Supplementary Information for
Using a Gaussian Process Emulator to approximate the 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]Corndel, 53-79 Highgate Road, London, NW5 1TL, United Kingdom.

[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)

Text S1. Linearity of Main Effects

To calculate the main effect for a single pollutant, we vary the pollutant perturbation from the minimum to the maximum and run 200 different realizations of this where the other input parameters are sampled from a normal distribution centered at the present day levels with a standard deviation ¼ x the input range presented in Table 1. This is to ensure the average behavior of all simulations does not deviate significantly from present-day levels. The mean behavior of each parameter perturbation is calculated by averaging over these different realizations and can be found **in Supplementary Figure S6** for each key region.

We find that the main effect is linear for all pollutants, both GHGs and aerosols, over the ranges explored here, consistent with the finding that a linear Gaussian process emulator performs similarly well (i.e. linear regression, Supplementary Figure S3). This is intuitive under small forcing perturbations, as we know that the relationship between emissions and response is fairly linear at low forcing levels, leading to its use in estimating climate response given forcing, via the

equilibrium climate sensitivity, for GHGs (Boer & Yu, 2003; Gregory et al., 2004; Hansen et al., 1997) and aerosols, using an adjustment based on individual aerosol efficacy (Collins et al., 2013; Hansen et al., 2005; Richardson et al., 2019). Although many of these studies assume the equilibrium climate sensitivity to be constant to the first order, other studies have demonstrated that it can change under a changing climate (Friedrich et al., 2016; Gregory & Andrews, 2016) or under larger magnitude forcings (Knutti et al., 2017; Rugenstein et al., 2020; Zhu & Poulsen, 2020). One may not expect the emulator to capture large non-linearities due to a changing climate (or only to the degree that they manifest themselves within the first five years) as it is trained on abrupt perturbations from a fixed initial climate. However, based on past studies, we may expect to see non-linearities caused by particularly large forcings in this study. The CO_2 perturbation range selected here gives the largest range of forcing perturbations (up to ~2xCO_2). However, it does not extend as far as the forcings explored in previous studies where the relationship between forcing and response is found to deviate from linearity (up to 4xCO_2 from pre-industrial levels) (Knutti et al., 2017; Meraner et al., 2013; Rugenstein et al., 2020; Zhu & Poulsen, 2020). Furthermore, many of the non-linearities are explained by transient effects, that would not be captured in the short 5 year simulations used to train the emulator. Knutti et al. (2017) find that the first few years of a model simulation gives a linear relationship between forcing and temperature response and that the slope of this relationship changes at longer timescales of 50-100 year, due to slow climate adjustments, feedbacks and as warming patterns change over time. For instance, CO_2 sinks, such as the ocean, become less efficient as they become saturated (Gregory et al., 2015), which is an effect that would not been seen in the short 5 year timescales predicted here but could lead to increased response at high forcings on longer timescales (Ceppi et al., 2017). Other feedbacks include the positive water vapor feedback in which increasing temperatures allow the atmosphere to hold more water vapor, a greenhouse gas (Meraner et al., 2013), and the negative lapse rate feedback in which surface warming caused by increased CO_2 weakens the lapse rate, making the climate less sensitive to further warming (Colman & McAvaney, 2009). These feedbacks appear after the climate has already warmed significantly, which requires some time even after an abrupt forcing perturbation. The main effects of emulators built on longer timescales may reveal these non-linearities in the long-lived climate pollutants.

Past studies also suggest that we would expect a saturation effect of temperature response as the short-lived pollutants increase. Firstly, the SO_2 perturbations made here are a precursor to the sulfate aerosol SO_4. SO_4 formation begins with nucleation where the precursor gas forms clusters of sulfuric acid (Curtius, 2006). This is more efficient in cleaner air, where new clusters form as opposed to existing clusters growing in size (Carslaw et al., 2013). Therefore increases in SO_2 in regions with lower pollution levels directly leads to a larger increase in SO_4 compared regions with highly polluted air. Secondly, a greater increase in concentration of cloud condensation nuclei (CCN) also leads to a greater increase in cloud droplet number because higher droplet concentrations suppress supersaturation within clouds. Finally, increased water droplet number enhances the cloud albedo effect more when there are fewer water droplets, with the sensitivity of cloud albedo falling with $1/N$ for N cloud droplets.

As a consequence of this, several studies have found that regions with lower SO_2 levels experience greater climate response to increases in SO_2 per unit emission, (Collins et al., 2013; Kasoar et al., 2018; Liu et al., 2018). Following this reasoning, as SO_2 emissions are increased, we would expect the decrease in temperature to decelerate, which would lead to a decreasing gradient at high SO_2 perturbations in Figure S6 (Carslaw et al., 2013). Unlike the GHG perturbations, these effects are not limited by the short 5 year timescale of the output. However, the main effect of each perturbation appears linear across the entire perturbation range, over all regions. The fact that these non-linearities are not learned by the emulator may explain why the Example B in Figure 1 was poorly predicted. These non-linearities may be too small to be realized by the emulator given the reduced number of training points at high perturbations and the significant uncertainty associated with the internal variability. Increasing the number of training points at large perturbations and increasing the input ranges to include more extreme forcing may be beneficial for better constraining the response under strong aerosol perturbations.

Text S2. Sensitivity Indices

The first order sensitivity indices in CO_2, CH_4 and the aerosols sum to 1 almost exactly, leaving $<10^{-4}$ down to higher order terms, which would describe sensitivity due to interactions between multiple perturbations. This falls in line with aerosol perturbation studies that have found the effects of multiple perturbations across different regions to be approximately additive at low levels (Kasoar et al., 2018), but at high levels we may expect the atmosphere to become more saturated with aerosol particles, reducing the efficiency of subsequent perturbations as noted above. Furthermore, previous studies have found non-linear effects when perturbing GCMs with both GHGs and aerosols (Feichter et al., 2004; Marvel et al., 2015; Ming & Ramaswamy, 2009). The lack of interaction terms between these perturbations indicates these are not significant on the timescales performed here. As with any non-linearities due to GHGs alone, these would take longer than 5 years to affect the temperature response, as climate feedbacks caused by temperature change takes time to be realized.

Supporting Figures

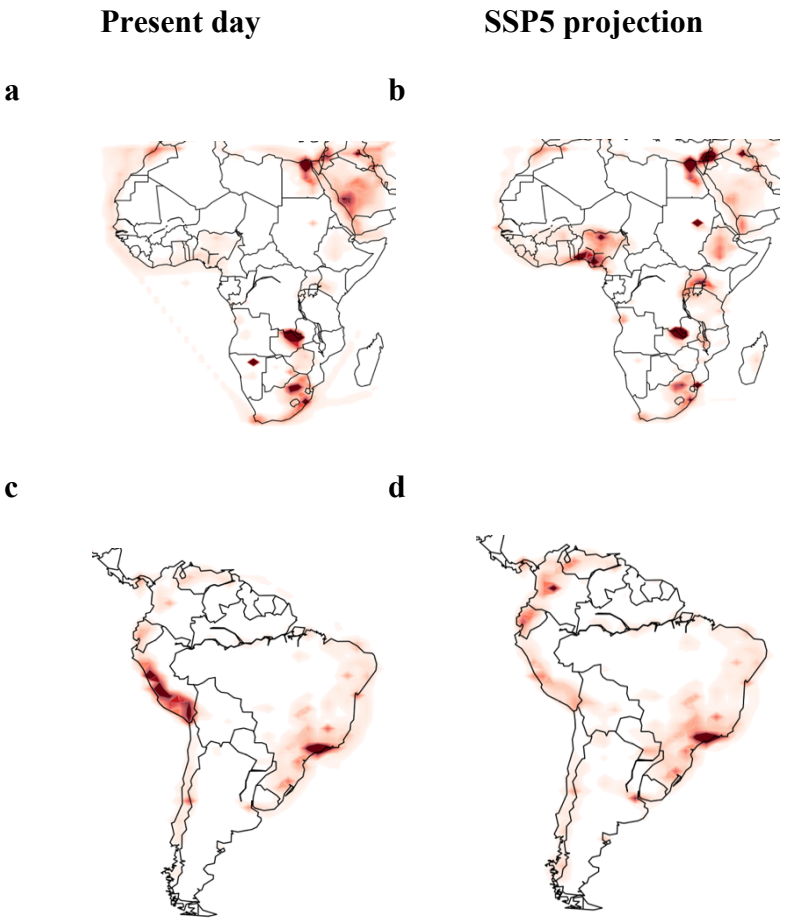


Figure S1. Distribution of SO_2 emissions for (a,b) Africa and (c,d) South America using (a,c) present day estimates and (b,d) SSP5 projections. Arbitrary colour scale. The present day emissions show highly localized emissions dominate in a few grid cells whereas the SSP5 projections show more evenly distributed emissions over more grid cells. The SSP5 projections were used in the emulator as we aim to project different scenarios which includes a rapid growth in new African cities and megacities (Liousse et al., 2014). This does not have much effect on the response to present day or reduced emissions (since these perturbations are weak). For strong positive perturbations, simulations performed with the same total emissions, but emissions more evenly distributed over a region (e.g. the SSP5 projection) will give a stronger response compared to a simulation with more concentrated emissions over a given region (e.g. the present day emissions). This is because adding sulfate aerosol to an already dirty atmosphere leads to a saturation of sulfate aerosol and therefore a saturation of response (e.g. Carslaw et al., 2013).

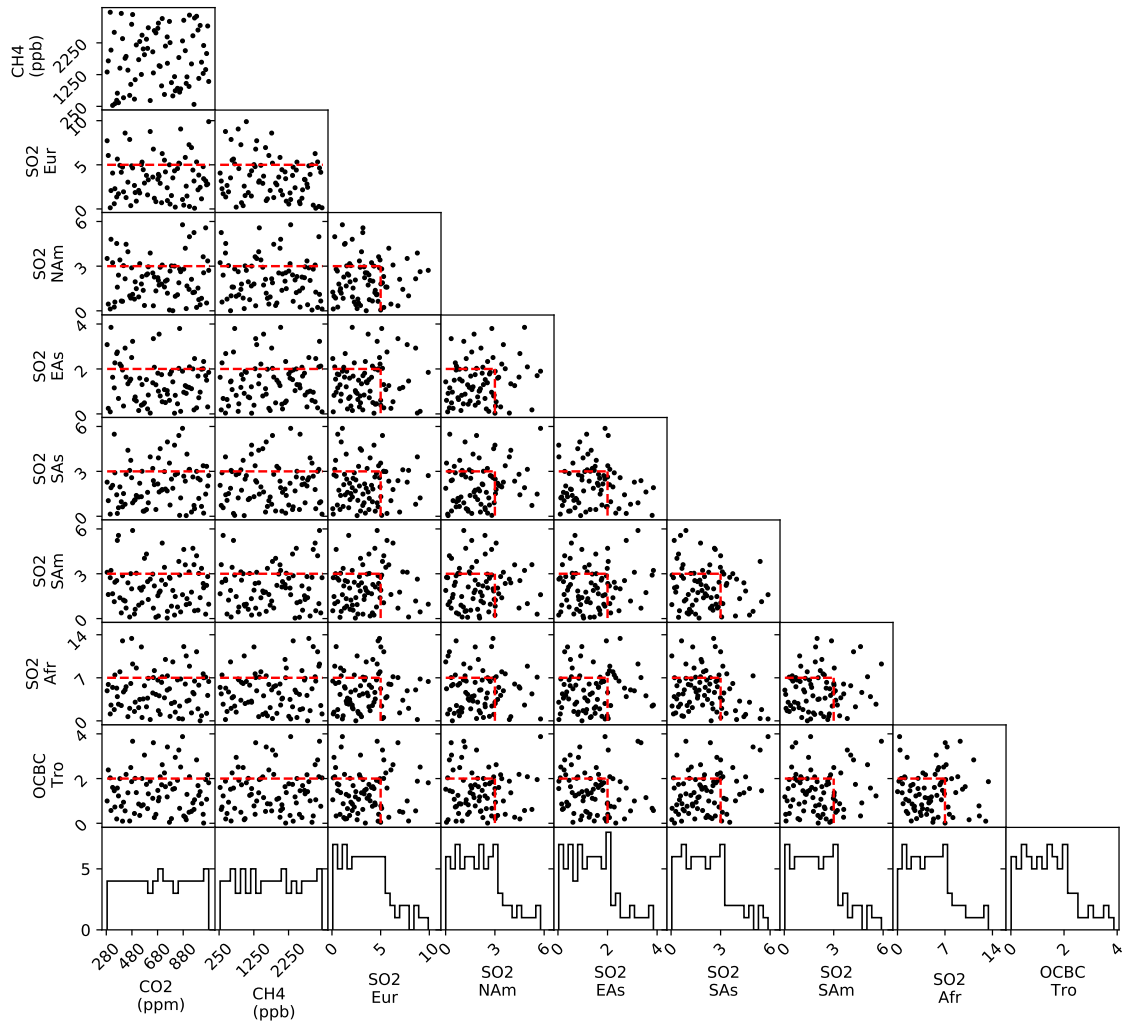


Figure S2. Training data generated from Latin Hypercube Sampler for each input against every other input, demonstrating the coverage of the input parameter space. The red lines show the feasible ranges (Table 1). For the aerosol pollutants, we include some additional simulations beyond this (25% of the total) to increase the signal-to-noise ratio and better confine the emulator. The bottom row shows the overall histogram for each input parameter independently.

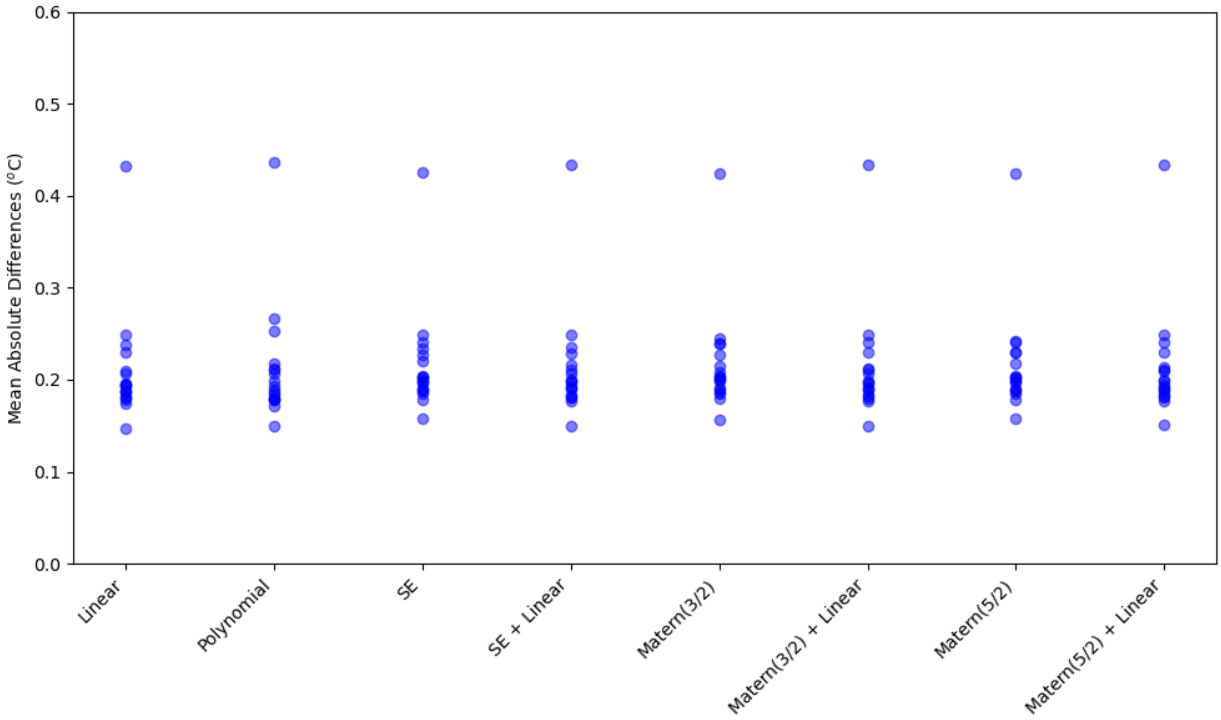
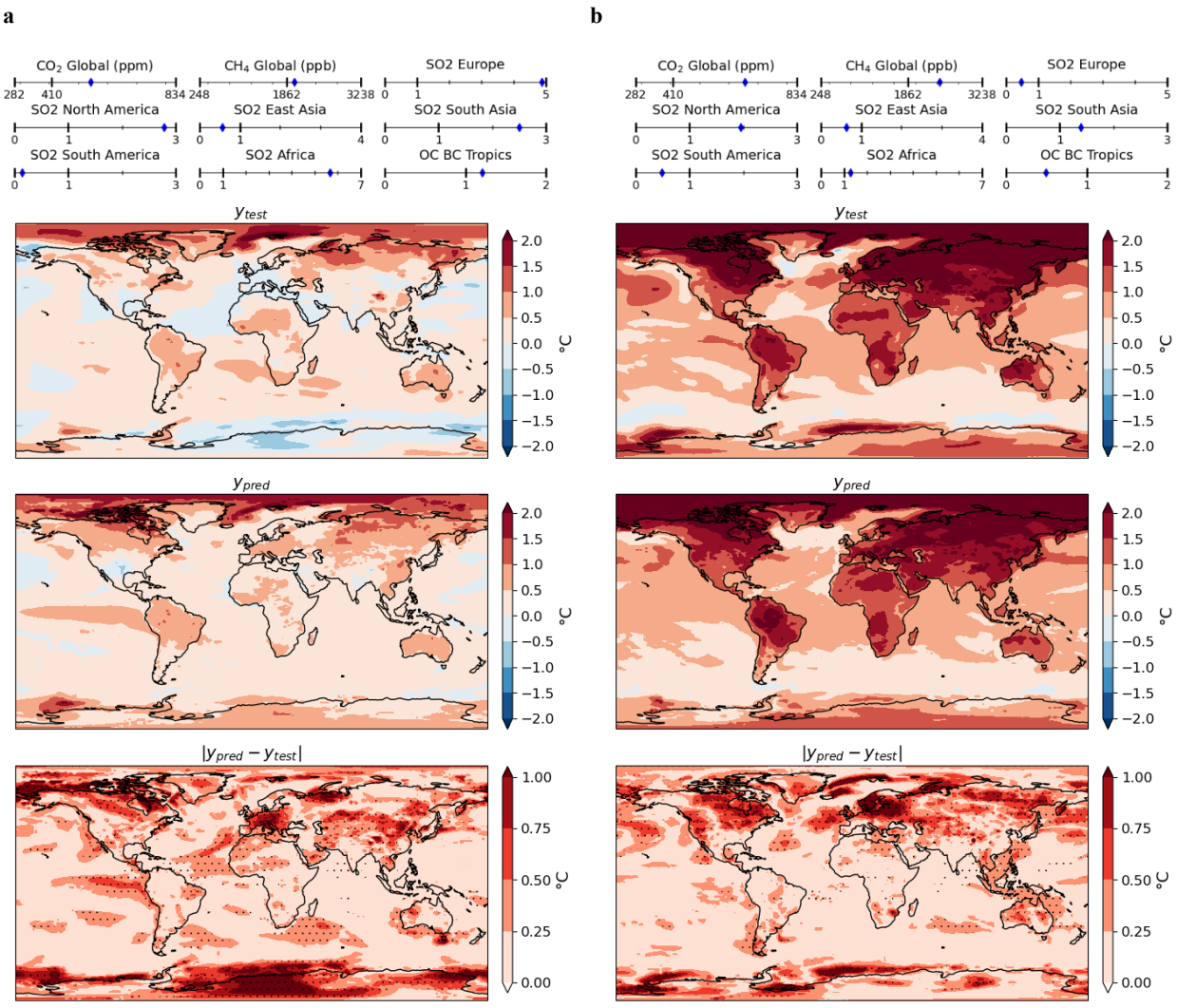
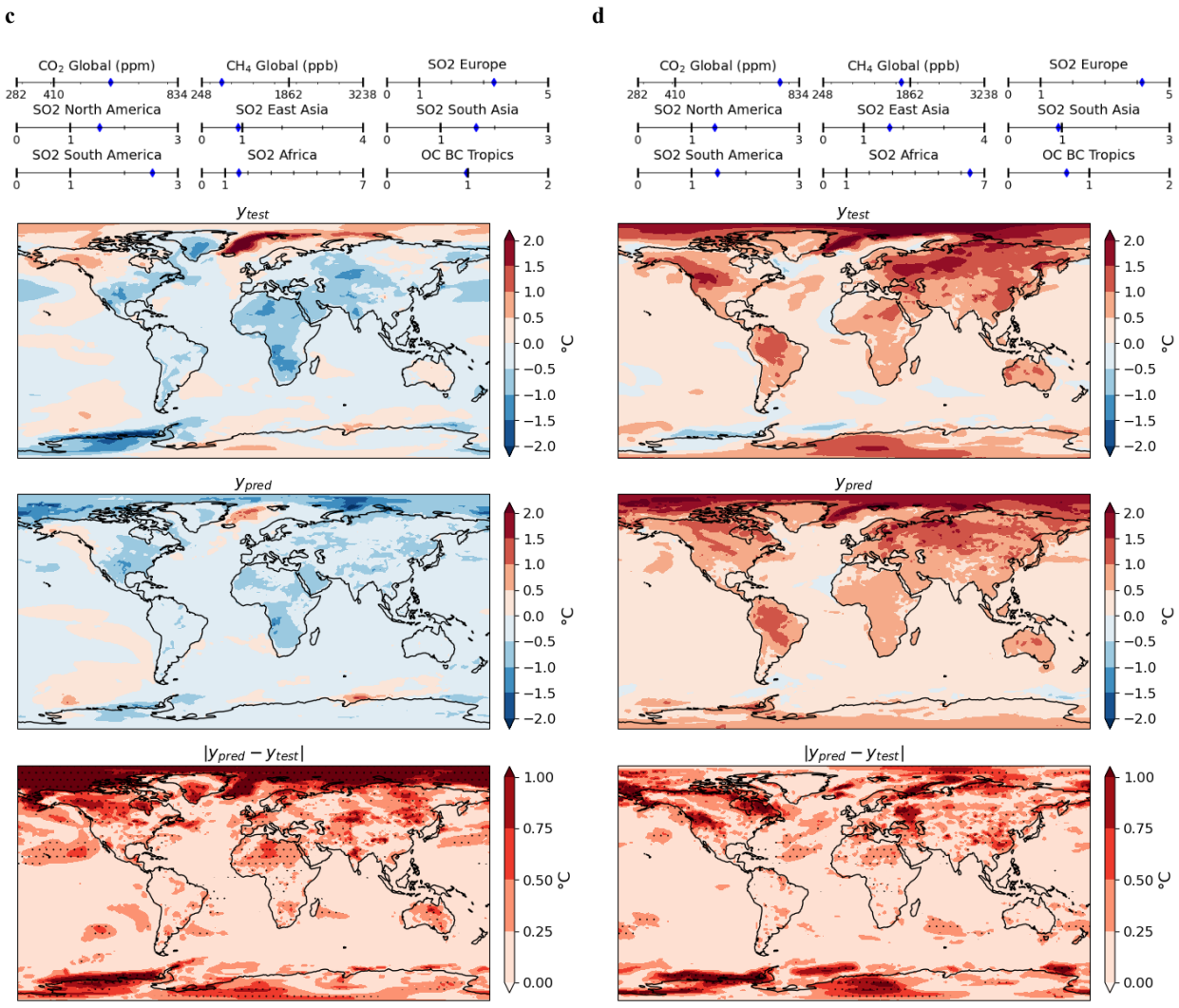
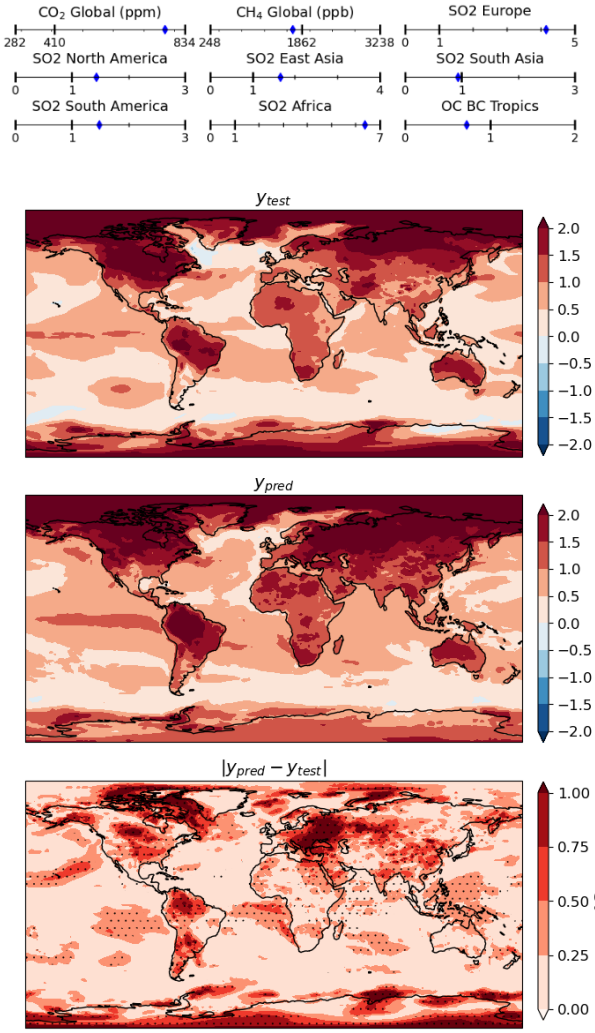


Figure S3. Emulator performance in terms of mean absolute difference for all grid cells, where each point indicates the mean absolute difference for one of the 18 test scenarios, for a range of kernel choices: linear, polynomial, squared exponential (SE), SE + Linear, Matern(3/2), Matern(3/2) + Linear, Matern(5/2) and Matern(5/2) + Linear. Details of kernels can be found in Williams & Rasmussen (2006).

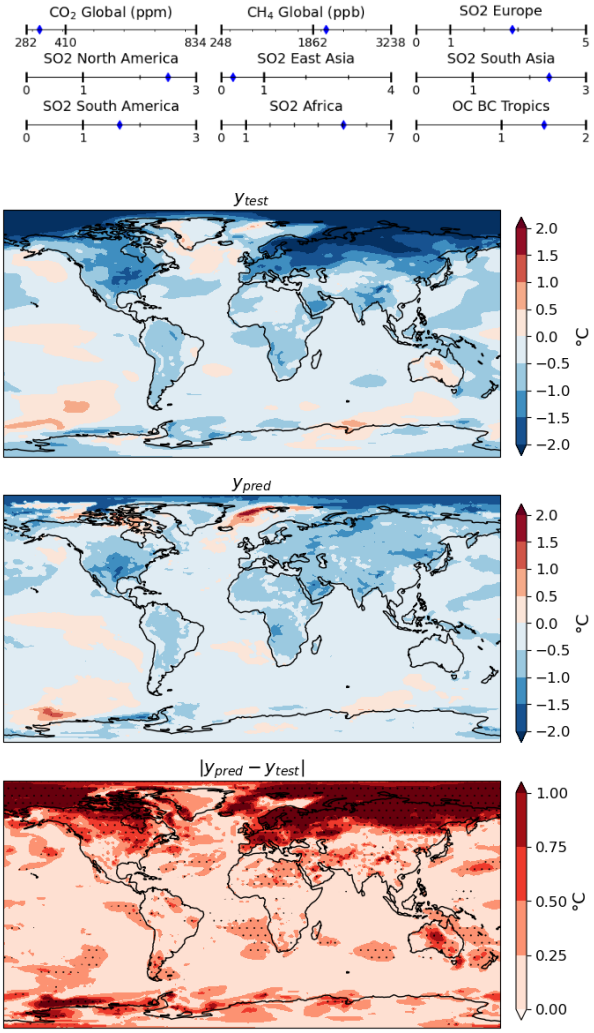

a
b
CO_2 Global (ppm)
282 410 834
CH_4 Global (ppb)
248 1862 3238
SO2 Europe
0 1 5
SO2 North America
0 1 3
SO2 East Asia
0 1 4
SO2 South Asia
0 1 3
SO2 South America
0 1 3
SO2 Africa
0 1 7
OC BC Tropics
0 1 2
y_{test}
y_{pred}
$|y_{pred} - y_{test}|$
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
°C
1.00
0.75
0.50
0.25
0.00


c
d
CO_2 Global (ppm)
282 410 834
CH_4 Global (ppb)
248 1862 3238
SO2 Europe
0 1 5
SO2 North America
0 1 3
SO2 East Asia
0 1 4
SO2 South Asia
0 1 3
SO2 South America
0 1 3
SO2 Africa
0 1 7
OC BC Tropics
0 1 2
y_{test}
y_{pred}
$|y_{pred} - y_{test}|$
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
°C
1.00 0.75 0.50 0.25 0.00
°C

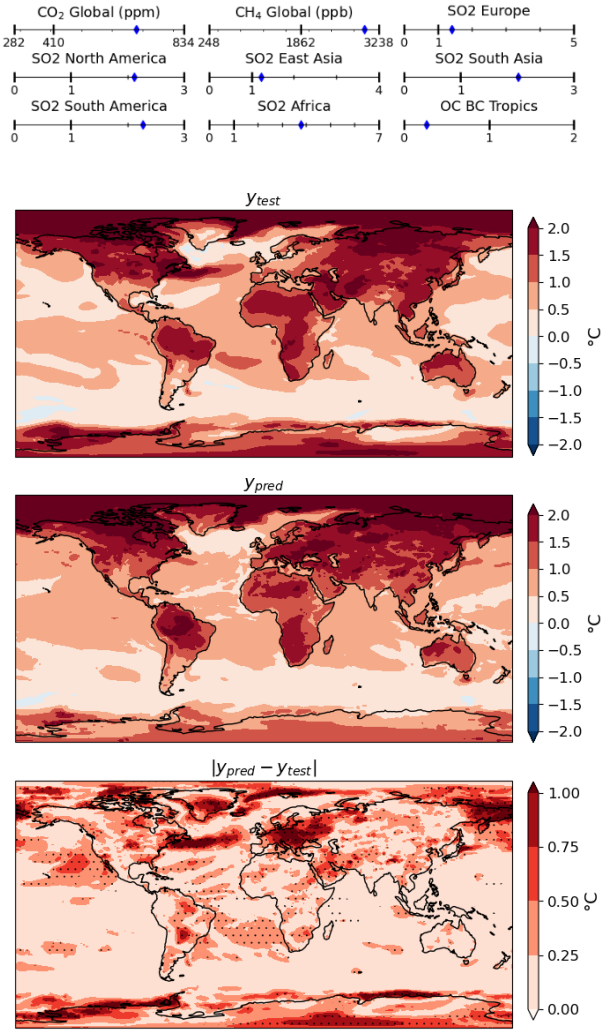
e


CO2 Global (ppm)
282 410 834
CH4 Global (ppb)
248 1862 3238
SO2 Europe
0 1 5
SO2 North America
0 1 3
SO2 East Asia
0 1 4
SO2 South Asia
0 1 3
SO2 South America
0 1 3
SO2 Africa
0 1 7
OC BC Tropics
0 1 2
ytest
ypred
|ypred − ytest|
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
°C
1.00 0.75 0.50 0.25 0.00
°C

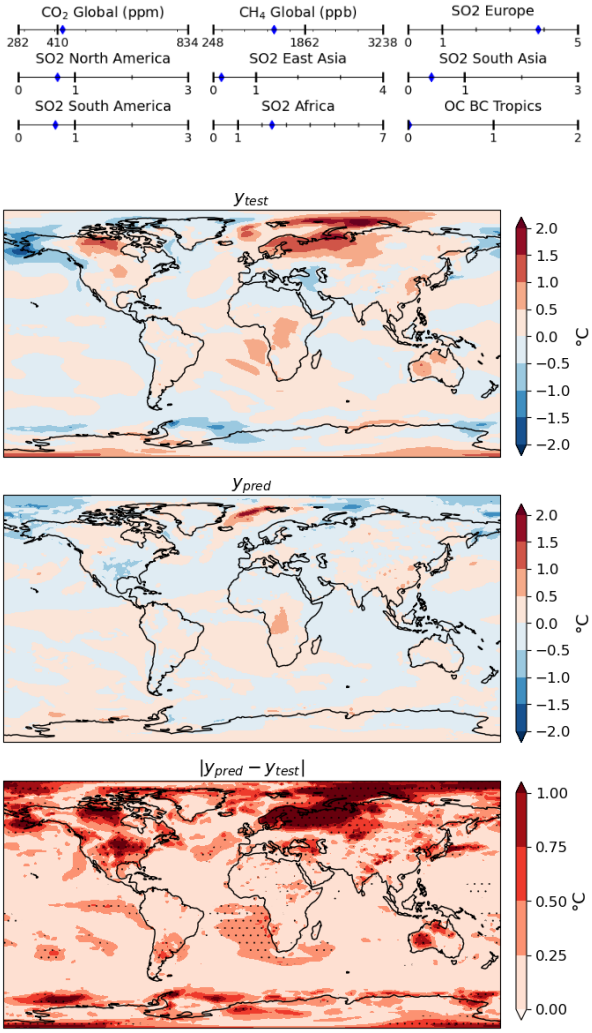
f

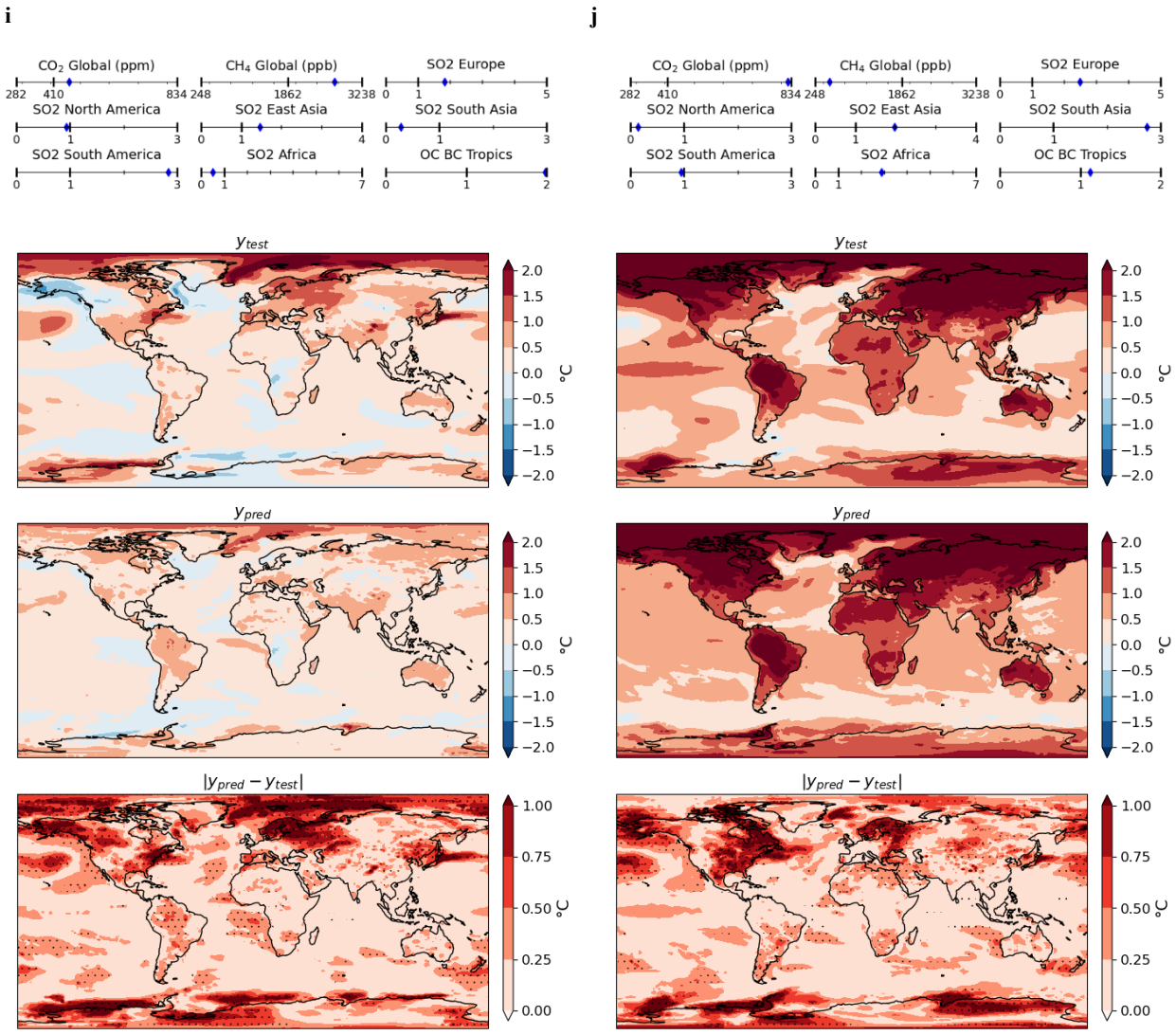

CO2 Global (ppm)
282 410 834
CH4 Global (ppb)
248 1862 3238
SO2 Europe
0 1 5
SO2 North America
0 1 3
SO2 East Asia
0 1 4
SO2 South Asia
0 1 3
SO2 South America
0 1 3
SO2 Africa
0 1 7
OC BC Tropics
0 1 2
ytest
ypred
|ypred − ytest|
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
°C
1.00 0.75 0.50 0.25 0.00
°C

g

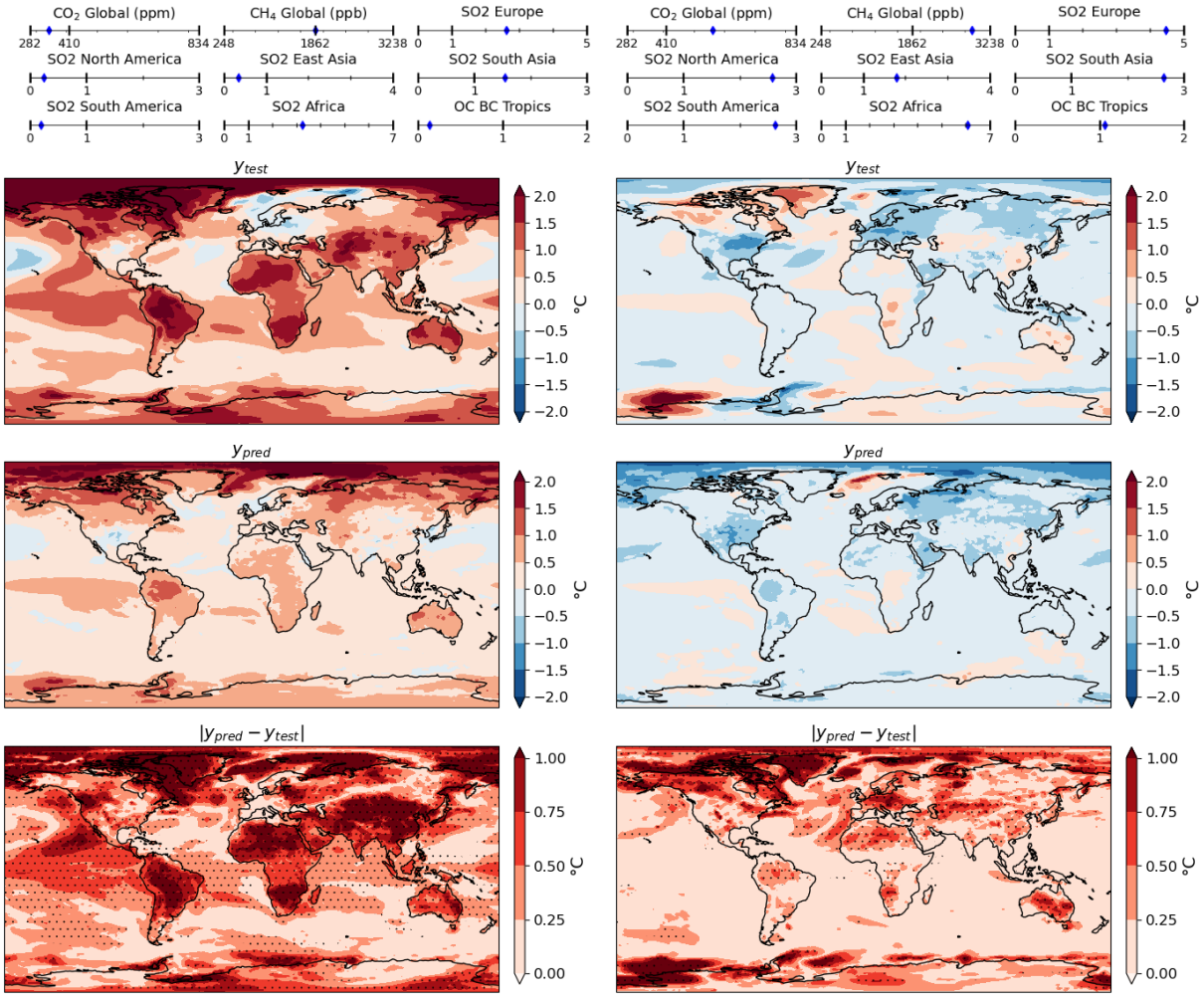

CO2 Global (ppm)
282 410 834
CH4 Global (ppb)
248 1862 3238
SO2 Europe
0 1 5
SO2 North America
0 1 3
SO2 East Asia
0 1 4
SO2 South Asia
0 1 3
SO2 South America
0 1 3
SO2 Africa
0 1 7
OC BC Tropics
0 1 2
y_{test}
y_{pred}
$|y_{pred} - y_{test}|$
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
°C
1.00 0.75 0.50 0.25 0.00
°C

h

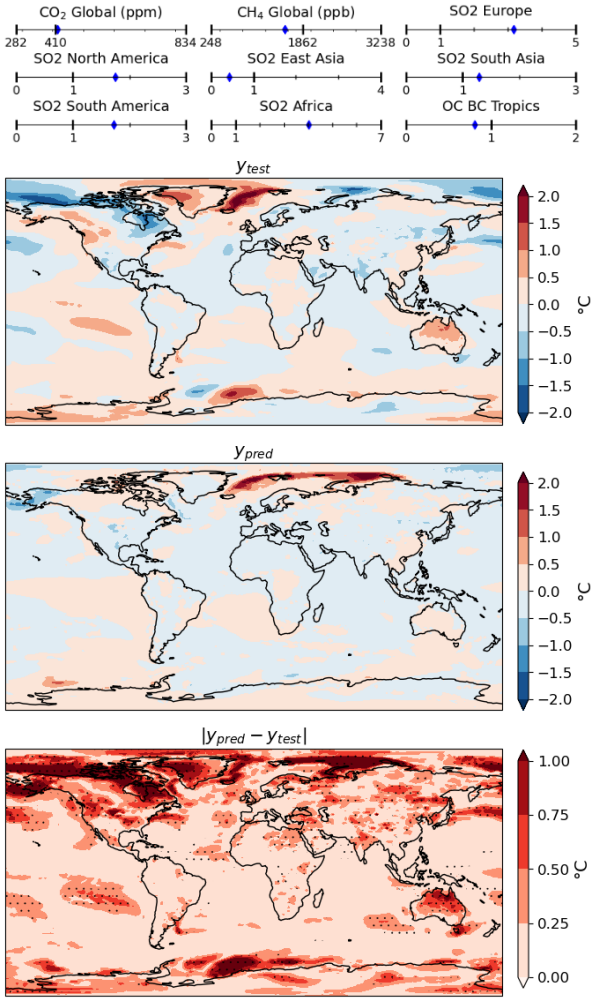

CO2 Global (ppm)
282 410 834
CH4 Global (ppb)
248 1862 3238
SO2 Europe
0 1 5
SO2 North America
0 1 3
SO2 East Asia
0 1 4
SO2 South Asia
0 1 3
SO2 South America
0 1 3
SO2 Africa
0 1 7
OC BC Tropics
0 1 2
y_{test}
y_{pred}
$|y_{pred} - y_{test}|$
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
°C
1.00 0.75 0.50 0.25 0.00
°C


i
j
CO_2 Global (ppm)
CH_4 Global (ppb)
SO2 Europe
SO2 North America
SO2 East Asia
SO2 South Asia
SO2 South America
SO2 Africa
OC BC Tropics
y_{test}
y_{pred}
$|y_{pred} - y_{test}|$
°C

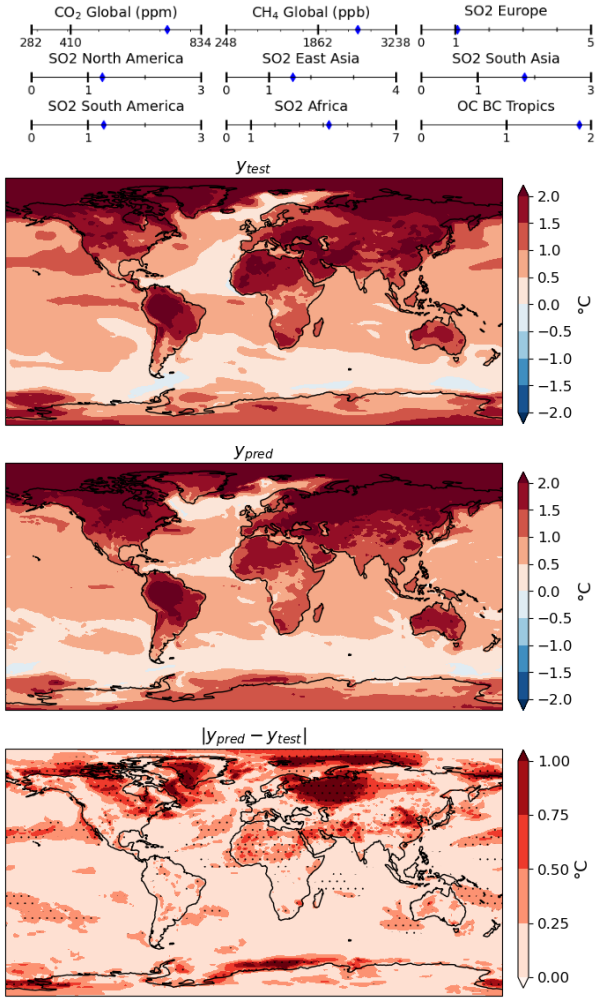
k **l**

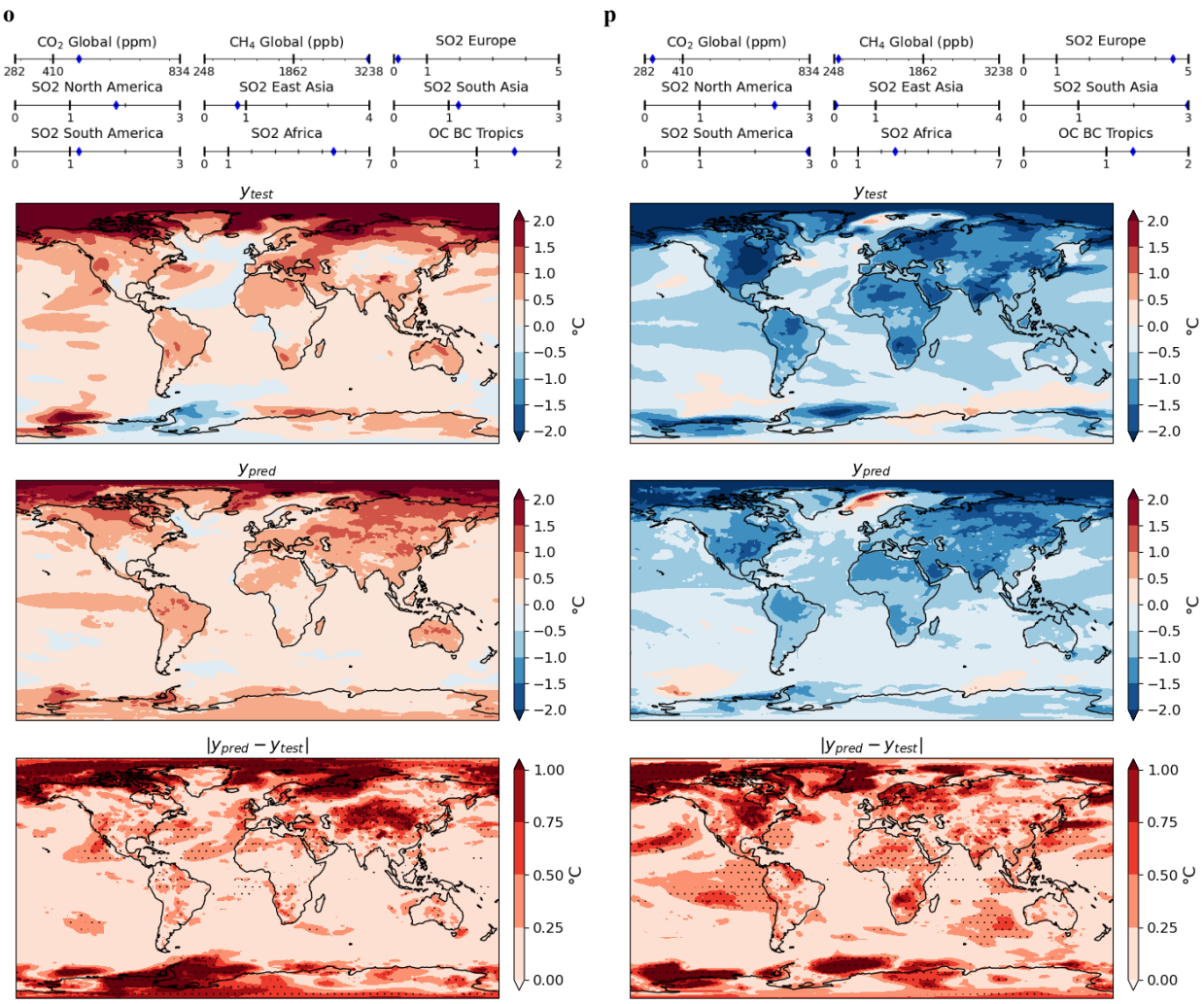

CO2 Global (ppm)
282 410 834
CH4 Global (ppb)
248 1862 3238
SO2 Europe
0 1 5
SO2 North America
0 1 3
SO2 East Asia
0 1 4
SO2 South Asia
0 1 3
SO2 South America
0 1 3
SO2 Africa
0 1 7
OC BC Tropics
0 1 2
y_{test}
y_{pred}
$|y_{pred} - y_{test}|$
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
°C
1.00 0.75 0.50 0.25 0.00
°C

m


CO2 Global (ppm)
282 410 834
CH4 Global (ppb)
248 1862 3238
SO2 Europe
0 1 5
SO2 North America
0 1 3
SO2 East Asia
0 1 4
SO2 South Asia
0 1 3
SO2 South America
0 1 3
SO2 Africa
0 1 7
OC BC Tropics
0 1 2
y_{test}
y_{pred}
$|y_{pred} - y_{test}|$
°C
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
1.00
0.75
0.50
0.25
0.00

n


CO2 Global (ppm)
282 410 834
CH4 Global (ppb)
248 1862 3238
SO2 Europe
0 1 5
SO2 North America
0 1 3
SO2 East Asia
0 1 4
SO2 South Asia
0 1 3
SO2 South America
0 1 3
SO2 Africa
0 1 7
OC BC Tropics
0 1 2
y_{test}
y_{pred}
$|y_{pred} - y_{test}|$
°C
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
1.00
0.75
0.50
0.25
0.00


o
p
CO_2 Global (ppm)
282 410 834
CH_4 Global (ppb)
248 1862 3238
SO2 Europe
0 1 5
SO2 North America
0 1 3
SO2 East Asia
0 1 4
SO2 South Asia
0 1 3
SO2 South America
0 1 3
SO2 Africa
0 1 7
OC BC Tropics
0 1 2
y_{test}
y_{pred}
$|y_{pred} - y_{test}|$
2.0 1.5 1.0 0.5 0.0 −0.5 −1.0 −1.5 −2.0
°C
1.00 0.75 0.50 0.25 0.00
°C

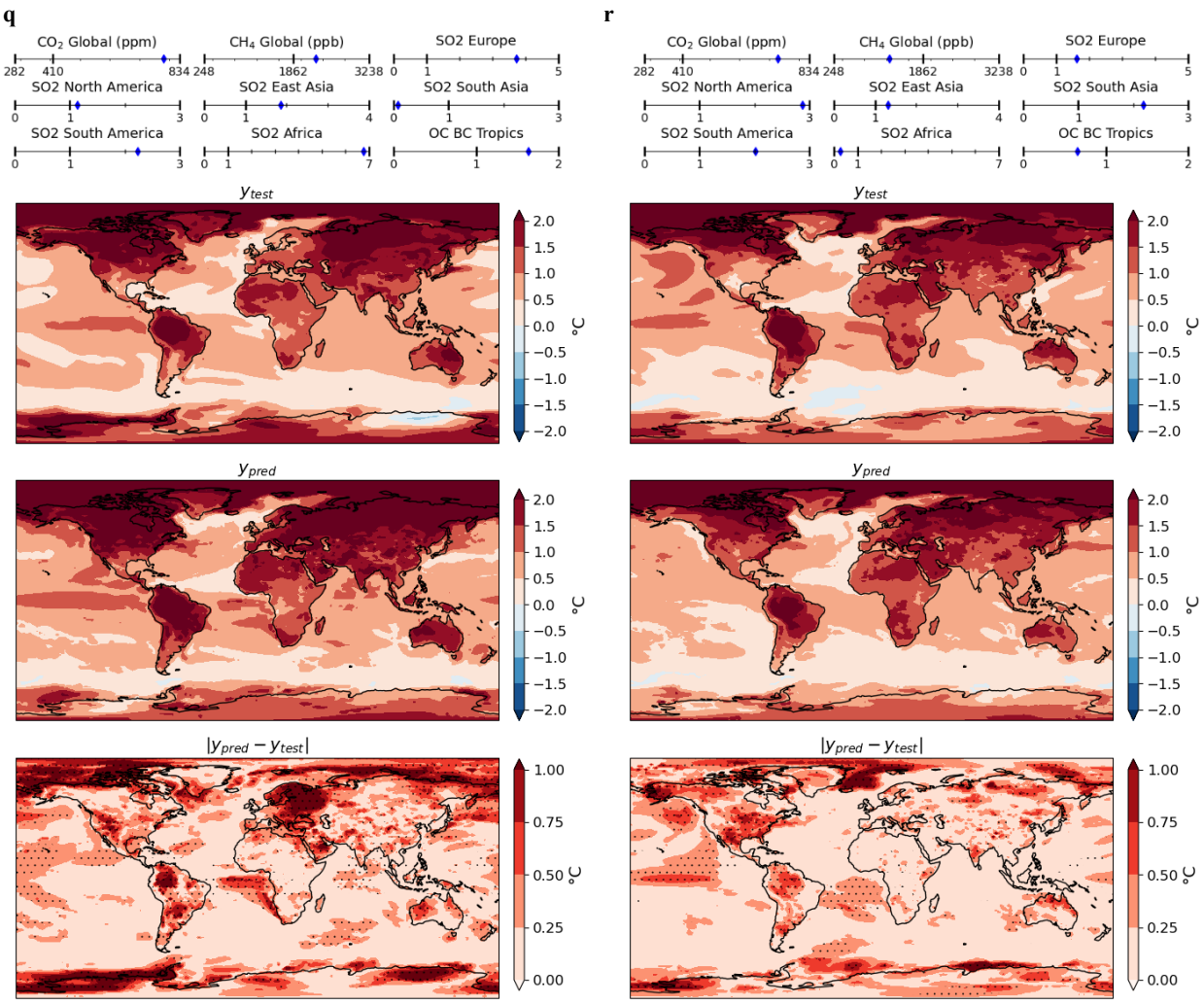


Figure S4. Absolute error maps for all example test scenarios, where top panel shows inputs values for the two scenarios shown as well as the baseline, minimum and maximum possible values. The maps show, from top to bottom, the GCM output for the test scenario, y_{test}, the GP prediction, y_{pred}, and the absolute errors, $|y_{pred} - y_{test}|$, where in the latter the stippling indicates regions where the absolute error exceeds 1 s.d. predicted by the GP emulator, $|y_{pred} - y_{test}| > \sigma_{total}$.

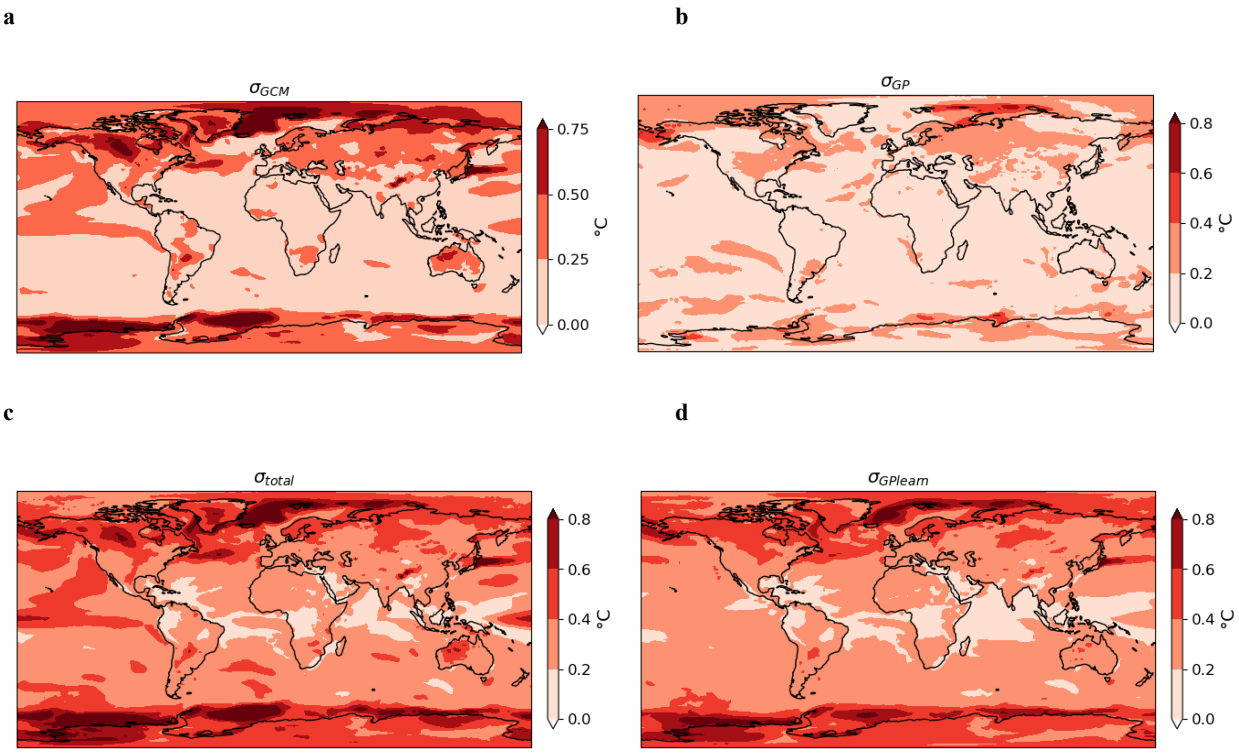


Figure S5. Different Gaussian process emulator uncertainty measures. **a)** GCM contribution to uncertainty($\boldsymbol{\sigma}_{\mathrm{GCM}}$) treated as the GCM internal variability, calculated as the average variance across 14 simulations of 5 year segments. This contribution to the uncertainty is assumed to be fixed for all possible inputs. **b)** GP contribution to uncertainty ($\boldsymbol{\sigma}_{\mathrm{GP}}$) calculated as a mean over the test dataset, **c)** Total GP uncertainty ($\boldsymbol{\sigma}_{\mathrm{total}}$) which includes contributions from the first two plots related via $\boldsymbol{\sigma}_{\mathrm{total}}^{2} = \boldsymbol{\sigma}_{\mathrm{GCM}}^{2} + \boldsymbol{\sigma}_{\mathrm{GP}}^{2}$ and **d)** GP uncertainty when we build a new emulator without specifying a fixed component from GCM uncertainty ($\boldsymbol{\sigma}_{\mathrm{GPlearn}}$). This emulator learns the uncertainty for new data points based on the distance to the training data but appears remarkably similar to the internal variability derived from the GCM, suggesting this is the main contributor to uncertainty in the GP and that is can be learned without specifying a fixed contribution term. Note that while **a)** is fixed based on all GCM simulations, **(b-d)** depend on the prediction scenario, although these were very similar across all 18 test simulations and here we show a mean across these.

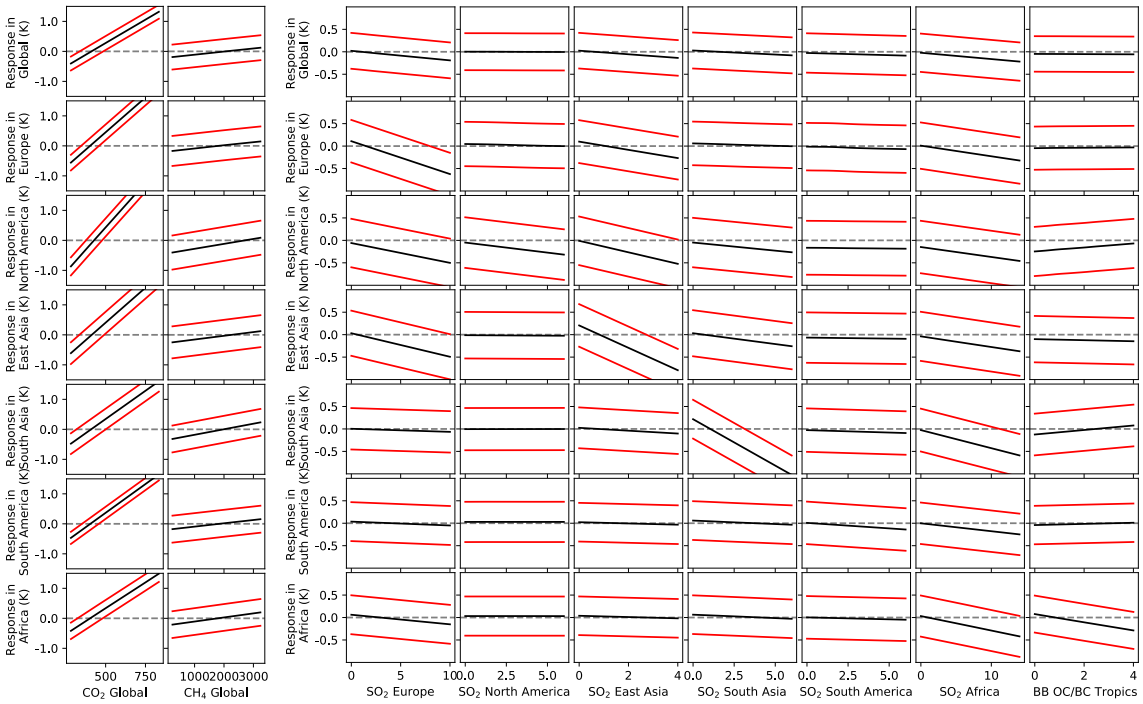


Figure S6. Main effects of each pollutant on x-axis on each region of interest on y-axis, averaging over all other pollutants. Here, 200 different realizations are run with input parameters sampled from a normal distribution centered at the present day levels with a standard deviation ¼ x the input range presented in Table 1. The black line shows the mean and the red lines show 1 standard deviation across these realizations. across these realizations and the red lines show 1 standard deviation across these different realizations. A steeper gradient indicates stronger change due to that pollutant. A discussion on the linearity can be found in Supplementary Text S1.

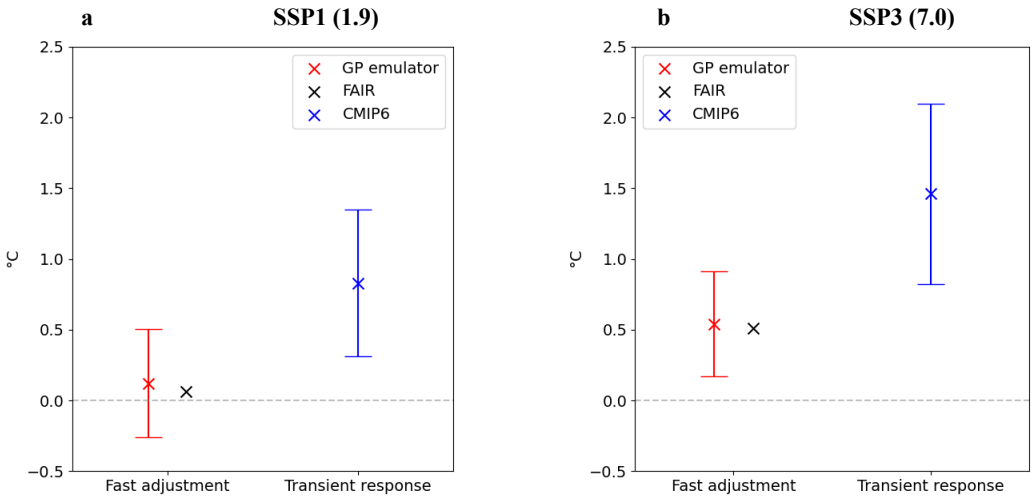


Figure S7. Global mean estimates of fast adjustment to the two different SSP scenarios, **a)** SSP1 (1.9) and **b)** SSP3 (7.0). Estimates made using the GP emulator presented here in red, including the 1 s.d. uncertainty bars. For comparison, the FAIR model is also used to predict global mean using the same set up, with a sudden transition to the 2050 emission levels from the equilibrium state set up with 2014 conditions. This shows the fast adjustment to these new 2050 conditions are in agreement with the GP emulator built here. For reference, we also show the global mean temperature relative to 1995-2014 mean estimated by the CMIP6 models in blue, where the mean response is shown by the cross with 1 s.d. uncertainty bars across all CMIP6 models. These are run under transient conditions from historical emissions, which therefore includes the response to past emissions that takes some time to be realized. Furthermore, the perturbations include all forcings, not just CO_2, CH_4, regional SO_2 and tropical biomass burning, which contribute to ~75% of the total forcing in SSP1 (1.9) and ~80% of the total forcing in SSP3 (7.0) (Gidden et al., 2019). The remaining forcing comes from other sources such as tropospheric ozone, N_2O and Montreal Protocol gases, most of which are positive forcings and would lead to a slight additional warming effect.