From emission scenarios to spatially resolved projections with a chain of computationally efficient emulators: MAGICC (v7.5.1) – MESMER (v0.8[..*].3) coupling

Lea Beusch[1], Zebedee Nicholls[2,3,4], Lukas Gudmundsson[1], Mathias Hauser[1], Malte Meinshausen[2,3,4], and Sonia I. Seneviratne[1]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[2]Climate and Energy College, The University of Melbourne, Parkville, Victoria, Australia
[3]School of Geography, Earth and Atmospheric Sciences, The University of Melbourne, Parkville, Victoria, Australia
[4]Climate Resource, Northcote, Victoria, Australia

Correspondence: Lea Beusch (lea.beusch@env.ethz.ch)

Abstract. Producing targeted climate information at the local scale, including major sources of climate change projection uncertainty for diverse emissions scenarios, is essential to support climate change mitigation and adaptation efforts. Here, we present the first chain of computationally efficient Earth System Model (ESM) emulators allowing to [..[2]]translate any greenhouse gas emission [..[3]]pathway into spatially resolved [..[4]]annual mean temperature anomaly field time series, accounting for both forced climate response and natural variability uncertainty at the local scale. By combining the [..[5]]global mean, emissions-driven emulator MAGICC with the spatially resolved emulator MESMER, ESM-specific as well as constrained probabilistic emulated ensembles can be derived. This [..[6]]emulator chain can hence build on and extend large multi-ESM ensembles such as the ones produced within the [..[7]]sixth phase of the Coupled Model Intercomparison Project (CMIP6). The main extensions are threefold[..[8]]: (i) A more thorough sampling of the forced climate response and the natural variability uncertainty is possible with millions of emulated realizations being readily created. (ii) The same uncertainty space can be sampled for any emission pathway, which is not the case in CMIP6, where only a limited number of scenarios have been explored and where some of the most societally relevant strong mitigation scenarios have been run by only a small number of ESMs. (iii) Other lines of evidence to constrain future projections, including observational constraints, can be introduced, which helps to refine projected [..[9]]ranges beyond the multi-ESM ensemble's estimates. In addition to presenting results from the coupled MAGICC-MESMER emulator chain, we carry out an extensive validation of MESMER, which is trained on and

*removed: .1
[2]removed: rapidly translate
[3]removed: pathways
[4]removed: annual-mean
[5]removed: global-mean
[6]removed: emulation
[7]removed: 6th
[8]removed: .
[9]removed: future

applied to multiple emission pathways for the first time in this study. [..[10]]By coupling MAGICC and MESMER, we pave the way for rapid assessments of any emission pathway's regional climate change consequences and the associated uncertainties.

1 Introduction

Earth System Models (ESMs) are the primary tools to study the impact of greenhouse gas emissions on [..[11]]regional climate change (IPCC, 2013, 2021). While the insights they provide are invaluable to advance our understanding of the coupled Earth [..[12]]system to external influences, their projections are affected by three major sources of uncertainty: (i) internal variability uncertainty, i.e., unforced natural climate variability; (ii) forced climate response uncertainty, i.e., uncertainty in the response of the climate system to both forced natural (solar and volcanic) and anthropogenic (greenhouse gases, aerosols, land-use change, etc.) influences; and (iii) emission scenario uncertainty, i.e., which emission pathway the world chooses (Hawkins and Sutton, 2009; Lehner et al., 2020). Each of these uncertainty classes again encompass a myriad of different contributions to the total uncertainty, e.g., carbon cycle uncertainty, aerosol forcing uncertainty, and climate sensitivity uncertainty are all captured within the climate response uncertainty in the above categorization. Due to their high computational cost, ESMs can only sparsely explore the full uncertainty phase space.

This sparse exploration is problematic [..[13]]since targeted climate information accounting for all major sources of climate change uncertainty is urgently needed, especially [..[14]]because both Earth's climate [..[15]](IPCC, 2013, 2018, 2021) and the future emission pathway the world's nations have pledged to follow [..[16]](CAT, 2019, 2021a, b) are changing rapidly. When assessing the implications of a large number of emission pathways for future climate, it is neither computationally feasible nor efficient to create full ESM ensembles for each emission pathway, especially ESM ensembles which thoroughly sample the natural variability and climate response uncertainty space. Instead, computationally [..[17]]inexpensive ESM emulators can be useful tools to provide targeted climate information for a few key variables, such as surface air temperatures.

[..[18]]

[10]removed: The newly developed MAGICC-MESMER coupled emulator will allow unprecedented assessments of the implications of manifold emissions pathways at regional scale

[11]removed: our climate (IPCC, 2013)

[12]removed: System

[13]removed: because

[14]removed: given that

[15]removed: (IPCC, 2013, 2018)

[16]removed: (CAT, 2019, 2021a)

[17]removed: efficient ESM emulators could

[18]removed: Thus far, ESM emulators have primarily been employed to swiftly translate emission pathways into global-mean climate projections, most prominently forced global temperature change (Meinshausen et al., 2009; Clarke et al., 2014; Rogelj et al., 2016, 2018; Nicholls et al., 2021c; CAT, 2021a). Regional climate information would constitute a valuable addition, since it is more directly related to climate change impacts and the climate that people experience (Seneviratne et al., 2016). Spatially resolved climate information accounting for both climate response uncertainty and natural climate variability for a given emission pathway would be especially useful for policy makers to understand the implications of mitigation efforts for their own country and to assess what climate change adaptation measures need to be implemented.

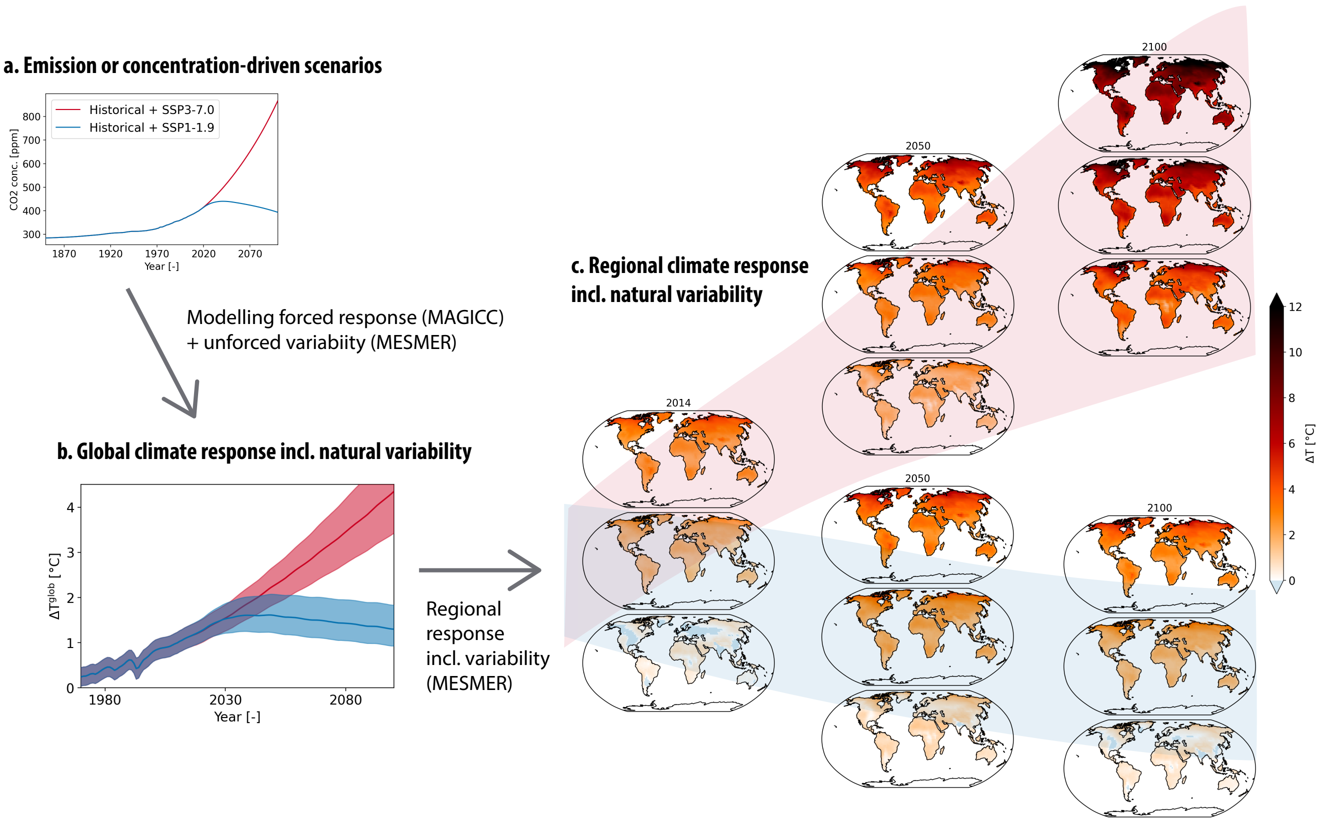

Figure 1. Illustration of the MAGICC-MESMER emulator chain. (a) The sequence of the analysis starts from either an emission or a concentration [..[19]]scenario, as highlighted here for the CO_2 concentration time series of the historical time period and two Shared Socioeconomic Pathway (SSP) scenarios (O'Neill et al., 2017), namely SSP1-1.9 and SSP3-7.0. (b) The probabilistic global mean temperature change distributions – whose medians and 90 % ranges (5th – 95th percentile) are shown – consist of realizations which are a combination of the scenario-specific forced global [..[20]]warming from MAGICC and emulated global natural variability from MESMER. (c) Based on this information, MESMER can derive the associated spatially resolved temperature change distributions, with the maps shown here representing the 5th percentile, the median, and the 95th percentile (each map trio from bottom to top) for 2014 (first column of maps), for 2050 for both SSP scenarios (second column of maps), for and 2100 for both SSP scenarios (third column of maps).

Here, we present the first chain of computationally efficient ESM emulators able to translate user-defined emission or concentration scenarios into spatially and temporally resolved temperature anomalies with respect to a pre-industrial baseline accounting for all major sources of climate change uncertainty (Fig. 1). [..[21]]In this study, the global MAGICC emulator [..[22]](Meinshausen et al., 2009, 2011, 2020) is used to turn greenhouse gas [..[23]]concentration pathways into constrained

[21]removed: The
[22]removed: (Meinshausen et al., 2009, 2011)
[23]removed: emission – or concentration –

probabilistic forced global temperature change time series by taking multiple lines of evidence into account. These are then translated into temperature anomaly field time series, accounting for both regional forced climate response and internal natural variability uncertainty, with the spatially resolved MESMER emulator (Beusch et al., 2020a), calibrated on a [..[24]]set of ESMs from the [..[25]]sixth phase of the Coupled Model Intercomparison Project (CMIP6; Eyring et al., 2016) ensemble.

Figure 2 illustrates how the different sources of uncertainty accounted for in the MAGICC-MESMER chain accumulate at the local scale by presenting time series for a single grid point in [..[32]]eastern North America as well as maps of example realizations in 2030 for a low emission scenario. In the most basic setup of the MAGICC-MESMER chain, a single forced global mean temperature change time series can be coupled to a single set of local [..[33]]forced response parameters, resulting in a single [..[34]]spatially resolved forced warming time series (Fig 2a). When accounting for global climate response uncertainty by using MAGICC's probabilistic distribution of forced global mean temperature time series in combination with the single local [..[35]]forced response parameter set, [..[36]]a range of realizations are obtained but [..[37]]they all share the same spatial pattern (Fig 2b). Different spatial patterns become available once the uncertainty in the regional forced response is included and the forced global temperature time series are combined with each of the different available ESM-specific local forced response parameter sets (Fig. 2c). The last source of uncertainty is ESM-specific natural climate variability, which is added on top of the local forced response patterns (Fig. 2d). In this low emission scenario, natural variability accounts for roughly half of the uncertainty at the local scale even at the end of the 21st century. [..[38]]For a high emission scenario[..[39]], the overall spread at the end of the century in terms of regional forced response would increase, and thus, natural variability would be less important compared to the other sources of uncertainty.

While this study is, to our knowledge, the first to combine all of these uncertainties in a single emulator chain, rich background literature exists on emulating each individual uncertainty aspect. The uncertainty in the global climate response to greenhouse gas emissions is modeled by a variety of global emulators, which usually have a physical core and use different statistical approaches during calibration (Nicholls et al., 2021c). The uncertainty in the local forced response to global temperature is most frequently accounted for through different flavours of pattern scaling (Tebaldi and Arblaster, 2014; Lynch et al., 2017; Beusch et al., 2020a). The last element of our emulation chain, local-scale natural variability, has been stochastically created through re-sampling of either individual temperature field realizations (Alexeeff et al., 2018) or of principal components with perturbed phases (Link et al., 2019), and by sampling from autoregressive processes with spatially correlated innovation terms (Beusch et al., 2020a). Additionally, there are two studies which directly translate emissions into spatially resolved temperature change realizations (Goodwin et al., 2020; Yuan et al., 2021). However, both of the available approaches lack one

[24]removed: range
[25]removed: 6th
[32]removed: Eastern North-America
[33]removed: mean
[34]removed: realization
[35]removed: mean
[36]removed: different
[37]removed: each emulation ensemble member exhibits
[38]removed: If
[39]removed: were studied instead

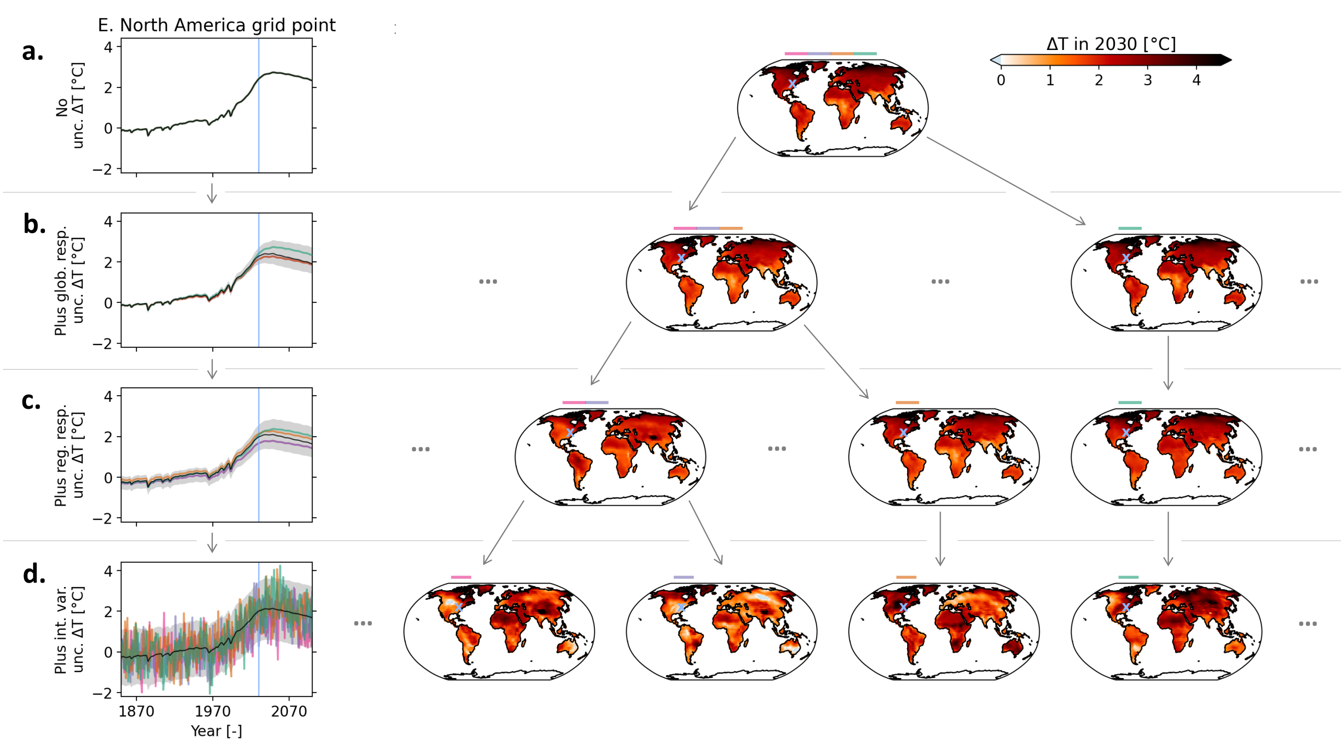

Figure 2. Visualization of the cumulative contribution of different uncertainty sources accounted for in the MAGICC-MESMER emulator chain with time series for a single grid point in [..[26]]eastern North America and with maps of example realizations [..[27]]in the year 2030 [..[28]]for the low emission scenario SSP1-1.9. The example emulations [..[29]]shown in the maps are also depicted in the grid-point-level time series by using the [..[30]]colors displayed on top of the maps. Note that whenever several colors are depicted on top of a single map, the associated emulations all exhibit an identical temperature change trajectory. Hence, the individual example emulations fully overlap both in the map and in the time series plot. Light blue markers are employed to indicate the year 2030 in the time series plots and to show the location the time series belong to on the maps. From top to bottom, the rows exhibit an increasing reflection of uncertainties: (a) Accounts for no uncertainty since a single MAGICC forced global warming time series is combined with a single MESMER forced response pattern and thus, all emulations are identical. (b) Accounts for the uncertainty in the global response by combining all of MAGICC's forced global warming time series with a single MESMER forced response pattern. In the time series plot, the solid black line indicates the median of the temperature change distribution for that grid point and the gray shading the 90 % range (5th – 95th percentile). (c) Adds uncertainty in the regional response by combining all of MAGICC's forced global warming time series with all of MESMER's forced response patterns. (d) Adds natural internal variability [..[31]]provided by MESMER on top of the forced response patterns.

major source of uncertainty. Goodwin et al. (2020) combine the global emulator WASP with a pattern scaling approach to obtain probabilistic mean warming realizations but do not account for internal climate variability. Yuan et al. (2021), on the other hand, use a statistical emulator calibrated on a single ESM to create spatially resolved temperature realizations directly from CO_2 equivalence concentrations and thus neither account for forced climate response uncertainty nor uncertainty in the representation of unforced natural climate variability.

2 Data

2.1 Earth System Model data

We use climate information from 25 ESMs participating in the Scenario Model Intercomparison Project (ScenarioMIP; O'Neill et al., 2016) of CMIP6 (Eyring et al., 2016) listed in Appendix Table A1. While our ESM ensemble does not contain all of the ESMs available within CMIP6, it does cover a wide range of temperature realizations and is broadly representative of the overall warming spread and the relative fraction of simulations available for each scenario.

Data from 1850 – 2100 are employed, covering the historical time period (1850 – 2014) and the ScenarioMIP's future (2015 – 2100) emission and [..[40]]land-use change scenarios, the Shared Socioeconomic Pathways (SSPs), i.e., SSP1-1.9, SSP1-2.6, SSP4-3.4, SSP5-3.4-over, SSP2-4.5, SSP4-6.0, SSP3-7.0, SSP5-8.5. In this study, a special focus is set on SSP1-1.9, the most ambitious mitigation scenario of CMIP6, and SSP3-7.0, a physically plausible high-end emission scenario (Hausfather and Peters, 2020).

We primarily employ annual 2-m air temperature (T) data but ocean heat uptake (OHU) is additionally used for some analyses. Both variables are re-gridded onto a common 2.5° x 2.5° spatial grid according to Brunner et al. (2020a). Anomalies with respect to 1850 – 1900 are obtained at the [..[41]]grid-point level and used for all analyses.

Regional averages refer to area-weighted averages and regional land averages to area- and land-fraction-weighted averages. Grid cells with more than one third land fraction according to the land-sea mask of the regionmask package v0.8.8 (Hauser et al., 2021b) are considered to be land grid points. The global land average does not include Antarctica, since no emulations are created for Antarctica. The [..[42]]eastern North America region frequently employed [..[43]]throughout this study is one of the updated Intergovernmental Panel on Climate Change (IPCC) climate reference regions (Iturbide et al., 2020) and is extracted from the gridded fields with the help of the regionmask package [..[44]]v0.8.0 (Hauser et al., 2021b).

2.2 Observational data

To calibrate MESMER, in addition to temperature data, a proxy for volcanic activity is needed (Beusch et al., 2020a). Here, we use the globally averaged stratospheric aerosol optical depth time series employed in the recently published sixth assessment report of the IPCC (Forster et al., 2021; Smith et al., 2021b) for this purpose. The time series is available in monthly resolution (Smith et al., 2021a) and thus, annual averages need to be created before using it.

[40] removed: land use
[41] removed: grid point
[42] removed: Eastern North-America
[43] removed: in
[44] removed: (?)

For a qualitative visual validation of the MAGICC-MESMER emulations, annual blended temperature anomalies from the Berkeley Earth data set are employed [..[45]](Rohde and Hausfather, 2020), which consist of land and sea ice 2-m air temperature anomalies and ocean surface temperature anomalies. They are interpolated onto the same spatial grid as the ESM data, also following the approach described in Brunner et al. (2020a). To account for differences in the global average of blended temperature anomalies and [..[46]]2-m air temperature anomalies, the observational global blended averages $\Delta T^{glob,blend}$ are transformed into 2-m air global averages $\Delta T^{glob,air}$ at every point in time t via [..[47]]$\Delta T_t^{glob,air} = 1.098 \cdot \Delta T_t^{glob,blend} - 0.001\,°\mathrm{C}$, a relationship derived by Beusch et al. (2020b) based on CMIP6 data. Additionally, the observational data [..[48]]need to be shifted from their native 1951 – 1980 baseline to the 1850 – 1900 baseline employed within this study. However, during the 1850 – 1900 time period, the observational data are not spatially complete and the quality of the available data is lower. Hence, the MAGICC-MESMER emulated regional median warming between the 1850 – 1900 baseline and the native baseline of the observational data is added to the observational data as a constant offset during plotting.

3 Methods

3.1 MAGICC (v7.5.1)

The Model for the Assessment of Greenhouse gas Induced Climate Change (MAGICC) is a reduced complexity climate model, which calculates – among other quantities – forced global warming and global ocean heat uptake. Its hemispherically-resolved, multi-layer upwelling-entrainment-diffusion ocean and climate core (based on the energy balance equation with state- and forcing-dependent climate sensitivity) are described in Meinshausen et al. (2011) and Meinshausen et al. (2020). It also includes representations of the carbon cycle, non-CO2 greenhouse gas cycles and the relationship between aerosol prescursor species emissions, and aerosol effective radiative forcing (Meinshausen et al., 2011, 2020) alongside a parameterization of the response of permafrost to global heating (Schneider von Deimling et al., 2012).

The MAGICC output employed throughout this paper is the same as presented in Nicholls et al. (2020) and Nicholls et al. (2021c). We use greenhouse gas concentration-driven simulations, following the CMIP6 ScenarioMIP approach (O'Neill et al., 2016). A slightly wider output temperature distribution would be achieved, if MAGICC were run in emission-driven mode instead (Nicholls et al., 2021c).

[45]removed: (Rohde et al., 2013)

[46]removed: and

[47]removed: the following formula $\Delta T_t^{glob,air} = 1.098 \cdot \Delta T_t^{glob,blend} - 0.001°\mathrm{C}$

[48]removed: needs

3.2 MESMER (v0.8[..[49]].3)

3.2.1 Default configuration – Local temperature anomalies as a function of global temperature anomalies

MESMER, a Modular Earth System Model Emulator with spatially Resolved output, was introduced [..[50]]in Beusch et al. (2020a) to emulate multi-ESM initial-condition ensembles for a specific emission pathway. This first MESMER [..[51]]configuration has thus far been successfully applied to emulate temperature [..[52]]anomalies for the highest emission scenarios of both the CMIP5 (Beusch et al., 2020a) and the CMIP6 (Beusch et al., 2020b) ensembles.

MESMER is an ESM-specific emulator which needs to be calibrated for each [..[53]]emulated ESM separately, to capture the unique characteristics of that ESM, both in terms of local forced warming and local variability around the forced warming (Beusch et al., 2020a). Within MESMER, local temperature anomalies ΔT for a specific climate model m at every point in space s and time t are emulated as follows:

$$\Delta T_{m,s,t} = f(\Delta T^{glob}_{m,t}) + \eta_{m,s,t} = \beta^{fr}_{m,s} \cdot \Delta T^{glob,fr}_{m,t} + \beta^{iv}_{m,s} \cdot \Delta T^{glob,iv}_{m,t} + \beta^{intercept}_{m,s} + \eta_{m,s,t}. \tag{1}$$

The local temperature anomaly is a direct function of global mean temperature change ΔT^{glob} and a spatio-temporally correlated noise term η. At every point in space, a multiple linear regression is employed to relate ΔT^{glob} information to local temperature anomalies. Forced global temperature change $\Delta T^{glob,fr}$ and internal global temperature variability $\Delta T^{glob,iv}$ serve as predictors. The associated regression coefficients are β^{fr} and β^{iv}[..[54]]. $\beta^{intercept}$ constitutes the intercept term. $\Delta T^{glob,fr}$ is obtained by first applying locally weighted scatterplot smoothing (LOWESS) to the ΔT^{glob} time series and subsequently adding volcanic spikes to the smooth forced temperature time series via a linear regression of the global temperature anomaly residuals to stratospheric aerosol optical depth. The remaining internal global variability $\Delta T^{glob,iv}$ and the residual local variability η are both modeled as autoregressive processes. For [..[55]]η, spatially correlated innovation terms are drawn to account for local to regional cross-correlations [..[56]]between grid points. Beusch et al. (2020a) describe the full algorithm and additionally carry out a thorough evaluation of the resulting emulations. In short, their evaluation highlights that MESMER reliably emulates grid-cell-level forced warming and internal climate variability. However, they also show that MESMER's emulations are increasingly underdispersive for larger regional averages since MESMER's local residual variability module reduces the magnitude of the empirical covariance between grid cells as a function of their geographical distance. At the global land level, MESMER emulations become reliable again, due to the global scale variability captured through the $\beta^{iv}_{m,s} \cdot \Delta T^{glob,iv}_{m,t}$ term.

---

[49] removed: .1
[50] removed: by
[51] removed: variant has
[52] removed: changes
[53] removed: ESM it emulates
[54] removed: , and
[55] removed: the
[56] removed: . The full algorithm is described in Beusch et al. (2020a)

In this work, MESMER is for the first time trained on and applied to the full range of emission scenarios covered within the CMIP6 ScenarioMIP projections instead of a single emission pathway. The main additional assumption when extending MESMER from emulating a single emission pathway to emulating a range of scenarios is that [..[57]]the estimated parameters are scenario-independent. This assumption of universal scenario applicability of the calibrated parameters is also regularly made in the well established pattern scaling literature (Tebaldi and Arblaster, 2014), although care is nonetheless required, particularly for strong mitigation scenarios (Goodwin et al., 2020).

To obtain robust MESMER parameter estimates for each ESM, MESMER is trained on all available ensemble members of each available scenario and equal weight is given to each scenario. The scenarios employed for training are – subject to availability – SSP5-8.5, SSP3-7.0, SSP4-6.0, SSP2-4.5, SSP5-3.4-over, SSP4-3.4, SSP1-2.6, SSP1-1.9, and Historical. Note that we consider the historical time period as its own scenario during the MESMER training. Since [..[58]]projections for different SSPs usually branch from the same historical members, the historical years would receive more weight if the historical and the future time period were concatenated during training.

Many ESMs have a different number of ensemble members (n_e) for each scenario, and the scenarios (historical and projections) have a different number of time steps (n_t), resulting in a different number of samples ($n_s = n_e \cdot n_t$) per scenario.

Therefore, when estimating the coefficients of the multiple linear regression, all simulations from all scenarios are pooled and each sample (e.g., the temperature anomaly for a specific location, year, and scenario) is weighted by $1/n_s$, i.e., by the inverse number of samples available for the respective scenario. The same approach is applied when estimating the empirical spatial covariance matrix of the residual local variability.

For the autoregressive processes, the coefficients are determined for each ensemble member individually and are subsequently averaged for each scenario before averaging across all scenarios. For the autoregressive process describing the global variability, the order of the process must additionally be selected before the parameters are fit. The order is first chosen for each ensemble member individually based on the Bayesian Information Criterion and then, the median order is identified for each scenario. Lastly, the median order over all scenarios is selected to describe the global variability of the ESM at hand.

The ESM-specific forced global warming time series for each scenario is obtained by using solely data from that scenario, since the forced global warming module relies on a simple statistical smoothing and volcanic spikes approach (see Beusch et al., 2020a, for details). To facilitate the task of the forced global warming module, it operates directly on the average global warming across all ensemble members for a given ESM and scenario, rather than on individual ensemble members.

3.2.2 Additional predictors configuration – Local temperature anomalies as a function of global temperature anomalies and global ocean heat uptake

To account for potential non-linearities in the climate system (Mitchell, 2003) and for changing forced response warming patterns when moving from a transient to an equilibrium climate (King et al., 2020), we introduce two additional predictors

[57] removed: , once calibrated, the ESM-specific MESMER parameters can be used to emulate any emission scenario
[58] removed: different future SSP projections

into the grid-point-level multiple linear regressions of MESMER (Eq. 1)[..[59]]: (i) Squared forced global temperature change $(\Delta T^{glob,fr})^2$ is used to represent non-linear feedbacks and (ii) the forced trend in global ocean heat uptake $\Delta OHU^{glob,fr}$ as a proxy for how close the climate system is to equilibrium. $\Delta OHU^{glob,fr}$ is obtained with the same LOWESS approach used to derive $\Delta T^{glob,fr}$ but no volcanic spikes are added, since they are already accounted for in the global temperature response. Hence, the grid-point-level temperature anomalies are emulated as follows:

$$\Delta T_{m,s,t} = \beta^{fr}_{m,s} \cdot \Delta T^{glob,fr}_{m,t} + \beta^{fr2}_{m,s} \cdot (\Delta T^{glob,fr}_{m,t})^2 + \beta^{fr,ohu}_{m,s} \cdot \Delta OHU^{glob,fr}_{m,t} + \beta^{iv}_{m,s} \cdot \Delta T^{glob,iv}_{m,t} + \beta^{intercept}_{m,s} + \eta_{m,s,t}, \tag{2}$$

where β^{fr2} and $\beta^{fr,ohu}$ are the newly introduced regression coefficients.

3.2.3 Evaluating both MESMER configurations

Before continuing with the methods, we evaluate the two proposed MESMER configurations. Specifically, we assess if the additional predictors configuration offers considerable benefits over the default configuration. When analyzing climate information, it is often helpful to distinguish between forced response and natural variability. [..[60]]Here, we evaluate both MESMER configurations in terms of their performance in emulating single ESMs with respect to these two quantities. The two can be separated in a straightforward fashion within the additive MESMER framework ([..[61]]Eqs. 1 and 2).

The forced local temperature anomaly ΔT^{fr} of the default MESMER configuration consists of the following terms:

$$\Delta T^{fr}_{m,s,t} = \beta^{fr}_{m,s} \cdot \Delta T^{glob,fr}_{m,t} + \beta^{intercept}_{m,s}. \tag{3}$$

In the additional [..[62]]predictors configuration of MESMER, the contribution of the additional predictors enters the forced local temperature anomaly too, resulting in:

$$\Delta T^{fr}_{m,s,t} = \beta^{fr}_{m,s} \cdot \Delta T^{glob,fr}_{m,t} + \beta^{fr2}_{m,s} \cdot (\Delta T^{glob,fr}_{m,t})^2 + \beta^{fr,ohu}_{m,s} \cdot \Delta OHU^{glob,fr}_{m,t} + \beta^{intercept}_{m,s}. \tag{4}$$

Figure 3 visualizes the latitudinally-averaged mean of the local forced warming over the last 30 years for each individual ESM and scenario for both MESMER configurations. For a given scenario, the ESMs' latitudinal patterns exhibit considerable differences. For example, the magnitude of the Arctic amplification (Serreze and Barry, 2011) differs [..[69]]clearly between the ESMs. These regional differences are additionally visualized in maps for select high and low emission scenarios for each ESM in the [..[70]]supplement (Figs. S1 – S5). Overall, the patterns of forced warming are very similar in the default and additional [..[71]]predictors configurations (Fig. 3) and inter-ESM differences are considerably larger than inter-MESMER-configuration differences.

[59] removed: .
[60] removed: In this section
[61] removed: Eq
[62] removed: predictor
[69] removed: starkly
[70] removed: supplementary material
[71] removed: predictor

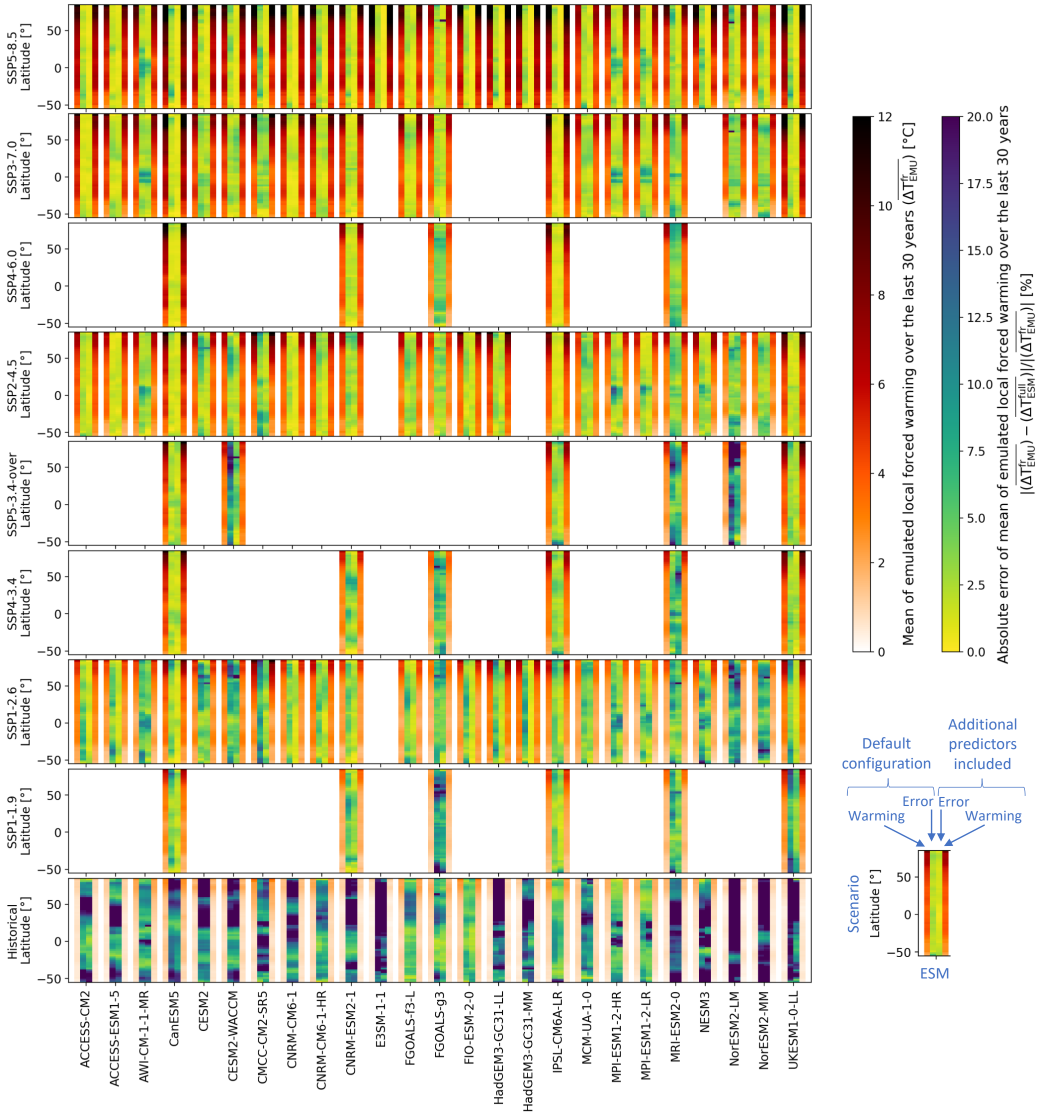

Figure 3. Latitudinally-averaged grid-point-level mean of emulated local forced warming and performance with respect to ESM simulations for both MESMER configurations [..[63]]over the last 30 years of each scenario. The [..[64]]error shown here represents the absolute deviation [..[65]]between the [..[66]]mean of the emulated forced warming [..[67]]and the [..[68]]mean temperature anomaly value across all available ESM initial-condition members for the scenario at hand and is normalized by the absolute value of the mean of the emulated forced warming.

Nevertheless, [..[72]]with the additional predictors configuration, consistent but small improvements are achieved [..[73]]for most ESMs (Fig. 3). The magnitude of the improvements depends on the ESM, the scenario, and the latitude. The improve-

[72]removed: small but consistent

[73]removed: in most ESMs with the additional predictor configuration

ments mostly occur in the highest emission scenario SSP5-8.5 and in the strong mitigation scenarios SSP5-3.4-over, SSP1-2.6, and SSP1-1.9. This is an expected consequence of the additional predictors targeting non-linearities and the transition to an equilibrium climate.

Overall, the performance is excellent for both MESMER configurations (Fig. 3): the [..[74]]latitudinally-averaged absolute error with respect to the local warming signal rarely reaches 10 % in the future projections. The SSP1 scenarios have a tendency for higher [..[75]]percentage errors, because of their lower overall forced warming signal. This is amplified in the historical period[..[76]], since its low regional warming leads to small absolute [..[77]]deviations translating into large [..[78]]percentage errors. In the [..[79]]supplement (Figs. S1 – S5), [..[80]]error maps can be found which further highlight regional differences in the errors for selected high and low emission scenarios.

Sensitivity experiments are additionally carried out to quantify the impact on emulation performance when using a reduced number of scenarios during training. For MESMER's default configuration, a generally very similar although slightly reduced performance is achieved by only training on a single high (SSP5-8.5) and low (SSP1-2.6) emission scenario (Fig. S6). If only a single future scenario and the historical time period are used for training, a high emission future scenario results in a better overall performance. This is because training on a low emission scenario requires MESMER to extrapolate to warm climates when emulating higher emission scenarios. Overall, the SSP5-8.5 trained default MESMER configuration often performs comparably to MESMER trained on all scenarios but experiences minor performance reductions in strong mitigation scenarios whose climate moves towards equilibrium conditions[..[81]]. The SSP1-2.6 trained emulator, on the other hand, struggles to emulate high emission scenarios (Fig. S6). Hence, for MESMER's default configuration even a single (high emission) scenario largely suffices to emulate a wide range of different emission pathways as long as their emissions do not exceed the ones used for training. The best emulation performance across all scenarios is, however, achieved if both high and low emission scenarios are included in the training.

If the additional predictors configuration of MESMER is applied instead, it is more important that at least [..[82]]one high and one low emission scenario are available during training because training on a single emission pathway leads to overfitting on the scenario type and thus poor emulation skill in other scenarios (Fig. S7). This mainly occurs because the cross-correlations between the three predictors ($T_{m,t}^{glob,fr}$,$(\Delta T_{m,t}^{glob,fr})^2$, and [..[83]]$\Delta OHU_{m,t}^{glob,fr}$) are fundamentally different in the high emission transient climate change scenarios compared to the low emission equilibrium-approaching scenarios.

---

[74] removed: latitudinal

[75] removed: relative

[76] removed: because

[77] removed: errors

[78] removed: relative

[79] removed: supplementary material

[80] removed: average error maps are found to

[81] removed: (King et al., 2020)

[82] removed: a high and a

[83] removed: $\Delta OHU_{m,t}^{glob,fr}$

After considering the forced response in detail, we now turn our attention to the local temperature variability [..[84]]for MESMER trained on all available scenarios. The ESM-specific emulated local temperature variability around the forced warming consists of the combination between the local response to the global variability and the residual local variability for both MESMER configurations:

$$\Delta T^{iv}_{m,s,t} = \beta^{iv}_{m,s} \cdot \Delta T^{glob,iv}_{m,t} + \eta_{m,s,t}. \tag{5}$$

Figure 4 shows the latitudinally-averaged standard deviation of the emulated local variability, analogously to Fig. 3 for each ESM and scenario available for that ESM. However, the standard deviations are computed over the full scenario length instead of only over the last 30 years, to obtain more robust estimates. Additionally, the standard deviation of the emulated variability is identical for every scenario since no scenario dependence is integrated in the variability emulation. Differences in the error between different scenarios for individual ESMs hence solely occur because the ESMs' variability differs from simulation to simulation.

Generally, the emulated variability is smallest in the tropics and largest in the northern high latitudes, but considerable inter-ESM differences in the magnitude and spatial distribution exist (Fig. 4). In [..[103]]supplement Figs. S8 – S12, spatially resolved maps of the standard deviation of the emulated local variability are shown, which further highlight region-specific characteristics.

Inter-MESMER-configuration differences are [..[104]]minimal, also in terms of improvements with respect to capturing the characteristics of the local variability of the true ESM simulations (Fig. 4). However, the spatially resolved maps can help pinpoint differences in the errors of the two configurations in some regions (Figs. S8 – S12).

The [..[105]]percentage errors in the emulated local variability (Fig. 4) are generally a bit larger than the forced local warming ones (Fig. 3). In part, the larger errors are caused by the fact that percentage differences are considered and [..[106]]that standard deviation of local variability is mostly a small quantity. This impression is reinforced by the fact that the largest percentage differences occur most often in the low variability tropics. A possible physical explanation of the remaining errors is revealed by consulting maps of the [..[107]]error in the standard deviation (Figs. S8 – S12). For certain ESMs and regions, the standard deviations are overestimated in the high-end scenario but underestimated in the low-end scenario (or the other way round), indicative of changes in local variability across climate states (Olonscheck and Notz, 2017), which are not accounted for in [..[108]]either MESMER configuration. In line with this finding, the highest deviations are generally observed for SSP5-8.5 which experiences the most extreme climatic change and thus also the largest changes in variability (Fig. 4). Given [..[109]]that SSP5-8.5 is designed to represent an unlikely high-risk future (Hausfather and Peters, 2020), one could [..[110]]justify

[84] removed: when training MESMER
[103] removed: supplementary material
[104] removed: only
[105] removed: relative
[106] removed: the
[107] removed: average
[108] removed: MESMER
[109] removed: the physical implausibility of this high emission scenario
[110] removed: consider

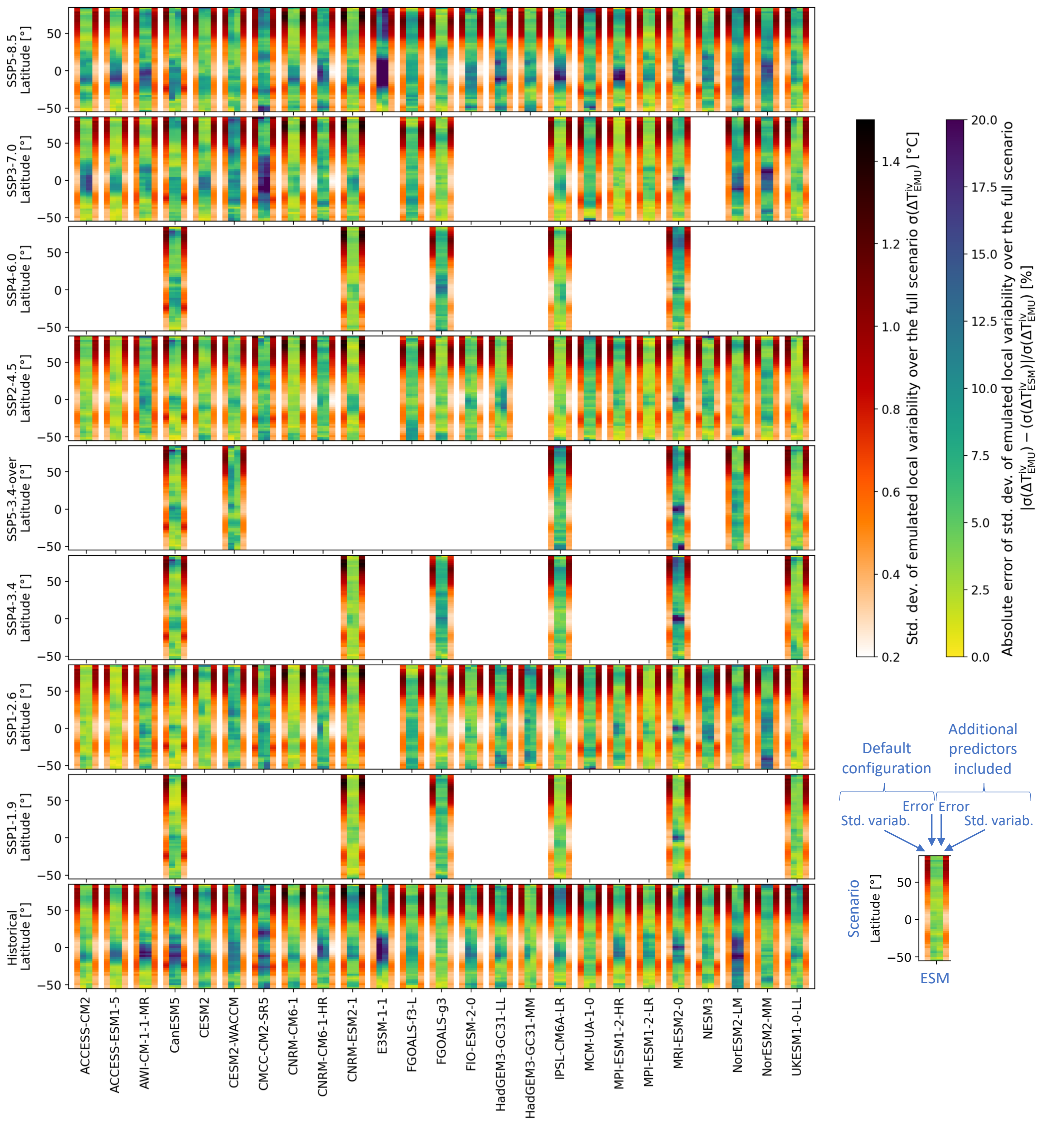

Figure 4. Latitudinally-averaged grid-point-level standard [..[85]]dev/iation of emulated local variability and performance with respect to ESM simulations [..[86]]for both MESMER configurations [..[87]]over the full scenario period. The standard deviation of the emulated local variability [..[88]]is computed based on 600 variability emulations for each ESM. To obtain local variability from ESM simulations, the emulated local forced warming is subtracted from every ESM simulation. Subsequently, the standard deviation of [..[89]]these estimates of [..[90]]ESM [..[91]]simulations' [..[92]]local variability is computed while employing all ESM initial-condition ensemble members available for the scenario at [..[93]]hand. The [..[94]]error shown here represents the absolute deviation [..[95]]between the [..[96]]standard [..[97]]deviation of the emulated variability and the [..[98]]standard deviation [..[99]]of the [..[100]]ESMs' variability normalized by the standard deviation [..[101]]of the [..[102]]emulated variability.

considering excluding it from the local variability training to further improve the local variability [..[111]]emulations for the other scenarios. However, since the absolute differences in standard deviations are nevertheless rather small, we continue to use all scenarios for training.

This detailed MESMER evaluation reveals that the additional predictors [..[112]]bring systematic but small improvements. Nevertheless, the main forced warming signal can already be successfully extracted based on forced global temperature change alone and local variability is generally emulated similarly well in both MESMER configurations. In the following, we thus use the default MESMER configuration instead of the additional predictors one. This minimizes the risk of poorly calibrated parameters when only a limited number of scenarios is available for training in which different predictors are strongly correlated.

3.3 Coupling MAGICC and MESMER

After determining the MESMER configuration to use throughout the rest of the paper, we now describe the employed MAGICC-MESMER coupling approach. MAGICC and MESMER are calibrated individually before using them jointly to create ensembles of spatially resolved emulations. In the coupled MAGICC-MESER emulation mode, MESMER's own statistical estimates of $\Delta T^{glob,fr}$ in Eq. 1 are replaced with MAGICC's $\Delta T^{glob,fr}$ estimates (Sect. 3.2.1), making it possible to also provide spatially resolved emulations for emission scenarios which were not available during the MESMER training.

3.3.1 [..[113]]Earth-System-Model-specific emulations

When emulating a specific ESM with MAGICC, a single [..[114]]forced global temperature anomaly $\Delta T^{glob,fr}$ time series is obtained for every emission scenario. Here, we employ ESM-specific MAGICC output published as part of the Reduced Complexity Model Intercomparison Project (RCMIP) Phase 1 (Nicholls et al., 2020) for two ESMs, CanESM5 and CNRM-CM6-1. Note that this output was generated with MAGICC v7.1.0.-beta but that very similar results would be obtained with MAGICC v7.5.1, which is employed throughout all other parts of our study.

To obtain full global realizations ΔT^{glob} for these ESMs, the stochastically-generated, ESM-specific global variability of MESMER $\Delta T^{glob,iv}$ is added to MAGICC's ESM-specific $\Delta T^{glob,fr}$ time series. For this study, 600 $\Delta T^{glob,iv}$ emulations are created with MESMER for each ESM. Hence, the MAGICC-MESMER ΔT^{glob} ensemble [..[115]]contains 600 realizations per ESM.

The spatially resolved forced warming fields ΔT^{fr} are obtained by combining MAGICC's $\Delta T^{glob,fr}$ time series with the associated ESM's local forced response parameters provided by MESMER (Eq. 3). The resulting ΔT^{fr} field time series is combined with 600 local variability emulations ΔT^{iv} for that ESM provided by MESMER (Eq. 5), leading to a 600 member ESM-specific MAGICC-MESMER ensemble of ΔT (Eq. 1).

[111]removed: representation in

[112]removed: do bring subtle but systematic improvements. Nonetheless

[113]removed: ESM-specific

[114]removed: global forced

[115]removed: also

The ESM-specific MAGICC-MESMER ensembles can be regarded as a direct approximation of very large [..[116]]ESM initial-condition ensembles (Deser et al., 2012, 2020), which can be provided at a negligible computational cost for any emission scenario of interest.

3.3.2 Globally-constrained probabilistic emulations

To derive trustworthy probabilistic climate projections, which thoroughly sample the climate response and natural variability [..[117]]uncertainty space for any emission scenario, observational constraints are often required. At each point of the MAGICC-MESMER emulation chain, an observational constraint could theoretically be introduced.

Several studies employing fundamentally different approaches have all demonstrated that the CMIP6 ESMs which exhibit the strongest [..[118]]future forced global warming are not consistent with observationally-constrained forced global warming estimates (Forster et al., 2019; Beusch et al., 2020b; Brunner et al., 2020b; Tokarska et al., 2020; Ribes et al., 2021; Nicholls et al., 2021c). In terms of regional forced warming response to the global [..[119]]forced warming, i.e., in terms of regionally-averaged $\beta^{fr}_{m,s}$ (see Eq. 1), however, most CMIP6 ESMs perform in an observationally-consistent manner in most regions (Beusch et al., 2020b). [..[120]]Therefore, the MAGICC-MESMER [..[121]]constrained probabilistic projections are constrained solely at the global level [..[122]]and span the full regional ESM response uncertainty [..[123]]in this study.

The probabilistic MAGICC output used in this study follows the HadCRUT5 (Morice et al., 2021) calibration of MAGICC presented in Nicholls et al. (2021c) and consists of 600 forced global temperature change $\Delta T^{glob,fr}$ time series per scenario. [..[124]]It is assumed that global- and regional-scale performance are sufficiently decoupled to allow for combining MAGICC's probabilistic output with each of MESMER's ESM-specific local parameter sets. This assumption is supported by Beusch et al. (2020b) demonstrating that there is no direct relation between an ESM's performance skill for global-scale forced warming response to emissions and regional-scale forced warming response to forced global warming. Hence, similarly to the ESM-specific emulation approach, a full [..[125]]initial-condition ensemble is created for each of the 600 $\Delta T^{glob,fr}$ time series and each of the 25 ESM-specific parameter sets. For the global temperature realizations ΔT^{glob}, this means that each of the 600 MAGICC $\Delta T^{glob,fr}$ time series are combined with 600 global variability $\Delta T^{glob,iv}$ realizations of 25 ESMs resulting in a total number of nine million ΔT^{glob} realizations. For the local temperature anomaly field realizations ΔT, each of MAGICC's $\Delta T^{glob,fr}$ time series is combined with the ESM-specific local parameter sets of MESMER (Eq. 1). This results in 600 local forced temperature change ΔT^{fr} field time series for each ESM-specific parameter set (Eq. 3). Each of these local

[116] removed: ESM-initial condition
[117] removed: climate change
[118] removed: forced global future
[119] removed: warming,
[120] removed: In line with these findings
[121] removed: probabilistic constrained
[122] removed: but
[123] removed: range
[124] removed: Beusch et al. (2020b) demonstrated that there is no direct relation between an ESM's performance skill for global-scale response to emissions and regional-scale response to global temperatures. Therefore it
[125] removed: initial condition

forced temperature field time series is subsequently combined with 600 local variability ΔT^{iv} realizations from MESMER (Eq. 5) yielding 600 different [..[126]]initial-condition ensembles with 600 members each (Eq. 1). Thus, 360'000 probabilistic ΔT realizations are contributed from each set of ESM-specific local parameters and the full probabilistic ensemble contains nine million emulations.

The complete-sampling approach we employ here ensures a thorough sampling of the full uncertainty space, but may quickly lead to computer memory issues. Alternatively, a broad – albeit sparser – sampling of the full climate response and natural variability uncertainty space could also be achieved by randomly combining single MAGICC time series with single ESM-specific parameter sets.

4 Results

4.1 Earth-System-Model-specific temperature projections

To illustrate the MAGICC-MESMER chain's ability to capture the distinct behavior of individual ESMs in its ESM-specific configuration, we approximate the behavior of two example ESMs, CanESM5 and CNRM-CM6-1, for two example scenarios, SSP3-7.0 and SSP1-1[..[130]].9 (Fig. 5). In both scenarios, the emulations serve to fill the gaps in the natural variability uncertainty space, by successfully capturing the characteristic ESM-dependent forced warming and climate variability around that warming for the different spatial scales. The increasing magnitude of natural variability for smaller spatial scales results in an increasing overlap between the [..[131]]temperature anomaly distributions spanned by the two ESMs. In the low emission scenario, the emulations additionally fill a gap in the climate response uncertainty space, since no SSP1-1.9 simulations are available for CNRM-CM6-1. Hence, in this configuration, our emulator chain approximates the full climate change uncertainty phase space spanned by the considered ESMs since it can create very large emulated intial-condition ensembles for any given scenario.

It should be noted that the success of these ESM-specific emulations is strongly dependent on MAGICC's ability to match a given ESM's behavior over a range of scenarios. While we [..[132]]satisfactorily emulate the example ESMs shown here, there are other ESMs which are less well-captured by MAGICC in this emulation mode [..[133]](see Nicholls et al., 2020). On the other hand, the regional features of single ESMs given the [..[134]]forced global warming trajectories are generally well captured by MESMER (Sect. 3.2.3). Overall, in the context of ESM-specific emulation, more work on emulating ESMs' forced global temperature change is required compared to the relative success seen in emulating an ESM's natural variability and regional response to forced global warming.

[126]removed: initial condition
[130]removed: .0
[131]removed: two
[132]removed: successfully
[133]removed: (see e.g., Nicholls et al., 2020)
[134]removed: global forced

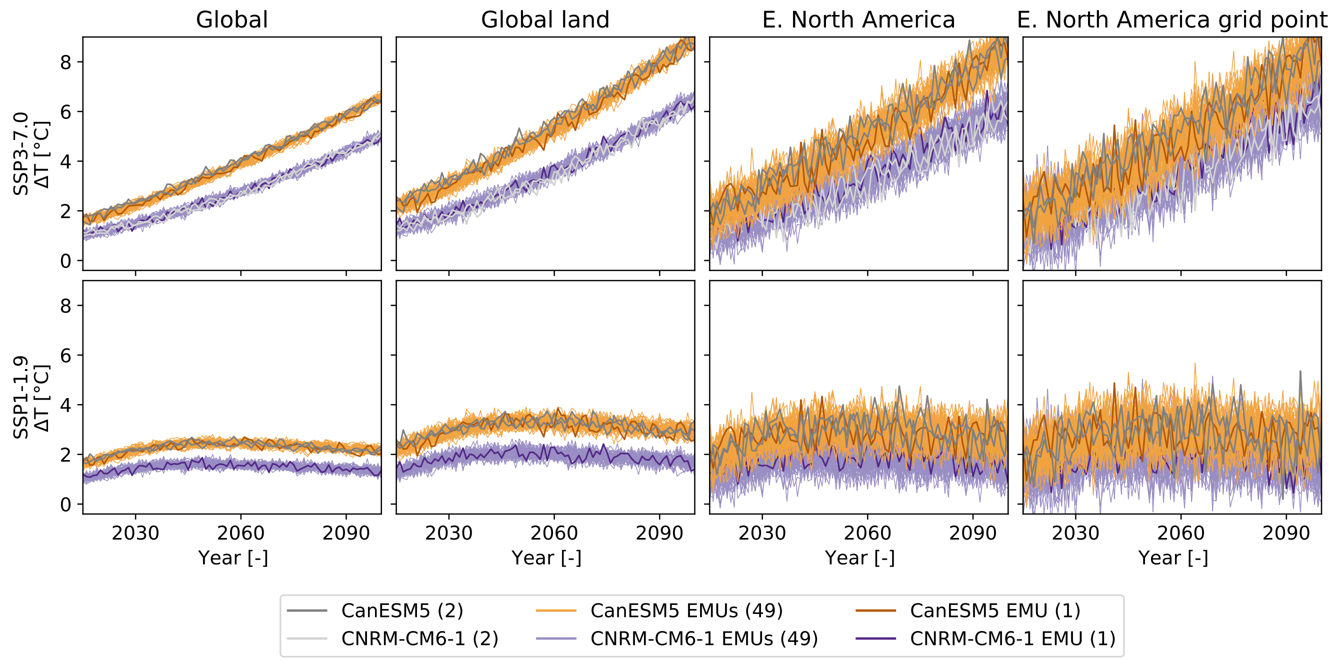

Figure 5. Temperature anomaly time series of ESM-specific MAGICC-MESMER emulations (in color) and actual ESM simulations (in gray), for CanESM5 and CNRM-CM6-1, averaged across different spatial scales: global, global land, [..[127]]eastern North America, and a single grid point within [..[128]]eastern North America for a high (SSP3-7.0) and a low [..[129]](SSP1-1.9) emission scenario. For illustrative purposes, 50 out of the 600 available emulations are shown for both ESMs in color, and a single emulation is highlighted in darker color. For the ESMs, two initial-condition ensemble members are shown for each scenario for which simulations are available.

4.2 Globally-constrained probabilistic temperature projections

When sampling the full globally-constrained climate response and internal variability uncertainty space with the probabilistic MAGICC-MESMER ensemble, the emulations no longer coincide with individual ESM simulations (Fig. 6). Instead, the globally-constrained emulated ensemble encompasses a smaller range of temperature anomalies than the raw CMIP6 ensemble, and also samples this space much more thoroughly with nine million emulations. This thorough sampling means that even extreme quantiles at individual grid cells can be statistically robustly estimated [..[138]]in any year. An additional advantage with respect to the CMIP6 ensemble is that the same forced climate response and natural variability uncertainty space can be sampled for each scenario, whereas there are [..[139]]strong differences in the number of ESMs providing simulations for each scenario as well as in the number of available simulations of a single ESM for a specific scenario. This is especially relevant, because SSP1-1.9, which is of great interest for society as it comes closest to the 1.5 °C Paris Agreement [..[140]]temperature

[138]removed: for
[139]removed: stark
[140]removed: target

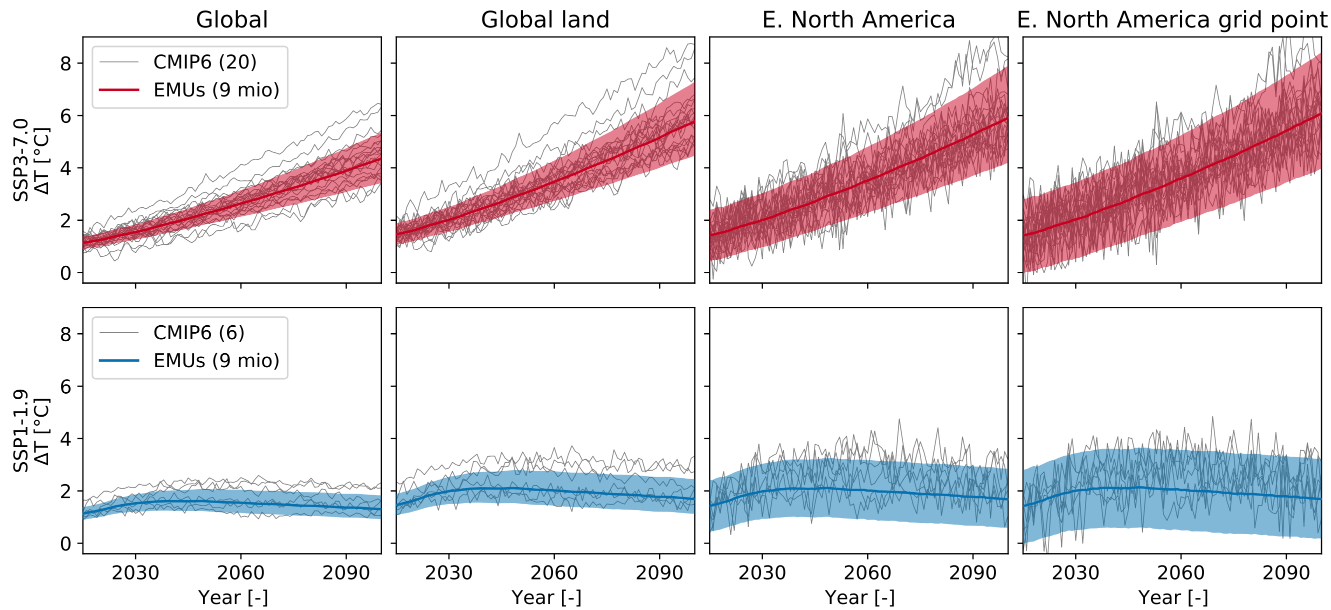

Figure 6. Temperature anomaly projection distributions for the emulated globally-constrained probabilistic MAGICC-MESMER ensembles and actual ESM simulations averaged across different spatial scales: global, global land, [..[135]]eastern North America, and a single grid point within [..[136]]eastern North America for a high (SSP3-7.0) and a low (SSP1-1.9) emission scenario. For the MAGICC-MESMER ensemble, the median and the 90 % range (5th – 95th percentile) of the temperature anomaly distribution are shown in color. For the ESMs, a single simulation per available ESM for that [..[137]]emission pathway is shown in gray.

limit (UNFCCC, 2015), is one of the scenarios which has only been run by a rather small – and warm – subselection of ESMs. In our CMIP6 ensemble, only six ESMs provide SSP1-1.9 simulations two of which warm unrealistically fast during the historical period in terms of global mean and continue to do so in the future (see e.g., Tokarska et al., 2020; Beusch et al., 2020b). The global warming of these two ESM simulations is clearly incompatible with the globally-constrained emulated ensemble (Fig. 6). The two ESMs’ global and global land mean temperature anomalies are almost constantly above the 95th percentile of the emulated ensemble. Also at the regional and the grid-point level, the constrained ensemble warms distinctly less than these high warming simulations, but the larger internal variability result in a partial overlap between those [..[141]]ESMs’ realizations and the climate change uncertainty space (5th – 95th percentile) of the emulated ensemble.

To qualitatively validate the probabilistic MAGICC-MESMER ensemble and to highlight further differences to the raw CMIP6 ensemble, we additionally turn back to the time period covered by observations (Fig. 7). For the spatial scales and locations shown here, the MAGICC-MESMER ensemble captures the key characteristics of the observations both in terms of forced warming and [..[144]]spatial-scale-specific variability around the forced warming. In line with observations and [..[145]]in

[141]removed: ESM’ s
[144]removed: scale-specific
[145]removed: opposed

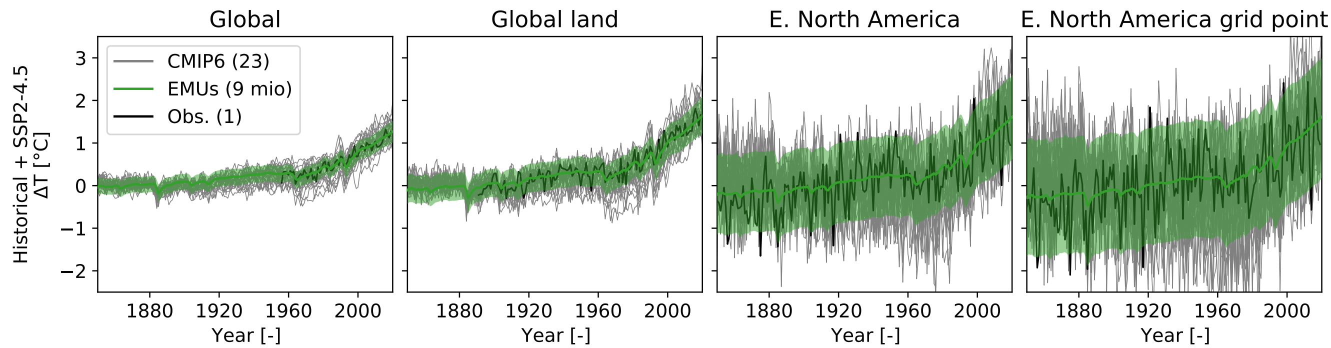

Figure 7. Temperature anomaly distribution for the emulated globally-constrained probabilistic MAGICC-MESMER ensemble, actual CMIP6 simulations, and observations averaged across different spatial scales: global, global land, [..[142]]eastern North America, and a single grid point within [..[143]]eastern North America for the period covered by observations, i.e., the historical time period extended with the middle of the road future scenario (SSP2-4.5). For the MAGICC-MESMER ensemble, the median and the 90 % range (5th – 95th percentile) of the temperature anomaly distribution is shown in color. For the ESMs, a single simulation per available ESM for that scenario is depicted in gray. The observations are shown in black.

contrast to CMIP6, the emulated ensemble does not contain any extreme outlier realizations which exhibit a drastic cooling after the 1950s. Hence, the emulated ensemble filters out physically implausible ESM simulations which affect the overall distribution of the CMIP6 ensemble.

5 Discusssion: Potential further extensions

5.1 Going beyond global mean temperature as a predictor for [..[146]]spatially resolved forced response

The careful evaluation in Sect. 3.2.3 showed that even when emulating multiple emission scenarios, a representation of local [..[147]]forced warming as a linear function of [..[148]]forced global warming and global natural variability is sufficient. Nevertheless, the current MESMER implementation can ingest additional predictors, as highlighted for the squared forced global temperature and [..[149]]forced global ocean heat uptake in Sect. 3.2.2. [..[150]]Future work could furthermore explore the added value of employing vertically resolved global heat uptake from MAGICC to separately represent mixed-layer and deeper-layer processes. Alternative predictors could be emission time series of short lived climate forcers, such as aerosols and greenhouse gases, which have been documented to influence regional warming too (Persad and Caldeira, 2018; Lund

[146] removed: the regional scale

[147] removed: temperature realizations

[148] removed: global forced warming and

[149] removed: smoothed

[150] removed: This ability to include multiple predictors is expected to become especially useful once MESMER will be extended to model variables other than annual-mean

et al., 2020). Also these predictors could readily be integrated into the MAGICC-MESMER emulator chain, since greenhouse gas and aerosol concentrations are direct outputs of the MAGICC emulator. However, such additional predictors are expected to be most useful when emulating variables whose regional forced response signal can be less successfully approximated as a function of solely global mean temperature than it is the case for annual mean temperature. For example, [..[151]]annual mean forced precipitation changes are known to clearly depend on the greenhouse gas and aerosol compositions of the emission pathway, both at global [..[152]](Ramanathan et al., 2001; Frieler et al., 2012; Pendergrass et al., 2015; Richardson et al., 2018) and at regional [..[153]](Frieler et al., 2012; Samset et al., 2016) scales.

However, great care should be taken when deciding to introduce additional predictors into MESMER. The additional predictors [..[154]]need to be sufficiently decoupled from the original predictors in the range of emission scenarios used for training to avoid artifacts in the calibrated parameters stemming from cross-correlations in the training data and leading to poor emulation skill in different scenarios. For example, if ocean heat uptake is included as a predictor, it is vital that the training data contains both a high-end and low-end emission scenario. If MESMER [..[155]]is trained only on a high emission scenario, the calibrated parameters [..[156]]are not well defined because global temperature and ocean heat uptake are strongly correlated throughout the full scenario. Hence, [..[157]]the emulation skill is poor if this MESMER calibration [..[158]]is used in a low emission scenario, in which the two predictors are no longer correlated (see Fig. S7). The more predictors are included, the higher the risk that the current approach, in which the linear regression parameters are fit for each grid cell independently, could result in noisy calibrated parameter fields. This would in turn negatively impact the residual local variability module, since the residual local variability fields employed for training would exhibit less spatial coherence. Hence, if MESMER were to move towards a larger set of predictors, additional strategies would need to be explored to encourage physically meaningful spatial coherence in the regression coefficients.

Additionally, when using the full MAGICC-MESMER emulator chain, instead of MESMER alone, it should be verified that MAGICC successfully emulates each of the additional predictors and that [..[159]]MAGICC's internal cross-correlations between the predictors are similar to the ones found within the ESMs used to train MESMER. Naturally, deviations from these inter-predictor relationships are acceptable for physically-justified reasons, such as too strong feedbacks within individual ESMs. However, this would raise the question of whether these ESMs' regional forced climate response uncertainty should be excluded from the climate change uncertainty space sampled by the MAGICC-MESMER ensemble.

Lastly, predictors do not need to be limited to the forced response module. They could additionally be introduced in the local variability module to account for non-stationarities in internal climate variability. Since the overall emulation

[151]removed: annual-mean

[152]removed: (Ramanathan et al., 2001; Frieler et al., 2012; Pendergrass et al., 2015)

[153]removed: (Frieler et al., 2012) scales. This documented emission pathway dependence could readily be integrated into the MAGICC-MESMER emulator chain since greenhouse gas and aerosol concentrations are direct outputs of the MAGICC emulator.

[154]removed: should

[155]removed: were

[156]removed: would not be

[157]removed: poor emulation skill would be expected

[158]removed: were

[159]removed: its

performance for annual mean temperature anomalies is satisfactory in most regions and for most ESMs with MESMER's stationary emulated internal variability approach (Sect. 3.2.3), no other approaches have been explored in this study. However, the monthly version of MESMER, MESMER-M, has already been extended to capture non-stationarities in month-to-month temperature variability (Nath et al., 2021). MESMER-M's local variability module starts with a month-specific, yearly-temperature-dependent power transformation of the non-stationary and skewed monthly temperature variability to obtain a Gaussian distribution. Subsequently, the same sampling approach as in MESMER is employed to create additional realizations before backtransforming the variability emulations into the original distribution. Such an approach could also be of interest for annual mean quantities of variables like precipitation which exhibit more pronounced skewness and non-stationarities across different climate states than annual mean temperature (Pendergrass et al., 2017).

5.2 Constraints on regional scale

In this study, we solely constrain the quantity best documented to be in need of a constraint [..[160]]in CMIP6: the forced global temperature response to changes in greenhouse gas concentrations and anthropogenic aerosol precursor emissions (see Sect. 3.3.2). In Sect. 4.2, we show that the resulting emulated MAGICC-MESMER ensemble exhibits a smaller spread than the raw CMIP6 ensemble, especially reducing the high-end global warming estimates in line with published literature (e.g., Brunner et al., 2020b; Tokarska et al., 2020), and that it can successfully approximate observed warming at various spatial scales.

However, to further improve the regional accuracy of the emulated ensemble's projections, a regional constraint could be applied in addition to the global constraint. Beusch et al. (2020b) [..[161]]have proposed a first regional-scale observational constraint by identifying ESMs which have a regional response to forced global warming which can be regarded as consistent with observations and by only including those ESMs' climate response and natural variability uncertainty when deriving regionally-optimized projections. To account for inter-ESM dependencies (Abramowitz et al., 2019; Knutti, 2010), [..[162]]Beusch et al. (2020b) have chosen the simplest possible way and only [..[163]]considered one ESM per "ESM name family", e.g., only one of the CNRM ESMs. Their constraint could be further refined to account for inter-ESM dependencies and consistency with observations in a more elaborate way, potentially moving towards a Bayesian constraining framework and introducing Monte-Carlo sampling of parameters like in MAGICC (Meinshausen et al., 2009). Additional thoughts should be put into explicitly constraining natural variability, which differs [..[164]]substantially between different ESMs (Beusch et al., 2020a; Deser et al., 2020) but has thus far received considerably less attention than the forced response.

[160]removed: : the global
[161]removed: propose
[162]removed: they choose
[163]removed: consider
[164]removed: starkly

5.3 Exploring regional climate change uncertainty beyond the MAGICC-MESMER coupling

With the MAGICC-MESMER coupling we are able to thoroughly sample [..[165]]climate response and natural variability uncertainty for arbitrary emission scenarios from global to spatially resolved local [..[166]]scales. However, the MAGICC-MESMER coupling is still confined in its representation of local-scale climate change uncertainty by MAGICC's representation of the forced global response to greenhouse gas [..[167]]emission uncertainty and by MESMER's representation of the forced local response to global climate information [..[168]]and its local variability uncertainty. A straightforward way to address this issue, would be to additionally combine different global and [..[169]]spatially resolved ESM emulators with each other to ultimately create [..[170]]multi-emulator-based probabilistic emulations.

On the global emulator side, first progress towards this direction has already been achieved with the Open Simple Climate Models (OpenSCM) initiative, which aims to bring different global emulators together and to provide standardized output for them. A uniform interface for emissions-driven runs, [..[171]]OpenSCM Runner (Nicholls et al., 2021b), is available [..[172]]and the implementations for MAGICC (Meinshausen et al., 2011, 2020), FaIR (Smith et al., 2018; Leach et al., 2021), and CICERO-SCM (Skeie et al., 2017, 2021) are [..[173]]fully functional. Additionally, more global emulators have expressed their interest in joining [..[174]]this initiative. In this study, MAGICC output was accessed through OpenSCM tools and thus, from a technical viewpoint, it would be straightforward to assess the implications of [..[175]]differences in forced global warming [..[176]]distributions from multiple global emulators for regional-scale temperature change realizations.

For [..[177]]spatially resolved emulators, no such common framework exists to date. However, in addition to MESMER, the fldgen emulator (Link et al., 2019; Snyder et al., 2019) is publicly available and could be coupled to the same global emulators. With this, the [..[178]]local-scale uncertainty introduced by different regional [..[179]]emulators' representation of regional climate change could be quantified.

Depending on the scientific question, different emulation strategies are called for. If the aim is to sample as broad an uncertainty space as possible, several global and [..[180]]spatially resolved emulator combinations should be included when deriving probabilistic emulated ensembles. On other occasions, it may be more beneficial to try and identify the best performing em-

[165] removed: the
[166] removed: scale
[167] removed: emissions
[168] removed: uncertainty
[169] removed: regional
[170] removed: multi-emulator based
[171] removed: OpenSCM-Runner
[172] removed: . The implementation for MAGICC is fully functioning and employed throughout this study. FaIR (Smith et al., 2018; Leach et al., 2021)
[173] removed: two other global emulators which are actively working towards being integrated into the OpenSCM initiative and
[174] removed: . Once additional global emulators are available within the OpenSCM framework, assessing
[175] removed: their
[176] removed: distribution differences for regional scale realizationswill be straightforward
[177] removed: regional-scale
[178] removed: local scale
[179] removed: emulator' s representation of local-scale
[180] removed: regional

ulator pair for the question at hand. After all, each [..[181]]emulator comes with its own sets of assumptions and thus most appropriate domain of applicability. For example, the MAGICC-MESMER emulator chain presented in this paper is a suitable tool to quantify grid-point-level to regional temperature changes for any emission pathway but – due to the design of MESMER's currently implemented internal climate variability module (Beusch et al., 2020a) – cannot be used to reliably assess the risks associated with concurrent warm anomalies across the globe.

6 Conclusions

In this paper, we present the MAGICC-MESMER emulator chain. To the best of our knowledge, this is the first emulator chain that can be directly used to rapidly assess the implications of different [..[182]]emission scenarios for annual temperature changes at global to local spatial scales while accounting for both forced climate response and natural variability uncertainty. With illustrative examples, it is demonstrated that the MAGICC-MESMER chain is able to successfully provide ESM-specific as well as globally-constrained probabilistic emulations, for a range of emission pathways. While MAGICC is run here in greenhouse gas concentration-driven mode for different SSP scenarios, following the CMIP6 ScenarioMIP approach, MAGICC could alternatively directly translate [..[183]]any emission pathway into the predictors needed by MESMER. The default configuration of MESMER, which is employed for most of our analyses, only requires forced global [..[184]]temperature change as a predictor from MAGICC. However, [..[185]]MESMER can also ingest additional predictors. This feature could become especially useful in future work aiming to extend the MAGICC-MESMER chain to emulate additional variables such as precipitation.

To increase the accessibility of our [..[186]]emulator chain and thus the use of such targeted climate information, MESMER (https://github.com/MESMER-group/mesmer) and the MAGICC-MESMER coupler (https://github.com/MESMER-group/mesmer-openscmrunner) are already publicly available under the GNU General Public License version 3 (GPL-3.0), while MAGICC is in the process of becoming open source. In the meantime, MAGICC output for various emission scenarios has been published as part of RCMIP Phase 1 (Nicholls et al., 2020) and 2 (Nicholls et al., 2021c) and could be translated to the local scale with the currently publicly available parts of the MAGICC-MESMER chain.

Code and data availability. MESMER (https://github.com/MESMER-group/mesmer) and the MAGICC-MESMER coupler MESMER-OpenSCM Runner (https://github.com/MESMER-group/mesmer-openscmrunner) are publicly available on GitHub and the versions employed in this study are archived on Zenodo (Hauser et al., 2021a; Nicholls et al., 2021a). The documentation for both repositories is hosted on Read the Docs (https://mesmer-emulator.readthedocs.io/en/stable/; https://mesmer-openscm-runner.readthedocs.io/en/main/). The scripts

[181]removed: emulators
[182]removed: emissions scenarios for
[183]removed: emissions
[184]removed: temperatures
[185]removed: the local forced warming module of MESMER is also able to
[186]removed: emulation

Table A1. List of the 25 employed CMIP6 models and the modeling groups providing them.

Model	Modeling Center (or Group)
ACCESS-CM2	Commonwealth Scientific and Industrial Research Organisation, Australia; Australian Research Council Centre of Excellence for Climate System Science, Australia
ACCESS-ESM1-5	Commonwealth Scientific and Industrial Research Organisation, Australia
AWI-CM-1-1-MR	Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Germany
CanESM5	Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change, Canada
CESM2	National Center for Atmospheric Research, Climate and Global Dynamics Laboratory, USA
CESM2-WACCM	National Center for Atmospheric Research, Climate and Global Dynamics Laboratory, USA
CMCC-CM2-SR5	Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici, Italy
CNRM-CM6-1	Centre National de Recherches Météorologiques, France / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique, France
CNRM-CM6-1-HR	Centre National de Recherches Météorologiques, France / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique, France
CNRM-ESM2-1	Centre National de Recherches Météorologiques, France / Centre Européen de Recherche et Formation Avancée en Calcul Scientifique, France
E3SM-1-1	E3SM-Project and RUBISCO National Laboratories, USA
FGOALS-f3-L	Chinese Academy of Sciences, China
FGOALS-g3	Chinese Academy of Sciences, China
FIO-ESM-2-0	First Institute of Oceanography, Ministry of Natural Resources, China; Qingdao National Laboratory for Marine Science and Technology, China
HadGEM3-GC31-LL	Met Office Hadley Centre, UK; Natural Environment Research Council, UK
HadGEM3-GC31-MM	Met Office Hadley Centre, UK
IPSL-CM6A-LR	Institut Pierre Simon Laplace, France
MCM-UA-1-0	Department of Geosciences, University of Arizona, Tucson, USA
MPI-ESM1-2-HR	Max Planck Institute for Meteorology, Germany; Deutscher Wetterdienst, Germany; Deutsches Klimarechenzentrum, Germany
MPI-ESM1-2-LR	Max Planck Institute for Meteorology, Germany; Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Germany
MRI-ESM2-0	Meteorological Research Institute, Japan
NESM3	Nanjing University of Information Science and Technology, China
NorESM2-LM	NorESM Climate modeling Consortium, Norway
NorESM2-MM	NorESM Climate modeling Consortium, Norway
UKESM1-0-LL	Met Office Hadley Centre, UK; Natural Environment Research Council, UK; National Institute of Meteorological Sciences/Korea Meteorological Administration, Climate Research Division, Republic of Korea; National Institute of Water and Atmospheric Research, New Zealand

to create the figures in this paper can be found on GitHub (https://github.com/MESMER-group/Beusch_et_al_GMD_2021_MAGICC-MESMER_coupling) and are additionally archived on Zenodo (Beusch, 2021). The MAGICC data used in this study is available via Nicholls et al. (2020) and Nicholls et al. (2021c). The stratospheric aerosol optical depth data employed during MESMER training can be obtained from Smith et al. (2021a). The CMIP6 data are available from the public CMIP archive at https://esgf-node.llnl.gov/projects/esgf-llnl/.

Appendix A

Author contributions. L.B. initiated the study, wrote the initial implementation of MESMER, carried out all analyses, and wrote the first draft of the paper. Z.N. provided all the MAGICC data, supported the use of the data, wrote the initial implementation of the MAGICC-MESMER coupler, and wrote the first draft of the MAGICC description section. L.B., M.H., and Z.N. currently co-develop the MESMER and MAGICC-MESMER code bases. All authors designed the study together, discussed the results together, and contributed to improving the manuscript.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We acknowledge the World Climate Research Program's Working Group on Coupled Modelling, which is responsible for the Coupled Model Intercomparison Project (CMIP), and we thank the climate modelling groups (listed in Appendix Table A1) for producing and making available their model output. Furthermore, we are indebted to Urs Beyerle, Lukas Brunner, and Ruth Lorenz for pre-processing the CMIP6 data. We additionally thank Shruti Nath for her comments on an earlier version of this manuscript. L.B. acknowledges support from SNF grant P1EZP2_195662. S.I.S. acknowledges partial support from the ERC Proof-of-Concept Grant 964013 "MESMER-X". Lastly, we would like to thank Christopher Smith and Ben Sanderson, whose constructive reviews helped us to further improve this study.

References

Abramowitz, G., Herger, N., Gutmann, E., Hammerling, D., Knutti, R., Leduc, M., Lorenz, R., Pincus, R., and Schmidt, G. A.: ESD Reviews: Model dependence in multi-model climate ensembles: Weighting, sub-selection and out-of-sample testing, Earth System Dynamics, 10, 91–105, https://doi.org/10.5194/esd-10-91-2019, 2019.

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, Climatic Change, 146, 319–333, https://doi.org/10.1007/s10584-016-1809-8, 2018.

Beusch, L.: Code for the study: "From emission scenarios to spatially resolved projections with a chain of computationally efficient emulators: MAGICC (v7.5.1) – MESMER (v0.8.3) coupling", https://doi.org/10.5281/zenodo.5109674, 2021.

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, 2020a.

Beusch, L., Gudmundsson, L., and Seneviratne, S. I.: Crossbreeding CMIP6 Earth System Models with an emulator for regionally-optimized land temperature projections, Geophysical Research Letters, 47, e2019GL086 812, https://doi.org/10.1029/2019GL086812, 2020b.

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

Brunner, L., Pendergrass, A., Lehner, F., Merrifield, A., Lorenz, R., and Knutti, R.: Reduced global warming from CMIP6 projections when weighting models by performance and independence, Earth System Dynamics, 11, 995–1012, https://doi.org/10.5194/esd-11-995-2020, 2020b.

CAT: Governments still showing little sign of acting on climate crisis, p. (last accessed 03 January 2022), https://climateactiontracker.org/publications/governments-still-not-acting-on-climate-crisis/, 2019.

CAT: Climate summit momentum: Paris commitments improved warming estimate to 2.4°C, p. (last accessed 03 January 2022), https://climateactiontracker.org/publications/global-update-climate-summit-momentum/, 2021a.

CAT: Glasgow's one degree 2030 credibility gap: net zero's lip service to climate action, p. (last accessed 03 January 2022), https://climateactiontracker.org/press/Glasgows-one-degree-2030-credibility-gap-net-zeros-lip-service-to-climate-action/, 2021b.

Clarke, L., Jiang, K., Akimoto, K., Babiker, M., Blanford, G., Fisher-Vanden, K., Hourcade, J.-C., Krey, V., Kriegler, E., Löschel, A., McCollum, D., Paltsev, S., Rose, S., Shukla, P. R., Tavoni, M., van der Zwaan, B. C. C., and van Vuuren, D.: Assessing Transformation Pathways, in: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., von Stechow, C., Zwickel, T., and Min, J., pp. 413–510, IPCC, https://doi.org/10.1017/cbo9781107415416.012, 2014.

Deser, C., Phillips, A., Bourdette, V., and Teng, H.: Uncertainty in climate change projections: The role of internal variability, Climate Dynamics, 38, 527–546, https://doi.org/10.1007/s00382-010-0977-x, 2012.

Deser, C., Lehner, F., Rodgers, K. B., Ault, T., Delworth, T. L., DiNezio, P. N., Fiore, A., Frankignoul, C., Fyfe, J. C., Horton, D. E., Kay, J. E., Knutti, R., Lovenduski, N. S., Marotzke, J., McKinnon, K. A., Minobe, S., Randerson, J., Screen, J. A., Simpson, I. R., and Ting, M.: Insights from Earth System Model initial-condition large ensembles and future prospects, Nature Climate Change, 10, 277–286, https://doi.org/10.1038/s41558-020-0731-2, 2020.

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

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J. L., Frame, D., Lunt, D. J., Mauritsen, T., Palmer, M. D., Watanabe, M., Wild, M., and Zhang, H.: The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity, 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, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., chap. 7, Cambridge University Press, 2021.

Forster, P. M., Maycock, A. C., McKenna, C. M., and Smith, C. J.: Latest climate models confirm need for urgent mitigation, Nature Climate Change Comment, https://doi.org/10.1038/s41558-019-0660-0, 2019.

Frieler, K., Meinshausen, M., Mengel, M., Braun, N., and Hare, W.: A scaling approach to probabilistic assessment of regional climate change, Journal of Climate, 25, 3117–3144, https://doi.org/10.1175/JCLI-D-11-00199.1, 2012.

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, Geoscientific Model Development, 13, 5389–5399, https://doi.org/10.5194/gmd-13-5389-2020, 2020.

Hauser, M., Beusch, L., Nicholls, Z., and Schwaab, J.: MESMER-group/mesmer: version 0.8.3, https://doi.org/10.5281/zenodo.5802054, 2021a.

Hauser, M., Spring, A., and Busecke, J.: Regionmask: version 0.8.0, https://doi.org/10.5281/zenodo.5532848, 2021b.

Hausfather, Z. and Peters, G. P.: Emissions - the 'business as usual' story is misleading, Nature, 577, 618–620, https://doi.org/10.1038/d41586-020-00177-3, 2020.

Hawkins, E. and Sutton, R.: The potential to narrow uncertainty in regional climate predictions, Bulletin of the American Meteorological Society, 90, 1095–1107, https://doi.org/10.1175/2009BAMS2607.1, 2009.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

IPCC: Summary for Policymakers, in: 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, edited by Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J., Chen, Y., Zhou, X., Gomis, M., Lonnoy, E., Maycock, T., Tignor, M., and Waterfield, T., 2018.

IPCC: 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, 2021.

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cimadevilla, E., Cofiño, A., Cerezo-Mota, R., Di Luca, A., Faria, S. H., Gorodetskaya, I., Hauser, M., Herrera, S., Hewitt, H., Hennessy, K., Jones, R., Krakovska, S., Manzanas, R., Marínez-Castro, D., Narisma, G. T., Nurhati, I., Pinto, I., Seneviratne, S., van den Hurk, B., and Vera, C.: An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets, Earth System Science Data, 12, 2959–2970, https://doi.org/10.5194/essd-12-2959-2020, 2020.

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

Knutti, R.: The end of model democracy?, Climatic Change, 102, 395–404, https://doi.org/10.1007/s10584-010-9800-2, 2010.

Leach, N. J., Jenkins, S., Nicholls, Z., Smith, C. J., Lynch, J., Cain, M., Walsh, T., Wu, B., Tsutsui, J., and Allen, M. R.: FaIRv2.0.0: a generalized impulse response model for climate uncertainty and future scenario exploration, Geoscientific Model Development, 14, 3007–3036, https://doi.org/10.5194/gmd-14-3007-2021, 2021.

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E., Brunner, L., Knutti, R., and Hawkins, E.: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6, Earth System Dynamics, 11, 491–508, https://doi.org/10.5194/esd-11-491-2020, 2020.

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, Geoscientific Model Development, 12, 1477–1489, https://doi.org/10.5194/gmd-12-1477-2019, 2019.

Lund, M. T., Aamaas, B., Stjern, C. W., Klimont, Z., Berntsen, T. K., and Samset, B. H.: A continued role of short-lived climate forcers under the Shared Socioeconomic Pathways, Earth System Dynamics, 11, 977–993, https://doi.org/10.5194/esd-11-977-2020, 2020.

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

Meinshausen, M., Meinshausen, N., Hare, W., Raper, S. C. B., Frieler, K., Knutti, R., Frame, D. J., and Allen, M. R.: Greenhouse-gas emission targets for limiting global warming to 2°C, Nature, 458, 1158–1162, https://doi.org/10.1038/nature08017, 2009.

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

Meinshausen, M., Nicholls, Z., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., Beyerle, U., Gessner, C., Nauels, A., Bauer, N., Canadell, J. G., Daniel, J. S., John, A., Krummel, P. B., Luderer, G., Meinshausen, N., Montzka, S. A., Rayner, P. J., Reimann, S., Smith, S. J., Van Den Berg, M., Velders, G. J., Vollmer, M. K., and Wang, R. H.: The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500, Geoscientific Model Development, 13, 3571–3605, https://doi.org/10.5194/gmd-13-3571-2020, 2020.

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

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J., Osborn, T. J., Jones, P. D., and Simpson, I. R.: An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set, Journal of Geophysical Research: Atmospheres, 126, e2019JD032 361, https://doi.org/10.1029/2019JD032361, 2021.

Nath, S., Lejeune, Q., Beusch, L., Gudmundsson, L., Schleussner, C.-F., and Seneviratne, S. I.: MESMER-M: an Earth System Model emulator for local monthly temperature, Earth System Dynamics Discussions, https://doi.org/10.5194/esd-2021-59, 2021.

Nicholls, Z., Meinshausen, M., Lewis, J., Gieseke, R., Dommenget, D., Dorheim, K., Fan, C.-S., Fuglestvedt, J. S., Gasser, T., Goluke, U., Goodwin, P., Hartin, C., Hope, A. P., Kriegler, E., Leach, N. J., Marchegiani, D., McBride, L. A., Quilcaille, Y., Rogelj, J., Salawitch, R. J., Samset, B. H., Sandstad, M., Shiklomanov, A. N., Skeie, R. B., Smith, C. J., Smith, S., Tanaka, K., Tsutsui, J., and Xie, Z.: Reduced Complexity Model Intercomparison Project Phase 1: Introduction and evaluation of global-mean temperature response, Geoscientific Model Development, 13, 5175–5190, https://doi.org/10.5194/gmd-13-5175-2020, 2020.

Nicholls, Z., Beusch, L., and Hauser, M.: MESMER-group/mesmer-openscmrunner: version 0.1.0, https://doi.org/10.5281/zenodo.5094380, 2021a.

Nicholls, Z., Lewis, J., Smith, C. J., Sandstad, M., Kikstra, J., Gieseke, R., and Willner, S.: OpenSCM-Runner: Thin wrapper to run simple climate models (emissions driven runs only), https://github.com/openscm/openscm-runner, 2021b.

Nicholls, Z., Meinshausen, M., Lewis, J., Rojas Corradi, M., Dorheim, K., Gasser, T., Gieseke, R., Hope, A. P., Leach, N., McBride, L. A., Quilcaille, Y., Rogelj, J., Salawitch, R. J., Samset, B. H., Sandstad, M., Shiklomanov, A. N., Skeie, R. B., Smith, C. J., Smith, S. J., Su, X., Tsutsui, J., Vega-Westhoff, B., and Woodard, D. L.: Reduced Complexity Model Intercomparison Project Phase 2: Synthesising Earth system knowledge for probabilistic climate projections, Earth's Future, 9, e2020EF001 900, https://doi.org/10.1029/2020EF001900, 2021c.

Olonscheck, D. and Notz, D.: Consistently estimating internal climate variability from climate model simulations, Journal of Climate, 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, Geoscientific Model Development, 9, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016, 2016.

O'Neill, B. C., Kriegler, E., Ebi, K. L., Kemp-Benedict, E., Riahi, K., Rothman, D. S., van Ruijven, B. J., van Vuuren, D. P., Birkmann, J., Kok, K., Levy, M., and Solecki, W.: The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century, Global Environmental Change, 42, 169–180, https://doi.org/10.1016/j.gloenvcha.2015.01.004, 2017.

Pendergrass, A. G., Lehner, F., Sanderson, B. M., and Xu, Y.: Does extreme precipitation intensity depend on the emissions scenario?, Geophysical Research Letters, 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, Scientific Reports, 7, 17 966, 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, Nature Communications, 9, 3289, https://doi.org/10.1038/s41467-018-05838-6, 2018.

Ramanathan, V., Crutzen, P. J., Kiehl, J. T., and Rosenfeld, D.: Atmosphere: Aerosols, climate, and the hydrological cycle, Science, 294, 2119–2124, https://doi.org/10.1126/science.1064034, 2001.

Ribes, A., Qasmi, S., and Gillett, N. P.: Making climate projections conditional on historical observations, Science Advances, 7, 1–10, https://doi.org/10.1126/sciadv.abc0671, 2021.

Richardson, T. B., Forster, P. M., Andrews, T., Boucher, O., Faluvegi, G., Fläschner, D., Hodnebrog, Kasoar, M., Kirkevåg, A., Lamarque, J. F., Myhre, G., Olivié, D., Samset, B. H., Shawki, D., Shindell, D., Takemura, T., and Voulgarakis, A.: Drivers of precipitation change: An energetic understanding, Journal of Climate, 31, 9641–9657, https://doi.org/10.1175/JCLI-D-17-0240.1, 2018.

Rogelj, J., Den Elzen, M., Höhne, N., Fransen, T., Fekete, H., Winkler, H., Schaeffer, R., Sha, F., Riahi, K., and Meinshausen, M.: Paris Agreement climate proposals need a boost to keep warming well below 2 °c, Nature, 534, 631–639, https://doi.org/10.1038/nature18307, 2016.

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshgi, H., Kobayashi, S., Kriegler, E., Mundaca, L., Séférian, R., and Vilariño, M.: Mitigation Pathways Compatible with 1.5°C in the Context of Sustainable Development., in: 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, edited by Masson-Delmotte, V., Zhai, P., Pörtner, H.-O., Roberts, D., Skea, J., Shukla, P. R., Pirani, A., Moufouma-Okia, W., Péan, C., Pidcock, R., Connors, S., Matthews, J., Chen, Y., Zhou, X., Gomis, M. I., Lonnoy, E., Maycock, T., Tignor, M., and Waterfield, T., chap. 2, pp. 93–174, IPCC, 2018.

Rohde, R. and Hausfather, Z.: The Berkeley Earth Land/Ocean Temperature Record, Earth System Science Data, 12, 3469–3479, https://doi.org/10.5194/essd-12-3469-2020, 2020.

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., and Mosher, S.: Berkeley Earth Temperature Averaging Process, Geoinformatics \& Geostatistics: An Overview, 1, 1–13, https://doi.org/10.4172/2327-4581.1000103, 2013.

Samset, B. H., Myhre, G., Forster, P. M., Hodnebrog, Andrews, T., Faluvegi, G., Fläschner, D., Kasoar, M., Kharin, V., Kirkevåg, A., Lamarque, J. F., Olivié, D., Richardson, T., Shindell, D., Shine, K. P., Takemura, T., and Voulgarakis, A.: Fast and slow precipitation responses to individual climate forcers: A PDRMIP multimodel study, Geophysical Research Letters, 43, 2782–2791, https://doi.org/10.1002/2016GL068064, 2016.

Schneider von Deimling, T., Meinshausen, M., Levermann, A., Huber, V., Frieler, K., Lawrence, D. M., and Brovkin, V.: Estimating the near-surface permafrost-carbon feedback on global warming, Biogeosciences, 9, 649–665, https://doi.org/10.5194/bg-9-649-2012, 2012.

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.

Serreze, M. C. and Barry, R. G.: Processes and impacts of Arctic amplification: A research synthesis, Global and Planetary Change, 77, 85–96, https://doi.org/10.1016/j.gloplacha.2011.03.004, 2011.

Skeie, R. B., Fuglestvedt, J., Berntsen, T., Peters, G. P., Andrew, R., Allen, M., and Kallbekken, S.: Perspective has a strong effect on the calculation of historical contributions to global warming, Environmental Research Letters, 12, 024 022, https://doi.org/10.1088/1748-9326/aa5b0a, 2017.

Skeie, R. B., Peters, G. P., Fuglestvedt, J., and Andrew, R.: A future perspective of historical contributions to climate change, Climatic Change, 164, 1–13, https://doi.org/10.1007/s10584-021-02982-9, 2021.

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, Geoscientific Model Development, 11, 2273–2297, https://doi.org/10.5194/gmd-11-2273-2018, 2018.

Smith, C. J., Forster, P. M., Berger, S., Collins, W., Hall, B., Lunt, D., Palmer, M. D., Watanabe, M., Cain, M., Harris, G., Leach, N. J., Ringer, M., and Zelinka, M.: Figure and data generation for Chapter 7 of the IPCC's Sixth Assessment Report, Working Group 1 (plus assorted other contributions). Version 1.0, https://doi.org/10.5281/zenodo.5211357, 2021a.

Smith, C. J., Nicholls, Z., Armour, K., Collins, W., Forster, P., Meinshausen, M., Palmer, M. D., and Watanabe, M.: The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity Supplementary Material, 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, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., chap. S7, Cambridge University Press, 2021b.

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, e0223 542, https://doi.org/10.1371/journal.pone.0223542, 2019.

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

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., and Knutti, R.: Past warming trend constrains future warming in CMIP6 models, Science Advances, 6, eaaz9549, https://doi.org/10.1126/sciadv.aaz9549, 2020.

UNFCCC: Adoption of the Paris Agreement., p. FCCC/CP/2015/10/Add.1, 2015.

Yuan, X.-C., Zhang, N., Wang, W.-Z., and Wei, Y.-M.: Large-scale emulation of spatio-temporal variation in temperature under climate change, Environmental Research Letters, 16, https://doi.org/10.1088/1748-9326/abd213, 2021.