Understanding pattern scaling errors across a range of emissions pathways

Christopher D. Wells[1,2], Lawrence S. Jackson[1,2], Amanda C. Maycock[1,2], Piers M. Forster[1,2]

[1]Institute of Climate and Atmospheric Science, University of Leeds, Leeds, UK
[2]Priestley International Centre for Climate, School of Earth and Environment, University of Leeds, Leeds, UK

Correspondence to: Christopher D. Wells (c.d.wells@leeds.ac.uk)

Abstract. The regional climate impacts of hypothetical future emission scenarios can be estimated by combining Earth System Model simulations with a linear pattern scaling model such as MESMER, which uses estimated patterns of local response per degree global temperature change. Here we use the mean trend component of MESMER to emulate the regional pattern of surface temperature response based on historical single-forcer and future Shared Socioeconomic Pathway (SSP) CMIP6 simulations. Errors in the emulations for selected target scenarios (SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-8.5) are decomposed into two components: differences in scaling patterns between scenarios as a consequence of varying combinations of external forcings, and intrinsic timeseries differences between the local and global responses in the target scenario. The timeseries error is relatively small for high-emissions scenarios, contributing around 20% of the total error, but is similar in magnitude to the pattern error for lower-emission scenarios. This irreducible timeseries error limits the efficacy of linear pattern scaling for emulating strong mitigation pathways and reduces the dependence on the predictor pattern used. The results help guide the choice of predictor scenarios for simple climate models and where to target introducing other dependent variables beyond global surface temperature into pattern scaling models.

1 Introduction

Anthropogenic climate change has already driven significant impacts throughout the globe, and these will continue to become more severe (IPCC, 2021). Estimates of the climatic impacts of future emissions depend on several sources of uncertainty – internal variability, model structural uncertainty, and unknowns in the emissions themselves (Hawkins & Sutton, 2009). The first two sources of uncertainty are often explored using multi-member ensembles of Earth System Models (ESMs). The emissions uncertainty can only be explored by constructing multiple hypothetical future emissions scenarios, and investigating their respective impacts (Riahi et al., 2022).

The most recent generation of Integrated Assessment Model (IAM)-simulated scenarios are the Shared Socioeconomic Pathways (SSPs) (Gidden et al., 2019). These have been used in ESMs to assess the future climate response in the contribution of Working Group 1 to the 6th Assessment Report of the Intergovernmental Panel on Climate Change (Lee et al., 2021). ESMs are computationally expensive to run, so in general they can only simulate a handful of future emission scenarios (O'Neill et al., 2016, Tebaldi et al., 2021). The small number of scenarios used for ESM simulations can mask uncertainties in future

pathways as the full range of plausible emissions parameter space is not adequately sampled (Grubler et al., 2018; Otero et al., 2020; Partanen et al., 2018). Accurate simplified methods which allow a good understanding of the regional impacts of novel emissions scenarios are therefore highly motivated, to explore a broader range of emissions pathways than possible with comprehensive ESMs.

Both the IPCC WGI and WGIII AR6 reports used simplified physical climate model emulators trained on more complex ESM simulations to assess possible future projections of global surface temperature (Lee et al., 2021; Forster et al., 2021; Riahi et al., 2022; Kikstra et al., 2022). However, the assessment of regional climate projections relied largely on simulations from ESMs and Regional Climate Models (RCMs). Since the UNFCCC 2015 Paris Agreement, which frames international climate goals in terms of global temperature targets, there has been growing emphasis on examining climate change impacts as a function of global warming level (IPCC, 2018). For example, in IPCC AR6 future scenarios simulated by ESMs were compared at different global mean temperature levels (Lee et al., 2021; Seneviratne et al., 2021). The suitability of this approach depends on the extent to which regional changes scale with global temperature change, which in turn depends on the variable of interest and the details of the GHG and aerosols scenario. Regional statistical emulation tools operate on a similar basis, using techniques such as pattern scaling to translate simple large-scale information such as global mean temperature into estimates of the spatially-resolved climate response for a broad range of scenarios (James et al., 2017; J. F. B. Mitchell et al., 1999; Osborn et al., 2018).

Pattern scaling approaches can be limited by systematic response variations between and within scenarios. The radiative forcing from long-lived species is independent of the emission location, but the impacts of short-lived species such as aerosols strongly depend on the pattern of emissions, with typically larger effects locally (L. Liu et al., 2018; Persad & Caldeira, 2018). The patterns of climate change impacts also differ between transient and equilibrium climates – the transient temperature response is typically larger over land than over ocean, due to the greater thermal inertia of the latter (Herger et al., 2015; Huang et al., 2020; King et al., 2020; T. D. Mitchell, 2003). These two factors – variations in the forcing pattern and the level of disequilibrium – can be expected to break the linearity assumed within pattern scaling if they differ between and within scenarios (Goodwin et al., 2020; Herger et al., 2015; James et al., 2017; T. D. Mitchell, 2003; Osborn et al., 2018). Pattern scaling has been modified to partially address these issues using patterns varying between forcers (Kravitz et al., 2017; Xu & Lin, 2017; Schlesinger et al., 2000) and response timescale (Zappa et al., 2020).

Other nonlinearities in climate response can occur, especially under higher emissions scenarios (Lopez et al., 2014); for example, the removal of sea ice in the Arctic will saturate the sea ice albedo feedback and reduce the local temperature sensitivity (Huang et al., 2020; Lynch et al., 2017), though higher sensitivity may initially occur due to sea ice thinning (Ishizaki et al., 2012). Some responses of the climate system to external forcing such as sea ice retreat and intertropical convergence zone shifts move geographically and will therefore be poorly represented by pattern scaling (Herger et al., 2015).

Despite the potential limitations of pattern scaling for climate emulation, the technique has been shown to work well across a range of scenarios (Alexeeff et al., 2018; Beusch et al., 2020; J. F. B. Mitchell et al., 1999). A linear approximation has been found to be reasonable for regional temperature changes within and between sets of scenarios (Seneviratne et al., 2016;

Seneviratne & Hauser, 2020), with similar response patterns found within different emissions scenarios in the near-term (Lee et al., 2021). The variation in response patterns between scenarios is typically less than the variation between models, indicating that the errors arising from pattern scaling are smaller than model uncertainty (Goodwin et al., 2020; Herger et al., 2015; Osborn et al., 2016, 2018; Tebaldi & Arblaster, 2014; Tebaldi & Knutti, 2018). Pattern scaling errors are generally substantially larger when "extrapolating" (projecting a scenario with higher forcing than the data used to generate the pattern) than when interpolating (Herger et al., 2015; T. D. Mitchell, 2003; Tebaldi et al., 2020; Beusch et al., 2022a), with smaller errors for more modest forcing differences between the predictor and target scenarios (Osborn et al., 2018).

Pattern scaling has been used to emulate regional changes in temperature (Beusch et al., 2020; Link et al., 2019), to forecast temperature and precipitation simultaneously (Snyder et al., 2019), and to study changes in extreme precipitation (Thackeray et al., 2022). The application of pattern scaling to precipitation is complicated relative to temperature by the larger role of internal variability (Hawkins & Sutton, 2011), the presence of strong nonlinearities and local factors (G. Liu et al., 2022), and the role of forcing-specific adjustments (Myhre et al., 2018). However, extreme precipitation is more closely constrained by moisture availability and may be more successfully emulated through pattern scaling (Pendergrass et al., 2015; Sillmann et al., 2017). Pattern scaling has also been incorporated with Earth system components such as land surface models to make faster projections (Zelazowski et al., 2018). It has been applied to estimate the regional effects of single-country greenhouse gas emissions in order to attribute their local temperature impacts (Beusch et al., 2022b), and to estimate the country-level economic impacts attributable to each other country's CO_2 emissions (Callahan & Mankin, 2022). These exercises would require vast computer resources and time if they were attempted with ESMs.

The scenarios to which pattern scaling has generally been successfully applied typically have smaller inter-scenario variation in aerosol emissions and warming rates than more recent, and likely future, scenarios of interest. Many historical studies applied pattern scaling to the CMIP5-era Representative Concentration Pathway (RCP) scenarios (Alexeeff et al., 2018; Goodwin et al., 2020; Herger et al., 2015; Ishizaki et al., 2012; Kravitz et al., 2017; Lynch et al., 2017; Osborn et al., 2018; Tebaldi et al., 2020; Tebaldi & Arblaster, 2014; Xu & Lin, 2017), which all exhibit similar decreases in anthropogenic aerosol emissions into the future (Gidden et al., 2019), a much narrower range than projected among the newer SSP scenarios used in CMIP6. This may make the SSP scenarios less amenable to pattern scaling than prior scenarios (Goodwin et al., 2020). The SSP scenarios include SSP-1.19, approaching the 1.5ºC level under the Paris Agreement (Meinshausen et al., 2020), with stronger mitigation than the RCPs. Many low-emissions Paris agreement-consistent scenarios were assessed as part of IPCC AR6 WGIII (IPCC, 2022), but relatively few have been systematically studied in multi-ESM projects. Low-emission scenarios may also encounter issues due to contamination by internal variability in the pattern-generation regression, and stabilisation scenarios may be more susceptible to physical nonlinearities (Osborn et al., 2018). Indeed, many studies that find pattern scaling to be accurate with earlier scenarios note that under stronger mitigation or wider ranges in aerosol emissions, the technique would be less effective (Alexeeff et al., 2018; J. F. B. Mitchell et al., 1999; Tebaldi & Arblaster, 2014).

While tools such as MESMER (Modular Earth System Model Emulator with spatially Resolved output) have been developed to implement pattern scaling using the SSP scenarios (Beusch et al., 2020), particularly studying the reproduction of scenarios

from their own pattern (self-emulation), a systematic analysis of the range of errors associated with the application of pattern scaling to temperature within the SSPs remains to be done. Multi-model studies analysing pattern scaling efficacy for low-emissions scenarios are also lacking (Tebaldi & Knutti, 2018). The effect of the choice of predictor data used to generate the pattern utilised in pattern scaling has also not been fully explored. This paper takes steps to address these gaps, through a novel decomposition of pattern scaling errors into their relation to pattern scaling assumptions.

This paper studies the effects of the pattern scaling assumptions on regional temperature projections, decomposing the pattern scaling error into two components relating to space and time (see Section 2.3):

i) The Pattern Assumption: Pattern Scaling assumes that the pattern of change is constant between all scenarios, regardless of the mix and level of forcers within them.

ii) The Timeseries Assumption: Pattern Scaling assumes that the timeseries response at each location follows the same shape as the global response, simply modified by the local sensitivity (i.e. the pattern), thus allowing the local change timeseries to be estimated by scaling the global timeseries by a constant local pattern coefficient. This can be thought of as assuming the pattern is constant in time within a given scenario.

We explore how these errors vary when projecting different emissions scenarios, and the effect of different choices of (one or more) predictor scenario(s), to determine the impacts of this decision on emulation accuracy. Inter-model variation is investigated for all the impacts studied.

Section 2 sets out the ESM data utilised and the model used to perform the pattern scaling analysis, and shows how to diagnose the error associated with each pattern scaling assumption. Section 3 presents the key results, Section 4 explores the implications for the application of pattern scaling, and Section 5 provides discussion and conclusions.

2 Methods

2.1 Earth System Model Data

Two sets of emissions pathways are the focus of this paper. To understand the effect of different forcers on the warming pattern and pattern scaling errors, two historical (1850-2020) scenarios from the Detection and Attribution Model Intercomparison Project (DAMIP) (Gillett et al., 2016) project are used: hist-aer, which includes only anthropogenic aerosol emissions, and hist-GHG, which includes only greenhouse gas emissions. This allows for an idealised comparison of the different patterns attributable to historical levels of different forcers, although neither represents a realistic emissions pathway due to the co-emission of aerosols and GHGs in reality. To determine the difference between warming patterns amongst coherent emissions pathways, the SSP scenarios are used, examining several SSPs but focussing on SSP1-1.9 and SSP5-8.5, with data for both taken from 2015-2100.

Data for all scenarios were taken for annual mean temperatures from the cmip6-ng database (Brunner et al., 2020), which re-grids all data to a common 2.5°x2.5° latitude-longitude grid to allow for inter-model comparison. For each of the two sets of

emissions pathways, all models with at least one member of each experiment were used: 10 ESMs for the two DAMIP scenarios and eight for the five SSPs. The models and number of members for each scenario are given in Supplementary Tables S1 and S2. Inter-model results are averaged first over each model ensemble, with this model average then compared, to avoid weighting by the ensemble size of each model.

2.2 Pattern Scaling Methodology

This study utilises the mean response component of the MESMER model (Beusch et al., 2020), implementing pattern scaling to emulate the spatial annual mean temperature response in a scenario. While both pattern scaling and the timeshift method – which selects a window of data centred around the year in which the global average reaches a desired Global Warming Level – generate accurate emulations out-of-sample mean emulations, pattern scaling has been found to perform slightly better in some metrics (Tebaldi et al., 2020). The pattern is derived from linear regression of the local response timeseries on the global response at each gridcell; the full timeseries is regressed for each pattern (i.e. 1850-2020 for hist-aer and hist-GHG, and 2015-2100 for SSP1-1.9 and SSP5-8.5), with anomalies calculated relative to the first 50 years of each experiment. This linear regression approach to pattern scaling has been shown to provide more accurate patterns than the alternative "delta" method, whereby the average climate towards the end of a scenario is subtracted from that in the early period (Lynch et al., 2017; T. D. Mitchell, 2003). In the default configuration of MESMER, the raw annual gridcell-level data is regressed against smoothed global temperatures, but here this is modified to use the same smoothing on both local and global temperatures. This smoothing prior to regression is performed to ensure that the global average scaling factor is very close to unity (1 K/K), when applying the regression to an individual low-emission scenario such as SSP1-1.9. The weighted global average regression parameter should be 1 by construction when using the global mean of a variable to target the local response of the same variable (in this case temperature), since the global mean is simply the weighted average of the local values. When regressing unsmoothed local data against smoothed global temperatures in a low-emission scenario that exhibits a peak and decline in temperatures, the regression parameter can be artificially enhanced due to a smoothing of the peak. We therefore use the same smoothing on local and global temperatures before applying the regression. The smoothing performed is Locally Weighted Scatterplot Smoothing (LOWESS), which takes the weighted average of the timeseries across a moving window. The weighting is tricube, and the window fraction set to the default MESMER value of 50 divided by the number of timesteps as in the default case – when using the annual mean SSP data here, this is 50/85 = 0.6. MESMER's default version only includes land gridpoints, to focus on land impacts, but here all gridpoints were used, to study the broader response. The intercept of the linear regression is zero in theory, but while generally small in practice is non-zero and method-dependent (Beusch et al., 2020), and is added to the emulation in MESMER. For calculation of the pattern in a model with multiple ensemble members, MESMER applies the regression across the data from all the members simultaneously. The full MESMER emulator includes a representation of internal variability at global and local scales, but since this paper is focuses on the long-term response, only the pattern scaling ("local trends") component is used (Beusch et al., 2020).

A given emulation consists of two components: the predictor pattern and the target scenario. The predictor pattern can be derived from one scenario or many, with the regression applied across the full dataset. The target is a single scenario in each case. The pattern is derived via the linear regression of the local on global temperature anomalies, relative to the first 50 years of each predictor scenario. This pattern is then multiplied by the global temperature timeseries of the target scenario to generate the emulation. The difference between this emulation and the actual ESM pattern is defined as the pattern scaling error. The first 50 years of a given scenario are used as the baseline. The pattern scaling error is zero in the global mean by design – since the pattern (with average value 1) is simply scaled by the global temperature in the ESM – but errors occur regionally, and the global average of the local absolute error will therefore not be zero.

2.3 Decomposing pattern scaling errors

As described in Section 1, the pattern scaling error can be thought of as deviations from two key assumptions: the pattern and timeseries assumptions. Short-term inter-annual variability is dampened via the LOWESS smoothing, though decadal-scale variability will also be present.

Figure 1 illustrates the decomposition of pattern scaling errors into the pattern and timeseries errors. Pattern scaling determines the local parameter from the regression of the local on global predictor data, and then scales the target global temperature by this value. The pattern scaling error is then the difference between this projection and the actual local response in the target scenario.

Figures 1a-c show “perfect emulation”, whereby all four of these timeseries are identical. The scaling parameter is therefore equal between the scenarios (Figure 1a), and the shape of the emulated response is identical to the target dataset (Figure 1b), with zero error at all times (Figure 1c).

The second row in Fig. 1 shows the effect of altering the shape of the local timeseries, but keeping the warming parameter (i.e. the pattern) the same. The scenarios are still identical - the global response is the same for both, as is the local response - but the shape of the local and global responses within each scenario now differs. In this case, the pattern is “correct”, and the error is conceptualised as the “timeseries error”; that is, the error due to the differing local and global timeseries response within the scenario. The time-mean is not zero, as timeseries variability within the scenario modifies the pattern, leading to a non-zero regression intercept and overall emulation errors.

In Figures 1g-i the local and global responses are now identical within each timeseries, but the scenarios themselves are different - the local parameter of the predictor is larger than in the target. Since the local and global timeseries in the target scenario follow the same shape, the error is purely due to differences in the scaling parameter, i.e. the local value of the pattern. This error is therefore the “pattern error”.

Finally, Figures 1j-l apply both of these changes simultaneously; the local parameter differs between the scenarios, and the local and global timeseries differ within each scenario, as expected in real-world and ESM data due to nonlinearities within the climate system. The total error is comprised of both pattern and timeseries errors, but the split between them is not clear from the timeseries.

Figure 1 illustrates that the general total error, from using one scenario to emulate another, can be decomposed into pattern and timeseries associated components. The timeseries error, in row two of Figure 1, was generated by using the same scenario as the predictor and the target, termed "self-emulation". The error is due to the internal dynamics of the response; specifically, the difference between the shape of the local and global temperature timeseries. This error is therefore intrinsic to the target scenario. For a given predictor-target pair, the timeseries error can be found by calculating the target-target pattern scaling error, i.e. the error upon "self-emulation" of the target. This can then be subtracted from the original predictor-target pattern scaling error to determine the pattern error. The contribution of each error to the total error can then be studied; the error timeseries in the bottom row of Figure 1 shows this decomposition applied to the idealised scenarios.

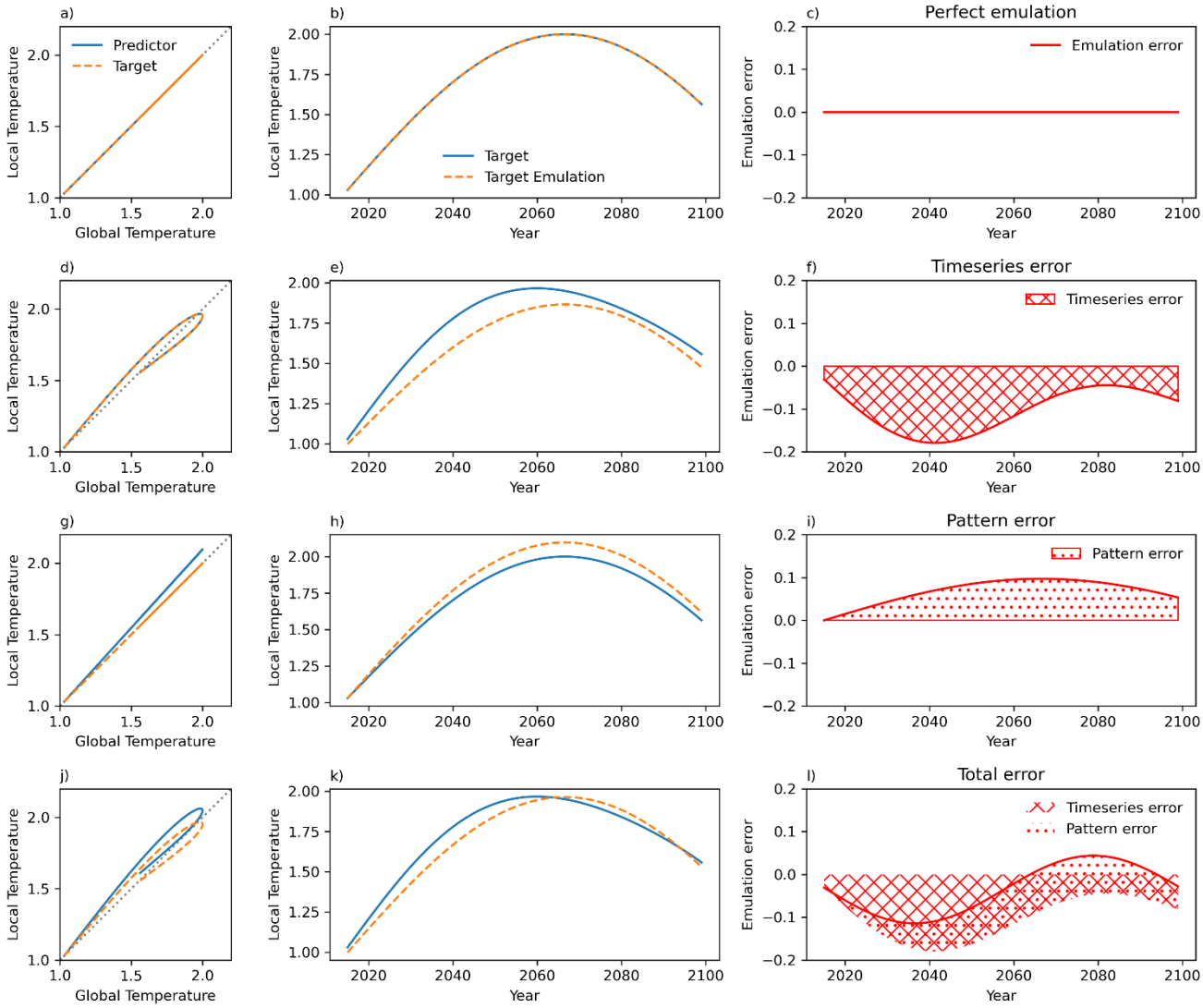

Figure 1: Demonstrating the pattern scaling error decomposition for several idealised scenarios (see text for details). An idealised scenario is shown whereby temperatures relative to pre-industrial times rise from 1 K in 2015 to 2 K by 2070, and fall to nearly 1.5 K by 2100. Each row represents a different relationship between the global and local temperature for an arbitrary location, with the left panel indicating the regression of local onto global temperatures, the middle panel the actual and emulated local temperature trajectories, and the right panel the emulation error at this location. Left: local temperature timeseries against global. Middle: timeseries of the target scenario local response and the emulated local projection. Right: pattern scaling error timeseries, i.e. the difference between the emulation and the target scenario in the middle column.

3 Results

3.1 Effect of pattern error

In this section, the assumption that the pattern is independent of the predictor scenario is investigated, first in the DAMIP experiments and then in the SSP scenarios.

Figures 2a and b shows the multi-model mean hist-aer and hist-GHG response pattern based on regression across the whole period (1850-2020). Note that while aerosols drive a cooling, since the local response is regressed against the global mean the sign cancels, with the pattern giving the sensitivity under warming or cooling. The dominant canonical pattern of the hist-GHG response is greater sensitivity over land than ocean, expected due to the lower heat capacity of the land surface and lower capacity for evaporative cooling (Byrne & O'Gorman, 2018; Lee et al., 2021). The Arctic exhibits a strong amplification (Holland & Bitz, 2003). In hist-aer, the land-ocean distinction is still clear, but the Northern Hemisphere land exhibits a particularly strong response relative to the global mean, due to the historical concentration of aerosol emissions within this region. The difference between the patterns (Figure 2c) is over 0.5 K in the Northern Hemisphere mid-latitudes (NHMLs), and 1 K over high-aerosol parts of Asia. Because the pattern averages to 1 globally, the larger parameter in the NHMLs in hist-aer leads to relatively weaker parameters more remotely, over the Southern Ocean and Antarctica. The strongest differences are larger than those typical between patterns found in prior work, which are usually around 0.4 K or less (Huang et al., 2020; Ishizaki et al., 2012; Lynch et al., 2017; J. F. B. Mitchell et al., 1999; T. D. Mitchell, 2003); this is to be expected due to the complete separation of forcers and their associated patterns. There is substantial inter-model variability, but broad areas still see a difference larger than one inter-model standard deviation (Figure 2d). There is inter-model agreement on a larger response over parts of the NHMLs in hist-aer, including the USA, Europe, and east Asia, with the mean pattern difference larger than the inter-model deviation. There is also agreement on the consequently weaker Southern Hemisphere ocean response. Despite the Indian subcontinent experiencing a large magnitude difference (i.e. more sensitive to aerosols historically), the large variability in the aerosol sensitivity (see Supplementary Figure S1), potentially due to model variation in the monsoon response to aerosols, leads to no inter-model agreement on this.

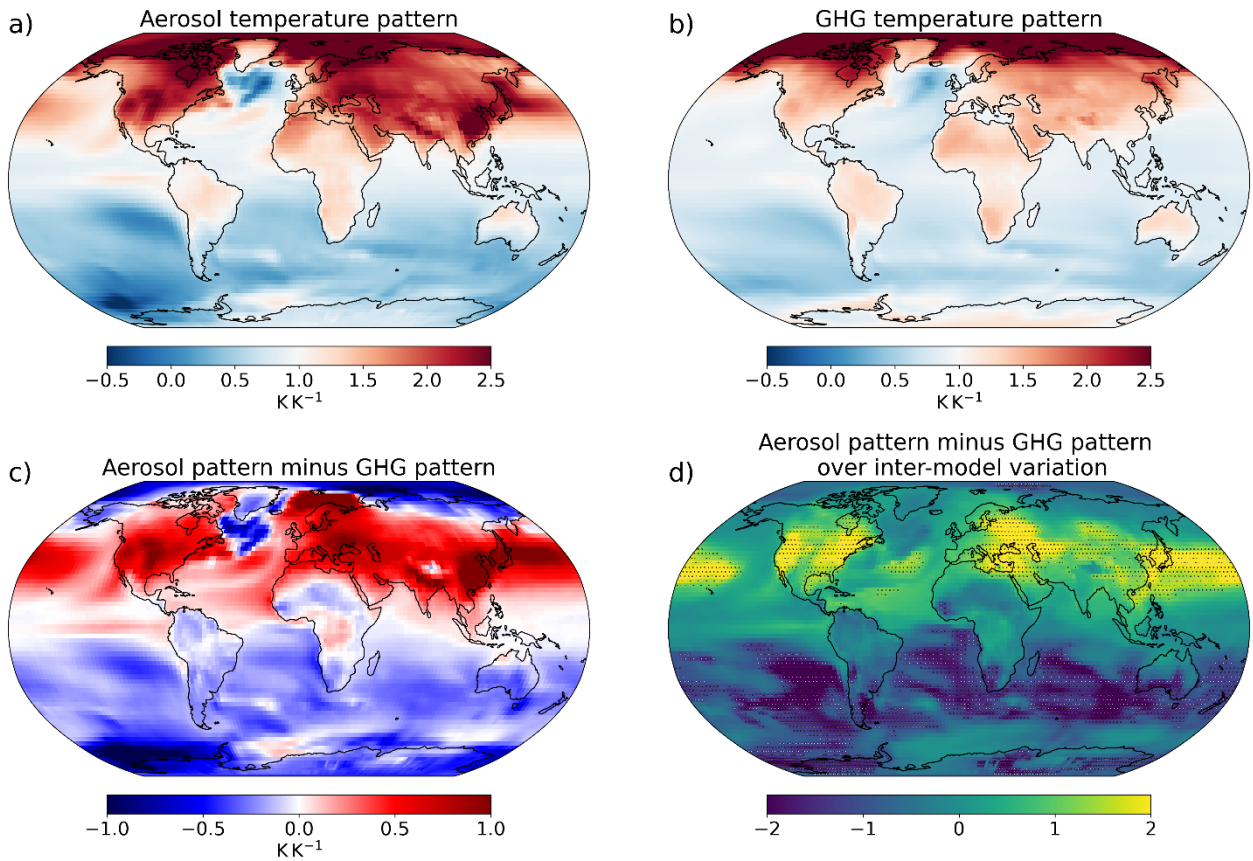

Figure 2: Mean warming patterns across 10 ESMs derived from historical GHG-only (hist-GHG; panel a) and aerosol-only (hist-aer; b) simulations, the aerosol pattern minus that from GHGs (c), and the magnitude of this difference divided by the inter-model standard deviation in this difference (d). Panel d shows stippling where the multi-model mean difference is greater than one inter-model standard deviation. The colour scheme in a) and b) diverges around 1 (the global average value), with regions which experiences temperature changes weaker than the global mean shown in blue, and areas with a stronger response in red.

Figure 3 applies the same analysis to SSP1-1.9 and SSP5-8.5; the local temperatures are again regressed onto the global response to generate the response pattern. The SSP5-8.5 derived pattern shows greater sensitivity over land, consistent with the higher warming rate maintained through the century in this scenario, and a less sensitive Arctic, likely due to a saturation of the Arctic sea ice feedback (Huang et al., 2020; Lynch et al., 2017). Overall the pattern difference is similar to the transient minus equilibrium patterns found in previous studies (Herger et al., 2015; Huang et al., 2020; King et al., 2020; T. D. Mitchell, 2003). This suggests that the difference between spatial patterns of warming in SSP1-1.9 and SSP5-8.5 may be driven primarily

by the differences in disequilibrium rather than aerosol emissions; aerosols can be expected to play a relatively larger role in the pattern difference between scenarios closer in radiative forcing and/or with larger aerosol emissions differences. The lower temperature sensitivity in East Asia under SSP5-8.5 may be linked to the weaker reduction in aerosols there than in SSP1-1.9, resulting in less "unmasking" of the cooling effect. Some features of the SSP5-8.5-SSP1-1.9 pattern difference vary from the RCP8.5-RCP2.6 differences found by Ishizaki et al. (2012), who found a more sensitive Arctic under the higher emissions scenario. This they suggested may be attributable to stronger ice melt overall under RCP8.5, due to thinning of the sea ice under warming. This highlights the contingency of the local sensitivity on the baseline climatology; further analysis of these background conditions in the ESMs may aid in explaining the differences (Lynch et al., 2017), but this is beyond the scope of this paper. Comparing to Figure 2, the pattern differences are typically smaller than those between hist-GHG and hist-aer, with most areas seeing differences less than 0.3 K, consistent with differences between scenario patterns in prior studies (Huang et al., 2020; Ishizaki et al., 2012; Lynch et al., 2017; J. F. B. Mitchell et al., 1999; T. D. Mitchell, 2003).

Compared to the DAMIP comparison, the SSP1-1.9 and SSP5-8.5 pattern difference is not as robust between models (Figure 3d compared to Figure 2d), reflecting both a similarity between the scenario patterns and the extent of inter-model variation. This is expected due to the narrower differences in forcers between scenarios and consistent with prior work on pattern differences between scenarios (Goodwin et al., 2020; Herger et al., 2015; Osborn et al., 2016, 2018; Tebaldi & Arblaster, 2014; Tebaldi & Knutti, 2018). Supplementary Figure S2 shows the analysis for all the combinations of the five SSPs analysed in this study; generally, differences between SSPs closer in radiative forcing show fewer coherent differences in their spatial patterns, in agreement with Osborn et al. (2018).

Clear differences, larger over broad areas than the inter-model standard deviation, are therefore found between the temperature response patterns attributable to different historical forcers, consistent with their different spatial patterns. These differences are less systematic across ESMs for the future scenarios analysed, which are likely relatively more affected by their differing warming rates.

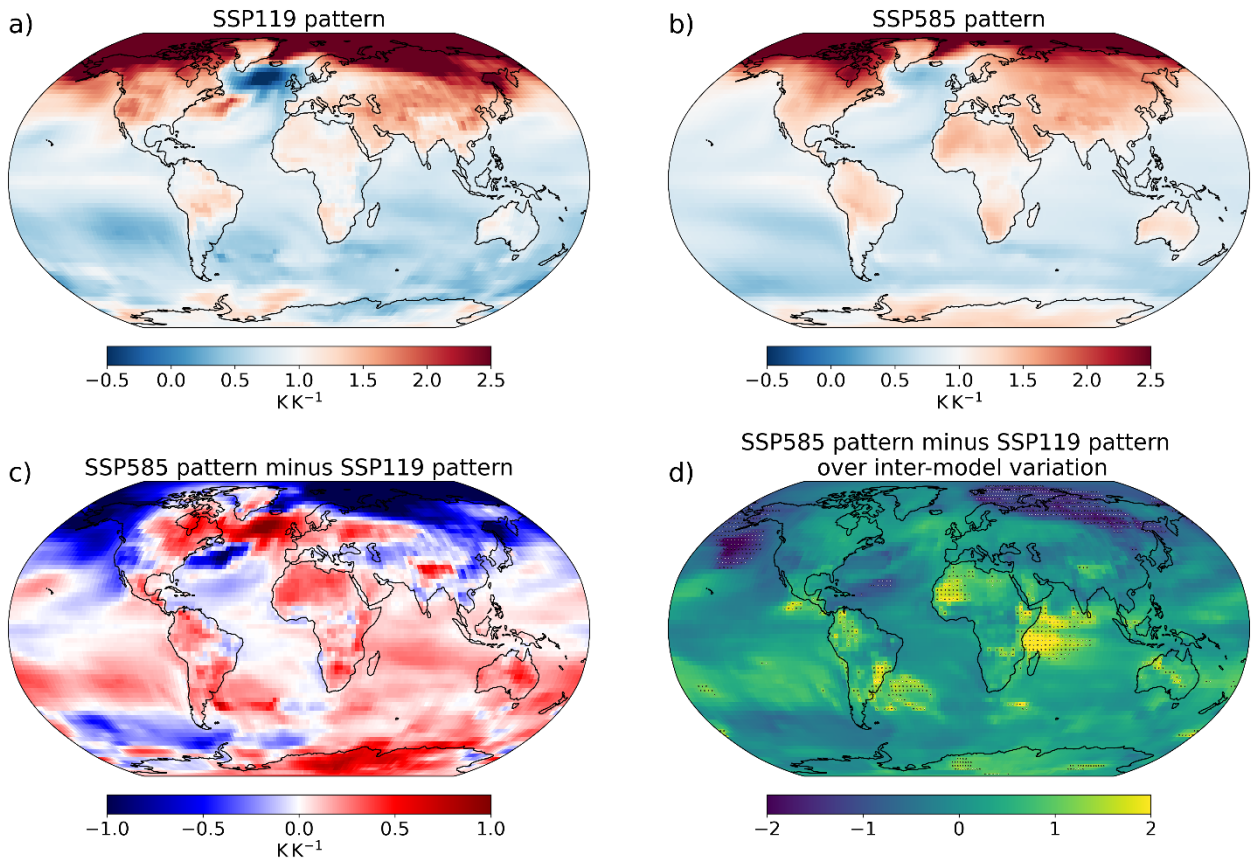

Figure 3: Mean warming patterns across 8 ESMs derived from SSP1-1.9 (top left) and SSP5-8.5 (top right) simulations, the SSP5-8.5 minus that from SSP1-1.9 (bottom left), and the magnitude of this difference divided by the inter-model standard deviation in this difference (bottom right). Panel d shows stippling where the multi-model mean difference is greater than one inter-model standard deviation.

3.2 Total pattern scaling error

The total emulation error arising from the use of the two DAMIP-derived patterns (hist-aer and hist-GHG, Fig. 2) to separately emulate both DAMIP experiments are displayed in Figure 4.

Out-of-sample errors are generally substantially larger than self-emulation (Osborn et al., 2018), since out-of-sample emulations introduce pattern errors in addition to the timeseries error. The out-of-sample emulations are both too warm over the NHMLs and too cool in the Southern Hemisphere, in keeping with the pattern differences. In the hist-aer:hist-GHG emulation (using the hist-aer pattern to emulate the hist-GHG response) (Fig. 4b), the warming in the NHMLs is overestimated,

since the aerosol pattern is stronger here than the GHG response, and the Southern Ocean is conversely weaker, due to the historical spatial pattern of aerosol forcing. In the hist-GHG:hist-aer emulation, although the anomaly is positive over the NHMLs, this represents an underestimation of the cooling (since both the emulation and ESM data are negative). This is due to the relatively weaker GHG response here.

Variation in the local and global timeseries shapes can be due to spatial variations in the forcing, or in the response. Self-emulation hist-GHG errors are small, indicating there is little internal timeseries variation in this experiment. This is consistent with the well-mixed nature of GHGs. There will still be physical non-linearities, in both the concentration-forcing and forcing-response mechanisms, and internal variability, which are reflected in the non-zero self-emulation errors, but these are small in magnitude. The largest feature is an oversensitivity in the Arctic, which may be due to a saturation of the ice albedo feedback. In the hist-aer self-emulation, by comparison, while still small compared to the out-of-sample errors, some coherent errors occur. Negative anomalies (overestimated cooling) occur in the NHMLs, with positive ones (underestimated cooling) over the tropics and south Asia. This indicates that the sensitivity of the NHML temperature to the global change is lower in the period 1990-2020 than the average across the full timeseries, since the local parameters calculated across the entire timeseries are too strong at this time. This is consistent with a shift in aerosol emissions from the NHMLs to Asia over the last three decades, and explains the positive anomalies over Asia – the emulation is under-sensitive here in this period since aerosol emissions are more concentrated in Asia in 1990-2020 than on average through the period. This is validated by Supplementary Figure S3 showing the reverse effect in the mid-20th century (as NHML aerosol emissions were historically concentrated locally in this period) and the earlier peaking of NHML temperatures than the global average in hist-aer.

Pattern scaling errors in a given period can therefore be related to the differences in the average pattern between the predictor and target scenarios, and the internal dynamics of the target scenario, particularly due to spatial variations in the forcing. Note that the timeseries self-emulation scaling error is also included in the total out-of-sample error, though the relative size indicates the pattern error is the dominant factor – these relative sizes are explored in more detail in Section 3.4.

The pattern scaling errors in Figure 4 are typically less than 0.2 K for self-emulation, and over 0.5 K for out-of-sample projection in the NHMLs. This out-of-sample error is substantial compared to the simulated temperature change of around 1.5 K/ -0.5 K globally in hist-GHG/hist-aer, and 2 K/-1 K in the NHMLs. As for the pattern differences, these out-of-sample errors are larger than typically found under pattern scaling (Alexeeff et al., 2018; Herger et al., 2015; J. F. B. Mitchell et al., 1999; T. D. Mitchell, 2003; Osborn et al., 2018; Tebaldi & Knutti, 2018), especially considering the magnitude of global warming in the scenarios, due to the starker differences between the forcing patterns in these idealised experiments. Since the out-of-sample pattern error scales with the pattern difference, panels b) and c) display very similar patterns, with the same sign of error over essentially all gridpoints (Figure S4).

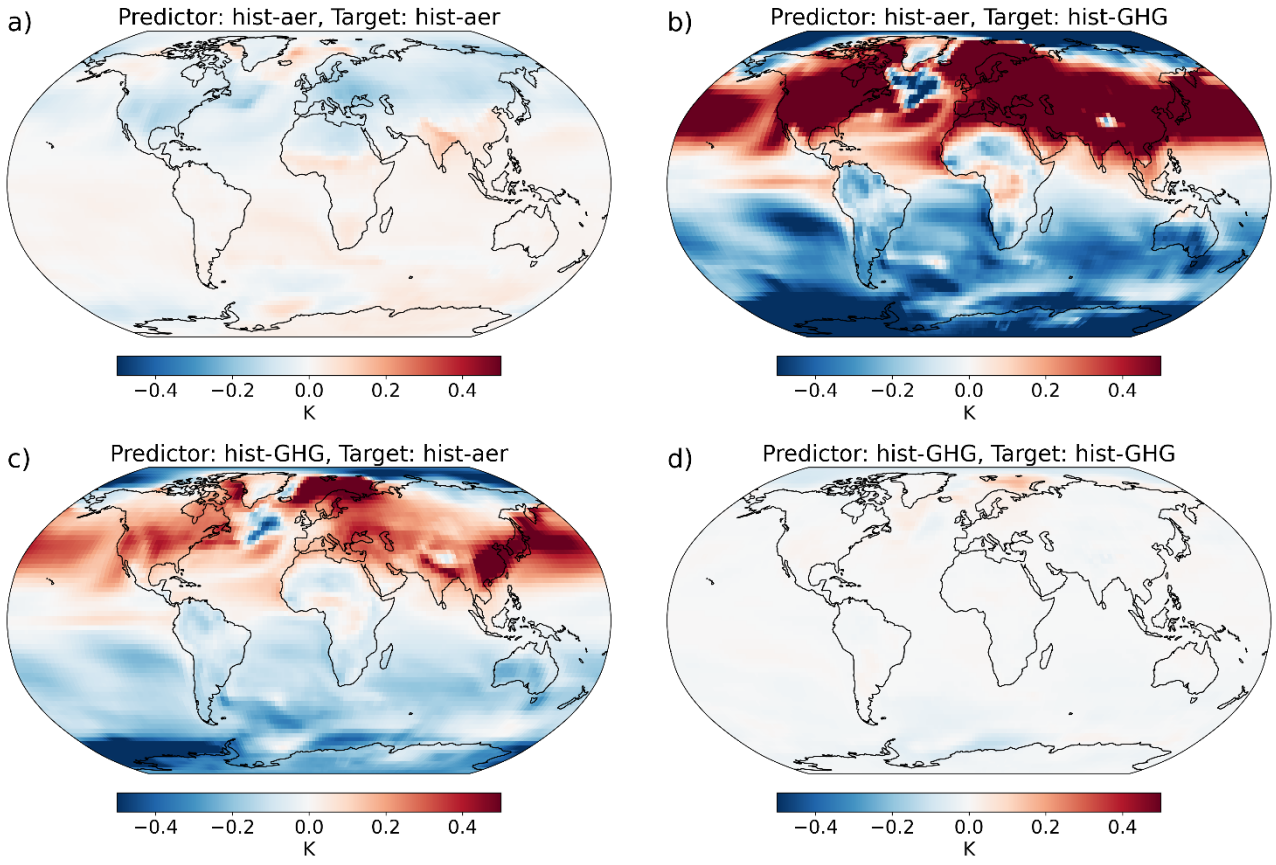

Figure 4: 1990-2020 pattern scaling errors (emulation minus ESM) averaged across 10 ESMs when predicting with historical GHGs and aerosols separately, and targeting the historical GHG and aerosol response separately; four combinations in total. The ESM and emulation data are taken relative to the first 50 years of the scenarios (i.e. 1850-1900). The last 30 years are shown to indicate the errors that arise once a substantial forcing has been applied, and to study the self-emulation scaling error within the period, as self-emulation scaling errors cancel over the whole period.

The errors in Figure 4 divided by the inter-model standard deviation in the error are shown in Supplementary Figure S5. Errors are substantial in the out-of-sample emulations, over both the NHMLs and the Southern Ocean, consistent with the large pattern differences in Figure 2. The hist-GHG self-emulation shows no coherent errors, due to the lack of pattern error and coherent timeseries variation, but the NHML errors in the hist-aer self-emulation are consistent between the models, indicating agreement on the timeseries error found here.

Similarly, Figure 5 shows the 2070-2100 multi-model mean pattern scaling errors under the four combinations of the SSP1-1.9 and SSP5-8.5 scenarios (self-emulation and cross emulation), and Supplementary Figure S6 shows these divided by the inter-model standard deviation. The out-of-sample errors are again larger than those attributable to self-emulation alone,

indicating a substantial role of the pattern difference. They are consistent with the pattern differences in Figure 3. However, similarly to the pattern differences, the inter-model standard deviation is generally larger than the magnitude, with errors greater than the inter-model standard deviation generally only in areas with similarly large difference in the patterns themselves. The timeseries (self-emulation) error is small in SSP5-8.5, similarly to hist-GHG, but SSP1-1.9 shows larger errors, with a pattern consistent with the pattern differences. This timeseries error, as with hist-aer, must be driven by the internal characteristics of SSP1-1.9. The period 2070-2100 exhibits less positive warming trends than the timeseries average, with peak warming occurring in mid-century on average. Thus, the parameters derived from the timeseries average will be too sensitive over land and tropical oceans, as the climate is experiencing less positive forcing than average, similar to the SSP5-8.5-SSP1-1.9 pattern difference. This leads here to broad, coherent self-emulation scaling errors over 2070-2100, but as for the out-of-sample cases there is little inter-model agreement. Consistent results are found for each pair of the five SSPs used here, shown in Supplementary Figures S7 and S8; extrapolation to project higher-forcing scenarios using lower-forcing patterns is found to introduce substantial errors, with interpolation to lower-forcing scenarios generating smaller errors, though still larger than self-emulation due to the additional effect of pattern errors (Herger et al., 2015; T. D. Mitchell, 2003; Tebaldi et al., 2020). The errors under self-emulation are typically less than 0.3 K, with strong interpolation less than 0.5 K but strong extrapolation over 0.5 K in broad areas. This compares to around 1.4 K (4.4 K) warming relative to pre-industrial times in SSP1-1.9 (SSP5-8.5) over 2081-2100 (IPCC, 2021). These self-emulation and interpolation errors are consistent with those found in prior work (Alexeeff et al., 2018; Herger et al., 2015; J. F. B. Mitchell et al., 1999; T. D. Mitchell, 2003; Osborn et al., 2018; Tebaldi & Knutti, 2018), while the extrapolation errors are larger due to the extreme case study presented in this section.

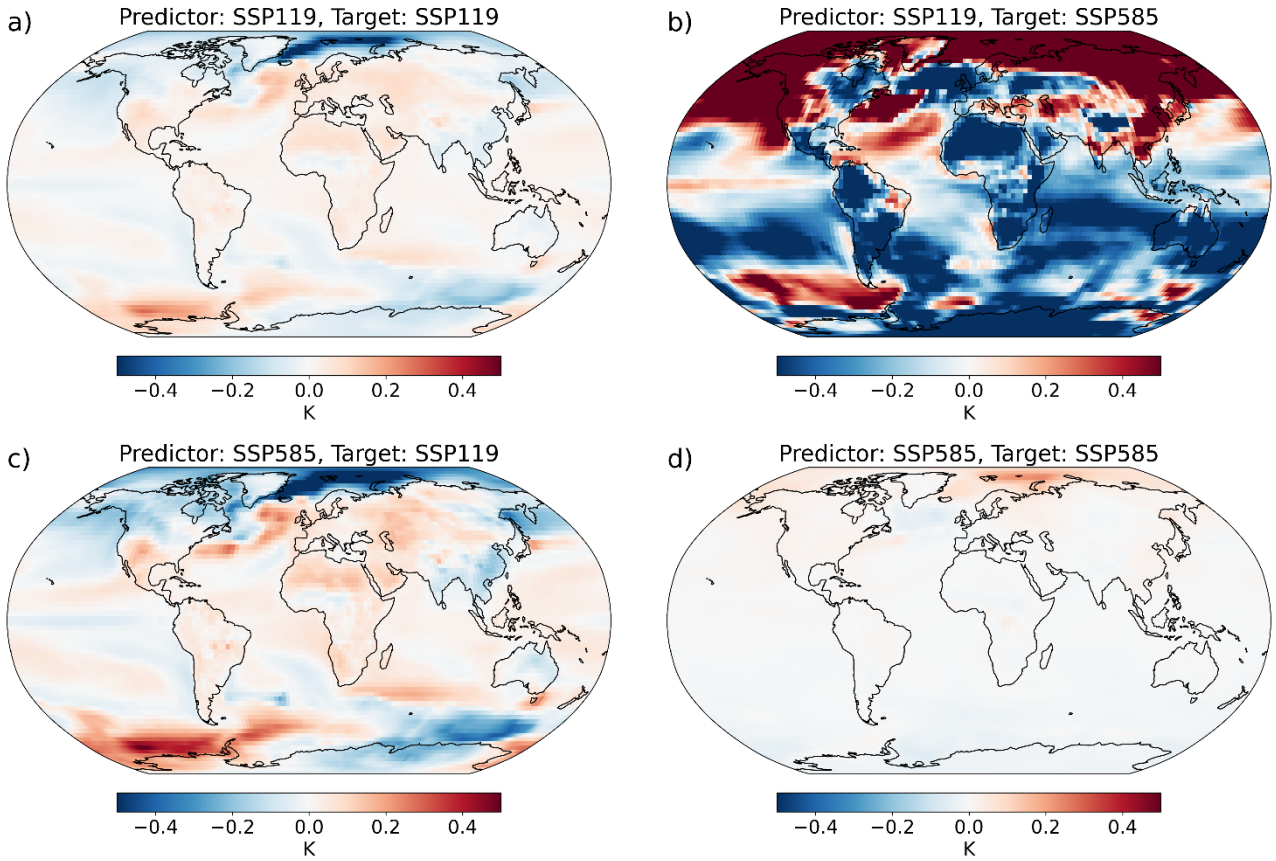

Figure 5: 2070-2100 pattern scaling errors (emulation minus ESM) averaged across eight ESMs when predicting with SSP1-1.9 and SSP5-8.5 separately, and targeting the SSP1-1.9 and SSP5-8.5 responses separately; four combinations in total. The ESM and emulation data are taken relative to the first 50 years of the scenarios (i.e. 2015-2065).

3.3 Relative importance of the pattern and timeseries errors

As highlighted above, it is important to understand how the magnitude and relative size of the two types of pattern scaling error – pattern and timeseries – depend on the target and predictor dataset.

Pairwise predictor-target emulations are performed for the 25 combinations of the five SSPs analysed here. Maps of the timeseries and pattern errors are both calculated in each year of each simulation. Pattern scaling errors are zero on the global average by design – as the pattern is scaled by the global mean response – so to analyse the size of the errors, the global average of the absolute error magnitude is taken for each. The magnitude of the total error is also taken – this is equal to the timeseries error for self-emulation, but for out-of-sample emulations, local cancellations from opposite sign timeseries and pattern errors

cause this total error to be less than the sum of the two. This sum of the two – termed the sum error – is also calculated to allow for comparisons between the two magnitudes.

Figure 6 shows the 2015-2100 timeseries of the pattern and timeseries errors for the four combinations of SSP1-1.9 and SSP5-8.5. Supplementary Figure S9 gives the results for each of the 25 SSP combinations, with a fixed scale.

The timeseries error, dependent only on the target scenario, varies relatively little between scenarios, while the pattern error is substantially larger for extrapolation cases. Even for adjacent extrapolation – e.g. using SSP1-2.6 to emulate SSP2-4.5 – the pattern error becomes increasingly large by 2100.

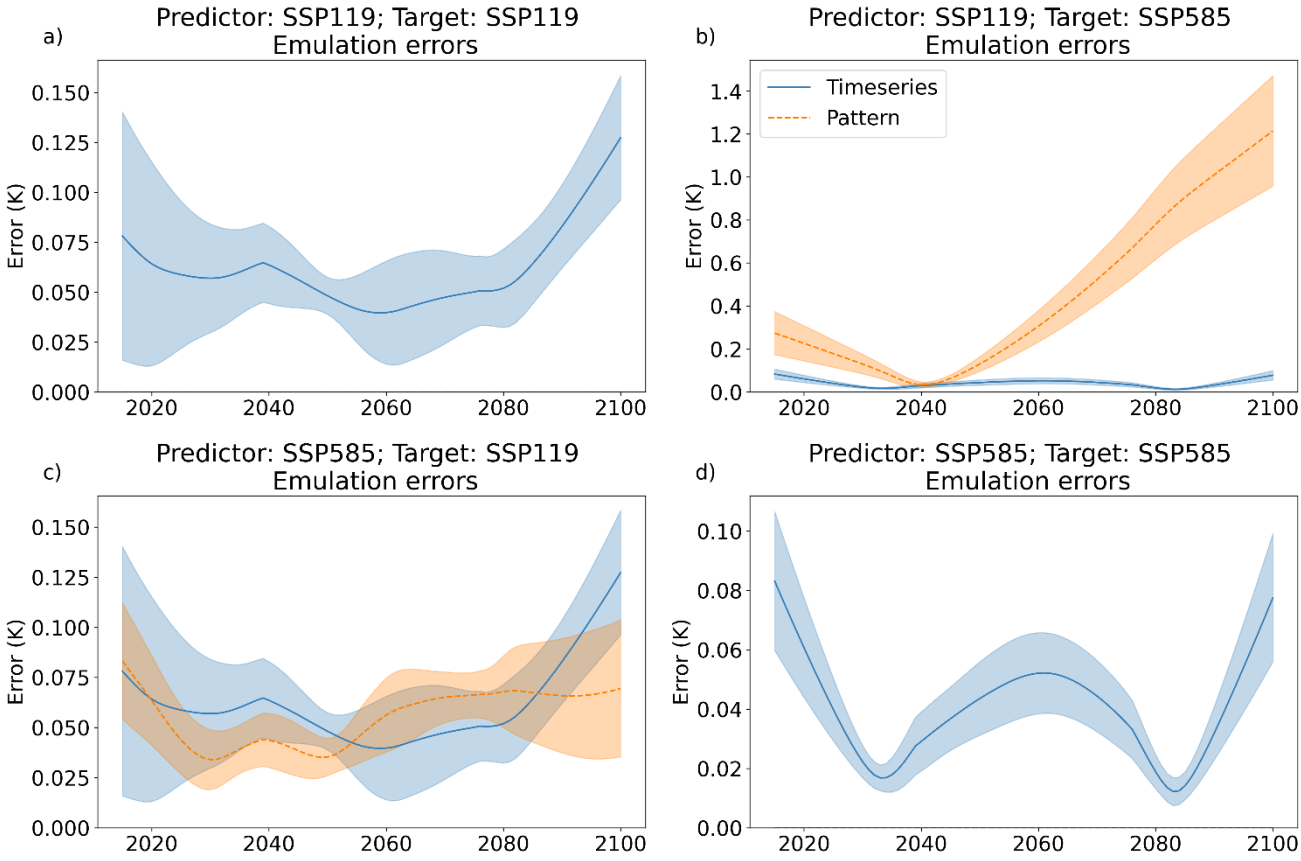

Figure 6: Timeseries of the size of the global average pattern scaling error attributed to pattern errors and timeseries errors. Each line gives the multi-model mean, with shading indicating plus and minus one inter-model standard deviation. Note the varying vertical scales.

The magnitude of the time-averaged pattern, timeseries, and total errors in each pair is shown in Figure 7 for the 25 scenario combinations, along with the percentage of the sum error (the sum of the magnitudes of each component) given by the pattern error. The sum of the error magnitudes is used for comparison, as opposing sign errors will partially cancel in the total. The magnitude of the timeseries error is less than 0.1 K in all scenarios, but largest in SSP1-1.9, a scenario in which this error can represent a substantial fraction of the mean response. The pattern error, zero for self-emulation by definition, is systematically greater for extrapolation, likely strongly influenced by the scaling of the pattern difference by the target scenario global mean timeseries. Global and time-averaged pattern error magnitudes can reach almost 0.5 K under the highest extrapolation (SSP1-1.9:SSP5-8.5), but are still around 0.2 K for slighter extrapolations and 0.1 K under interpolation. The total error is therefore highest for extrapolation. There is less dependence of the total error on the predictor scenario when targeting a low emissions scenario than targeting a high one, due to the greater role of the timeseries error – note the small variations in the SSP1-1.9 column compared to the SSP5-8.5 one.

Pattern errors represent a different proportion of the sum errors under different pairs, accounting for over 80% under high extrapolation, but only around half for projecting SSP1-1.9. The intrinsic timeseries error, irreducible under this methodology, accounts for a much larger fraction of the error under low-emissions scenarios than the pattern error. This larger role of the timeseries error is consistent with the lower correlation between local and global temperatures under low emissions scenarios found by Lynch et al. (2017).

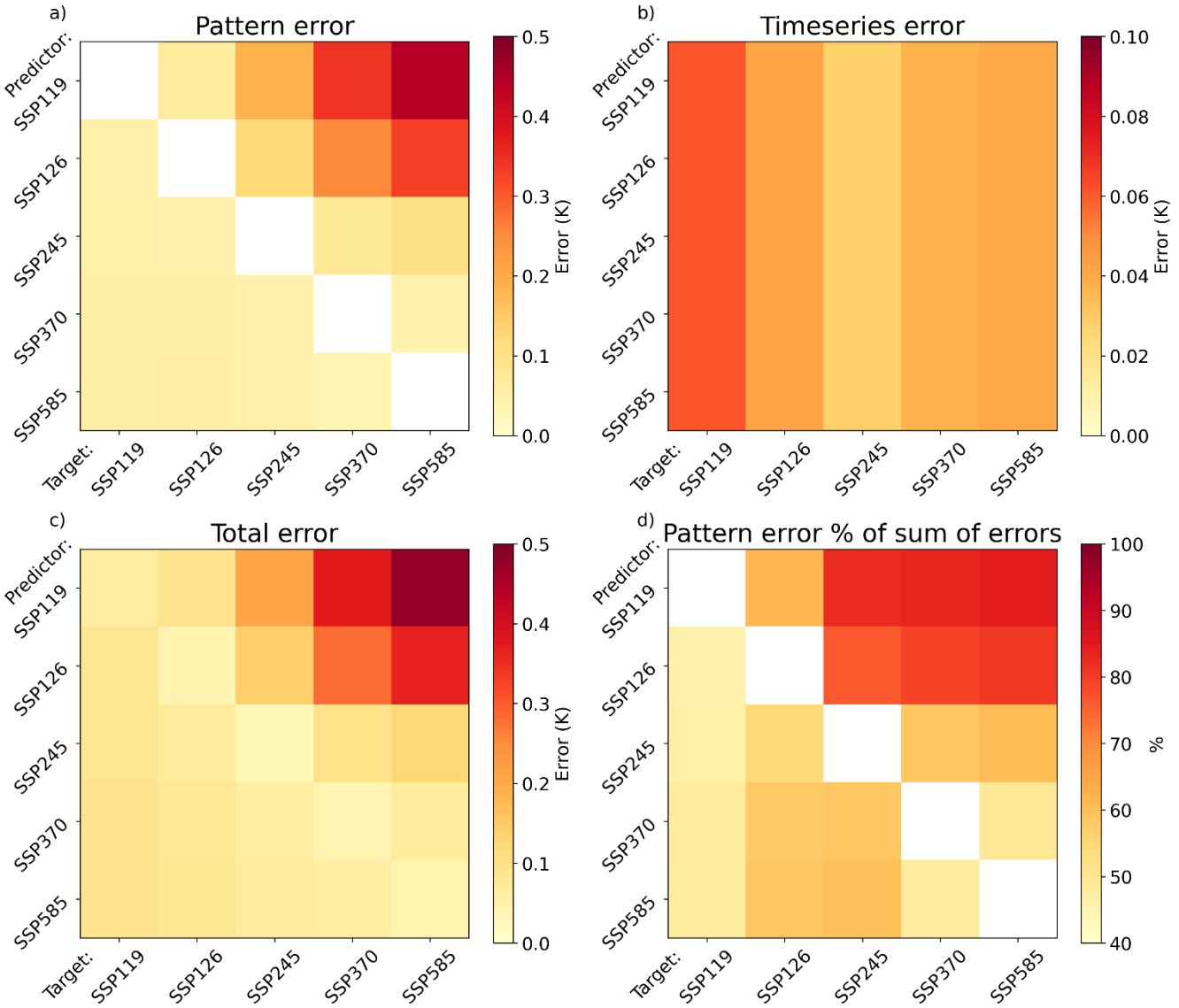

Figure 7: Pattern scaling errors averaged over the scenario time period for each predictor-target pair; errors are calculated annually, the absolute value taken, and then averaged across time and models. Top left shows the pattern error; top right the timeseries error (with a smaller scale); and bottom left the total error. Bottom right gives the percentage of the absolute total error (the sum of pattern and timeseries) attributed to the pattern error. Note the smaller scale on the timeseries error plot.

3.4 Effect of peak warming on timeseries error

The year of peak warming is increasingly used to classify emission scenarios (Riahi et al. 2022). One implication of simple pattern scaling approaches tied to global warming level is that if, in a low-emission scenario, the global warming timeseries

peaks in a particular year, then by construction the emulated temperature peaks in this same year in every gridpoint. The spatial pattern of this peak warming year is then homogeneous by design. Any spatial structure in the peak warming pattern of the actual ESM target data will be missed, leading to pattern scaling errors.

The effect of this can be tested by exploring the peak warming simulated in ESM simulations of low-emission scenarios. Supplementary Figure S10 shows the multi-model mean of the deviation in the local year of peak warming from the global average, along with the magnitude of this deviation minus the inter-model standard deviation, for both SSP1-1.9 and SSP1-2.6. The local year of peak warming is shown for each model and the multi-model mean in Supplementary Figure S11. Generally, tropical land and oceans peak earlier than average, and the Arctic and Southern Ocean later, consistent with the inertia of the system as seen in Figure 3. The patterns are similar between SSP1-1.9 and SSP1-2.6, indicating some consistency between scenarios in this effect. Few areas see inter-model agreement, however, with agreement on earlier peaking over some tropical oceans and land, and later peaking over the east of the Southern Ocean. By 2100, higher forcing scenarios are typically still warming everywhere, except for some models that show a pronounced "North Atlantic Warming Hole" (NAWH), the area of the North Atlantic where the general warming has been masked by circulation change-induced local cooling historically (Huang et al., 2020) (not shown).

Figure 8 shows results from three region-model-scenario combinations, chosen to demonstrate the different effects this can have on emulation errors. In the "Arctic in EC-Earth3 under SSP1-1.9" case (Figure 8a), the ESM global temperature peaks in mid-century, but the Arctic continues to warm to 2100. Since the pattern scaling projection is simply scaled by the global mean, however, the Arctic emulation peaks with the global temperature, diverging from the ESM to the end of the century, and projecting the wrong sign of trend from mid-century onwards.

Figure 8b applies this to the NAWH in MRI-ESM2-0 under SSP1-2.6. In the ESM, the NAWH cools throughout the century, while global temperatures peak around 2070. The pattern scaling parameter here is negative, as local cooling is regressed onto overall global warming. The projected NAWH response therefore reaches a minimum when the global mean peaks, and warms from there to 2100 as global temperatures reduce.

Finally, Figure 8c applies this to Europe in MRI-ESM2-0 under SSP1-2.6. In the ESM data, European and global temperatures rise, stabilise, and fall, but as a land region near the NAWH, Europe peaks in temperature several decades before the global mean. As shown in Figure 8d, these different regional and global shapes happen to produce a linear fit with almost exactly zero gradient – although Europe sees substantial temperature change through the century, the sensitivity averages to zero due to the differing global peak time. The resultant emulation thus has essentially zero amplitude, deviating strongly from the substantial changes modelled over Europe in the ESM.

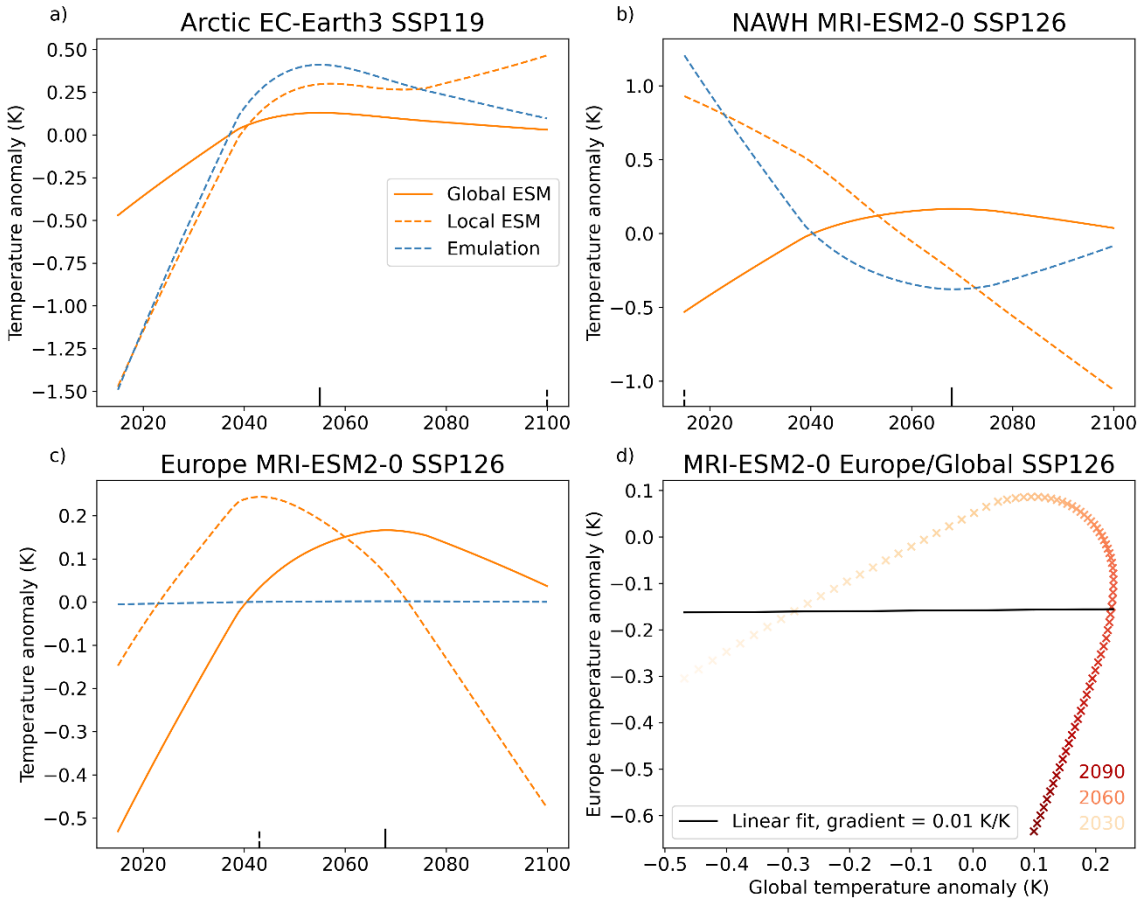

Figure 8: Timeseries of global and regional ESM annual temperature changes (relative to the first 50 years of a scenario), both annual and LOWESS smoothed, and the projected regional emulation from fitting a linear regression to the local and global smoothed temperatures, for three region-model-scenarios: Arctic (60° N to 90° N) in EC-Earth3 under SSP1-1.9, and the North Atlantic (40° N to 50° N, 40° W to 20° W) and Europe (35° N to 70° N, 10° W to 40° E) in MRI-ESM2-0 in SSP1-2.6. Perfect emulation would occur if the orange and blue dashed lines - the local ESM data and the local emulation - overlapped; deviations from this are the pattern scaling error. The local and global peak warming years are indicated with short dashed and solid vertical black lines respectively, projecting from the time axis. Also shown is the regression curve for the third example (panel d).

4 Implications for the use of pattern scaling

In this section, the implications of the prior findings for the application of pattern scaling are investigated. When emulating a given scenario, a choice must be made for the predictor dataset – i.e. the scenario(s) used to generate the pattern. The effect of this choice on pattern scaling errors is important; it won't affect the timeseries error by definition, but will modify the pattern error and hence the overall emulation efficacy. Four options are studied here, for targeting both SSP1-1.9 and SSP5-8.5 separately; some include the target scenario in the predictor dataset, thus not being entirely "out-of-sample", but multiple predictor scenarios are used in these cases, thus not being purely self-emulation. The "envelope" option uses both SSP1-1.9 and SSP5-8.5, to construct a pattern using information from the extremes of the available dataset. The "furthest" option uses the most different scenario – SSP5-8.5 to emulate SSP1-1.9 and vice versa – an unlikely choice, but a test that can give information about the effect of a wide range of choices. "all others" utilises all of the available scenarios except the target one, and "nearest" uses the closest scenario in nominal 2100 radiative forcing. The 2070-2100 multi-model mean errors associated with these eight cases are shown in Figure 9; the errors divided by the inter-model standard deviation are shown in Supplementary Figure S12.

The SSP1-1.9 error maps are remarkably insensitive to the choice of predictor datasets; the timeseries error intrinsic to SSP1-1.9, a substantial fraction of the error as shown in Figure 7, ensures a baseline error. Since the disequilibrium is qualitatively similar between early-late SSP1-1.9 (responsible for the timeseries error) and other scenarios to SSP1-1.9 (giving the pattern error), the distributions of the pattern and timeseries errors are similar. When targeting SSP5-8.5, however, a clear difference in the error magnitude is found when using different predictor datasets. Low errors occur with the envelope method, with similarities to the self-emulation pattern, suggesting SSP5-8.5 drives the pattern generated from the combination. The "furthest" case gives large errors as expected; "all others" and "nearest" give relatively similar, and smaller, errors.

The patterns of comparison between the error magnitude and the inter-model deviation (Supplementary Figure S12) are very similar, with few areas of agreement between models. This is unsurprising for SSP1-1.9 given the similarity in the error, but for SSP5-8.5 the substantially different error magnitudes and patterns result in similar, and small, areas of agreement in the pattern scaling error. This is presumably as the inter-model spread scales with the magnitude of the error, reflecting disagreement in the driving processes of the pattern scaling errors. This is again consistent with prior findings that inter-model variation is larger than pattern scaling errors (Goodwin et al., 2020; Herger et al., 2015; Osborn et al., 2016, 2018; Tebaldi & Arblaster, 2014; Tebaldi & Knutti, 2018).

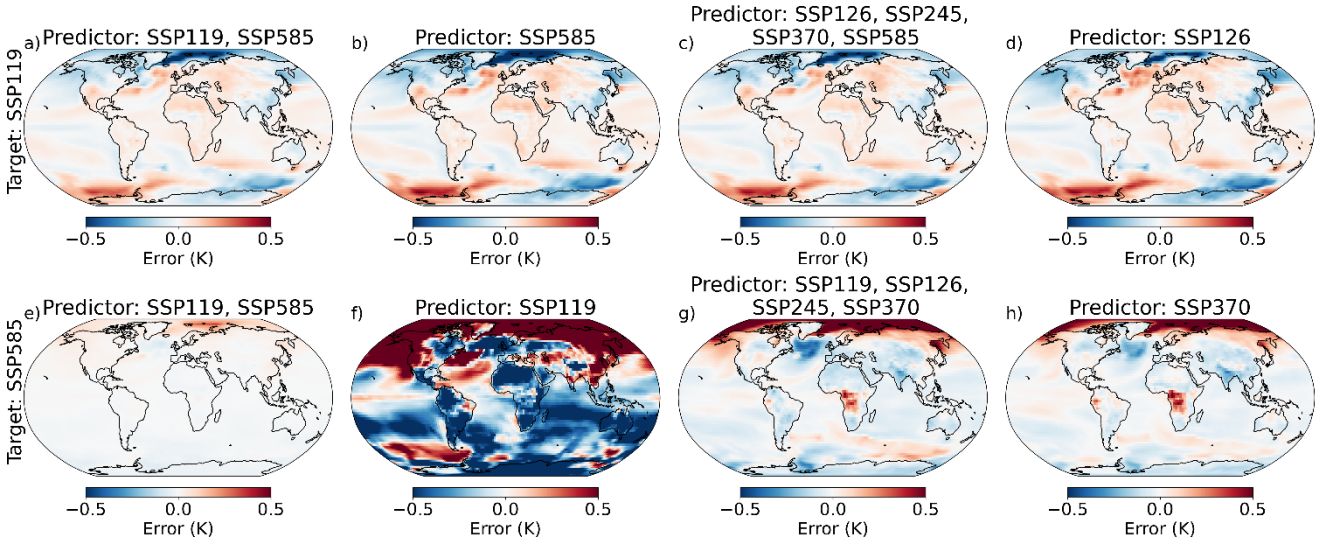

Figure 9: 2070-2100 pattern scaling errors when projecting SSP1-1.9 (top) and SSP5-8.5 (bottom) for patterns using four sets of predictors: envelope (left; SSP1-1.9 and SSP5-8.5), opposite (2nd column; SSP5-8.5 for targeting SSP1-1.9 and vice versa); all others (3rd column; i.e. the four scenarios other than that being targeted); and nearest (4th column; SSP1-2.6 to target SSP1-1.9 and SSP3-7.0 to target SSP5-8.5 - the nearest scenario to each target in RF terms). The patterns are calculated via the regression of local on global temperatures as described in Section 2.2, and the ESM and emulated data are also baselined to the first 50 years of the scenario (i.e. 2015-2065).

5 Discussion and Conclusions

This study presents a decomposition of pattern scaling errors into two components: one due to differences in the pattern between the predictor and target datasets (the pattern error), and one due to internal nonlinearities in the target scenario (the timeseries error). The differences in warming patterns between pairs of single-forcer experiments and plausible future scenarios, causing the pattern error, were also investigated, along with case studies of the impact of the timeseries error, and the total impact on the application of pattern scaling to the SSPs was tested.

Self-emulation uses the "correct" pattern – i.e. that of the scenario being emulation – and is conceptualised to therefore have zero pattern error. Errors then occur due to differences in the temporal shape of the local and global temperature. These differences are intrinsic to the scenario, and irreducible under simple pattern scaling. Here, spatial differences in the peak warming year in low emissions pathways were found to manifest in substantial emulation errors across regions and models.

When emulating out-of-sample, pattern errors are introduced due to the pattern differences between the scenarios, combining with the timeseries error in the target scenario. Robust differences were found between temperature change patterns under historical GHG and aerosol forcings, with the NHMLs more sensitive under aerosol forcing due to the historical predominance

of aerosol emissions there. Differences between temperature change patterns under future scenarios were less clear between models, as found in prior work studying future emissions scenarios (Goodwin et al., 2020; Herger et al., 2015; Osborn et al., 2016, 2018; Tebaldi & Arblaster, 2014; Tebaldi & Knutti, 2018). However, the difference resembled differences between transient and equilibrium patterns in prior work (Herger et al., 2015; Huang et al., 2020; King et al., 2020; T. D. Mitchell, 2003) – higher sensitivity over tropical land and lower over high-latitude oceans in SSP5-8.5 – indicating the different warming rates in the scenarios are an important cause of difference between these scenario patterns. Aerosol concentrations are also substantially different between these scenarios, potentially causing a less-sensitive East Asian response under SSP5-8.5, though this wasn't robust between models.

The pattern error drives over 80% of the overall pattern scaling error when emulating a high-emissions scenario using a low-emissions pattern, causing pattern scaling errors to be strongly dependent on the predictor dataset used. In contrast, the timeseries error contributes around half the error for emulating low-emissions scenarios, rendering the choice of pattern less important, though choosing scenarios closer in radiative forcing to the target still reduces the overall error.

Splitting the total error into these components allows for an understanding of the relative importance of the limitations of the assumptions which generate the errors. Understanding which source drives the error of a given pattern scaling application can guide efforts to reduce these uncertainties.

The errors associated with differing aerosol emissions and differing levels of warming, including stabilisation and relative cooling, will presumably be more important for the SSPs analysed here than the prior RCPs, which saw a narrower range in aerosol emissions and levels of warming. The tighter range in CO_2/non-CO_2 forcings ensured pattern scaling worked well under the RCPs (Goodwin et al., 2020), but variations in the aerosol pattern will lead to greater pattern scaling errors (Xu & Lin, 2017). Projecting existing and future Paris Agreement-consistent scenarios, i.e. those which stabilise temperatures below 2ºC and reach net-zero GHG emissions by 2100 (Schleussner et al., 2022), such as the C1 and C2 scenarios in IPCC AR6 WGIII (Kikstra et al., 2022), will lead to issues related to equilibrium and transient pattern differences (King et al., 2020).

The efficacy of pattern scaling is constrained by the choices of patterns available, i.e. the dataset of scenarios simulated in multi-model ESM ensembles, which is itself determined by the trajectories chosen under projects such as ScenarioMIP. These might not cover the full relevant range of scenario attributes (Guivarch et al., 2022); there is a lack of stabilising and cooling scenarios in the extant datasets (Tebaldi et al., 2022), with a recognised need for more equilibrium experiments in the future (King et al., 2021). However, this work demonstrates that scenarios with these properties are less amenable to emulation via pattern scaling than higher-emissions ones.

These results suggest that caution should be taken when applying simple linear pattern scaling to emulate low emission scenarios, as these are intrinsically less amenable to emulation via pattern scaling. Large differences in forcing pattern and rate of warming between predictor and target scenarios also lead to substantial emulation errors.

This paper focused on annual mean temperature, but it would be useful to determine the relative roles of the pattern and timeseries errors for other variables, to determine the extent to which their emulation is limited by nonlinearities in the target scenario. The distribution of temperature variability is also crucial for impact analysis, can change under external forcing

(Olonscheck & Notz, 2017; Pendergrass et al., 2017), and has been incorporated into emulation tools such as MESMER (Beusch et al., 2020).

Only simple pattern scaling – using one predictor, global temperature, to emulate the local temperature response using a single pattern – was studied here. These errors may be mitigated to an extent by other methods, such as using patterns dependent on forcer (Xu & Lin, 2017; Kravitz et al., 2017; Schlesinger et al., 2000) or response timescale (Zappa et al., 2020). Improvements have also been found by adding extra predictors in addition to global temperature, such as land sea contrast (Herger et al., 2015) and ocean heat uptake (Beusch et al., 2020).

Regional climate changes are key for understanding the impacts of different policy choices, but the links between global mean temperatures and their regional impacts haven't been fully explored within the IPCC framework (Kikstra et al., 2022). Emulation of regional impacts via tools such as pattern scaling, in a consistent framework such as MESMER (Beusch et al., 2020), provides a crucial means by which to estimate the local impacts of new emissions scenarios without the need to perform expensive, time-consuming ESM simulations.

This paper furthers the understanding of pattern scaling by decomposing the errors into those attributable to different assumptions. Further research to reduce these errors where possible will be crucial to enhance our understanding of future climate change.

Data availability

The CMIP6 data utilised in this study are available for download at https://aims2.llnl.gov/search

Author contributions

All authors designed the study and reviewed the manuscript. CDW carried out the analysis and led the preparation of the manuscript.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors would like to thank Lea Beusch and Mathias Hauser for their useful discussions and input on MESMER and its implementation. The authors would also like to thank the modelling teams who designed and ran the models (see Supplementary Tables S1 and S2) and made the data available, and the World Climate Research Program for facilitating the CMIP6 process.

Financial support

CDW, ACM, PMF, and LSJ were supported by the European Union's Horizon 2020 programme under grant agreement No 820829 (CONSTRAIN). PMF was supported by the Vol-Clim NERC grant NE/S000887/1.

References

Alexeeff, S. E., Nychka, D., Sain, S. R., and Tebaldi, C.: Emulating mean patterns and variability of temperature across and within scenarios in anthropogenic climate change experiments, Clim Change, 146, 319–333, https://doi.org/10.1007/s10584-016-1809-8, 2018.

Beusch, L., Gudmundsson, L., and Seneviratne, S. I.: Emulating Earth system model temperatures with MESMER: From global mean temperature trajectories to grid-point-level realizations on land, Earth System Dynamics, 11, 139–159, https://doi.org/10.5194/esd-11-139-2020, 2020.

Beusch, L., Nicholls, Z., Gudmundsson, L., Hauser, M., Meinshausen, M., and Seneviratne, S. I.: 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), Geosci Model Dev, 15, 2085–2103, https://doi.org/10.5194/gmd-15-2085-2022, 2022a.

Beusch, L., Nauels, A., Gudmundsson, L., Gütschow, J., Schleussner, C. F., and Seneviratne, S. I.: Responsibility of major emitters for country-level warming and extreme hot years, Commun Earth Environ, 3, https://doi.org/10.1038/s43247-021-00320-6, 2022b.

Brunner, L., Hauser, M., Lorenz, R., and Beyerle, U.: The ETH Zurich CMIP6 next generation archive : technical documentation, https://doi.org/10.5281/zenodo.3734128.Contact., 2020.

Byrne, M. P. and O'Gorman, P. A.: Trends in continental temperature and humidity directly linked to ocean warming, Proc Natl Acad Sci U S A, 115, 4863–4868, https://doi.org/10.1073/PNAS.1722312115, 2018.

Callahan, C. and Mankin, J. S.: National attribution of historical climate damages, Clim Change, In Press, https://doi.org/10.1007/s10584-022-03387-y, 2022.

Gidden, M. J., Riahi, K., Smith, S. J., Fujimori, S., Luderer, G., Kriegler, E., van Vuuren, D. P., van den Berg, M., Feng, L., Klein, D., Calvin, K., Doelman, J. C., Frank, S., Fricko, O., Harmsen, M., Hasegawa, T., Havlik, P., Hilaire, J., Hoesly, R., Horing, J., Popp, A., Stehfest, E., and Takahashi, K.: Global emissions pathways under different socioeconomic scenarios for use in CMIP6: a dataset of harmonized emissions trajectories through the end of the century, Geosci. Model Dev, 12, 1443–1475, https://doi.org/10.5194/gmd-12-1443-2019, 2019.

Gillett, N. P., Shiogama, H., Funke, B., Hegerl, G., Knutti, R., Matthes, K., Santer, B. D., Stone, D., and Tebaldi, C.: The Detection and Attribution Model Intercomparison Project (DAMIP v1.0) contribution to CMIP6, Geosci Model Dev, 9, 3685–3697, https://doi.org/10.5194/GMD-9-3685-2016, 2016.

Goodwin, P., Leduc, M., Partanen, A. I., Damon Matthews, H., and Rogers, A.: A computationally efficient method for probabilistic local warming projections constrained by history matching and pattern scaling, demonstrated by WASP-LGRTC-1.0, Geosci Model Dev, 13, 5389–5399, https://doi.org/10.5194/gmd-13-5389-2020, 2020.

Grubler, A., Wilson, C., Bento, N., Boza-Kiss, B., Krey, V., McCollum, D. L., Rao, N. D., Riahi, K., Rogelj, J., de Stercke, S., Cullen, J., Frank, S., Fricko, O., Guo, F., Gidden, M., Havlík, P., Huppmann, D., Kiesewetter, G., Rafaj, P., Schoepp, W., and Valin, H.: A low energy demand scenario for meeting the 1.5°C target and sustainable development goals without negative emission technologies, Nat Energy, 3, 515–527, https://doi.org/10.1038/s41560-018-0172-6, 2018.

Guivarch, C., le Gallic, T., Bauer, N., Fragkos, P., Huppmann, D., Jaxa-Rozen, M., Keppo, I., Kriegler, E., Krisztin, T., Marangoni, G., Pye, S., Riahi, K., Schaeffer, R., Tavoni, M., Trutnevyte, E., van Vuuren, D., and Wagner, F.: Using large ensembles of climate change mitigation scenarios for robust insights, Nat Clim Chang, 12, 428–435, https://doi.org/10.1038/s41558-022-01349-x, 2022.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bull Am Meteorol Soc, 90, 1095–1107, https://doi.org/10.1175/2009BAMS2607.1, 2009.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in projections of regional precipitation change, Clim Dyn, 37, 407–418, https://doi.org/10.1007/s00382-010-0810-6, 2011.

Herger, N., Sanderson, B. M., and Knutti, R.: Improved pattern scaling approaches for the use in climate impact studies, Geophys Res Lett, 42, 3486–3494, https://doi.org/10.1002/2015GL063569, 2015.

Holland, M. M. and Bitz, C. M.: Polar amplification of climate change in coupled models, Clim Dyn, 21, 221–232, https://doi.org/10.1007/S00382-003-0332-6/FIGURES/8, 2003.

Huang, D., Dai, A., and Zhu, J.: Are the transient and equilibrium climate change patterns similar in response to increased CO2?, J Clim, 33, 8003–8023, https://doi.org/10.1175/JCLI-D-19-0749.1, 2020.

IPCC: Global warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty [V. Masson-Delmotte, P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, T. Waterfield (eds.)]. In Press, 2018.

IPCC: Summary for Policymakers in:Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 3–32 pp., 2021.

IPCC: Summary for Policymakers. In: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [P.R. Shukla, J. Skea, R. Slade, A. Al Khourdajie, R. van Di, Cambridge University Press, Cambridge, UK and New York, NY, USA., 2022.

Ishizaki, Y., Shiogama, H., Emori, S., Yokohata, T., Nozawa, T., Ogura, T., Abe, M., Yoshimori, M., and Takahashi, K.: Temperature scaling pattern dependence on representative concentration pathway emission scenarios: A letter, Clim Change, 112, 535–546, https://doi.org/10.1007/s10584-012-0430-8, 2012.

James, R., Washington, R., Schleussner, C.-F., Rogelj, J., and Conway, D.: Characterizing half-a-degree difference: a review of methods for identifying regional climate responses to global warming targets, WIREs Clim Change, 8, 457, https://doi.org/10.1002/wcc.457, 2017.

Kikstra, J. S., Nicholls, Z. R. J., Smith, C. J., Lewis, J., Lamboll, R. D., Byers, E., Sandstad, M., Meinshausen, M., Gidden, M. J., Rogelj, J., Kriegler, E., Peters, G. P., Fuglestvedt, J. S., Skeie, R. B., Samset, B. H., Wienpahl, L., van Vuuren, D. P., van der Wijst, K.-I., al Khourdajie, A., Forster, P. M., Reisinger, A., Schaeffer, R., and Riahi, K.: The IPCC Sixth Assessment Report WGIII climate assessment of mitigation pathways: from emissions to global temperatures, EGUsphere, 2022, 1–55, 2022.

King, A. D., Lane, T. P., Henley, B. J., and Brown, J. R.: Global and regional impacts differ between transient and equilibrium warmer worlds, Nat Clim Chang, 10, 42–47, https://doi.org/10.1038/s41558-019-0658-7, 2020.

King, A. D., Sniderman, J. M. K., Dittus, A. J., Brown, J. R., Hawkins, E., and Ziehn, T.: Studying climate stabilization at Paris Agreement levels, Nature Climate Change 2021 11:12, 11, 1010–1013, https://doi.org/10.1038/s41558-021-01225-0, 2021.

Kravitz, B., Lynch, C., Hartin, C., and Bond-Lamberty, B.: Exploring precipitation pattern scaling methodologies and robustness among CMIP5 models, Geosci Model Dev, 10, 1889–1902, https://doi.org/10.5194/gmd-10-1889-2017, 2017.

Lee, J.-Y., J. Marotzke, G. Bala, L. Cao, S. Corti, J.P. Dunne, F. Engelbrecht, E. Fischer, J.C. Fyfe, C. Jones, A. and Maycock, J. Mutemi, O. Ndiaye, S. Panickal, and T. Z.: Chapter 4: Future global climate: scenario-based projections and near-term information, in: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 195, https://doi.org/10.1017/9781009157896.006, 2021.

Link, R., Snyder, A., Lynch, C., Hartin, C., Kravitz, B., and Bond-Lamberty, B.: Fldgen v1.0: An emulator with internal variability and space-Time correlation for Earth system models, Geosci Model Dev, 12, 1477–1489, https://doi.org/10.5194/gmd-12-1477-2019, 2019.

Liu, G., Peng, S., Huntingford, C., and Xi, Y.: A new precipitation emulator (PREMU v1 . 0) for lower complexity models, 2022.

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., and Takemura, T.: A PDRMIP Multimodel study on the impacts of regional aerosol forcings on global and regional precipitation, J Clim, 31, 4429–4447, https://doi.org/10.1175/JCLI-D-17-0439.1, 2018.

Lopez, A., Suckling, E. B., and Smith, L. A.: Robustness of pattern scaled climate change scenarios for adaptation decision support, Clim Change, 122, 555–566, https://doi.org/10.1007/s10584-013-1022-y, 2014.

Lynch, C., Hartin, C., Bond-Lamberty, B., and Kravitz, B.: An open-Access CMIP5 pattern library for temperature and precipitation: Description and methodology, Earth Syst Sci Data, 9, 281–292, https://doi.org/10.5194/essd-9-281-2017, 2017.

Meinshausen, M., Nicholls, Z., Lewis, J., Gidden, M., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J., Daniel, J., John, A., Krummel, P., Luderer, G., Meinshausen, N., Montzka, S., Rayner, P., Reimann, S., Smith,

S., van den Berg, M., Velders, G., Vollmer, M., and Wang, H. J.: The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500, Geosci Model Dev, 13, 3571–3605, https://doi.org/10.5194/gmd-2019-222, 2020.

Mitchell, J. F. B., Johns, T. C., Eagles, M., Ingram, W. J., and Davis, R. A.: Towards the construction of climate change scenarios, Clim Change, 41, 547–581, https://doi.org/10.1023/a:1005466909820, 1999.

Mitchell, T. D.: Pattern Scaling. An Examination of the Accuracy of the Technique for Describing Future Climates, Clim Change, 60, 217–242, https://doi.org/10.1023/A:1026035305597, 2003.

Myhre, G., Kramer, R. J., Smith, C. J., Hodnebrog, Forster, P., Soden, B. J., Samset, B. H., Stjern, C. W., Andrews, T., Boucher, O., Faluvegi, G., Fläschner, D., Kasoar, M., Kirkevåg, A., Lamarque, J. F., Olivié, D., Richardson, T., Shindell, D., Stier, P., Takemura, T., Voulgarakis, A., and Watson-Parris, D.: Quantifying the Importance of Rapid Adjustments for Global Precipitation Changes, Geophys Res Lett, 45, 11,399-11,405, https://doi.org/10.1029/2018GL079474, 2018.

Olonscheck, D. and Notz, D.: Consistently estimating internal climate variability from climate model simulations, J Clim, 30, 9555–9573, https://doi.org/10.1175/JCLI-D-16-0428.1, 2017.

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., Knutti, R., Kriegler, E., Lamarque, J. F., Lowe, J., Meehl, G. A., Moss, R., Riahi, K., and Sanderson, B. M.: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6, Geosci Model Dev, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

Osborn, T. J., Wallace, C. J., Harris, I. C., and Melvin, T. M.: Pattern scaling using ClimGen: monthly-resolution future climate scenarios including changes in the variability of precipitation, Clim Change, 134, 353–369, https://doi.org/10.1007/s10584-015-1509-9, 2016.

Osborn, T. J., Wallace, C. J., Lowea, J. A., and Bernie, D.: Performance of pattern-scaled climate projections under high-end warming. Part I: Surface air temperature over land, J Clim, 31, 5667–5680, https://doi.org/10.1175/JCLI-D-17-0780.1, 2018.

Otero, I., Farrell, K. N., Pueyo, S., Kallis, G., Kehoe, L., Haberl, H., Plutzar, C., Hobson, P., García-Márquez, J., Rodríguez-Labajos, B., Martin, J. L., Erb, K. H., Schindler, S., Nielsen, J., Skorin, T., Settele, J., Essl, F., Gómez-Baggethun, E., Brotons, L., Rabitsch, W., Schneider, F., and Pe'er, G.: Biodiversity policy beyond economic growth, Conserv Lett, 13, https://doi.org/10.1111/conl.12713, 2020.

Partanen, A.-I., Landry, J.-S., and Matthews, D.: Climate and health implications of future aerosol emission scenarios, Environmental Research Letters, 13, 024028, https://doi.org/10.1088/1748-9326/aaa511, 2018.

Pendergrass, A. G., Lehner, F., Sanderson, B. M., and Xu, Y.: Does extreme precipitation intensity depend on the emissions scenario?, Geophys Res Lett, 42, 8767–8774, https://doi.org/10.1002/2015GL065854, 2015.

Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C., and Sanderson, B. M.: Precipitation variability increases in a warmer climate, Sci Rep, 7, 1–9, https://doi.org/10.1038/s41598-017-17966-y, 2017.

Persad, G. G. and Caldeira, K.: Divergent global-scale temperature effects from identical aerosols emitted in different regions, Nat Commun, 9, https://doi.org/10.1038/s41467-018-05838-6, 2018.

Schlesinger, M. E., Malyshev, S., Rozanov, E. v., Yang, F., Andronova, N. G., de Vries, B., Grübler, A., Jiang, K., Masui, T., Morita, T., Penner, J., Pepper, W., Sankovski, A., and Zhang, Y.: Geographical Distributions of Temperature Change for

Scenarios of Greenhouse Gas and Sulfur Dioxide Emissions, Technol Forecast Soc Change, 65, 167–193, https://doi.org/10.1016/S0040-1625(99)00114-6, 2000.

Schleussner, C.-F., Ganti, G., Rogelj, J., and Gidden, M. J.: An emission pathway classification reflecting the Paris Agreement climate objectives, Commun Earth Environ, 3, https://doi.org/10.1038/s43247-022-00467-w, 2022.

Seneviratne, S. I., Donat, M. G., Pitman, A. J., Knutti, R., and Wilby, R. L.: Allowable CO2 emissions based on regional and impact-related climate targets, Nature, 529, 477–483, https://doi.org/10.1038/nature16542, 2016.

Seneviratne, S. I. and Hauser, M.: Regional Climate Sensitivity of Climate Extremes in CMIP6 Versus CMIP5 Multimodel Ensembles, Earths Future, 8, e2019EF001474, https://doi.org/10.1029/2019EF001474, 2020.

Seneviratne, S.I., X. Zhang, M. Adnan, W. Badi, C. Dereczynski, A. Di Luca, S. Ghosh, I. Iskandar, J. Kossin, S. Lewis, F. Otto, I. Pinto, M. Satoh, S.M. Vicente-Serrano, M. Wehner, and B. Zhou, 2021: Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., 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, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1513–1766, doi:10.1017/9781009157896.013, 2021.

Sillmann, J., Stjern, C. W., Myhre, G., and Forster, P. M.: Slow and fast responses of mean and extreme precipitation to different forcing in CMIP5 simulations, Geophys Res Lett, 44, 6383–6390, https://doi.org/10.1002/2017GL073229, 2017.

Smith, C. J., Forster, P. M., Allen, M., Leach, N., Millar, R. J., Passerello, G. A., and Regayre, L. A.: FAIR v1.3: A simple emissions-based impulse response and carbon cycle model, Geosci Model Dev, 11, 2273–2297, https://doi.org/10.5194/gmd-11-2273-2018, 2018.

Snyder, A., Link, R., Dorheim, K., Kravitz, B., BondLamberty, B., and Hartin, C.: Joint emulation of Earth System Model temperature-precipitation realizations with internal variability and space-time and crossvariable correlation: Fldgen v2.0 software description, PLoS One, 14, 1–13, https://doi.org/10.1371/journal.pone.0223542, 2019.

Stjern, C. W., Lund, M. T., Samset, B. H., Myhre, G., Forster, P. M., Andrews, T., Boucher, O., Faluvegi, G., Fläschner, D., Iversen, T., Kasoar, M., Kharin, V., Kirkevåg, A., Lamarque, J. F., Olivié, D., Richardson, T., Sand, M., Shawki, D., Shindell, D., Smith, C. J., Takemura, T., and Voulgarakis, A.: Arctic Amplification Response to Individual Climate Drivers, Journal of Geophysical Research: Atmospheres, 124, 6698–6717, https://doi.org/10.1029/2018JD029726, 2019.

Tebaldi, C. and Arblaster, J. M.: Pattern scaling: Its strengths and limitations, and an update on the latest model simulations, Clim Change, 122, 459–471, https://doi.org/10.1007/s10584-013-1032-9, 2014.

Tebaldi, C. and Knutti, R.: Evaluating the accuracy of climate change pattern emulation for low warming targets, Environmental Research Letters, 13, https://doi.org/10.1088/1748-9326/aabef2, 2018.

Tebaldi, C., Armbruster, A., Engler, H. P., and Link, R.: Emulating climate extreme indices, Environmental Research Letters, 15, https://doi.org/10.1088/1748-9326/ab8332, 2020.

Tebaldi, C., Debeire, K., Eyring, V., Fischer, E., Fyfe, J., Friedlingstein, P., Knutti, R., Lowe, J., O'Neill, B., Sanderson, B., van Vuuren, D., Riahi, K., Meinshausen, M., Nicholls, Z., Tokarska, K., Hurtt, G., Kriegler, E., Meehl, G., Moss, R., Bauer, S., Boucher, O., Brovkin, V., Yhb, Y., Dix, M., Gualdi, S., Guo, H., John, J., Kharin, S., Kim, Y. H., Koshiro, T., Ma, L., Olivié, D., Panickal, S., Qiao, F., Rong, X., Rosenbloom, N., Schupfner, M., Séférian, R., Sellar, A., Semmler, T., Shi, X., Song, Z., Steger, C., Stouffer, R., Swart, N., Tachiiri, K., Tang, Q., Tatebe, H., Voldoire, A., Volodin, E., Wyser, K., Xin, X., Yang, S., Yu, Y., and Ziehn, T.: Climate model projections from the Scenario Model Intercomparison Project (ScenarioMIP) of CMIP6, Earth System Dynamics, 12, 253–293, https://doi.org/10.5194/esd-12-253-2021, 2021.

Tebaldi, C., Snyder, A., and Dorheim, K.: STITCHES: creating new scenarios of climate model output by stitching together pieces of existing simulations, Earth System Dynamics, 1–58, 2022.

Thackeray, C. W., Hall, A., Norris, J., and Chen, D.: Constraining the increased frequency of global precipitation extremes under warming, Nat Clim Chang, 12, 441–448, https://doi.org/10.1038/s41558-022-01329-1, 2022.

Xu, Y. and Lin, L.: Pattern scaling based projections for precipitation and potential evapotranspiration: sensitivity to composition of GHGs and aerosols forcing, Clim Change, 140, 635–647, https://doi.org/10.1007/s10584-016-1879-7, 2017.

Zappa, G., Ceppi, P., and Shepherd, T. G.: Time-evolving sea-surface warming patterns modulate the climate change response of subtropical precipitation over land, Proc Natl Acad Sci U S A, 117, 4539–4545, https://doi.org/10.1073/pnas.1911015117, 2020.

Zelazowski, P., Huntingford, C., Mercado, L. M., and Schaller, N.: Climate pattern-scaling set for an ensemble of 22 GCMs - Adding uncertainty to the IMOGEN version 2.0 impact system, Geosci Model Dev, 11, 541–560, https://doi.org/10.5194/gmd-11-541-2018, 2018.