



# ModE-Sim – a medium-sized atmospheric general circulation model (AGCM) ensemble to study climate variability during the modern era (1420 to 2009)

Ralf Hand<sup>1,2</sup>, Eric Samakinwa<sup>1,2</sup>, Laura Lipfert<sup>1,2</sup>, and Stefan Brönnimann<sup>1,2</sup>

<sup>1</sup>Institute of Geography, University of Bern, Bern, Switzerland

<sup>2</sup>Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

**Correspondence:** Ralf Hand (ralf.hand@giub.unibe.ch)

Received: 8 February 2023 – Discussion started: 21 April 2023

Revised: 27 July 2023 – Accepted: 11 August 2023 – Published: 28 August 2023

**Abstract.** We introduce ModE-Sim (Modern Era SIMulations), a medium-sized ensemble of simulations with the atmospheric general circulation model ECHAM6 in its LR (low-resolution) version (T63; approx. 1.8° horizontal grid width with 47 vertical levels). At the lower boundary we use prescribed sea surface temperatures and sea ice that reflect observed values while accounting for uncertainties in these. Furthermore we use radiative forcings that also reflect observed values while accounting for uncertainties in the timing and strength of volcanic eruptions. The simulations cover the period from 1420 to 2009. With 60 ensemble members between 1420 and 1850 and 36 ensemble members from 1850 to 2009, ModE-Sim consists of 31 620 simulated years in total. ModE-Sim is suitable for many applications as its various subsets can be used as initial-condition and boundary-condition ensembles to study climate variability. The main intention of this paper is to give a comprehensive description of the experimental setup of ModE-Sim and to provide an evaluation, mainly focusing on the two key variables, 2 m temperature and precipitation. We demonstrate ModE-Sim's ability to represent their mean state, to produce a reasonable response to external forcings, and to sample internal variability. Through the example of heat waves, we show that the ensemble is even capable of capturing certain types of extreme events.

## 1 Introduction

The use of large ensembles of climate model simulations enables the climate's response to external forcings to be separated from internal climate variability (Maher et al., 2019; Milinski et al., 2020; von Trentini et al., 2020). The individual realizations of a single-model ensemble can differ either in their boundary conditions, their initial conditions, or both.

If the ensemble size is large enough and capable of spanning the full range of physically plausible climate states, large ensembles are likely to include realizations that are close to the historically observed climate state within a reasonable range of uncertainty.

Some notable applications include but are not limited to the analysis of sea surface temperature (SST) trends (Olonschek et al., 2020), the production of initial state estimates for reanalyses datasets via data assimilation approaches (Bhend et al., 2012; Franke et al., 2017; Valler et al., 2022), and the modulation of global warming (Liguori et al., 2020) and extreme events (Landrum and Holland, 2020).

Before the widespread use of large ensembles, separating internal variability and external components of climate variability in model simulations usually was done by comparison of the statistics of a transient simulation with the statistics of a control simulation with climatological forcings. This approach does not fully allow inferences of the internal variability of such simulations, as it has limitations to capture changes of internal variability over time (Maher et al., 2015, 2019). To this end, large ensembles provide a more accurate method of separating the different components of climate variability and account for the changes over time but are

computationally demanding. With the increasing number of available resources, Single Model Initial-condition Large Ensembles (SMILEs hereafter) have become an important tool for understanding climate variability (e.g., Deser et al., 2020; Maher et al., 2021). Furthermore, large ensembles offer the opportunity to investigate the range of climate variability in a physically consistent framework designed to offer a representative sample of the range of climate states under the given forcing.

Here we present ModE-Sim (Modern Era Simulations), a medium-sized ensemble of simulations with an atmospheric model capturing the period 1420 to 2009 (later referred to as the “modern era”). Unlike some notable multi-model ensemble simulations offered by model intercomparison projects, our setup is similar to a SMILE setup using a single model, but it additionally accounts for uncertainties in the boundary conditions. We use the atmospheric general circulation model ECHAM6 with forcings that account for uncertainties in the SSTs and sea ice and for the effect of volcanic eruptions. The experimental designs of most SMILE simulations depend on its purpose (Maher et al., 2021). ModE-Sim aims to provide an initial state estimate for ModE-RA, a paleoclimate data assimilation approach, thereby producing a 3-dimensional reconstruction of the global climate for the according modern era (Valler et al., 2023). Focusing on the key variables necessary for our data assimilation procedure, namely surface air temperature and precipitation, we evaluate how well ModE-Sim represents the mean observed climate state and variability during the time for which reliable observations are available. We show that our ensemble size is sufficient for sampling the internal variability of the key variables and is also capable of capturing extreme events, such as heat waves. Providing boundary conditions for the early part of the modern era is challenging but necessary. Our ensemble simulations use HadISST2 SSTs and sea ice concentrations (Titchner and Rayner, 2014) in the later periods and SSTs and sea ice concentrations based on an ensemble multi-proxy temperature reconstruction over the modern era prior to 1850 (Neukom et al., 2019). These multi-proxy temperature reconstructions are selections based on the agreement between marine proxies and simulated SST from the past1000 of PMIP3 (Gómez-Navarro et al., 2017), resulting in physically consistent ensemble fields (Samakinwa et al., 2021).

We organize the remaining parts of this work as follows: in Sect. 2, we describe the initialization strategy and the experimental design of ModE-Sim. Section 3 gives explanations of the statistical methods used for the evaluation of the ensemble simulations, while Sect. 4 concludes the paper and provides a summary.

## 2 Model and experimental setup

### 2.1 The model: ECHAM6

We use ECHAM6, the atmospheric component of the Max Planck Institute’s Earth System Model (MPI-ESM), for all our simulations. ECHAM6 participates regularly in model intercomparison projects (MIPs), both as a submodel of MPI-ESM and as a stand-alone atmospheric model. We will only briefly describe the configuration used in this study; for further details on the model, please refer to Stevens et al. (2013). We use ECHAM version 6.3.5p2, which is the version that participated in the 6th Phase of the Coupled Model Intercomparison Project (CMIP6). This version is presumably the final development step of ECHAM. We use the low-resolution (LR) version of the model, with a horizontal resolution of T63, equivalent to a grid width of approximately  $1.8^\circ$ . In the vertical, the LR version uses 47 hybrid levels between the surface and 0.01 hPa.

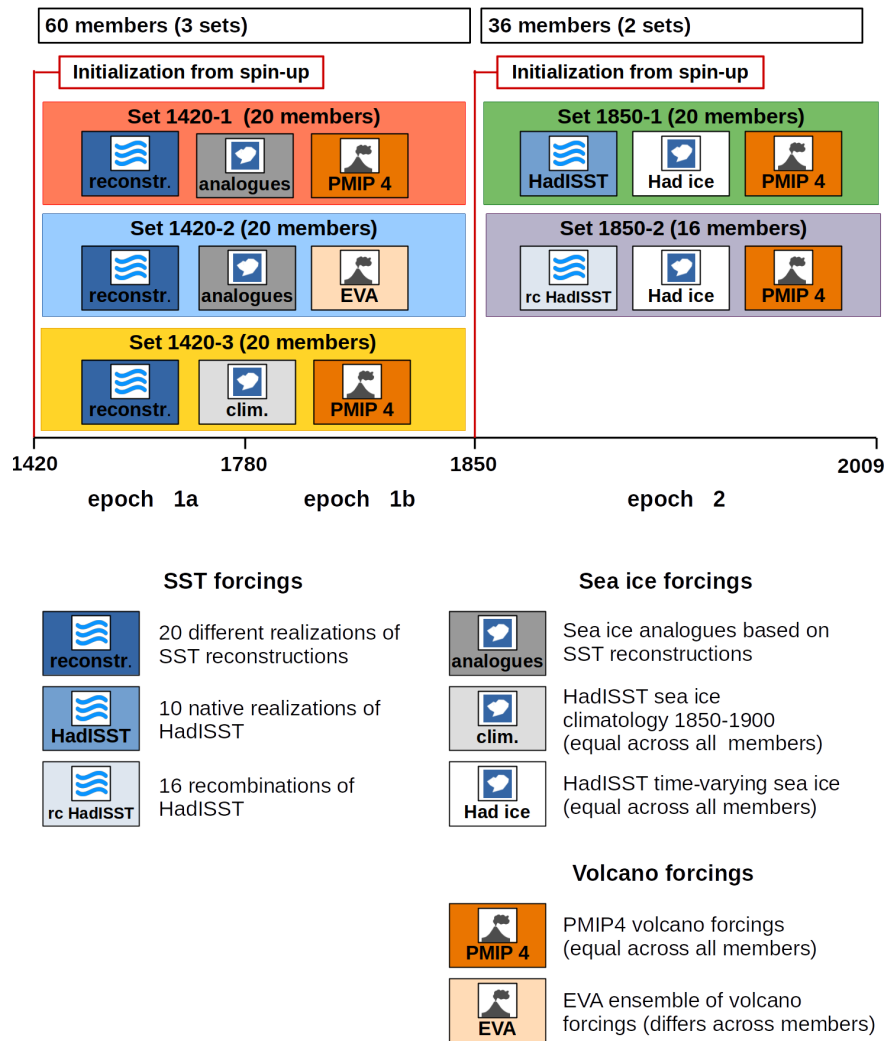
### 2.2 The experiments

Except for the differences described in the following paragraphs, our setup is designed to be close to the past2k simulations of the 4th edition of the Paleoclimate Modelling Intercomparison Project (PMIP4) performed at the Max Planck Institute for Meteorology (Jungclaus et al., 2017). Our simulations comprise three epochs, forced with slightly different ocean boundaries and radiative forcings. Figure 1 gives an overview of our experimental design.

As a main difference, in contrast to the past2k simulations that use the coupled version of the model, here we use the stand-alone atmospheric component of the model with prescribed SST and sea ice. This setup means that the ensemble spread does reflect internal variability in the ocean, but it ties the ocean to observed conditions. However, as there is high uncertainty in the SSTs, particularly in the early period, we account for these uncertainties by using ensembles of SSTs, with an individual realization of the SST resulting in the ocean forcing for each ensemble member of our simulations. In principle, earlier forcings include larger uncertainties. We account for the latter by choosing a larger ensemble size and a wider variety of forcings for the period prior to 1850. While date and strength of volcanic eruptions are relatively well constrained by observations in the time after 1850, there is more uncertainty in the prior period. We account for this fact by varying the volcanic forcings in one of our subsets for the time prior to 1850.

As another difference to the past2k simulations, we use prescribed land-cover maps rather than dynamic vegetation.

The initialization of our model simulation is in two steps: first, we forked an atmosphere-only spin-up simulation from the PMIP4 simulation of the coupled version of the model (i.e., MPI-ESM). These spin-up simulations span several decades until equilibrium is reached in atmosphere-only



**Figure 1.** Overview of the setup of the experiment sets in ModE-Sim and their forcings.

mode. They are forced with monthly varying boundary conditions and radiative forcings that repeat from year to year and that are representative of the starting year of transient simulations. As a second step of the initialization, we then forked the actual transient simulations from different time instances of this spin-up run.

We performed two spin-up simulations to generate a set of initial conditions for our transient simulations. The transient simulations include 36 ensemble members initialized in the year 1850 and 60 ensemble members initialized from 1420 conditions. We account for uncertainties of the lower boundary conditions by using different realizations of SSTs from the HadISST2 dataset for the runs starting in 1850 and an ensemble of novel SST reconstructions (Samakinwa et al., 2021) for the simulations from 1420. Furthermore, our setup accounts for uncertainties in the radiative forcing.

### 2.2.1 Spin-up runs and initialization strategy

Due to deficiencies in the representation of the large-scale ocean circulation in MPI-ESM, the ocean boundary conditions provided by the ocean component of the coupled system show distinct differences to observed SSTs in some regions (Müller et al., 2018, Fig. 2a therein). Therefore, an abrupt switch from MPI-ESM (coupled system) to the atmosphere-only setup might cause an initial shock to the atmospheric circulation.

For our transient runs, we performed two spin-up simulations initialized from the years 1420 and 1850 of the coupled PMIP4 simulations. These enable a smooth transition from the PMIP4 coupled simulations to our atmosphere-only ensemble. It also allows the use of slightly different initial conditions for the individual members. The spin-up runs use forcings and boundary conditions of the initialization year (monthly varying but no year-to-year variability) such that

each year of the spin-up run is a realization of the years 1420 and 1850, respectively. A potential caveat of this strategy is that we start our transient simulations from a world that has not experienced inter-annual variability of the forcing for several years. However, a realistic state of the atmosphere is achieved after a few years due to its short memory.

### 2.2.2 Transient simulations of epoch 1: 1420 to 1850

The simulations of epoch-1 consist of 60 simulations, divided into three subsets. The first subset (“set 1420-1”) consists of 20 simulations that were forced with 20 different ocean boundary conditions. The procedure used to generate the oceanic boundary conditions is described in Samakinwa et al. (2021). We initialize from 20 different time instances of the 1420 spin-up run. All these 20 simulations share the same radiative forcing that is identical to the standard PMIP4 setup.

The second subset (“set 1420-2”) consists of another 20 simulations that use another 20 realizations of the SST reconstructions and another 20 initializations. The difference to set 1 is in the radiative forcing: in contrast, the simulations in set 2 each have a different volcanic forcing. Consistent with the PMIP4 standard setup, these radiative forcings are outputs of the Easy Volcanic Aerosol model (EVA; Toohey et al., 2016) using the volcanic stratospheric sulfur injection (VSSI) reconstruction of Sigl et al. (2022). PMIP4 volcanic forcing also results from EVA, generated using the VSSI reconstruction of Toohey and Sigl (2017). The individual realizations of the volcanic forcing account for uncertainties in the timing and strength of the eruptions by varying the according quantities in the EVA input. Sets 1420-1 and 1420-2 use sea ice analogues selected from the HadISST2 dataset based on a pattern-matching algorithm applied to the SSTs (for details, please refer to Samakinwa et al., 2021).

The last subset (“set 1420-3”) uses the same 20 initializations as set 1420-1. While sets 1420-1 and 1420-2 use a preliminary version of the SST reconstructions, set 1420-3 utilizes the final version. Due to slight modifications in the algorithm that generates the SST reconstructions, the SST variability is reduced by up to approximately 20 % in set 1420-3 with respect to the first two sets of epoch 1 (see Fig. 3). A first analysis shows that the effect of this reduction seems to have a minor influence on the surface temperature variability over land but is limited to the affected ocean grid points. Another difference to the previous sets is that set 1420-3 uses HadISST2 historical sea ice climatology rather than the sea ice analogue approach used for sets 1420-1 and 1420-2. We found that the sea ice analogue approach shows a bias as almost all analogues were sampled from the late 20th century. This results in low sea ice concentration, particularly ice-free conditions in the marginal seas like the Labrador Sea and Sea of Okhotsk, which is unlikely to be realistic for the earlier periods. The effect of this seems to be limited to the direct surrounding where the sea ice cover changed, but sim-

ulations with climatological sea ice might be more suitable for certain analyses. Note that there are hardly any observations for the high latitudes prior to 1970, so any information on historical sea ice is subject to very high uncertainty. Our different sets can therefore be seen as the upper and lower bounds of a very conservative uncertainty range. Starting in 1780 (“epoch 1b”), the SST reconstruction assimilates large bodies of marine observations, pulling the ocean towards its true state. Because the differences between the SST and sea ice forcing between epoch 1a and epoch 1b are small, while uncertainty in the pre-1780 forcing is high, we did not perform an additional spin-up for the epoch 1b simulations but directly continued these runs from the epoch 1a simulations with only swapping to the new forcing.

### 2.2.3 Epoch 2: 1850 to 2009

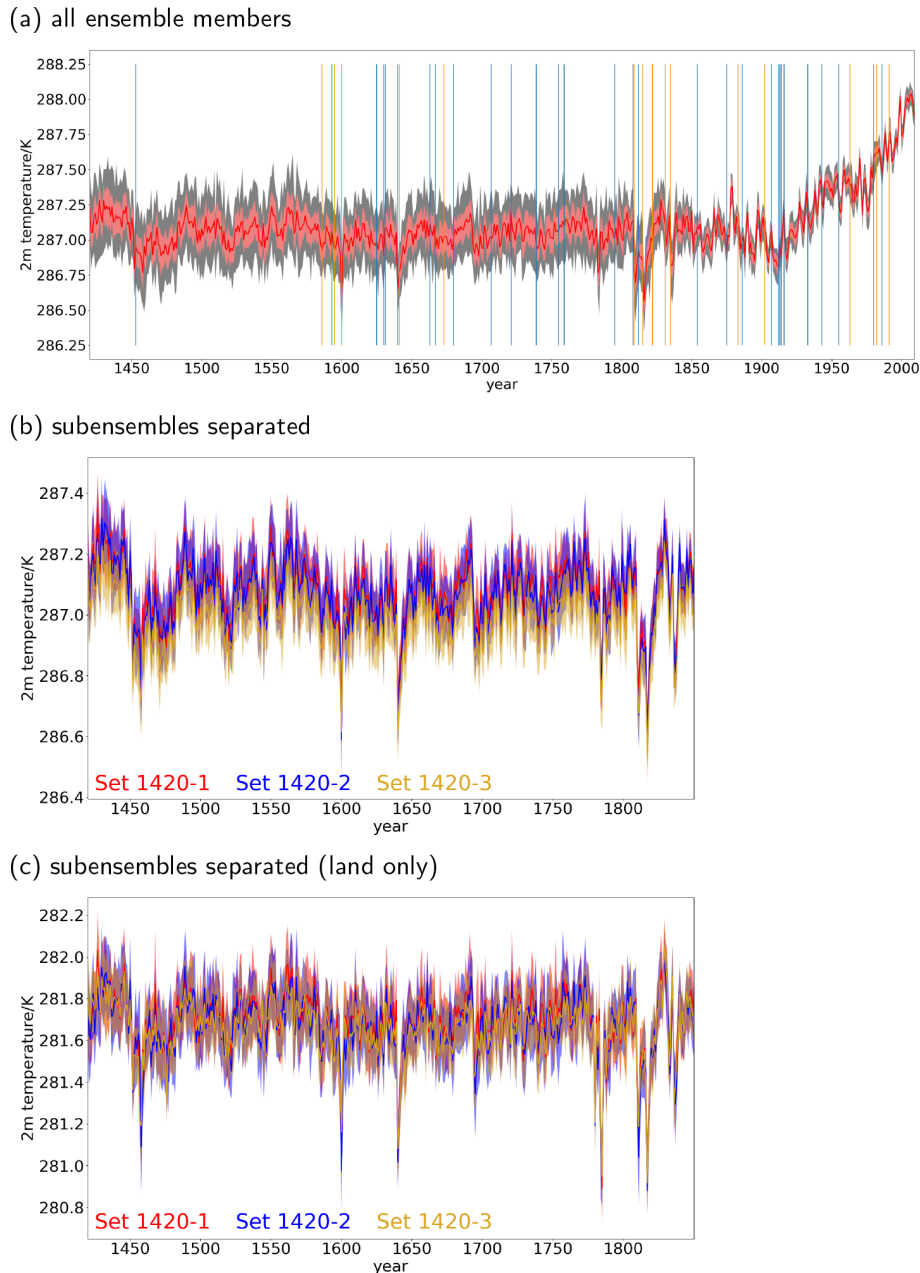
From 1850 onwards (“epoch 2”), uncertainties in the forcings become smaller. Most major volcanic eruptions are well documented, and ocean boundary conditions are more and more constrained by ship measurements and – from the end of the 20th century onwards – also satellite measurements. For epoch 2, our simulations consist of two subsets that differ only slightly in their ocean boundary forcings and their initial conditions. “Set 1850-1” was started from 20 different time instances of the spin-up simulation and uses 10 different realizations of HadISST2 as ocean boundary conditions (i.e., each HadISST2 realization is used with two different initializations). To expand the sample of SST forcings, the second set (“set 1850-2”) uses 16 different recombinations of the 10 available HadISST2 realizations as ocean forcing. These recombinations were formed as follows:

$$\text{SST}(i) = (\text{SST}(j)/\text{SST}(k)) \cdot \text{SST}(l), \text{ with } j \neq k \neq l,$$

where  $\text{SST}(i, x, t)$  is the SST anomaly in the  $i$ th recombination, and  $\text{SST}(j)$ ,  $\text{SST}(k)$ , and  $\text{SST}(l)$  are the according SST anomalies in the  $j$ th,  $k$ th, and  $l$ th realization of HadISST2. The sea ice concentrations were created by choosing analogues from the HadISST2 sea ice dataset. The radiative forcings, including volcanic forcings, are the PMIP4 input and are identical for all simulations and all sets of epoch 2.

## 3 Evaluation

To make an ensemble useful for studying climate variability, it is required to have information on how well its setup represents the mean state of the variables of interest and to what extent the ensemble can sample the forced and internal variability of the atmosphere. The following section will provide some analysis in response to these questions, focusing on the variables surface temperature and precipitation, as these are the variables of highest interest for the data assimilation in our project. Concerning variability, we consider different timescales, including annual, seasonal, and



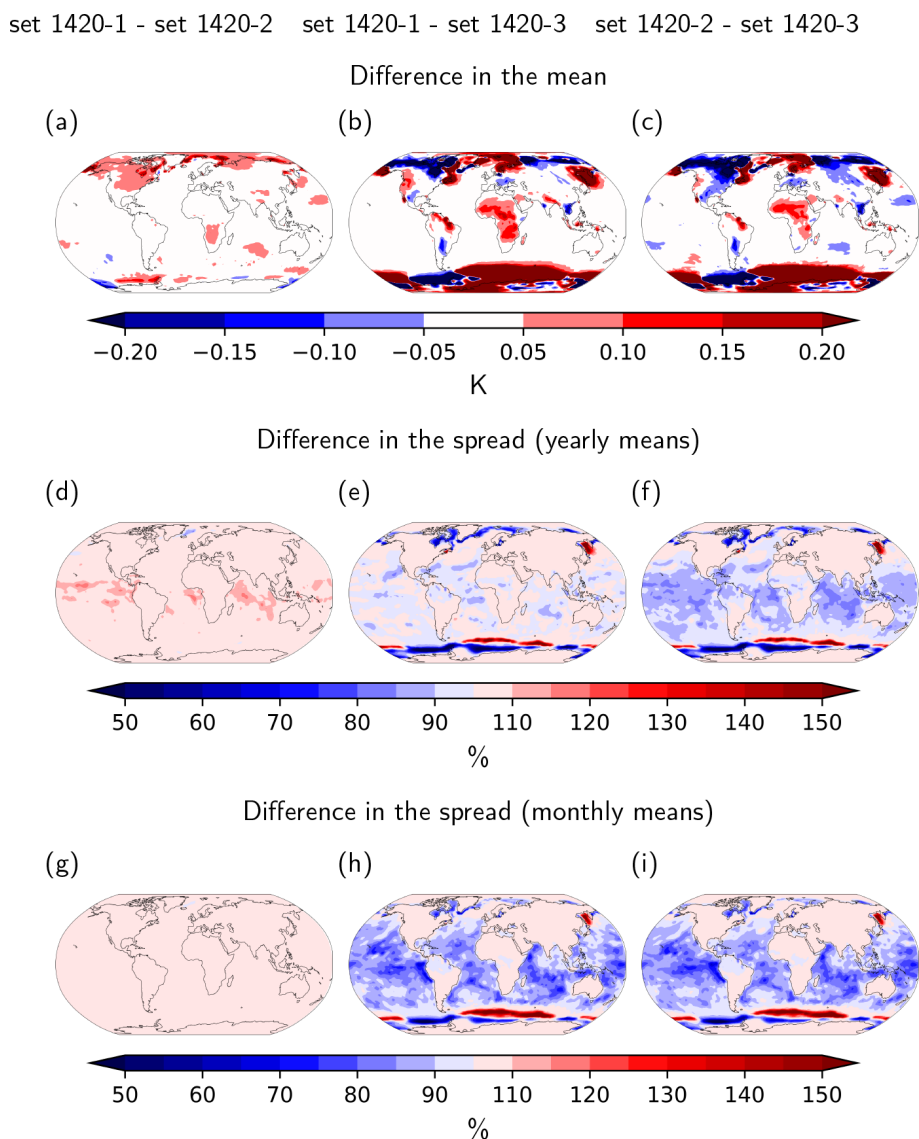
**Figure 2.** Time series of globally averaged 2 m temperature. **(a)** Ensemble mean (red line),  $\pm 1$  ensemble standard deviation (light-red shading), and ensemble minimum–maximum (grey shading) of all sets in epoch 1. Vertical orange lines indicate volcanoes used for the composites in Sect. 3.2, and vertical blue lines indicate additional volcanic eruptions with a volcanic explosivity index  $\geq 5$ . **(b)** Sets 1420-1 (red), 1420-2 (blue), and 1420-3 (dark yellow) separated. The colored lines indicate the ensemble mean of each set and the shadings the ensemble minimum–maximum range. **(c)** Same as **(b)** but limited to grid points over land.

monthly variability. As a reference dataset to compare with, we mainly use the Berkeley Earth dataset (Rohde and Hausfather, 2020), as it goes back to 1750 and therefore allows evaluation of our pre-industrial simulations. We evaluate two periods, 1780 to 1850 and 1950 to 2000. For precipitation, we use the Global Precipitation Climatology Centre (GPCC) dataset (Becker et al., 2013). We limit the evaluation to the

period from 1950 to 2000 because there are no reliable global precipitation datasets available for the earlier period.

### 3.1 Evaluation of mean state biases of precipitation and 2 m temperature

A detailed analysis of mean state biases of ECHAM6 is beyond the scope of this paper. For a more detailed mean state



**Figure 3.** Difference in the temporally averaged ensemble mean of 2 m temperature (in K; **a**, **b**, **c**) and the ratio between the temporally averaged ensemble standard deviations on annual (in %; **d**, **e**, **f**) and monthly (in %; **g**, **h**, **i**) timescales between the different sets in epoch 1a.

analysis, please refer to Stevens et al. (2013) and Giorgetta et al. (2013). ModE-Sim reproduces the known anomalies of ECHAM6 when used in AMIP mode (Fig. S1 in the Supplement). The temperature bias simulated for epoch 1b has the same spatial extent as for the period between 1950 and 2000. The main features are a warm bias in the Northern Hemisphere midlatitudes and over Australia and cold biases over South America, India, and the northern Rocky Mountains. We also found a wet bias over the Himalayas and the Andes. Our results are in agreement with the existing studies (Giorgetta et al., 2013).

### 3.2 Response to the external forcings

ModE-Sim reproduces the observed global mean near-surface temperature (Fig. 2). The time series shows a clear cooling imprint following volcanic eruptions and a warming trend in the 20th century. The forced signal related to the radiative forcing and the ocean boundary conditions can be detected from the subensemble means computed from each 20-member set separately (Fig. 2b and c), indicating that the ensemble size is clearly sufficient to separate forced signals from internal variability of the atmosphere and uncertainties related to forcing and boundary conditions. The ensemble spread declines towards the end of the simulations, indicating that for the earlier period, uncertainty in the forcings and

boundary conditions contributes to the ensemble spread to the same order of magnitude as internal variability does.

To determine the spacial manifestation of the volcanic signal, we computed composites of the ensemble mean anomalies of 2 m temperature, precipitation, and sea-level pressure for the first (Fig. 4) and second (Fig. S3) winter and summer after 15 major eruptions (Fischer et al., 2007, Table 1 therein), with the five summers/winters prior to each of the eruptions as the reference period. To test the response for statistical significance, we have additionally created 1000 surrogates, with each surrogate being created by picking 15 random years (with replacement) and computing the difference between these 15 random years and the according 5-year periods prior to them. We defined those parts of the response as statistically significant that fall outside the 5th–95th percentile range of these 1000 surrogates.

The spatial anomalies agree well with observations and previous modeling studies (e.g., Graft et al., 1993; Robock, 2000; Fischer et al., 2007; Sjolte et al., 2021). The most prominent feature in terms of 2 m temperature is a direct response to the negative radiative forcing that resembles a strong cooling over most continental regions that regionally exceeds 1 K and is statistically significant for most regions. The most prominent exception from this cooling response can be found over northern Eurasia in boreal winter but which is only statistically significant in its center. The underlying mechanisms of this winter warming is not fully understood yet and differs between different climate models (Driscoll et al., 2012) and forcing datasets (Zambri et al., 2017). Based on analysis of simulations with MPI-ESM, the coupled version of ECHAM6, Bittner et al. (2016) discuss that the warming may be caused by changes in midlatitude lower stratosphere zonal winds that lead to an equatorward deflection of planetary waves. This reduces higher-latitude wave breaking and hence disturbances of the polar vortex. In ModE-Sim we find an anomalously negative sea-level pressure (SLP) response in the northern polar latitudes in connection with an anomalous band of positive SLP that spans over the North Atlantic and northern Eurasia. This pattern, also known as the positive phase of the Arctic Oscillation (AO; or North Atlantic Oscillation (NAO) when restricted only to the Atlantic region), is a known mode of large-scale variability in the Northern Hemisphere, and its excitation in the first winter after volcanic eruptions is supported by observations (Christiansen, 2008; Fischer et al., 2007). A positive AO/NAO in winter typically leads to enhanced advection of marine air to the continent, resulting in the winter warming. While the warming is significant at least in its center, the SLP response itself is not, likely due to the high internal variability in SLP. The most prominent feature of the precipitation response is a shift of the intertropical convergence zone towards the summer hemisphere that is strongest over the central and western Pacific.

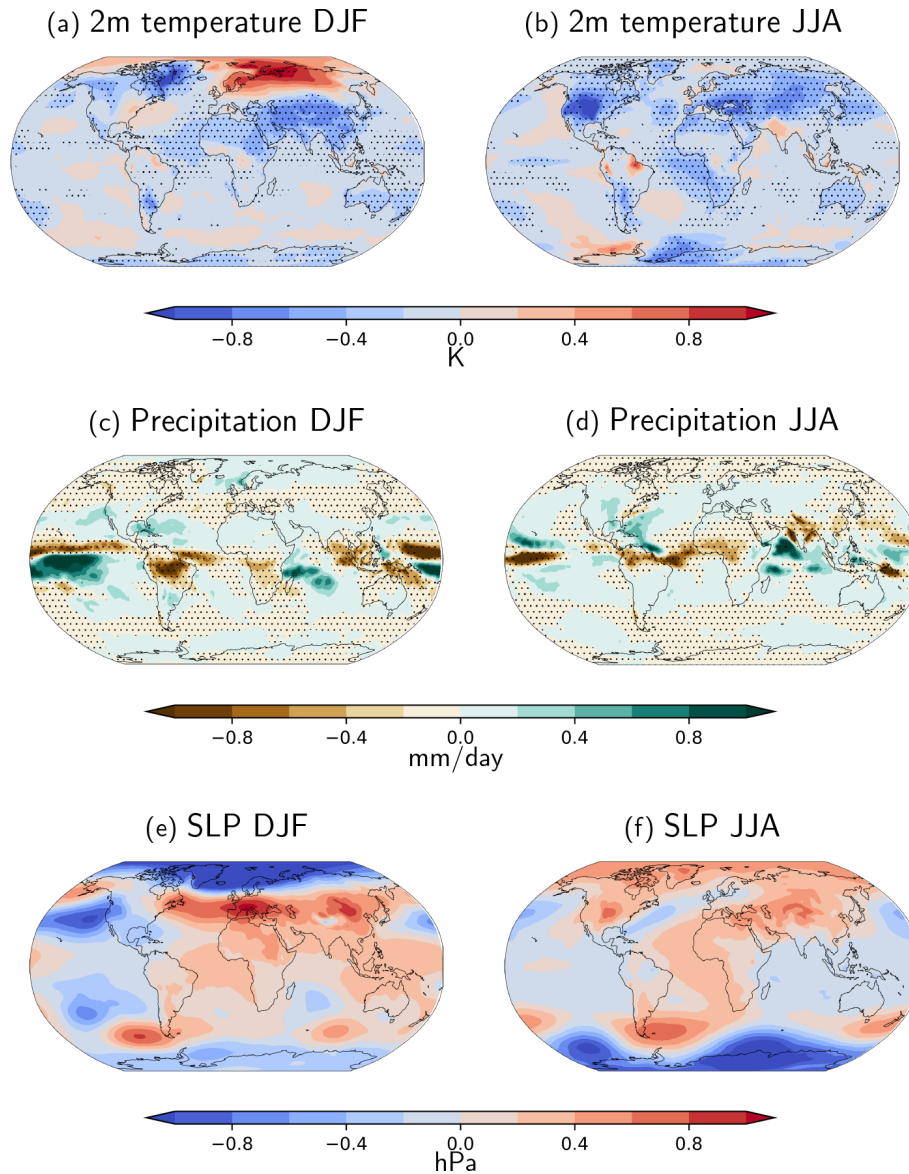
For an example of the atmospheric response to the ocean boundary conditions, we analyzed the response to the El

Niño–Southern Oscillation (ENSO) in boreal winter (DJF). We computed an ENSO index that basically follows the method provided in Trenberth (1997): we took the monthly deviations of SST from the climatology of epoch 1a, averaged these anomalies for the Niño 3.4 region (5° N to 5° S, 170 to 120° W), and applied a 5-month running mean. Afterwards, we computed regressions of atmospheric surface quantities in boreal winter on this index for all sets of epoch 1a. Due to the large ensemble size, the correlation is found to be highly significant in almost all regions of the world. ModE-Sim is able to reproduce the main features of the known ENSO teleconnection (Fig. 5): the SLP response shows good agreement with other models and reanalysis data (e.g., Döscher et al., 2022, Fig. 18 therein), with a strengthening of the Aleutian low, a weakening of the NAO, negative SLP anomalies in the eastern tropical Pacific, a band in the subpolar Southern Ocean, moderate positive SLP anomalies over the Arctic and the Antarctic, and a strongly positive SLP anomaly west over the Amundsen Sea. The observed 2 m temperature response (Brönnimann, 2007) is well captured around the Pacific and Europe, with a warming over northern Australia, a cooling over the US, a warming over Canada and Alaska connected to the weakened Aleutian Low, and a cooling over Europe induced by the weakened NAO. Concerning the precipitation response, the wettening of the US and central Brazil is captured, as well as the wettening of tropical South America and that over Australia and Malaysia.

### 3.3 Differences between the individual sets in epoch 1

For epoch 1, we computed three sets that differ in the ocean boundary conditions and the radiative forcing (see Sect. 2.2.2). A resulting question is whether these three sets show substantial differences in mean state and ensemble spread. Figure 2b shows that all sets have similar features in their global mean surface temperature for both externally forced signal and ensemble spread. Set 1420-3 shows an offset towards lower temperatures. These low temperatures are related to the enhanced sea ice extent when forced with climatological sea ice conditions. This offset vanishes when considering land grid points only.

The main differences in the mean state are found in the high latitudes and are a plausible response to the different sea ice forcing. In the lower latitudes, the main features are warm anomalies over subtropical northern Africa and along the Brazilian northeast coast. These warm anomalies are likely related to a slightly warmer state of the Atlantic in the forcing for set 1420-3. However, the amplitude of these anomalies is small compared to the expected variability of 2 m temperature (Fig. S3). Also, the differences in terms of the SST forcing are relatively small compared to the variations caused by different states of the ocean circulation.



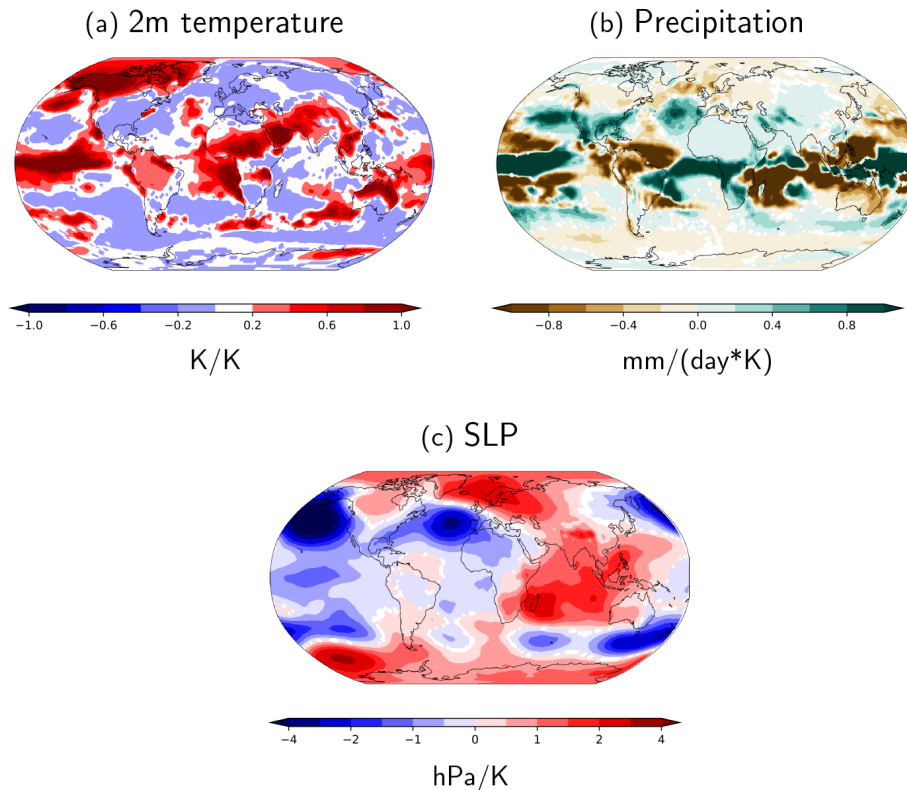
**Figure 4.** Response to volcanic forcing. Difference between the first winter (DJF; **a**, **c**, **e**) and summer (JJA; **b**, **d**, **f**) after an eruption for 2 m temperature (**a**, **b**), precipitation (**c**, **d**), and sea-level pressure (**e**, **f**) and the same quantity averaged over the five previous summers (left) and winters (right). The resulting response was then averaged over 15 major eruptions (same as in Fischer et al., 2007, Table 1 therein) and across all ensemble members. Only significant values are shown, i.e., grid points where the response is outside the 5th–95th percentile range of 1000 surrogates, with each surrogate being created by picking 15 random years and then computing the average difference between these 15 random years with respect to the 5 years before.

### 3.4 Ability of ModE-Sim to sample internal variability

MPI-ESM, the coupled version of ECHAM6, has been shown to be able to sample internal variability when using a 100-member ensemble of historic simulations (Suarez-Gutierrez et al., 2021). We apply the method proposed in the latter reference to ModE-Sim, to analyze whether the ability of capturing internal variability also holds in a stand-alone atmospheric mode and for our ensemble with fewer ensemble members that extend further into the past. The method

has the strength to evaluate internal variability in a model ensemble without making a priori assumptions. A detailed description of it can be found in the above reference; briefly summarized it works as follows: first, one calculates the ratio of time steps where the observations fall outside the ensemble minimum–maximum range in the same time step. If the ensemble captures the spread correctly, then such outliers should only occur very occasionally (how often exactly depends on the ensemble size; e.g., a 20-member ensemble with realistic internal variability should be likely to capture





**Figure 5.** Atmospheric response to ENSO. Regression of winter (DJF) means of (a) 2 m temperature, (b) precipitation, and (c) sea-level pressure on the Niño 3.4 index for all sets of epoch 1a. Values where the correlation is not significant at the 1 % confidence level are masked.

all events with a 20-year return period, so on average in 5 % of the time steps an observation should fall outside the ensemble minimum–maximum range). If the observation tends to be only above (only below) the ensemble maximum (minimum) for a disproportionately high number of time steps, then this indicates a negative (positive) mean state bias of the model. If outliers with respect to the ensemble minimum–maximum range occur frequently in both a positive and negative direction, this indicates that the model underestimates observed internal variability. Finally, the method also allows regions to be detected where the model overestimates internal variability with respect to observations. This is the case if a disproportionately high number of observations fall within the center of the ensemble spread. Consistent with Suarez-Gutierrez et al. (2021), here we use the 12.5th–87.5th quantile range, which – if the ensemble spread agrees with observed internal variability – by definition on average should include 75 % of the observations.

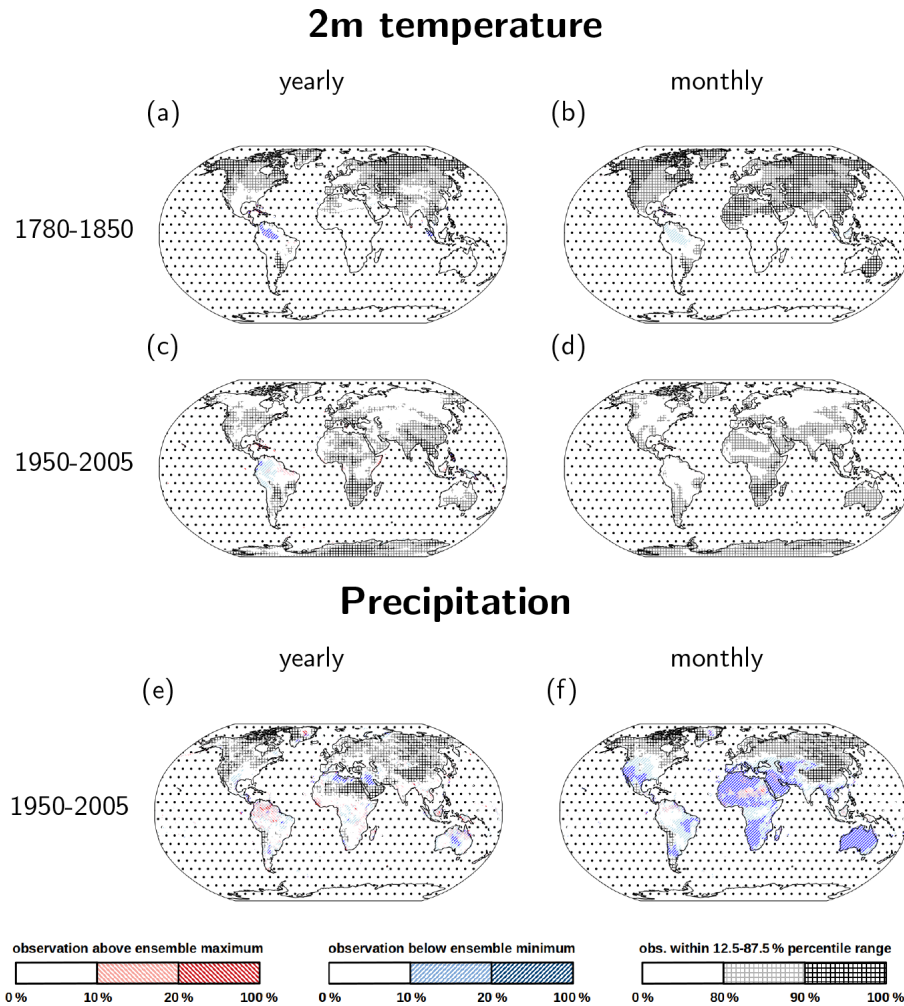
For 2 m temperature, ModE-Sim performs best over Eurasia and tropical South America, as well as over parts of North America (Fig. 6a–d). This performance holds for both monthly and yearly averaged anomalies. Generally, there is a better agreement for the period 1950 to 2005 than for 1780 to 1850. In most other regions, the ensemble spread tends to be too large, indicated by the hatched regions where the ob-

servations fall within the 12.5th–87.5th percentile range in a disproportionately high number of time steps. On seasonal timescales, it shows that the performance in boreal winter is better than in the summer season (Fig. S4). The results for temperature hold even when analyzing each set separately, indicating that only 20 ensemble members already give a reasonably good estimate of internal variability (Fig. S5).

For precipitation, we limit our analysis to the period 1950 to 2005 because of the lack of precipitation observations before the 20th century. During this period, ModE-Sim captures annual precipitation variability reasonably well in most parts of the world (Fig. 6e), but slightly too high an ensemble spread is found over parts of Europe and western Siberia and parts of North America. The ensemble spread is also large in northern Africa and over eastern Siberia and the Tibetan Plateau. On monthly timescales, ModE-Sim shows a tendency of being too wet in many time steps in the subtropical desert regions (Fig. 6f).

### 3.5 Heat waves as an example to demonstrate ModE-Sim’s ability to simulate extreme events

Beyond its ability to separate internal variability and externally forced events, a large ensemble can also be a useful tool for studying climate extremes. By definition, extreme events are rare, and therefore usually only a limited number

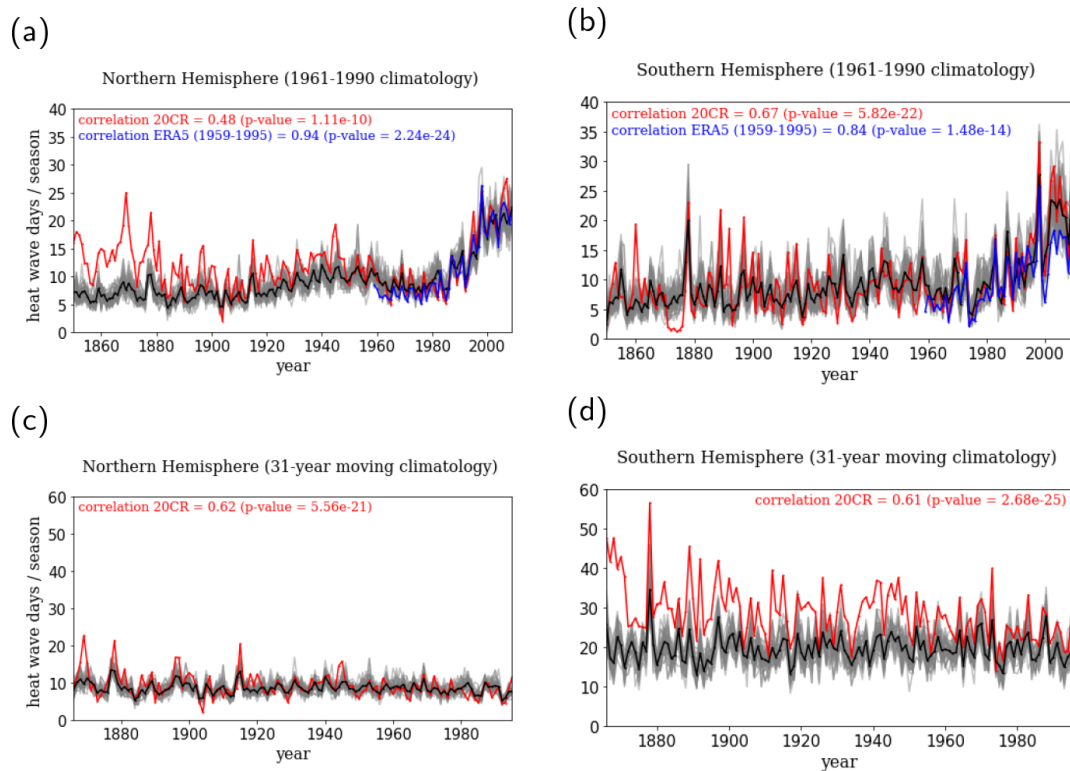


**Figure 6.** Ability of ModE-Sim to capture internal variability for 2 m temperature in the period 1780 to 1850 (a, b) and 1950 to 2005 (c, d) and precipitation (e, f) on yearly (a, c, e) and monthly (b, d, f) timescales. Light-red (dark-red) shadings indicate regions where Berkeley Earth (for temperature) and GPCP (for precipitation) observations lie below the ensemble maximum for more than 10% (20%) of the time steps. Light-blue and dark-blue shadings indicate regions where the observations are below the ensemble minimum accordingly. The grey hatching indicates regions where the ensemble overestimates internal variability, i.e., where more than 80% (light-grey hatching) and 90% (black hatching) of the time steps fall within the 12.5th–87.5th percentile range. Stippling indicates regions where observations are available for less than 10% of the time steps. For details on the method, refer to Sect. 3.4.

of observations is available. At a certain number of ensemble members, even extreme events may occur often enough to be able to perform a statistically robust analysis of the underlying mechanisms. Also, it is interesting to know to what extent these extremes are influenced by external forcing. A detailed study of extreme events is beyond the scope of this documentation paper, but we will show some first results that give evidence that ModE-Sim captures heat waves, making it a valuable tool for studying the underlying processes triggering climate extremes.

For our analysis, we computed the number of heat waves per season for the model simulations and, as a reference, for 20CRv3 and ERA5 reanalysis. Based on daily 2 m temperature data, we defined heat waves as those days on which

temperature exceeds the 90th percentile of a reference period at the according grid point. We used two different definitions of the reference period: in the first approach, the percentiles were computed from the 1961–1990 period. This fixed-climatology approach has the advantage that the threshold for a day being labeled as a heat wave day does not change during time. The obvious disadvantage is that the number of heat wave days is likely to increase with rising mean temperature. Therefore we additionally provide a second version of the analysis where the percentiles are computed from a 31-year running window. This adapts the temperature threshold for heat waves to the anthropogenic warming, and heat waves are restricted to days that are extremely warm with respect to other days in the according



**Figure 7.** Ability of ModE-Sim to capture late 19th- and 20th-century heat waves for (a) the Northern Hemisphere (10 to 80° N) in boreal summer (May to September) and (b) the Southern Hemisphere (65° S–0°) in austral summer (November to March). Shown is the number of heat wave days per season defined as days when 2 m temperature exceeds the 90th percentile of the 1961–1990 reference period. Light colors show the individual ModE-Sim ensemble members, the black line the ModE-Sim ensemble mean, the blue line ERA5, and the red line the 20CRv3 ensemble mean. Panels (c) and (d) are the same as (a) and (b) but with using the 90th percentile of a 31-year running climatology.

31-year window. To correct for model biases, the thresholds were computed separately for each dataset. To compare the datasets, we computed the Spearman rank correlation between the model and the reanalysis and according  $p$  values.

Figure 7 shows that ModE-Sim is able to produce a reasonable number of heat waves on a hemispheric scale. While the occurrence of heat waves is slightly underestimated for the Northern Hemisphere before 1950, the number of heat wave days almost perfectly matches that in 20CRv3 when considering the ensemble mean in the Southern Hemisphere. The high correlation between the number of observed heat waves and the ModE-Sim ensemble means indicates that a large number of the heat waves can be linked to external forcing, particularly for the Southern Hemisphere. External forcing in this case does not necessarily restrict to a direct radiative response but may also include dynamical components that are a result of the SST forcing and/or the volcano forcing. High correlations are not only resulting from the strong temperature trend in the late 20th century, but also hold with the 31-year moving climatology approach (lower row of Fig. 7). Discrepancies with 20CR in the Southern Hemisphere in the early 20th century are likely a result of missing data and high uncertainties in 20CR during that period.

A more detailed study of heat waves in ModE-Sim is currently in preparation.

#### 4 Conclusions

In this paper, we presented ModE-Sim, a medium-sized ensemble of atmospheric general circulation model (AGCM) simulations covering the period from 1420 to 2009. To our knowledge, it is the first ensemble of comparable size that covers such a long period. ModE-Sim has to be seen as an ensemble of opportunities that combines different, partly inhomogeneous setups. These reflect the very different levels of uncertainty in the forcing and boundary conditions for the different periods. Due to the re-initialization of the model in 1850, the switch from SST reconstructions to HadISST forcing, and the change in ensemble size, we strongly recommend analyzing epoch 1 and epoch 2 separately when using the dataset. Except for localized effects in the high latitudes when switching between climatological sea ice forcing and analogues, we only found minor differences between the individual sets in epoch 1. Therefore we can treat all three sets as one 60-member ensemble when the data are used for analysis that benefits from a large ensemble size. However,

the more conservative and accurate way would be to analyze each set separately.

Nevertheless, we show that ModE-Sim is a useful tool to study climate and its variability in the past 600 years. While sharing the previously known mean state biases of ECHAM in stand-alone mode, our ensemble performs well in sampling internal variability, particularly for near-surface temperature. Another interesting finding is that ModE-Sim has the ability to capture extreme events, such as heat waves.

Beside its usefulness for pure model studies, the original motivation for ModE-Sim was to create the input for an offline data assimilation approach. The according reanalysis product, ModE-RA (data available at [https://doi.org/10.26050/WDCC/ModE-RA\\_s14203-18501](https://doi.org/10.26050/WDCC/ModE-RA_s14203-18501), Valler et al., 2023), is published in a consistent data structure to easily allow direct comparison between the AGCM ensemble and the climate reconstruction based on it.

**Code availability.** ECHAM6 was published by the Max Planck Institute for Meteorology (MPIMET) under an institutional licence that guarantees access to the ECHAM6 source code to the scientific community. Accessing the ECHAM6 source code requires contacting the modeling department of MPIMET: <https://mpimet.mpg.de/en/research/modelling> (Max Planck Institute for Meteorology, 2023). Additionally, for documentation of the ECHAM6 setup that was used to create ModE-Sim, ECHAM6 example run scripts are provided as additional information with the data through the World Data Center for Climate: <https://www.wdc-climate.de/ui/entry?acronym=ModE-Sim> (Hand et al., 2023f). Code used to create EVA inputs and generate perturbed volcanic forcing can be found at <https://doi.org/10.5281/zenodo.7669569> (Samakinwa, 2023).

**Data availability.** A subset of ModE-Sim variables (including, but not limited to, those used in this paper), forcings and boundary conditions (unless standard PMIP4/HadISST2), and example run scripts are made available through the World Data Center for Climate at <https://www.wdc-climate.de/ui/entry?acronym=ModE-Sim> (Hand et al., 2023f). Individual ModE-Sim sets also can be accessed at [https://doi.org/10.26050/WDCC/ModE-Sim\\_s14201](https://doi.org/10.26050/WDCC/ModE-Sim_s14201) (Hand et al., 2023a) (set 1420-1), [https://doi.org/10.26050/WDCC/ModE-Sim\\_s14202](https://doi.org/10.26050/WDCC/ModE-Sim_s14202) (Hand et al., 2023b) (set 1420-2), [https://doi.org/10.26050/WDCC/ModE-Sim\\_s14203](https://doi.org/10.26050/WDCC/ModE-Sim_s14203) (Hand et al., 2023c) (set 1420-3), [https://doi.org/10.26050/WDCC/ModE-Sim\\_s18501](https://doi.org/10.26050/WDCC/ModE-Sim_s18501) (Hand et al., 2023d) (set 1850-1), and [https://doi.org/10.26050/WDCC/ModE-Sim\\_s18502](https://doi.org/10.26050/WDCC/ModE-Sim_s18502) (Hand et al., 2023e) (set 1850-2). Further variables are available upon request by contacting the authors.

**Supplement.** The supplement related to this article is available online at: <https://doi.org/10.5194/gmd-16-4853-2023-supplement>.

**Author contributions.** RH led the writing of the manuscript, performed the AGCM experiments, and did most of the presented evaluation. ES provided the ocean boundary conditions and generated the perturbed volcanic forcing. LL did the analysis on the heat wave part. All authors contributed to the discussion on the experimental setup, the analysis, and the manuscript.

**Competing interests.** The contact author has declared that none of the authors has any competing interests.

**Disclaimer.** Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Acknowledgements.** We would like to thank Doris Folini, Sylvaine Ferrachat, and the climate modeling group at ETH Zurich and Johann Jungclaus, Sebastian Rast, and their colleagues at the Max Planck Institute for Meteorology for assistance with installing the model and the selection and configuration of the forcings. We acknowledge technical support by the Swiss National Supercomputing Center (CSCS) and the University of Bern high-performance computing team. Furthermore, we would like to thank Eileen Herwig, Heinke Höck, and the data management team at the German Climate Computing Centre (DKRZ) for comprehensive support with publishing the datasets through the WDCC platform.

**Financial support.** This research has been supported by the European Research Council, H2020 European Research Council (grant no. PALAEO-RA 787574).

**Review statement.** This paper was edited by Charles Onyutha and reviewed by two anonymous referees.

## References

- Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Schamm, K., Schneider, U., and Ziese, M.: A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present, *Earth Syst. Sci. Data*, 5, 71–99, <https://doi.org/10.5194/essd-5-71-2013>, 2013.
- Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S.: An ensemble-based approach to climate reconstructions, *Clim. Past*, 8, 963–976, <https://doi.org/10.5194/cp-8-963-2012>, 2012.
- Bittner, M., Timmreck, C., Schmidt, H., Toohey, M., and Krüger, K.: The impact of wave-mean flow interaction on the Northern Hemisphere polar vortex after tropical volcanic eruptions, *J. Geophys. Res.-Atmos.*, 121, 5281–5297, <https://doi.org/10.1002/2015JD024603>, 2016.
- Brönnimann, S.: Impact of El Niño–Southern Oscillation on European climate, *Rev. Geophys.*, 45, RG3003, <https://doi.org/10.1029/2006RG000199>, 2007.

- Christiansen, B.: Volcanic Eruptions, Large-Scale Modes in the Northern Hemisphere, and the El Niño–Southern Oscillation, *J. Climate*, 21, 910–922, <https://doi.org/10.1175/2007JCLI1657.1>, 2008.
- 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, *Nat. Clim. Change*, 10, 277–286, <https://doi.org/10.1038/s41558-020-0731-2>, 2020.
- Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L.-P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., Fuentes-Franco, R., Gröger, M., v. Hardenberg, J., Hieronymus, J., Karami, M. P., Keskinen, J.-P., Koenigk, T., Makkonen, R., Massonnet, F., Ménégoz, M., Miller, P. A., Moreno-Chamarro, E., Nieradzick, L., van Noije, T., Nolan, P., O'Donnell, D., Olinaho, P., van den Oord, G., Ortega, P., Prims, O. T., Ramos, A., Reerink, T., Rousset, C., Ruprich-Robert, Y., Le Sager, P., Schmith, T., Schrödner, R., Serva, F., Sicardi, V., Sloth Madsen, M., Smith, B., Tian, T., Tourigny, E., Uotila, P., Vancoppenolle, M., Wang, S., Wärlind, D., Willén, U., Wyser, K., Yang, S., Yepes-Arbós, X., and Zhang, Q.: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, *Geosci. Model Dev.*, 15, 2973–3020, <https://doi.org/10.5194/gmd-15-2973-2022>, 2022.
- Driscoll, S., Bozzo, A., Gray, L. J., Robock, A., and Stenchikov, G.: Coupled Model Intercomparison Project 5 (CMIP5) simulations of climate following volcanic eruptions, *J. Geophys. Res.-Atmos.*, 117, D17105, <https://doi.org/10.1029/2012JD017607>, 2012.
- Fischer, E. M., Luterbacher, J., Zorita, E., Tett, S. F. B., Casty, C., and Wanner, H.: European climate response to tropical volcanic eruptions over the last half millennium, *Geophys. Res. Lett.*, 34, L05707, <https://doi.org/10.1029/2006GL027992>, 2007.
- Franke, J., Brönnimann, S., Bhend, J., and Brugnara, Y.: A monthly global paleo-reanalysis of the atmosphere from 1600 to 2005 for studying past climatic variations, *Sci. Data*, 4, 170076, <https://doi.org/10.1038/sdata.2017.76>, 2017.
- Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J., and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, *J. Adv. Model. Earth Sy.*, 5, 572–597, <https://doi.org/10.1002/jame.20038>, 2013.
- Graft, H.-F., Kirchner, I., Robock, A., and Schult, I.: Pinatubo eruption winter climate effects: model versus observations, *Clim. Dynam.*, 9, 81–93, <https://doi.org/10.1007/BF00210011>, 1993.
- Gómez-Navarro, J. J., Zorita, E., Raible, C. C., and Neukom, R.: Pseudo-proxy tests of the analogue method to reconstruct spatially resolved global temperature during the Common Era, *Clim. Past*, 13, 629–648, <https://doi.org/10.5194/cp-13-629-2017>, 2017.
- Hand, R., Brönnimann, S., Samakinwa, E., and Lipfert, L.: ModE-Sim – A medium size AGCM ensemble to study climate variability during the modern era (1420 to 2009): Set 1420-1, World Data Center for Climate (WDCC) at DKRZ [data set], [https://doi.org/10.26050/WDC/ModE-Sim\\_s14201](https://doi.org/10.26050/WDC/ModE-Sim_s14201), 2023a.
- Hand, R., Brönnimann, S., Samakinwa, E., and Lipfert, L.: ModE-Sim – A medium size AGCM ensemble to study climate variability during the modern era (1420 to 2009): Set 1420-2, World Data Center for Climate (WDCC) at DKRZ [data set], [https://doi.org/10.26050/WDC/ModE-Sim\\_s14202](https://doi.org/10.26050/WDC/ModE-Sim_s14202), 2023b.
- Hand, R., Brönnimann, S., Samakinwa, E., and Lipfert, L.: ModE-Sim – A medium size AGCM ensemble to study climate variability during the modern era (1420 to 2009): Set 1420-3, World Data Center for Climate (WDCC) at DKRZ [data set], [https://doi.org/10.26050/WDC/ModE-Sim\\_s14203](https://doi.org/10.26050/WDC/ModE-Sim_s14203), 2023c.
- Hand, R., Brönnimann, S., Samakinwa, E., and Lipfert, L.: ModE-Sim – A medium size AGCM ensemble to study climate variability during the modern era (1420 to 2009): Set 1850-1, World Data Center for Climate (WDCC) at DKRZ [data set], [https://doi.org/10.26050/WDC/ModE-Sim\\_s18501](https://doi.org/10.26050/WDC/ModE-Sim_s18501), 2023d.
- Hand, R., Brönnimann, S., Samakinwa, E., and Lipfert, L.: ModE-Sim – A medium size AGCM ensemble to study climate variability during the modern era (1420 to 2009): Set 1850-2, World Data Center for Climate (WDCC) at DKRZ [data set], [https://doi.org/10.26050/WDC/ModE-Sim\\_s18502](https://doi.org/10.26050/WDC/ModE-Sim_s18502), 2023e.
- Hand, R., Brönnimann, S., Samakinwa, E., and Lipfert, L.: ModE-Sim – A medium size AGCM ensemble to study climate variability during the modern era (1420 to 2009), World Data Center for Climate (WDCC) at DKR [code and data set], <https://www.wdc-climate.de/ui/entry?acronym=ModE-Sim>, last access: 22 August 2023f.
- Jungclaus, J. H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., Egorova, T., Evans, M., González-Rouco, J. F., Goosse, H., Hurrell, G. C., Joos, F., Kaplan, J. O., Khodri, M., Klein Goldewijk, K., Krivova, N., LeGrande, A. N., Lorenz, S. J., Luterbacher, J., Man, W., Maycock, A. C., Meinshausen, M., Moberg, A., Muscheler, R., Nehrbass-Ahles, C., Otto-Bliesner, B. I., Phipps, S. J., Pongratz, J., Rozanov, E., Schmidt, G. A., Schmidt, H., Schmutz, W., Schurer, A., Shapiro, A. I., Sigl, M., Smerdon, J. E., Solanki, S. K., Timmreck, C., Toohey, M., Usoskin, I. G., Wagner, S., Wu, C.-J., Yeo, K. L., Zanchettin, D., Zhang, Q., and Zorita, E.: The PMIP4 contribution to CMIP6 – Part 3: The last millennium, scientific objective, and experimental design for the PMIP4 past1000 simulations, *Geosci. Model Dev.*, 10, 4005–4033, <https://doi.org/10.5194/gmd-10-4005-2017>, 2017.
- Landrum, L. and Holland, M. M.: Extremes become routine in an emerging new Arctic, *Nat. Clim. Change*, 10, 1108–1115, <https://doi.org/10.1038/s41558-020-0892-z>, 2020.
- Liguori, G., McGregor, S., Arblaster, J. M., Singh, M. S., and Meehl, G. A.: A joint role for forced and internally-driven variability in the decadal modulation of global warming, *Nat. Commun.*, 11, 3827, <https://doi.org/10.1038/s41467-020-17683-7>, 2020.

- Maher, N., McGregor, S., England, M. H., and Gupta, A. S.: Effects of volcanism on tropical variability, *Geophys. Res. Lett.*, 42, 6024–6033, <https://doi.org/10.1002/2015GL064751>, 2015.
- Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Kornblueh, L., Takano, Y., Kroeger, J., Ghosh, R., Hedemann, C., Li, C., Li, H., Manzini, E., Notz, D., Putrasahan, D., Boysen, L., Claussen, M., Ilyina, T., Olonscheck, D., Raddatz, T., Stevens, B., and Marotzke, J.: The Max Planck Institute Grand Ensemble – Enabling the Exploration of Climate System Variability, *J. Adv. Model Earth Sy.*, 2019.
- Maher, N., Milinski, S., and Ludwig, R.: Large ensemble climate model simulations: introduction, overview, and future prospects for utilising multiple types of large ensemble, *Earth Syst. Dynam.*, 12, 401–418, <https://doi.org/10.5194/esd-12-401-2021>, 2021.
- Max Planck Institute for Meteorology: <https://mpimet.mpg.de/en/research/modelling>, last access: 22 August 2023.
- Milinski, S., Maher, N., and Olonscheck, D.: How large does a large ensemble need to be?, *Earth Syst. Dynam.*, 11, 885–901, <https://doi.org/10.5194/esd-11-885-2020>, 2020.
- Müller, W., Jungclaus, J., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornblueh, L., Li, H., Modali, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., and Marotzke, J.: A higher-resolution version of the Max Planck Institute Earth System Model (MPI-ESM 1.2-HR), *J. Adv. Model Earth Sy.*, 10, 1383–1413, <https://doi.org/10.1029/2017MS001217>, 2018.
- Neukom, R., Steiger, N., Gómez-Navarro, J. J., Wang, J., and Werner, J. P.: No evidence for globally coherent warm and cold periods over the preindustrial Common Era, *Nature*, 571, 550–554, <https://doi.org/10.1038/s41586-019-1401-2>, 2019.
- Olonscheck, D., Rugenstein, M., and Marotzke, J.: Broad Consistency Between Observed and Simulated Trends in Sea Surface Temperature Patterns, *Geophys. Res. Lett.*, 47, e2019GL086773, <https://doi.org/10.1029/2019GL086773>, 2020.
- Robock, A.: Volcanic eruptions and climate, *Rev. Geophys.*, 38, 191–219, <https://doi.org/10.1029/1998RG000054>, 2000.
- Rohde, R. A. and Hausfather, Z.: The Berkeley Earth Land/Ocean Temperature Record, *Earth Syst. Sci. Data*, 12, 3469–3479, <https://doi.org/10.5194/essd-12-3469-2020>, 2020.
- Samakinwa, E.: shamakson/ModE-Sim-RA\_volc\_forcing: Generation of perturbed volcanic forcing using EVA, Zenodo [code], <https://doi.org/10.5281/zenodo.7669569>, 2023.
- Samakinwa, E., Valler, V., Hand, R., Neukom, R., Gómez-Navarro, J. J., Kennedy, J., Rayner, N. A., and Brönnimann, S.: An ensemble reconstruction of global monthly sea surface temperature and sea ice concentration 1000–1849, *Sci. Data*, 8, 261, <https://doi.org/10.1038/s41597-021-01043-1>, 2021.
- Sigl, M., Toohey, M., McConnell, J. R., Cole-Dai, J., and Severi, M.: Volcanic stratospheric sulfur injections and aerosol optical depth during the Holocene (past 11 500 years) from a bipolar ice-core array, *Earth Syst. Sci. Data*, 14, 3167–3196, <https://doi.org/10.5194/essd-14-3167-2022>, 2022.
- Sjolte, J., Adolphi, F., Guðlaugsdóttir, H., and Muscheler, R.: Major Differences in Regional Climate Impact Between High- and Low-Latitude Volcanic Eruptions, *Geophys. Res. Lett.*, 48, e2020GL092017, <https://doi.org/10.1029/2020GL092017>, 2021.
- Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E.: Atmospheric component of the MPI-M Earth System Model: ECHAM6, *J. Adv. Model Earth Sy.*, 5, 146–172, <https://doi.org/10.1002/jame.20015>, 2013.
- Suarez-Gutierrez, L., Milinski, S., and Maher, N.: Exploiting large ensembles for a better yet simpler climate model evaluation, *Clim. Dynam.*, 57, 2557–2580, <https://doi.org/10.1007/s00382-021-05821-w>, 2021.
- Titchner, H. A. and Rayner, N. A.: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations, *J. Geophys. Res.-Atmos.*, 119, 2864–2889, <https://doi.org/10.1002/2013JD020316>, 2014.
- Toohey, M. and Sigl, M.: Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE, *Earth Syst. Sci. Data*, 9, 809–831, <https://doi.org/10.5194/essd-9-809-2017>, 2017.
- Toohey, M., Stevens, B., Schmidt, H., and Timmreck, C.: Easy Volcanic Aerosol (EVA v1.0): an idealized forcing generator for climate simulations, *Geosci. Model Dev.*, 9, 4049–4070, <https://doi.org/10.5194/gmd-9-4049-2016>, 2016.
- Trenberth, K. E.: The Definition of El Niño, *B. Am. Meteorol. Soc.*, 78, 2771–2777, [https://doi.org/10.1175/1520-0477\(1997\)078<2771:TDOENO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2771:TDOENO>2.0.CO;2), 1997.
- Valler, V., Franke, J., Brugnara, Y., and Brönnimann, S.: An updated global atmospheric paleo-reanalysis covering the last 400 years, *Geosci. Data J.*, 9, 89–107, <https://doi.org/10.1002/gdj3.121>, 2022.
- Valler, V., Franke, J., Brugnara, Y., Samakinwa, E., Hand, R., Lundstad, E., Burgdorf, A.-M., and Brönnimann, S.: ModE-RA – A global monthly paleo-reanalysis of the modern era (1421 to 2008), *Sci. Data*, in review, 2023.
- von Trentini, F., Aalbers, E. E., Fischer, E. M., and Ludwig, R.: Comparing interannual variability in three regional single-model initial-condition large ensembles (SMILEs) over Europe, *Earth Syst. Dynam.*, 11, 1013–1031, <https://doi.org/10.5194/esd-11-1013-2020>, 2020.
- Zambri, B., LeGrande, A. N., Robock, A., and Slawinska, J.: Northern Hemisphere winter warming and summer monsoon reduction after volcanic eruptions over the last millennium, *J. Geophys. Res.-Atmos.*, 122, 7971–7989, <https://doi.org/10.1002/2017JD026728>, 2017.