014. The present-day simulation was run for 300 years as `spin-up', to reach an approximate equilibrium. The last 40-years of this simulation is treated as a control run and this was then used as the starting point for simulations that serve as training data for the emulator.

Before computing any perturbation runs, an additional six 5-year-long control simulations were performed starting from this equilibrium state, giving a total of 70-years of control simulations. We estimate the control surface temperature, by averaging over all years and all control simulations $\overline{\boldsymbol{T}}^{CTRL}$. All perturbation runs are initialized from the same state and also run for 5 years. The well-mixed GHGs are perturbed via their global concentrations, while for the short-lived pollutants we use scaling factors for emissions over broad regions, described further in Section 2.2. We define these as inputs to the GCM, denoted $\boldsymbol{x}^p$ for perturbation p. We estimate the perturbation surface temperatures by averaging over the 5 years and we denote this as $f_{GCM}(\boldsymbol{x}^p)$. We define the short-term surface temperature 'response' as

$$\boldsymbol{y}^p = f_{GCM}(\boldsymbol{x}^p) - \overline{\boldsymbol{T}}^{CTRL}$$

The surface temperature response is defined at every grid cell of the GCM, giving a complete spatial map, i.e., $\boldsymbol{y} = (y_1, y_2, \cdots, y_N)$ for all $N = 27{,}648$ grid cells.

2.2 Gaussian process emulator

Our perturbation simulations create $P = 80$ input-output pairs which we call our training data $D = (\boldsymbol{x}, \boldsymbol{y})$. We use a GP emulator to estimate the surface temperature response to a new emissions scenario, $\boldsymbol{x}^*$. The GP provides a probability distribution for each $\boldsymbol{x}^*$, from which we can sample, i.e.,

$$\boldsymbol{y}_{GP} \sim GP(\boldsymbol{x}^* | D)$$

This means we can also estimate the mean μ_{GP} and uncertainty σ_{GP}^2 for each $\boldsymbol{x}^*$. Note that there is internal variability overlaying this (i.e., noise), which makes weaker responses more difficult to predict. While we do not attempt to model this noise explicitly, we take advantage of the GP emulator's ability to estimate uncertainty about the prediction and assume that a portion of this uncertainty is due to the internal variability of the GCM. Note that $\boldsymbol{y}$ and $\boldsymbol{y}_{GP}$ are vectors containing the estimated surface temperature response at each grid cell. Here, we build separate GP emulators for each grid cell. We learn the function, f_{GP}, by training the GP emulator with input-output pairs $(\boldsymbol{x}^p, \boldsymbol{y}^p)$ obtained from GCM simulations. In the following section, we describe how input values were chosen, including the choice of pollutants that are varied and the range of possible values they can take.

2.3 Choice of inputs and ranges

Due to the high cost of running the GCM, we aim to minimize the number of simulations required to build and test this emulator. This can be done with a careful choice of inputs and a space-filling design of these inputs (Currin et al., 1991). First, we limit the number of possible inputs to the dominant climate forcers. These include well-mixed climate pollutants, specifically the two most important anthropogenic greenhouse gases carbon dioxide (CO_2) and methane (CH_4) (Stocker, 2013). For the short-lived pollutants, we perturb sulfur dioxide (SO_2, the precursor to sulfate aerosol (SO_4)) and biomass burning (BB) carbonaceous aerosols (black carbon (BC) and organic carbon (OC)). Sulfate aerosol is chosen because it has a strong cooling impact on the climate, particularly at a regional level, and because its precursor is emitted mostly from fossil fuel burning, making it highly relevant for scenario projection studies (Stocker, 2013). BB OC/BC are chosen as they strongly affect the climate through scattering and absorption of radiation and due to their dominance in the tropics (Boucher et al., 2013; Haywood & Boucher, 2000). In addition, BB is a particularly high source of scenario uncertainty, but has not been studied as extensively as other forcing perturbations (Lasslop et al., 2019; Tosca et al., 2013; Voulgarakis & Field, 2015; Ward et al., 2012). While other pollutants (e.g., NOx, NMVOCs, OC/BC from other sources) are also relevant for climate change, we keep these fixed at present-day levels because even for large perturbations, the effects are too small to exceed the regional climate

noise in any of our simulations and thus are less useful for training the emulator (Aamaas et al., 2019; Baker et al., 2015; ECLIPSE, 2014; D. Shindell & Faluvegi, 2009; Stohl et al., 2015).

For the greenhouse gases, we apply global perturbations to their concentrations in the GCM as they have long lifetimes and relatively homogeneous spatial forcing. For the short-lived pollutants, we perturb the emissions fields for the GCM. As the region of emission is paramount when considering the effect of short-lived pollutants, these perturbations are made over regional scales. We select broad continental regions in which short-lived emissions are scaled up or down (e.g., 2x, 0.5x), while keeping the spatial distribution of emissions within each region unchanged.

Following the procedure used in previous studies, we perturb SO_2 in North America, Europe, East Asia and South Asia (the same regions as (Kasoar et al., 2018; Lewinschal et al., 2019; Westervelt et al., 2020)), by scaling the regional present-day distribution of emissions. This approach allows for the entire region to be represented by a single scaling factor, but restricts the emulator to scenarios where the spatial distribution of emissions within each region does not change significantly. In other words, it does not account for the possibility of new emission hotspots within each region. This assumption is made because in these four regions, SO_2 emissions are predicted to decline in the future (Westervelt et al., 2020, Lund et al., 2019).

We further extend our study by perturbing SO_2 in Africa and South America, which currently have low levels of SO_2 relative to the rest of the world but are likely to see increasing emissions in the coming decades, due to changes in land use and industrialization in South America (Popp et al., 2017) and rapid growth of cities in Africa (Liousse et al., 2014). For these regions, the assumption of a constant spatial distribution of emissions may not be valid, as present day SO_2 emissions are low and originate from isolated hotspots (see Supplementary Fig. S1). We therefore use the emission distribution from the SSP5 baseline scenario with rapid growth (Riahi et al., 2017) to represent future conditions, while keeping total emissions fixed. (Supplementary Fig. S1).

Finally, we perturb BB OC/BC in the tropics (between 24 °S and 24 °N), which covers tropical regions of Africa, South America and South East Asia, where biomass burning due to fires is the largest source of OC/BC (Cachier et al., 1989; Pechony & Shindell, 2010; Tosca et al., 2013; Van der Werf et al., 2003). We perturb OC and BC in the same proportions, assuming that their ratio remains constant, following other studies (Hodnebrog et al., 2016; Tosca et al., 2013). This allows the emulator to be used to explore future scenarios under different levels of tropical biomass burning, whether it be due to anthropogenic suppression/ignition, a consequence of a warmer planet (e.g., alongside increasing GHGs) or under natural variability (Pechony & Shindell, 2010). We note that by using tropics-wide perturbations, the emulator may not capture highly regionalised forcing-response patterns (Westervelt et al., 2020; Williams et al., 2022). However, we do not consider OC/BC on smaller regional scales because previous studies have found that large, unrealistic OC/BC forcings are required to obtain a climate responses that exceeds the internal climate noise (e.g., Myhre et al., 2017; Baker et al., 2015; Stohl et al., 2015).

This gives a total of nine inputs, two of which are the global GHG concentrations and seven of which are regional scaling factors of short-lived pollutant emissions, as shown in Table 1. For each input, we define an input space that encompasses the range of possible scenarios of interest to policy-makers and climate scientists, based on the minimum and maximum plausible values estimated by previous studies, which include projections of the Shared Socioeconomic Pathways (SSPs) (Riahi et al., 2017; van Vuuren et al., 2017) and Evaluating the Climate and air quality Impacts of short-lived pollutants (ECLIPSE) project (Aamaas et al., 2017; Baker et al. 2015; ECLIPSE, 2014; Stohl et al., 2015), as well as other individual studies (Liousse et al., 2014; Pechony & Shindell, 2010). For the greenhouse gases, we take the pre-industrial levels as the minimum value and for the short-lived pollutants we take the minimum value as a scaling factor of zero, representing near-idealized anthropogenic pollution-free conditions. For all forcings, we determined the maximum levels to be the worst-case estimate from both future projections and historical scenarios (up to pre-industrial) in previous studies. Table 1 shows the ranges selected for each input and the reasoning behind the maximum values. Note that these `maximum' scaling factors do not reflect realistic scenarios, but are chosen as an appropriate upper limit for scenarios that one may wish to emulate.

The fact that we apply regional perturbations to the short-lived pollutants inevitably leads to small changes in abundances and weak forcings. This can give weaker grid-scale climate response signals that are difficult to separate from the internal noise (Kasoar et al., 2018; Shawki et al., 2018). For this reason, we extend the input space beyond the ranges described in Table 1 for these forcers, in order to create an emulator that gives less uncertain results. We expect that including simulations with stronger perturbations, which should have larger signal-to-noise ratios, could be beneficial for constraining the emulator prediction to these perturbations. We therefore extend the input space for the short-lived pollutants only so that 75% of simulations fall within the input range indicated in Table 1 and the remaining 25% lie outside this but within a range that is double this (see Supplementary Fig. S2). This makes our emulator more capable of capturing responses to large idealized perturbations in short-lived pollutants, which are common in scientific studies exploring the physical processes linking short-lived pollutant forcing and response (Liu et al., 2018; Myhre et al., 2017; Richardson et al., 2019).

Table 1. All input and the ranges used in the emulator, with associated reasoning for maximum value. The minimum values are taken to be pre-industrial concentrations for CO_2 and CH_4 and a scaling factor of zero for all short-lived pollutants.

Input	Range	Reasoning
Global CO_2 concentration	282 - 834 ppm	SSP3 baseline projection in 2100 (Riahi et al., 2017)
Global CH_4 concentration	248 - 3238 ppb	SSP3 baseline projection in 2100 (Riahi et al., 2017)
SO_2 over North America	0 – 3x	Historical data from 1990 (ECLIPSE, 2014)
SO_2 over Europe	0 – 5x	Historical data from 1990 (ECLIPSE, 2014)
SO_2 over East Asia	0 – 2x	Historical data from 2005 (ECLIPSE, 2014) and SSP1 projection in 2100 (Riahi et al., 2017)
SO_2 over South Asia	0 – 3x	ECLIPSE projection in 2030 (ECLIPSE, 2014)
SO_2 over Africa	0 – 7x	Inventory based projection in 2030 (Liousse et al., 2014)
SO_2 over South America	0 – 3x	SSP3 baseline projection in 2080 (Riahi et al., 2017)
BB OC/BC over Tropics	0 – 2x	Estimates of fire activity by 2100 (Pechony & Shindell, 2010)

2.4 Design of training data

We design a set of training simulations by selecting combinations of input values that will be inputs to the GCM. The best approach to do this is to generate sets of input values that are space-filling (e.g., Currin et al., 1991; Lee et al., 2016; O'Hagan, 1978; Ryan et al., 2018; Santner et al., 2003; Wild et al., 2020). This minimizes the number of samples needed to cover the input space efficiently, which is particularly important for this study as model simulations are computationally expensive (around 80 node hours per simulation on a Cray XC40 (Met Office, 2021)). An informal rule of thumb is that the number of runs for training should be around 10 times the number of input dimensions (Loeppky et al., 2009), and this approach has worked well in previous climate model emulators (e.g., Carslaw et al., 2013; Lee et al., 2012; Ryan et al., 2018). Given the nine inputs of interest in this study and the high computational cost of the simulations, we perform 80 simulations for training. These 80 training points are selected to be space-filling by using a maximin Latin Hypercube Sampler (LHS), which subdivides the input space evenly and samples each partition so that the distance between samples is maximized (e.g., Lee et al., 2011, 2012; Ryan et al., 2018). To generate the samples with the distribution described above, we first sample 80 points using a maximin LHS for values between 0 and 1 for all inputs. For the GHGs, we use a linear transform to scale the input range so that all values fall within the defined bounds. For the short-lived pollutants, we use a linear transform to scale the first 75% of samples to lie within the ranges defined in Table 1. We then scale the remaining 25% of samples to fall between 1 and 2X the maximum values in Table 1. This provides us with samples that are further separated for the larger, unrealistic perturbations

values, but more closely sampled in the region we are most interested in. The samples are shown in Supplementary Fig. S2. We also perform an additional 18 simulations for testing, which are sampled separately from the ranges in Table 1 using a standard LHS design.

225

2.5 Gaussian process emulator details

With the 80 training simulations available, we build an emulator to project the short-term surface temperature response based on the nine inputs that control the CO_2 and CH_4 concentrations and the short-lived emission scaling factors described above
230 (Table 1). Prior to building the emulator, the inputs and outputs are scaled to be centered around 0 with standard deviation of 1. Separate GP emulators are then applied to estimate the surface temperature response for each grid-cell independently. For each grid cell, the GP emulator is

$$f_{GP}(\boldsymbol{x}) = \mathcal{G}\big(m(\boldsymbol{x}), K(\boldsymbol{x}, \boldsymbol{x}')\big)$$

where $m(\boldsymbol{x})$ is the prior mean function and $K(\boldsymbol{x}, \boldsymbol{x}')$ is the prior covariance function or "kernel" that describes the covariance
235 between two inputs, $\boldsymbol{x}$ and $\boldsymbol{x}'$. The choice of $m(\boldsymbol{x})$ and $K(\boldsymbol{x}, \boldsymbol{x}')$ controls the structure of the GP emulator (Rasmussen & Williams, 2006). The prior mean function is often assumed to be zero so that all choices are determined by the covariance function, as done here. The kernel defines the similarity of two inputs $\boldsymbol{x}$ and $\boldsymbol{x}'$ and how this propagates through to the similarity of the outputs $f_{GP}(\boldsymbol{x})$ and $f_{GP}(\boldsymbol{x}')$. We use

$$K(\boldsymbol{x}, \boldsymbol{x}') = \sum_{d=1}^{9} \sigma_{Ld}^2 \, x_d \, x_d' + \sigma_{SE}^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{9} \frac{(x_d - x_d')^2}{l_d^2}\right) + \sigma_W^2$$

240

This describes an additive kernel consisting of three components. The first term represents a non-stationary linear kernel which learns independent hyperparameters, σ_{Ld}^2 , for each input parameter, d (known as Automatic Relevance Determination, ARD), which represent the rate of change of the output for each input parameter. This captures the first-order linear relationships between emission perturbations and temperature response, while the remaining terms aim to capture the local variability (Craig
245 et al., 1996; Edwards et al., 2019; Vernon et al., 2010). The second term is a squared exponential kernel which also uses ARD to learns independent lengthscales l_d for each input parameter, d. These lengthscales describe how rapidly the output changes with respect to varying input parameters. The hyperparameter σ_{SE}^2 is a scale that describes the total amount of variation in the output (Rasmussen & Williams, 2006). The last term is a white noise kernel which represents the internal variability of the GCM, assumed to be constant σ_W^2, for all predictions. We explored both fixing σ_W^2 at the internal variability derived from the
250 GCM (described further below and used throughout the paper unless stated otherwise), and allowing σ_W^2 to be learned as an additional hyperparameter (described in Section 3.3). We chose this kernel function because of its interpretability, however, we found the emulator to be robust to a range of kernel choices (Supplementary Fig. S3). We learn all kernel hyperparameters through maximum likelihood estimation, using the python library, GPy (GPy, 2012). Here, we use independent emulators for

each of the 27,648 grid cells, however, we also found similar results with GP emulators applied to the top principal components, a dimension reduction technique that also enforces spatial correlations between grid-cells (e.g., Salter et al., 2019; Salter & Williamson, 2022; Wilkinson, 2010). We also tested different methods such as random forests but we did not find improvements in performance, so we focus only on GP emulators here.

In this study, we consider two types of uncertainty. Firstly, we view the uncertainty associated with the chaotic nature of the Earth system, the aleatoric uncertainty (Shepherd, et al., 2018), as a limitation of the GCM prediction, and secondly, we diagnose the uncertainty due to the emulator prediction being an approximation of the GCM at new unseen scenarios not within the training set (which we call the predictive uncertainty). As a GP emulator is a Bayesian method and predicts a distribution of responses, the latter is a natural output of each prediction. Note that we do not consider the uncertainty caused by the difference in the GCM predictions and the truth, i.e., epistemic uncertainty or model inadequacy/discrepancy.

The uncertainty due the chaotic nature of the GCM is the internal variability, σ_{GCM}, and is estimated as the standard deviation between as many 5-year control runs as available. We have available 8 segments from one long, 40-year control run after equilibrium had been reached in the spin-up simulation, and 6 independent 5-year control runs that branched off from this. Ideally, all control runs would be strictly independent but the 8 segments of the longer control run are included to improve the sample size for estimating the standard deviation. The standard deviation across the 14 control runs is calculated for each grid point independently, giving a map of uncertainty associated with the internal noise. This is an uncertainty associated with any prediction from the GCM, and is partly due to the limited number of years we have averaged over. We expect this to be present for any prediction and therefore, we include this as a fixed noise term in our emulator, by setting $\sigma_W^2=\sigma_{\mathrm{GCM}}{}^2$ for the white noise covariance function of the GP emulator. When the emulator predicts the temperature response at new unseen input values, there is also predictive uncertainty provided by the GP emulator, which we label σ_{GP}. Note that this comes from the other covariance functions and is dependent on the inputs, x. The total uncertainty arising from both the GCM internal variability and the GP emulator prediction is denoted σ_{total} and these are related via $\sigma_{\mathrm{total}}{}^2 = \sigma_{\mathrm{GCM}}{}^2 + \sigma_{\mathrm{GP}}{}^2$, discussed later in Section 3.3.

3 Emulator Performance

To test the performance of the emulator, we predict the output for 18 test simulations that have not been used for emulator training, where the input values have been generated with a Latin Hypercube Sampler, with a uniform distribution across all input ranges of interest (Table 1).

3.1 Global response predictions

Figure 1 shows an example of typical emulator performance, with the input values shown along the top panel, relative to the baseline, minimum and maximum possible values. Note that these perturbations are not intended to be realistic perturbations over a five-year period, but rather they estimate the short-term response to abrupt perturbations, which at this stage would be used for scientific experimentation, with the view to develop the ideas further in the future for more direct policy relevance. The maps show, from top to bottom, the GCM output for the test scenario, y_{test}, the GP prediction, y_{pred}, and the absolute errors, $|y_{pred} - y_{test}|$, with the stippling indicating regions where this exceeds 1 s.d. predicted by the GP emulator. These examples are representative of the test simulations which are presented in Supplementary Fig. S4. Most test simulations are reasonably well predicted by the emulator and show patterns of warming or cooling generally in the correct spatial location. The absolute differences are generally larger at high latitudes, particularly around the Arctic, Greenland and Northern Europe, for instance, in Figure 1a where the emulator does not capture the cooling over Europe due to the strong SO2 Europe perturbation. Although the magnitude of σ_{total} is slightly larger in these regions than in the tropics, the stippling where $|y_{\text{pred}} - y_{\text{test}}| > \sigma_{\text{total}}$ is more prominent in these high latitude areas, indicating that prediction is poorer there compared to other regions worldwide. These regions are highly correlated with the regions that experienced larger errors in our previous emulation study that found these to be areas that experience both increased sensitivity to forcings and increased variability (Mansfield et al., 2020). Statistically, we would expect ~32% of predicted points to fall outside the 1σ prediction level, which is generally the case across all test scenarios, with the exception of Supplementary Fig. S4k. However, the predicted points that fall outside this accuracy threshold have a specific regional pattern and occur more often in the Northern Hemisphere, rather than being distributed randomly across the globe.

For the test scenario in Supplementary Fig. S4k, the emulator predicts the correct sign of response in most regions, but the prediction is too weak in magnitude compared to the GCM. We believe this to be because the inputs are all strongly positive relative to present-day levels meaning this scenario falls close to the corner of the input space, where there are fewer training data points to constrain the prediction. The emulator overestimates the cooling effect of strong aerosol perturbations, highlighting that caution should be taken when interpreting emulator predictions on very strong aerosol perturbation scenarios. Future studies may wish to refine the accuracy of the emulator in this region of the input space, by running additional training simulations with even stronger aerosol perturbations. The remainder of this study does not focus on the strong aerosol perturbation regime since we consider more realistic emission perturbations. Overall, however, the test simulations indicate that the GP emulator can predict the response to a range of scenarios under these nine inputs. Based on all 18 test simulations, 76.1% of grid points are predicted to within 1 σ of the true response and 95.0% are predicted to within 2 σ, which is roughly in line with expectations, assuming a Gaussian distribution of responses.

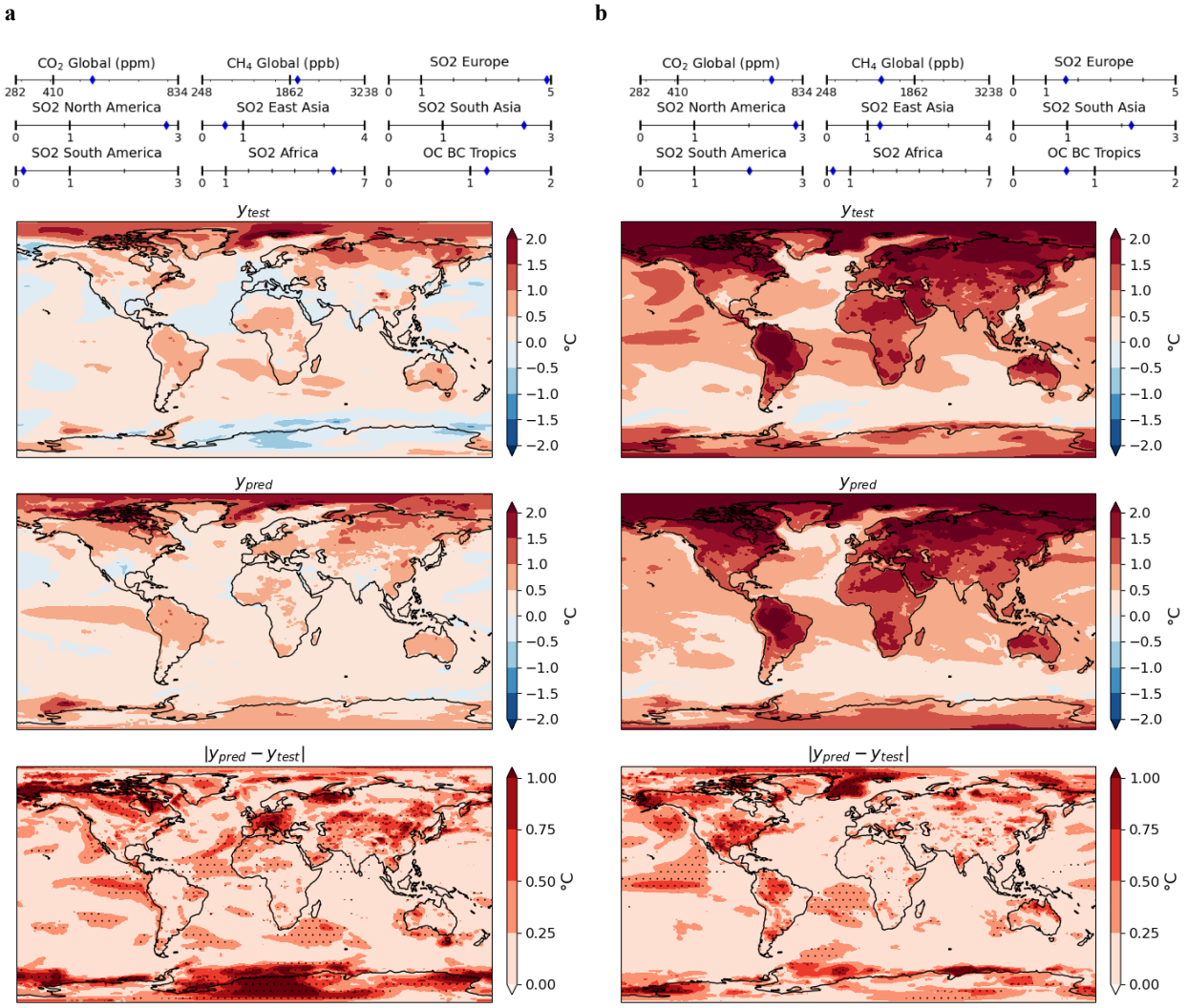


Figure 1. Two example test scenarios, where top panel shows inputs values, as well as the baseline, minimum and maximum possible values. The maps show, from top to bottom, the GCM output for the test scenario, y_{test} the GP prediction, y_{pred} and the absolute errors, $|y_{\text{pred}} - y_{\text{test}}|$, where in the latter the stippling indicates regions where this absolute error exceeds 1 s.d. predicted by the GP emulator, $|y_{\text{pred}} - y_{\text{test}}| > \sigma_{\text{total}}$.

3.2 Regional mean response predictions

We have also assessed the response for each region by taking broad averages over areas of interest. The response predicted by the GP emulator is plotted against the true GCM response for all test scenarios in Figure 2 where the error bars indicate the 1 σ uncertainty levels. A perfect prediction would correspond to the $y = x$ line (black dashed line), a coefficient of determination $R^2 = 1$ and a mean absolute difference (MAD) equal to zero. Most regional mean predictions fall very close to this line, or within 1 σ from it. The exception to this is the outlier with a positive true response of around 1°C in most regions, that is strongly underestimated, which corresponds to the test scenario in Supplementary Fig. S4k. Other than this scenario, the majority of responses are well predicted and cover a fairly broad range of responses including both cooling and warming of up to 1-2 °C (R^2 values are all above 0.87). These plots also highlight the increased uncertainty associated with some regions, namely North America and Europe, in comparison to the tropical regions such as South Asia, South America and Africa. As well has exhibiting increased uncertainties, these regions also show a wider range of temperature responses in the test data. This leads to slightly higher absolute errors (increased MAD values) in these regions, but overall each prediction is still close to the ground truth (GCM), as highlighted by the high R^2 values. This gives confidence to the use of an emulator in this way for quick predictions, for instance in policy-relevant studies.

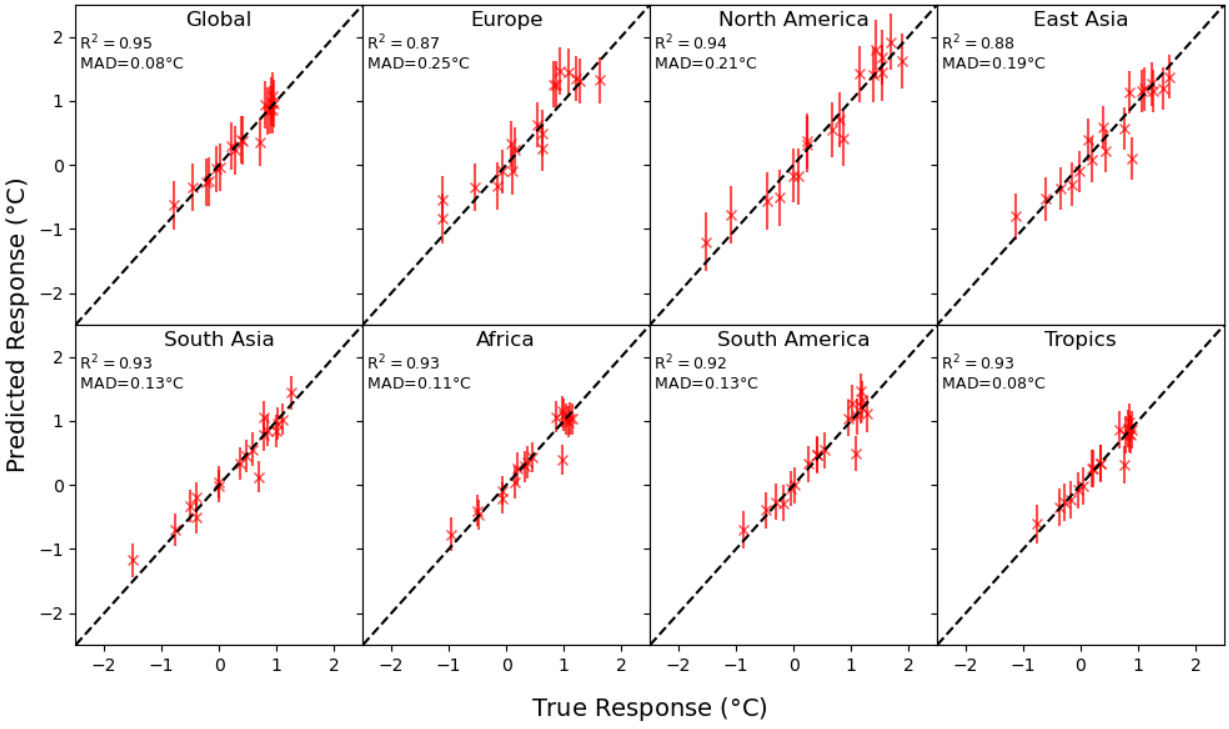


Figure 2. Regional mean predicted responses of test simulations against true GCM response, where error bars represent 1σ uncertainty level of Gaussian process prediction, for the main regions of interest, including the global mean. The coefficient of determination R^2 and the mean absolute difference (MAD) is also shown for each region.

3.3 Internal variability

We used the internal variability of the GCM as a fixed component of uncertainty when building the emulator where $\sigma_w^2 = \sigma_{GCM}^2$, in order to view it as a form of aleatoric uncertainty, in other words, caused by the unpredictability of the system and intrinsic to any GCM prediction. The internal variability is calculated over 14 control simulations of 5-year segments, and is shown in Supplementary Fig. S5a. The GP emulator provides an additional predictive uncertainty when predicting new scenarios, but this uncertainty is generally much smaller than the internal variability (Supplementary Fig. S5b). For comparison, we built an alternative version of the GP emulator, where rather than fixing the internal variability based on the GCM internal variability, $\sigma_w^2 = \sigma_{GCM}^2$, we treat the white noise covariance σ_w^2, as an additional hyperparameter to be optimized. This emulator performs virtually identically and reproduces almost exactly the same distribution of total uncertainty (Supplementary Fig. S5c,d). This result is encouraging, as it suggests that even without providing the GCM internal variability, the GP emulator is able to learn it as a form of uncertainty. This suggests that the accuracy of the emulator is limited primarily by GCM internal variability when predicting the short-term surface temperature response.

4 Global Sensitivity Analysis

The climate modelling community has benefited from a wide range of GP emulation studies concerned with uncertainty quantification and sensitivity analysis, which aim to characterize the uncertainty on the outputs and identify where it arises from. Due to the high computational cost of GCM simulations, GP emulators have been utilized to diagnose uncertainties and sensitivities in a wide range of GCM components and settings, including atmospheric chemistry models (Beddows et al., 2017; Lee et al., 2012; Stell et al., 2021; Wild et al., 2020), sea-ice models (Edwards et al., 2019; Urrego-Blanco et al., 2016), climate-vegetation models (Bounceur et al., 2015), convection parameterizations (Souza et al., 2020), and atmospheric gravity waves (King et al., 2024). The inputs to these are typically climate model parameters that are not known perfectly, where gaining an understanding of how sensitive the outputs are to each input parameter is beneficial for model development and calibration. In this study, the inputs are emission perturbations and we are primarily interested in exploring patterns of potential future climate responses to a range of scenarios. Sensitivity analysis can be used to deepen our understanding of the sensitivity of climate across all world regions to different competing emissions. In particular, we carry out a global sensitivity analysis, which quantifies sensitivities to each input averaging over the effects of all other inputs (Saltelli et al., 2019; Saltelli & Annoni, 2010).

4.1 Main Effects

Firstly, we explore the "main effect" of each pollutant, meaning the effect of each pollutant on the output independent of all other pollutants. This is done by using the emulator to predict surface temperatures across the full range of possible input values for the pollutant of interest, while randomly sampling all other inputs repeatedly. This generates a set of many independent realizations, where here we use 200 realizations for our sensitivity analysis. We average over all realizations to find the main effect attributed to the pollutant of interest. Here, we wish to assess the main effect of the pollutants relative to the present-day climate and therefore we sample all other inputs from independent normal distributions, centered at present day levels with standard deviations set to ¼× the input range presented in Table 1.

The mean behavior of each pollutant perturbation is calculated by averaging over the 200 different realizations and can be found in Supplementary Fig. S6 for each key region. The main effects of CO_2 are much stronger in all regions than the other pollutants. As expected, local aerosol perturbations give rise to stronger responses in the region of emission (particularly SO_2 released over Europe, East Asia and South Asia), and often the second most important contributor is an emission perturbation in an area that is closer to the area examined than others (e.g., Africa is sensitive to SO_2 from Europe, also noted in Dong et al., 2014; South Asia is influenced by East Asia, also see Shawki et al., 2018). Additionally, remote regions show reduced response, often near zero. We find that the main effect is approximately linear for all pollutants, both GHGs and aerosols, over the ranges explored here, which is intuitive under small forcing perturbations (Boer & Yu, 2003; Gregory et al., 2004; Hansen et al., 1997; Collins et al., 2013; Hansen et al., 2005; Richardson et al., 2019),. However, we would expect to see non-linearities in the response when extending to larger forcing perturbations due to the saturation of aerosol-cloud interactions (Carslaw et al., 2013; Kasoar et al., 2018), as discussed further in Supplementary Text S1. This likely explains the overestimation of cooling due to strong aerosol perturbations in test scenario in Supplementary Fig. S4k.

Figure 3 summarizes the main effect per unit Tg of emissions for the aerosol perturbations, in each emission region, estimated as the gradients in the main effects plots (Supplementary Fig. S6), with the error bars representing the 1 s.d. uncertainty propagated from the main effects. These demonstrate similar patterns to previous studies that explore the temperature change per unit Tg emissions, following an abrupt emission perturbation (Aamaas et al., 2017; Collins et al., 2013; Shindell & Faluvegi, 2009). For instance, both Aamaas et al. (2017) and Kasoar et al., (2018) find European SO_2 perturbations to consistently produce larger responses per unit Tg than East Asian SO_2. (Shindell & Faluvegi, 2009) also find that broad latitudinal SO_2 perturbations in the tropics give fairly similar responses globally, while Northern hemispheric perturbations responses are much stronger in the latitude band of emission relative to other regions. Figure 3 shows consistencies with this as the Northern hemispheric perturbations show much stronger regional effects in other Northern hemispheric response regions while tropical perturbations (SO_2 Africa, South Asia, and South America) show more evenly distributed responses after accounting for the localized effect. Note that here we focus only on the 5-year response, rather than 20 years or longer explored in other studies and also that we average over the effect of other pollutants, in contrast to previous studies that keep them fixed

at present-day levels based on simulations where one pollutant at a time is varied (Aamaas et al., 2017; Collins et al., 2013; Shindell & Faluvegi, 2009). This is also one of the first studies to explore the response to regional tropical perturbations (e.g., Wells et al., 2022; Westervelt et al., 2020) as well as extratropical (e.g., Persad & Caldeira, 2018; Westervelt et al., 2020), rather than broad latitudinal perturbations (e.g., Aamaas et al., 2017; Collins et al., 2013; Joshi et al., 2003;. Shindell & Faluvegi, 2009; D. T. Shindell, 2012). The main effects of the OC/BC tropical perturbation show more diversity between regions. There are local decreases in temperature over Africa, where the majority of the BB emissions are released. This is because, although the absorbing effect of BC aerosol increases temperatures globally (e.g., North America, South Asia), it reduces incident radiation to the surface, which is combined with the cooling effect of scattering from OC (Wells et al., 2022).

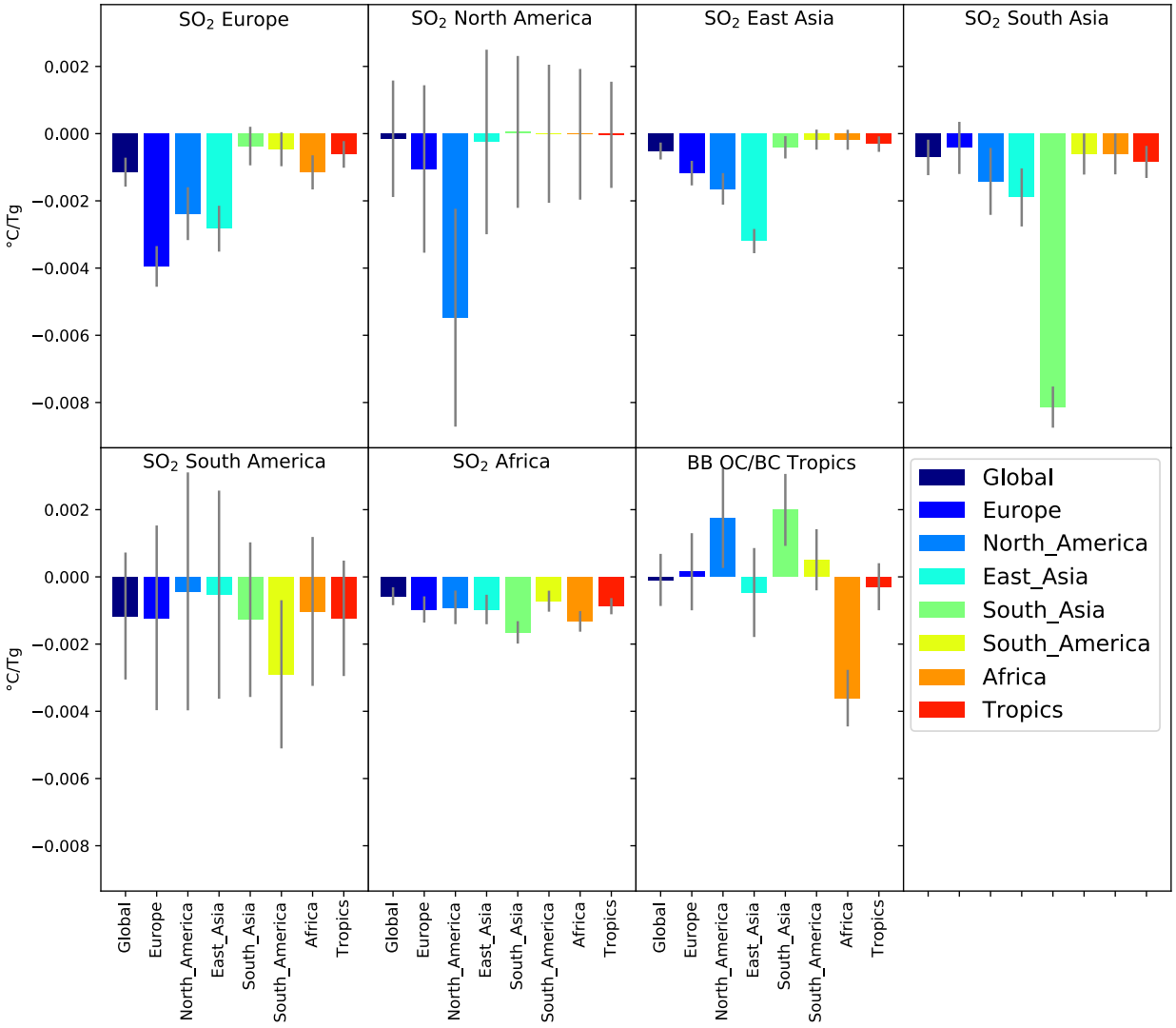


Figure 3. The expected change in temperature (°C) for every 1 Tg increase of each aerosol pollutant in each response region. The values are estimated from the gradients of the lines of each plot in Supplementary Fig. S6, based on the linear relationships found between the response in °C and the aerosol emission perturbation. The error bars represent 1σ uncertainty estimated by 200 different realizations.

4.2 Sensitivity Indices

We have also carried out a variance-based global sensitivity analysis, by estimating how much of the variance in the output can be attributed to the variance in the inputs via the Sobol method (Saltelli et al., 2008, 2010). This can be calculated for each output variable at each grid point. Figure 4a shows the global and regional means of first order sensitivity indices for CO_2, CH_4 and aerosols, with a more detailed breakdown for each aerosol pollutant in Figure 4b. Firstly, the variance due to the CO_2 concentration is responsible for the majority of the variance in the emulator (~90%) across all regions, which is not surprising given the wide input ranges covered by the emulator based on the high uncertainty of future CO_2 emissions. The aerosol contributions appear small relative to these large perturbations, and we expect this reflects the weaker range of aerosol perturbations in comparison to the large CO_2 perturbations. The first order sensitivity indices to CO_2, CH_4 and the aerosols sum to one almost exactly, leaving $< 10^{-4}$ down to higher order terms, indicating that there are only relatively weak interactions between the effects of the different inputs, described further in Supplementary Text S2. Figure 5 shows the sensitivity analyses for each grid-point, highlighting a fairly homogeneous spatial response for the greenhouse gases, but inhomogeneous spatial responses for the aerosol pollutants. Aside from the strong localized effect, the patterns of sensitivity to SO_2 from Europe and East Asia are remarkably similar; reminiscent of results from Kasoar et al. (2018), where removal of SO_2 from different Northern Hemisphere regions gave similar temperature response patterns. Additionally, the perturbations in tropical regions (SO_2 from South America, Africa and South Asia and OC/BC from the Tropics) give rise to similar distributions of sensitivity indices, with enhanced sensitivity over the tropical Atlantic and Pacific oceans. This mirrors the enhanced temperature sensitivities over the Northern hemispheric oceans due to the Northern hemispheric perturbations, an indication that the climate response is projected onto existing modes, in line with previous studies (Corti et al., 1999; Kasoar et al., 2018; Palmer, 1999; Ring & Plumb, 2008; Shindell et al., 1999).

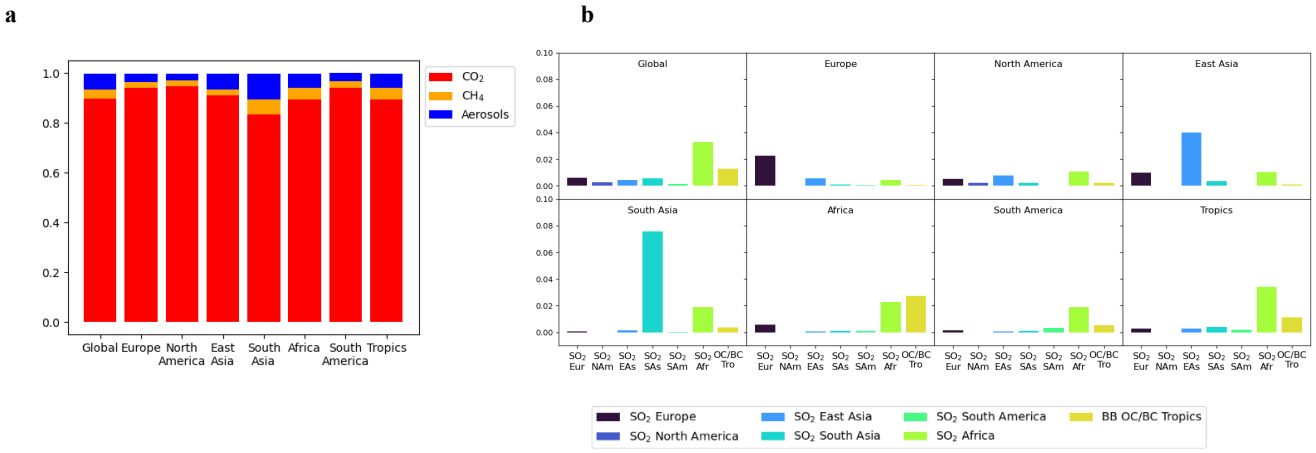


Figure 4. Regional mean first order sensitivity indices a) for CO_2, CH_4 and all aerosols in each region and b) breakdown of the aerosol sensitivity indices: the regional SO_2 perturbations and the tropical biomass burning aerosols (OC/BC).

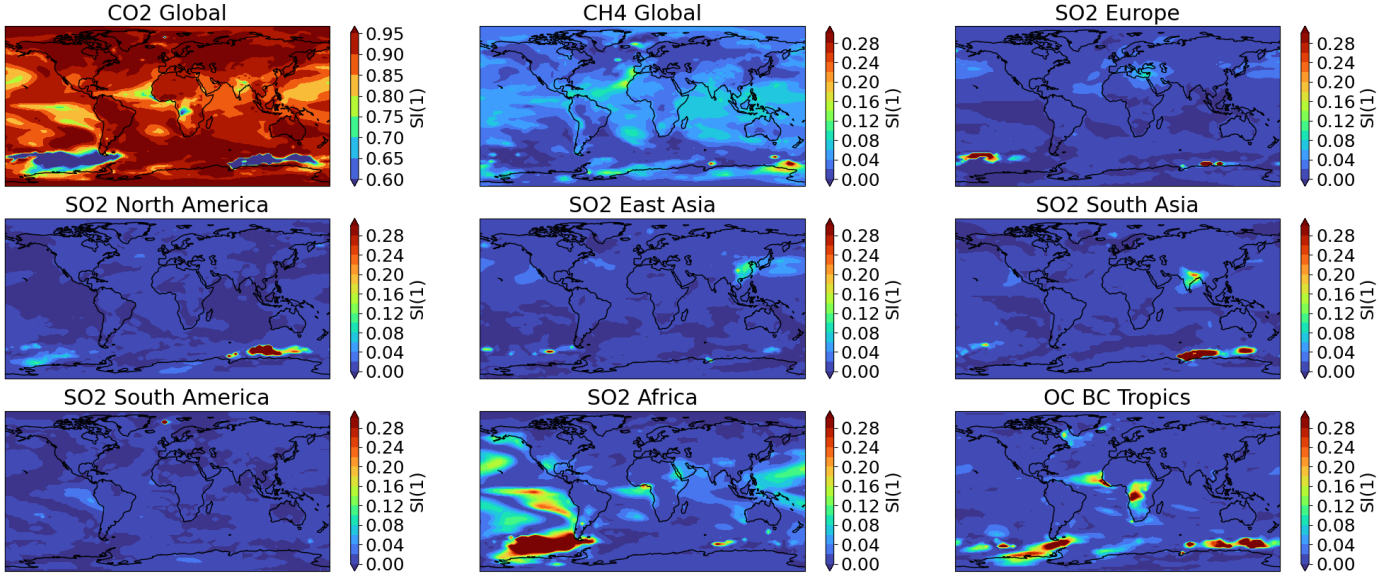


Figure 5. Global maps of first order sensitivity indices, SI(1), for each pollutant perturbation to emulator, where SI(1) is the fraction of contribution to the total sensitivity for each pollutant. Note the different color bar ranges on each subfigure.

5 Conclusions

We have presented a new Gaussian process emulator that estimates the *global pattern of short-term response* to a sudden change in emissions based on a carefully designed set of training data from the HadGEM3 coupled climate model. This emulator has the ability to predict the spatial pattern of responses, rather than just the global mean responses as previous studies have done (e.g., Castruccio et al., 2014; Meinshausen et al., 2011; Smith, Forster, et al., 2018). It predicts the short-term temperature response to a range of global and regional perturbations that include both long-lived greenhouse gases and short-lived aerosol pollutants.

The main source of uncertainty in the emulator predictions is the internal variability of the GCM. Since this is also a source of uncertainty in GCM runs on the same timescales, the choice of simulating climate response with an emulator is a sensible one. Naturally, one of the next steps would be to reduce this uncertainty, for example, simulations could be averaged over longer timescales (e.g., 10 or 20 years) to reduce the effect of year-to-year fluctuations, although this would require a substantial increase in runtime per simulation. Emulator prediction uncertainty could also be targeted by increasing the size of the training dataset. In addition, we assumed here that the internal variability is constant under different scenarios. In future work, it would be interesting to explore how well this assumption holds by estimating the internal variability for each training simulation independently, for example, by performing an ensemble of simulations each seeded with different initial conditions, for all 80 training simulations. Another extension of this work could involve building an additional GP emulator that explicitly models the internal variability (heteroscedasticity, e.g., Andrianakis et al., 2017; Binois et al., 2018; Gramacy, 2020). Another approach to reducing uncertainties could be develop multivariate GP emulators, which account for the spatial structure and may be able to share information between neighbouring grid cells (e.g., Conti & O'Hagan, 2010; Gu et al., 2019; Gu & Berger, 2016; Higdon et al., 2008; Rougier, 2009; Rougier et al., 2009).

The overarching goal of this study was to identify a framework for policy-relevant emulators. We made several choices when designing the training and test dataset using GCM simulations, from which future studies can learn. For instance, we chose a maximin Latin hypercube design to sample uniformly across the distribution of each parameter. We extended the parameter space for the aerosol forcings beyond their feasible ranges to account for the weaker aerosol signals compared to the greenhouse gas forcings. Still, we found that many simulations are dominated by the response to greenhouse gas forcings. Future studies may wish to compare the emulator performance under different ratios of extended parameters to realistic parameters. Alternative designs, such as selecting some simulations with restricted greenhouse gas perturbations, which lead to stronger aerosol signals could potentially improve performance in this range. Furthermore, Bayesian optimization techniques could be used to directly identify new parameter choices for additional GCM simulations that would reduce emulator error or uncertainty (Shahriari et al., 2016).

We also identify the opportunity of coupling this fast adjustment emulator with a long-term response emulator, such as that presented in Mansfield et al. (2020), which could lead to full long-term climate projections. Similarly, we see scope to extend on our work by generating similar emulators from the output of other GCMs, for example to characterize inter-model uncertainty in the relationships derived here. Finally, additional forcings, such as those from separate biomass burning aerosol components, ozone precursors and/or natural aerosols, could also be included in such an emulator. We thus highlight Gaussian process emulators as a useful method to accelerate and understand regional climate change projections under varying anthropogenic emissions scenarios, including not only responses to GHG forcings but also, importantly, to short-lived forcing agents such as aerosols.

Acknowledgments

There are no financial conflicts of interests for any author. LAM was funded through EPSRC grant EP/L016613/1. PJN was supported through an Imperial College Research Fellowship and UK Natural Environment Research Council (NERC) grant number NE/V012045/1. AV is partially funded by the Leverhulme Trust, grant RC-2018-023, AXA Research Fund (project 'AXA Chair in Wildfires and Climate') and by the Hellenic Foundation for Research and Innovation (Grant ID 3453). EMR and OW were partly supported through NERC grant NE/N003411/1. Simulations with HadGEM3 were performed using the MONSooN system, a collaborative facility supplied under the Joint Weather and Climate Research Programme, which is a strategic partnership between the Met Office and the Natural Environment Research Council. Analysis was carried out on Imperial College London's high performance computing cluster.

Code and Data Availability

Data was produced using HadGEM3, developed by the UK Met Office. Code and data used to train the emulator and produce all plots for this manuscript are available at https://doi.org/10.5281/zenodo.17814594 (Mansfield, 2025)

Author Contribution

Conceptualisation: AV, PJN, LAM. Supervision: AV, PJN. Formal analysis: LAM, PJN, AV. Software: LAM, Validation: LAM, PJN, EMR, OW, AV. Methodology: LAM, PJN, EMR, OW, AV. Writing (original draft): LAM. Writing (review and editing): LAM, PJN, EMR, OW, AV.

Competing Interests

The authors declare that they have no conflicts of interest.

References

Aamaas, B., Berntsen, T. K., Fuglestvedt, J. S., Shine, K. P., & Collins, W. J. (2017). Regional temperature change potentials for short-lived climate forcers based on radiative forcing from multiple models. *Atmospheric Chemistry and Physics*, *17*(17), 10795–10809. https://doi.org/10.5194/acp-17-10795-2017

Aamaas, B., Berntsen, T. K., & Samset, B. H. (2019). The regional temperature implications of strong air quality measures. *Atmospheric Chemistry and Physics*, *19*(24), 15235–15245. https://doi.org/10.5194/acp-19-15235-2019

Andrews, M. B., Ridley, J. K., Wood, R. A., Andrews, T., Blockley, E. W., Booth, B., Burke, E., Dittus, A. J., Florek, P., Gray, L. J., Haddad, S., Hardiman, S. C., Hermanson, L., Hodson, D., Hogan, E., Jones, G. S., Knight, J. R., Kuhlbrodt, T., Misios, S., … Sutton, R. T. (2020). Historical Simulations With HadGEM3-GC3.1 for CMIP6. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1029/2019MS001995

Andrianakis, I., Vernon, I., McCreesh, N., McKinley, T. J., Oakley, J. E., Nsubuga, R. N., Goldstein, M., & White, R. G. (2017). History matching of a complex epidemiological model of human immunodeficiency virus transmission by using variance emulation. *Journal of the Royal Statistical Society. Series C, Applied Statistics*, *66*(4), 717–740. https://doi.org/10.1111/rssc.12198

Baker, L. H., Collins, W. J., Olivié, D. J. L., Cherian, R., Hodnebrog, Ø., Myhre, G., & Quaas, J. (2015). Climate responses to anthropogenic emissions of short-lived climate pollutants. *Atmospheric Chemistry and Physics*, *15*(14), 8201–8216. https://doi.org/10.5194/acp-15-8201-2015

Baker, L. H., Collins, W. J., Olivié, D. J. L., Cherian, R., Hodnebrog, Ø., Myhre, G., Quaas, J., & Samset, B. H. (2015). *Climate responses to anthropogenic emissions of short-lived climate pollutants* [Preprint]. Aerosols/Atmospheric Modelling/Troposphere/Physics (physical properties and processes). https://doi.org/10.5194/acpd-15-3823-2015

Bao, J., McInerney, D. J., & Stein, M. L. (2016). A spatial-dependent model for climate emulation. *Environmetrics*, *27*(7), 396–408. https://doi.org/10.1002/env.2412

Beddows, A. V., Kitwiroon, N., Williams, M. L., & Beevers, S. D. (2017). Emulation and Sensitivity Analysis of the Community Multiscale Air Quality Model for a UK Ozone Pollution Episode. *Environmental Science and Technology*. https://doi.org/10.1021/acs.est.6b05873

Beusch, L., Nicholls, Z., Gudmundsson, L., Hauser, M., Meinshausen, M., & Seneviratne, S. I. (2022). From emission scenarios to spatially resolved projections with a chain of computationally efficient emulators: Coupling of MAGICC (v7.5.1) and MESMER (v0.8.3). *Geoscientific Model Development*, *15*(5), 2085–2103. https://doi.org/10.5194/gmd-15-2085-2022

Binois, M., Gramacy, R. B., & Ludkovski, M. (2018). Practical Heteroscedastic Gaussian Process Modeling for Large Simulation Experiments. *Journal of Computational and Graphical Statistics*, *27*(4), 808–821. https://doi.org/10.1080/10618600.2018.1458625

Boer, G. J., & Yu, B. (2003). Climate sensitivity and response. *Climate Dynamics*. https://doi.org/10.1007/s00382-002-0283-3

Bolton, T., & Zanna, L. (2019). Applications of Deep Learning to Ocean Data Inference and Subgrid Parameterization. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1029/2018MS001472

Boucher, O., Granier, C., Hoose, C., & Jones, A. (2013). IPCC 2013 Chapter 7—Clouds and aerosols. *Climate Change 2013 the Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.

Bounceur, N., Crucifix, M., & Wilkinson, R. D. (2015). Global sensitivity analysis of the climate-vegetation system to astronomical forcing: An emulator-based approach. *Earth System Dynamics*. https://doi.org/10.5194/esd-6-205-2015

Boussard, J., Nagda, C., Kaltenborn, J., Lange, C. E. E., Brouillard, P., Gurwicz, Y., Nowack, P., & Rolnick, D. (2023). *Towards Causal Representations of Climate Model Data* (arXiv:2312.02858). arXiv. https://doi.org/10.48550/arXiv.2312.02858

Cachay, S. R., Henn, B., Watt-Meyer, O., Bretherton, C. S., & Yu, R. (2024). *Probabilistic Emulation of a Global Climate Model with Spherical DYffusion* (arXiv:2406.14798). arXiv. https://doi.org/10.48550/arXiv.2406.14798

Cachier, H., Brémond, M. P., & Buat-Ménard, P. (1989). Carbonaceous aerosols from different tropical biomass burning sources. *Nature*. https://doi.org/10.1038/340371a0

Carslaw, K. S., Lee, L. A., Reddington, C. L., Mann, G. W., & Pringle, K. J. (2013). The magnitude and sources of uncertainty in global aerosol. *Faraday Discussions*. https://doi.org/10.1039/c3fd00043e

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., & Pierce, J. R. (2013). Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature*. https://doi.org/10.1038/nature12674

Castruccio, S., McInerney, D. J., Stein, M. L., Liu Crouch, F., Jacob, R. L., & Moyer, E. J. (2014). Statistical Emulation of Climate Model Projections Based on Precomputed GCM Runs. *Journal of Climate*, *27*(5), 1829–1844. https://doi.org/10.1175/JCLI-D-13-00099.1

Collins, W. J., Fry, M. M., Yu, H., Fuglestvedt, J. S., Shindell, D. T., & West, J. J. (2013). Global and regional temperature-change potentials for near-term climate forcers. *Atmospheric Chemistry and Physics*. https://doi.org/10.5194/acp-13-2471-2013

Conti, S., & O'Hagan, A. (2010). Bayesian emulation of complex multi-output and dynamic computer models. *Journal of Statistical Planning and Inference*. https://doi.org/10.1016/j.jspi.2009.08.006

Corti, S., Molteni, F., & Palmer, T. N. (1999). Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*. https://doi.org/10.1038/19745

Craig, P. S., Goldstein, M., Seheult, A. H., & Smith, J. A. (1996). Bayes Linear Strategies for Matching Hydrocarbon Reservoir History. In J. M. Bernardo, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 5: Proceedings of the Fifth Valencia International Meeting* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780198523567.003.0004

Currin, C., Mitchell, T., Morris, M., & Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.1991.10475138

Dewey, M., Hansson, H.-C., Watson-Parris, D., Samset, B. H., Wilcox, L. J., Lewinschal, A., Sand, M., Seland, Ø., Krishnan, S., & Ekman, A. M. L. (2025). AeroGP: Machine Learning How Aerosols Impact Regional Climate. *Journal of Geophysical Research: Machine Learning and Computation*, *2*(4), e2025JH000741. https://doi.org/10.1029/2025JH000741

Dong, B., Sutton, R. T., Highwood, E., & Wilcox, L. (2014). The impacts of European and Asian anthropogenic sulfur dioxide emissions on Sahel rainfall. *Journal of Climate*. https://doi.org/10.1175/JCLI-D-13-00769.1

Durack, P. J., Naik, V., Nicholls, Z., O'Rourke, E., Turner, B., Buontempo, C., Brookshaw, A., Goddard, C., MacIntosh, C., Hewitt, H., & Dunne, J. (2025). Earth System Forcing for CMIP7 and Beyond. *Bulletin of the American Meteorological Society*, *106*(8), E1580–E1588. https://doi.org/10.1175/BAMS-D-25-0119.1

ECLIPSE. (2014). *ECLIPSE V5 global emission fields*. International Institute for Applied Systems Analysis.

Edwards, T. L., Brandon, M. A., Durand, G., Edwards, N. R., Golledge, N. R., Holden, P. B., Nias, I. J., Payne, A. J., Ritz, C., & Wernecke, A. (2019). Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature*. https://doi.org/10.1038/s41586-019-0901-4

Foley, A. M., Holden, P. B., Edwards, N. R., Mercure, J. F., Salas, P., Pollitt, H., & Chewpreecha, U. (2016). Climate model emulation in an integrated assessment framework: A case study for mitigation policies in the electricity sector. *Earth System Dynamics*. https://doi.org/10.5194/esd-7-119-2016

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D. J., Mauritsen, T., Palmer, M. D., Watanabe, M., Wild, M., & Zhang, H. (2021). The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. B. R. Matthews, T. K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B.

Zhou (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 923–1054). Cambridge University Press. https://doi.org/10.1017/9781009157896.009

GPy. (2012). *GPy: A Gaussian process framework in python*. http://github.com/SheffieldML/GPy

Gramacy, R. B. (2020). *Surrogates: Gaussian Process Modeling, Design and Optimization for the Applied Sciences*. Chapman Hall/CRC.

Gregory, J. M., Ingram, W. J., Palmer, M. A., Jones, G. S., Stott, P. A., Thorpe, R. B., Lowe, J. A., Johns, T. C., & Williams, K. D. (2004). A new method for diagnosing radiative forcing and climate sensitivity. *Geophysical Research Letters*, *31*(3). https://doi.org/10.1029/2003GL018747

Gu, M., & Berger, J. O. (2016). Parallel Partial Gaussian Process Emulation for Computer Models with Massive Output. *The Annals of Applied Statistics*, *10*(3), 1317–1347.

Gu, M., Palomo, J., & Berger, J., O. (2019). RobustGaSP: Robust Gaussian Stochastic Process Emulation in R. *The R Journal*, *11*(1), 112. https://doi.org/10.32614/RJ-2019-011

Guan, H., Arcomano, T., Chattopadhyay, A., & Maulik, R. (2024). *LUCIE: A Lightweight Uncoupled ClImate Emulator with long-term stability and physical consistency for O(1000)-member ensembles* (arXiv:2405.16297). arXiv. https://doi.org/10.48550/arXiv.2405.16297

Hansen, J., Sato, M., & Ruedy, R. (1997). Radiative forcing and climate response. *Journal of Geophysical Research Atmospheres*. https://doi.org/10.1029/96JD03436

Hansen, J., Sato, M., Ruedy, R., Nazarenko, L., Lacis, A., Schmidt, G. A., Russell, G., Aleinov, I., Bauer, M., Bauer, S., Bell, N., Cairns, B., Canuto, V., Chandler, M., Cheng, Y., Del Genio, A., Faluvegi, G., Fleming, E., Friend, A., … Zhang, S. (2005). Efficacy of climate forcings. In

Journal of Geophysical Research D: Atmospheres (Vol. 110, Issue 18, pp. 1–45). https://doi.org/10.1029/2005JD005776

Haywood, J., & Boucher, O. (2000). Estimates of the direct and indirect radiative forcing due to tropospheric aerosols: A review. In *Reviews of Geophysics*. https://doi.org/10.1029/1999RG000078

Higdon, D., Gattiker, J., Williams, B., & Rightley, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*. https://doi.org/10.1198/016214507000000888

Hodnebrog, O., Myhre, G., Forster, P. M., Sillmann, J., & Samset, B. H. (2016). Local biomass burning is a dominant cause of the observed precipitation reduction in southern Africa. *Nature Communications*. https://doi.org/10.1038/ncomms11236

Joshi, M., Shine, K., Ponater, M., Stuber, N., Sausen, R., & Li, L. (2003). A comparison of climate response to different radiative forcings in three general circulation models: Towards an improved metric of climate change. *Climate Dynamics*, *20*(7), 843–854. https://doi.org/10.1007/s00382-003-0305-9

Kaltenborn, J., Lange, C. E. E., Ramesh, V., Brouillard, P., Gurwicz, Y., Nagda, C., Runge, J., Nowack, P., & Rolnick, D. (2023). *ClimateSet: A Large-Scale Climate Model Dataset for Machine Learning* (arXiv:2311.03721). arXiv. http://arxiv.org/abs/2311.03721

Kasoar, M., Shawki, D., & Voulgarakis, A. (2018). Similar spatial patterns of global climate response to aerosols from different regions. *Npj Climate and Atmospheric Science*, *1*(1), 12. https://doi.org/10.1038/s41612-018-0022-z

King, R. C., Mansfield, L. A., & Sheshadri, A. (2024). Bayesian History Matching Applied to the Calibration of a Gravity Wave Parameterization. *Journal of Advances in Modeling Earth Systems*, *16*(4), e2023MS004163. https://doi.org/10.1029/2023MS004163

Lasslop, G., Coppola, A. I., Voulgarakis, A., Yue, C., & Veraverbeke, S. (2019). Influence of Fire on the Carbon Cycle and Climate. In *Current Climate Change Reports*. https://doi.org/10.1007/s40641-019-00128-9

Lee, L. A., Carslaw, K. S., Pringle, K. J., & Mann, G. W. (2012). Mapping the uncertainty in global CCN using emulation. *Atmospheric Chemistry and Physics*. https://doi.org/10.5194/acp-12-9739-2012

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., & Spracklen, D. V. (2011). Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmospheric Chemistry and Physics*, *11*(23), 12253–12273. https://doi.org/10.5194/acp-11-12253-2011

Lee, L. A., Reddington, C. L., & Carslaw, K. S. (2016). On the relationship between aerosol model uncertainty and radiative forcing uncertainty. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1507050113

Lewinschal, A., Ekman, A. M. L., Hansson, H. C., Sand, M., Berntsen, T. K., & Langner, J. (2019). Local and remote temperature response of regional SO2 emissions. *Atmospheric Chemistry and Physics*. https://doi.org/10.5194/acp-19-2385-2019

Liousse, C., Assamoi, E., Criqui, P., Granier, C., & Rosset, R. (2014). Explosive growth in African combustion emissions from 2005 to 2030. *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/9/3/035003

Liu, L., Shawki, D., Voulgarakis, A., Kasoar, M., Samset, B. H., Myhre, G., Forster, P. M., Hodnebrog, Sillmann, J., Aalbergsjø, S. G., Boucher, O., Faluvegi, G., Iversen, T., Kirkevåg, A., Lamarque, J. F., Olivié, D., Richardson, T., Shindell, D., & Takemura, T. (2018). A PDRMIP Multimodel study on the impacts of regional aerosol forcings on global and regional precipitation. *Journal of Climate*. https://doi.org/10.1175/JCLI-D-17-0439.1

Loeppky, J. L., Sacks, J., & Welch, W. J. (2009). Choosing the sample size of a computer experiment: A practical guide. *Technometrics*. https://doi.org/10.1198/TECH.2009.08040

Mann, G. W., Carslaw, K. S., Spracklen, D. V., Ridley, D. A., Manktelow, P. T., Chipperfield, M. P., Pickering, S. J., & Johnson, C. E. (2010). Description and evaluation of GLOMAP-mode: A modal global aerosol microphysics model for the UKCA composition-climate model. *Geoscientific Model Development*. https://doi.org/10.5194/gmd-3-519-2010

Mansfield, L. A. (2025). *Lm2612/emulator* (Version First release (v1.0)) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.17814595

Mansfield, L. A., Nowack, P. J., Kasoar, M., Everitt, R. G., Collins, W. J., & Voulgarakis, A. (2020). Predicting global patterns of long-term climate change from short-term simulations using machine learning. *Npj Climate and Atmospheric Science*. https://doi.org/10.1038/s41612-020-00148-5

Meinshausen, M., Raper, S. C. B., & Wigley, T. M. L. (2011). Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6 – Part 1: Model description and calibration. *Atmospheric Chemistry and Physics*, *11*(4), 1417–1456. https://doi.org/10.5194/acp-11-1417-2011

Met Office. (2021). The Cray XC40 supercomputing system. *Https://Www.Metoffice.Gov.Uk/about-Us/What/Technology/Supercomputer. Date Accessed: 21/06/2021*.

Nath, S., Lejeune, Q., Beusch, L., Seneviratne, S. I., & Schleussner, C.-F. (2022). MESMER-M: an Earth system model emulator for spatially resolved monthly temperature. *Earth System Dynamics*, *13*(2), 851–877. https://doi.org/10.5194/esd-13-851-2022

Nicholls, Z. R. J., Meinshausen, M., Lewis, J., Gieseke, R., Dommenget, D., Dorheim, K., Fan, C.-S., Fuglestvedt, J. S., Gasser, T., Golüke, U., Goodwin, P., Hartin, C., Hope, A. P., Kriegler, E., Leach, N. J., Marchegiani, D., McBride, L. A., Quilcaille, Y., Rogelj, J., … Xie, Z. (2020). Reduced Complexity Model Intercomparison Project Phase 1: Introduction and evaluation of global-mean temperature response. *Geoscientific Model Development*, *13*(11), 5175–5190. https://doi.org/10.5194/gmd-13-5175-2020

Nowack, P., Braesicke, P., Haigh, J., Abraham, N. L., Pyle, J., & Voulgarakis, A. (2018). Using machine learning to build temperature-based ozone parameterizations for climate sensitivity simulations. *Environmental Research Letters*. https://doi.org/10.1088/1748-9326/aae2be

O'Hagan, A. (1978). Curve Fitting and Optimal Design for Prediction. *Journal of the Royal Statistical Society: Series B (Methodological)*. https://doi.org/10.1111/j.2517-6161.1978.tb01643.x

Palmer, T. N. (1999). A nonlinear dynamical perspective on climate prediction. *Journal of Climate*. https://doi.org/10.1175/1520-0442(1999)012%3C0575:ANDPOC%3E2.0.CO;2

Pechony, O., & Shindell, D. T. (2010). Driving forces of global wildfires over the past millennium and the forthcoming century. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1003669107

Persad, G. G., & Caldeira, K. (2018). Divergent global-scale temperature effects from identical aerosols emitted in different regions. *Nature Communications*, *9*(1), Article 1. https://doi.org/10.1038/s41467-018-05838-6

Popp, A., Calvin, K., Fujimori, S., Havlik, P., Humpenöder, F., Stehfest, E., Bodirsky, B. L., Dietrich, J. P., Doelmann, J. C., Gusti, M., Hasegawa, T., Kyle, P., Obersteiner, M., Tabeau, A., Takahashi, K., Valin, H., Waldhoff, S., Weindl, I., Wise, M., … Vuuren, D. P. van. (2017). Land-use futures in the shared socio-economic pathways. *Global Environmental Change*. https://doi.org/10.1016/j.gloenvcha.2016.10.002

Rasmussen, C. E., & Williams, C. K. I. (2006). *Gaussian processes for machine learning*. MIT Press.

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1810286115

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., … Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Richardson, T. B., Forster, P. M., Smith, C. J., Maycock, A. C., Wood, T., Andrews, T., Boucher, O., Faluvegi, G., Fläschner, D., Hodnebrog, Kasoar, M., Kirkevåg, A., Lamarque, J. F., Mülmenstädt, J., Myhre, G., Olivié, D., Portmann, R. W., Samset, B. H., Shawki, D., …

Watson-Parris, D. (2019). Efficacy of Climate Forcings in PDRMIP Models. *Journal of Geophysical Research: Atmospheres*. https://doi.org/10.1029/2019JD030581

Ridley, J. K., Blockley, E. W., Keen, A. B., Rae, J. G. L., West, A. E., & Schroeder, D. (2018). The sea ice model component of HadGEM3-GC3.1. *Geoscientific Model Development*, *11*(2), 713–723. https://doi.org/10.5194/gmd-11-713-2018

Ring, M. J., & Plumb, R. A. (2008). The response of a simplified GCM to axisymmetric forcings: Applicability of the fluctuation-dissipation theorem. *Journal of the Atmospheric Sciences*. https://doi.org/10.1175/2008JAS2773.1

Rougier, J. (2009). Efficient emulators for multivariate deterministic Functions. *Journal of Computational and Graphical Statistics*. https://doi.org/10.1198/106186008X384032

Rougier, J., Guillas, S., Maute, A., & Richmond, A. D. (2009). Expert Knowledge and Multivariate Emulation: The Thermosphere–Ionosphere Electrodynamics General Circulation Model (TIE-GCM). *Technometrics*, *51*(4), 414–424. https://doi.org/10.1198/TECH.2009.07123

Ryan, E., Wild, O., Voulgarakis, A., & Lee, L. (2018). Fast sensitivity analysis methods for computationally expensive models with multi-dimensional output. *Geoscientific Model Development*, *11*, 3131–3146. https://doi.org/10.5194/gmd-11-3131-2018

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., & Wu, Q. (2019). Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2019.01.012

Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling and Software*. https://doi.org/10.1016/j.envsoft.2010.04.012

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*. https://doi.org/10.1016/j.cpc.2009.09.018

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., & Tarantola, S. (2008). Global Sensitivity Analysis. The Primer. In *Global Sensitivity Analysis. The Primer*. https://doi.org/10.1002/9780470725184

Salter, J. M., & Williamson, D. (2016). A comparison of statistical emulation methodologies for multi-wave calibration of environmental models. *Environmetrics*. https://doi.org/10.1002/env.2405

Salter, J. M., & Williamson, D. B. (2022). EFFICIENT CALIBRATION FOR HIGH-DIMENSIONAL COMPUTER MODEL OUTPUT USING BASIS METHODS. *International Journal for Uncertainty Quantification*, *12*(6). https://doi.org/10.1615/Int.J.UncertaintyQuantification.2022039747

Salter, J. M., Williamson, D. B., Scinocca, J., & Kharin, V. (2019). Uncertainty Quantification for Computer Models With Spatial Output Using Calibration-Optimal Bases. *Journal of the American Statistical Association*. https://doi.org/10.1080/01621459.2018.1514306

Samset, B. H., Sand, M., Smith, C. J., Bauer, S. E., Forster, P. M., Fuglestvedt, J. S., Osprey, S., & Schleussner, C.-F. (2018). Climate Impacts From a Removal of Anthropogenic Aerosol Emissions. *Geophysical Research Letters*, *45*(2), 1020–1029.

Santner, T. J., Williams, B. J., Notz, W. I., & Williams, B. J. (2003). *The design and analysis of computer experiments* (Vol. 1). Springer.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., & De Freitas, N. (2016). Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE*, *104*(1), 148–175. https://doi.org/10.1109/JPROC.2015.2494218

Shawki, D., Voulgarakis, A., Chakraborty, A., Kasoar, M., & Srinivasan, J. (2018). The South Asian Monsoon Response to Remote Aerosols: Global and Regional Mechanisms. *Journal of Geophysical Research: Atmospheres*. https://doi.org/10.1029/2018JD028623

Shindell, D., & Faluvegi, G. (2009). Climate response to regional radiative forcing during the twentieth century. *Nature Geoscience*, *2*, 294 EP-. https://doi.org/10.1038/ngeo473

Shindell, D. T. (2012). Evaluation of the absolute regional temperature potential. *Atmospheric Chemistry and Physics*, *12*(17), 7955–7960. https://doi.org/10.5194/acp-12-7955-2012

Shindell, D. T., Miller, R. L., Schmidt, G. A., & Pandolfo, L. (1999). Simulation of recent northern winter climate trends by greenhouse-gas forcing. *Nature*. https://doi.org/10.1038/20905

Smith, C. J., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., & Regayre, L. A. (2018). FAIR v1.3: A simple emissions-based impulse response and carbon cycle model. *Geoscientific Model Development*. https://doi.org/10.5194/gmd-11-2273-2018

Smith, C. J., Kramer, R. J., Myhre, G., Forster, P. M., Soden, B. J., Andrews, T., Boucher, O., Faluvegi, G., Fläschner, D., Hodnebrog, Ø., Kasoar, M., Kharin, V., Kirkevåg, A., Lamarque, J.-F., Mülmenstädt, J., Olivié, D., Richardson, T., Samset, B. H., Shindell, D., … Watson-Parris, D.

(2018). Understanding Rapid Adjustments to Diverse Forcing Agents. *Geophysical Research Letters*, *45*(21), 12,023-12,031. https://doi.org/10.1029/2018GL079826

Souza, A. N., Wagner, G. L., Ramadhan, A., Allen, B., Churavy, V., Schloss, J., Campin, J., Hill, C., Edelman, A., Marshall, J., Flierl, G., & Ferrari, R. (2020). Uncertainty Quantification of Ocean Parameterizations: Application to the K-Profile-Parameterization for Penetrative Convection. *Journal of Advances in Modeling Earth Systems*. https://doi.org/10.1029/2020MS002108

Stell, A. C., Western, L. M., Sherwen, T., & Rigby, M. (2021). Atmospheric-methane source and sink sensitivity analysis using Gaussian process emulation. *Atmospheric Chemistry and Physics*. https://doi.org/10.5194/acp-21-1717-2021

Stocker, T. (2013). IPCC Summary for Policymakers in Climate Change 2013: The Physical Science Basis. In *Climate Change Lecture Series*.

Stohl, A., Aamaas, B., Amann, M., Baker, L. H., Bellouin, N., Berntsen, T. K., Boucher, O., Cherian, R., Collins, W., Daskalakis, N., Dusinska, M., Eckhardt, S., Fuglestvedt, J. S., Harju, M., Heyes, C., Hodnebrog, Ø., Hao, J., Im, U., Kanakidou, M., … Zhu, T. (2015). Evaluating the climate and air quality impacts of short-lived pollutants. *Atmospheric Chemistry and Physics*, *15*(18), 10529–10566. https://doi.org/10.5194/acp-15-10529-2015

Storkey, D., Blaker, A. T., Mathiot, P., Megann, A., Aksenov, Y., Blockley, E. W., Calvert, D., Graham, T., Hewitt, H. T., Hyder, P., Kuhlbrodt, T., Rae, J. G. L., & Sinha, B. (2018). UK Global Ocean GO6 and GO7: A traceable hierarchy of model resolutions. *Geoscientific Model Development*, *11*(8), 3187–3213. https://doi.org/10.5194/gmd-11-3187-2018

Tosca, M. G., Randerson, J. T., & Zender, C. S. (2013). Global impact of smoke aerosols from landscape fires on climate and the Hadley circulation. *Atmospheric Chemistry and Physics*. https://doi.org/10.5194/acp-13-5227-2013

Urrego-Blanco, J. R., Urban, N. M., Hunke, E. C., Turner, A. K., & Jeffery, N. (2016). Uncertainty quantification and global sensitivity analysis of the Los Alamos sea ice model. *Journal of Geophysical Research: Oceans*. https://doi.org/10.1002/2015JC011558

Van der Werf, G. R., Randerson, J. T., Collatz, G. J., & Giglio, L. (2003). Carbon emissions from fires in tropical and subtropical ecosystems. *Global Change Biology*. https://doi.org/10.1046/j.1365-2486.2003.00604.x

van Vuuren, D. P., Stehfest, E., Gernaat, D. E. H. J., Doelman, J. C., van den Berg, M., Harmsen, M., de Boer, H. S., Bouwman, L. F., Daioglou, V., Edelenbosch, O. Y., Girod, B., Kram, T., Lassaletta, L., Lucas, P. L., van Meijl, H., Müller, C., van Ruijven, B. J., van der Sluis, S., & Tabeau, A. (2017). Energy, land-use and greenhouse gas emissions trajectories under a green growth paradigm. *Global Environmental Change*. https://doi.org/10.1016/j.gloenvcha.2016.05.008

Vernon, I., Goldstein, M., & Bower, R. G. (2010). Galaxy Formation: A Bayesian Uncertainty Analysis. *Bayesian Analysis*, *05*(04). https://doi.org/10.1214/10-ba524

Voulgarakis, A., & Field, R. D. (2015). Fire Influences on Atmospheric Composition, Air Quality and Climate. In *Current Pollution Reports*. https://doi.org/10.1007/s40726-015-0007-z

Walters, D., Baran, A. J., Boutle, I., Brooks, M., Earnshaw, P., Edwards, J., Furtado, K., Hill, P., Lock, A., Manners, J., Morcrette, C., Mulcahy, J., Sanchez, C., Smith, C., Stratton, R., Tennant, W., Tomassini, L., Van Weverberg, K., Vosper, S., … Zerroukat, M. (2019). The Met Office

Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geoscientific Model Development*, *12*(5), 1909–1963. https://doi.org/10.5194/gmd-12-1909-2019

Ward, D. S., Kloster, S., Mahowald, N. M., Rogers, B. M., Randerson, J. T., & Hess, P. G. (2012). The changing radiative forcing of fires: Global model estimates for past, present and future. *Atmospheric Chemistry and Physics*. https://doi.org/10.5194/acp-12-10857-2012

Watson-Parris, D. (2021). Machine learning for weather and climate are worlds apart. In *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. https://doi.org/10.1098/rsta.2020.0098

Watt-Meyer, O., Dresdner, G., McGibbon, J., Clark, S. K., Henn, B., Duncan, J., Brenowitz, N. D., Kashinath, K., Pritchard, M. S., Bonev, B., Peters, M. E., & Bretherton, C. S. (2023). *ACE: A fast, skillful learned global atmospheric model for climate prediction* (arXiv:2310.02074). arXiv. http://arxiv.org/abs/2310.02074

Watt-Meyer, O., Henn, B., McGibbon, J., Clark, S. K., Kwa, A., Perkins, W. A., Wu, E., Harris, L., & Bretherton, C. S. (2025). ACE2: Accurately learning subseasonal to decadal atmospheric variability and forced responses. *Npj Climate and Atmospheric Science*, *8*(1), 205. https://doi.org/10.1038/s41612-025-01090-0

Wells, C. D., Kasoar, M., Bellouin, N., & Voulgarakis, A. (2022). *Local and remote climate impacts of future African aerosol emissions* [Preprint]. Aerosols/Atmospheric Modelling/Troposphere/Physics (physical properties and processes). https://doi.org/10.5194/egusphere-2022-919

Westervelt, D. M., Mascioli, N. R., Fiore, A. M., Conley, A. J., Lamarque, J. F., Shindell, D. T., Faluvegi, G., Previdi, M., Correa, G., & Horowitz, L. W. (2020). Local and remote mean and extreme temperature response to regional aerosol emissions reductions. *Atmospheric Chemistry and Physics*. https://doi.org/10.5194/acp-20-3009-2020

Wild, O., Voulgarakis, A., O'Connor, F., Lamarque, J. F., Ryan, E. M., & Lee, L. (2020). Global sensitivity analysis of chemistry-climate model budgets of tropospheric ozone and OH: Exploring model diversity. *Atmospheric Chemistry and Physics*. https://doi.org/10.5194/acp-20-4047-2020

Wilkinson, R. D. (2010). Bayesian Calibration of Expensive Multivariate Computer Experiments. In *Large-Scale Inverse Problems and Quantification of Uncertainty*. https://doi.org/10.1002/9780470685853.ch10

Williams, A. I. L., Stier, P., Dagan, G., & Watson-Parris, D. (2022). Strong control of effective radiative forcing by the spatial pattern of absorbing aerosol. *Nature Climate Change*, *12*(8), 735–742. https://doi.org/10.1038/s41558-022-01415-4

Williams, K. D., Copsey, D., Blockley, E. W., Bodas-Salcedo, A., Calvert, D., Comer, R., Davis, P., Graham, T., Hewitt, H. T., Hill, R., Hyder, P., Ineson, S., Johns, T. C., Keen, A. B., Lee, R. W., Megann, A., Milton, S. F., Rae, J. G. L., Roberts, M. J., … Xavier, P. K. (2018). The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *Journal of Advances in Modeling Earth Systems*, *10*(2), 357–380. https://doi.org/10.1002/2017MS001115

Williamson, D., Blaker, A. T., Hampton, C., & Salter, J. (2015). Identifying and removing structural biases in climate models with history matching. *Climate Dynamics*. https://doi.org/10.1007/s00382-014-2378-z