ModE-Sim - A medium size AGCM ensemble to study climate variability during the modern era (1420 to 2009)

Ralf Hand[1,2], Eric Samakinwa[1,2], Laura Lipfert[1,2], and Stefan Brönnimann[1,2]

[1]Institute of Geography, University of Bern, Switzerland.
[2]Oeschger Centre for Climate Change Research, University of Bern, Switzerland.

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

Abstract.

We introduce ModE-Sim, a medium size ensemble of simulations with the atmospheric general circulation model ECHAM6 in its LR version (T63/approx. 1.8°horizontal with 47 vertical levels) that covers 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 31620 simulated years in total. The dataset forms the input for a data assimilation procedure that combines historical climate informations with additional constraints from a climate model to produce a novel gridded 3-dimensional dataset of the modern era. Additionally, ModE-Sim on its own is also suitable for many other applications as its various subsets can be used as initial condition and boundary condition ensemble 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 of the two key variables 2m-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. At the example of heat waves we show that the ensemble is even capable of capturing extreme events.

1 Introduction

The use of large ensembles of climate model simulations enables to separate the climate's response to external forcings from
internal climate variability (Maher et al., 2019; Milinski et al., 2020; von Trentini et al., 2020). The individual realizations of
an ensemble can differ either in their boundary conditions, their initial conditions, or in both. Furthermore, they also allow
sampling of the supposed true state of simulated atmospheric variables within a reasonable range of uncertainty. Some notable
applications include but are not limited to the analysis of Sea Surface Temperature (SST) trends (Olonscheck et al., 2020), the
production of initial state estimates for reanalyses data sets via data assimilation approaches (Bhend et al., 2012; Franke et al.,
2017; Valler et al., 2022), 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 comparision with the statistics of a control simulation with climatological forc-
ings. This approach does not fully allow inferences about 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 Ensem-
bles (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 Modern Era SIMulations (ModE-Sim), a medium-sized ensemble of simulations with an atmospheric
model. Unlike some notable multi-model ensemble simulations offered by model intercomparison projects, our setup is sim-
ilar to a SMILEs setup using a single model but 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, sea ice, and
the effect of volcanic eruptions. The experimental design of most SMILE simulations depends 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 time period (Valler et al., 2023, submitted).
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 in sampling the internal variability of the key variables
and is also capable of capturing extreme events, such as heat waves. Providing boundary conditions for this timeslice is chal-
lenging but necessary. Our ensemble simulations use HadISST2 sea 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 reconstruc-
tion over the modern era (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 section 2, we describe the initialization and experimental de-
50 sign procedure of ModE-Sim. Section 3 give explanations of the statistical methods used for the evaluation of the ensemble simulations while section 4 concludes the manuscript and provides a summary.

2 Model & 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
55 simulations. MPI-ESM and ECHAM6 stand-alone models participate regularly in the Atmospheric/Coupled Model Intercomparison Project (AMIP/CMIP). 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, the CMIP6 version, and presumably the last development step of the ECHAM model family. Furthermore, 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°. In the vertical, the LR version uses 47 hybrid
60 levels between the surface and 0.01 hPa.

2.2 The experiments

In principle, our set-up is designed to be close to the PMIP4 *past2k* simulations performed at the Max Planck Institute for Meteorology (Jungclaus et al., 2017) with few exceptions. We use a stand-alone atmospheric component of the model with prescribed SST and sea ice. Furthermore, we use prescribed land-cover maps rather than dynamic vegetation. Our simulations
65 comprise three epochs, forced with slightly different ocean boundaries and radiative forcings. Figure 1 gives an overview of our experimental design. 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.

The initialization of our model simulation is in two steps: First, we forked an atmosphere-only spin-up simulation from
70 the PMIP4 simulation of the coupled version of the model (i.e. MPI-ESM). This spin-up simulation was forced with constant boundary conditions and radiative forcings spanning several decades until equilibrium is reached in an atmosphere-only mode. Secondly, 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.
75 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

80 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.

85 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

90 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 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 bound-

95 ary 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

100 each have a different volcanic forcing. Consistent with the PMIP4 standard setup, these radiative forcings are outputs of the Easy Volcanic Aerosol Model (Toohey et al., 2016) using Volcanic Stratospheric Sulfur Injections reconstruction of Sigl et al. (2022). PMIP4 volcanic forcing also results from EVA, generated using 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. Both sets 1420-1 and 1420-2 use sea ice analogs selected from the HadISST2 dataset

105 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 1430-2 w.r.t. the first

110   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 analog approach used for sets 1420-1 and 1420-2. We found that the sea ice analogs approach shows a bias as almost all analogs 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

115   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 simulations with climatological sea ice might be more suitable for certain analysis. 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 observation, pulling the ocean

120   towards its true state. Because the differences between the SST/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 docu-

125   mented 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 HadISST realization is used with two different initializations). To expand the sample of SST forcings the second set (**"set 1850-2"**) uses 16 different linear

130   combinations of the 10 available HadISST2 realizations as ocean forcing. These linear combinations were formed as follows:

$$SST(i,x,t) = (SST(j,x,t)/SST(k,x,t)) * SST(l,x,t), \quad with \quad j \neq k \neq l$$

where $SST(i,x,t)$ is the SST anomaly in the ith linear combination at gridpoint x and timestep t, and $SST(j,x,t)$, $SST(k,x,t)$ and $SST(l,x,t)$ are the according SST anomalies in the jth, k[th] and l[th] realization of HadISST2. The sea ice concentrations were created by choosing analogs from the HadISST2 sea ice data set. The radiative forcings, including volcanic forcings, are the PMIP4 input and are identical for all simulations and all sets of epoch 2.

3   Evaluation

135

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 of 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

140   consider different time scales, including annual, seasonal, and monthly variability. As reference data set 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 data sets available for the earlier period.

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

A detailed analysis of mean state biases of ECHAM6 is beyond the scope of this paper. For a more detailed mean state 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 (suppl. fig. S1). 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 mid-latitudes and over Australia and

150   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 the warming trend in the 20th century. The forced signal related to the radiative

155   forcing and the ocean boundary conditions can be detected from the subensemble means computed from each 20-member set

separately (Fig. 2b & c), indicating that the ensemble size is clearly sufficient to separate forced signals from internal variability

and uncertainties related to forcing and boundary conditions. The ensemble spread declines towards the end of the simulations,

indicating that for the earlier period, the forcing/boundary condition uncertainty contributes to the ensemble spread in the same

order of magnitude as internal variability does.

160   To determine the spacial manifestation of the volcanic signal we computed composites of the ensemble mean anomalies of

2m temperature, precipitation and sea level pressure for the first (fig. 4) and second (suppl. fig. S3) winter and summer after

15 maior eruptions (Fischer et al., 2007, table 1 therein) with the 5 summers/winters prior to each of the eruptions as reference

period. The spatial anomalies agree well with observations and previous modelling studies (e.g., Graft et al., 1993; Robock,

2000; Fischer et al., 2007; Sjolte et al., 2021). The most prominent feature in terms of 2m temperature is a direct response to the

165   negative radiative forcing that resembles a strong cooling over most continental regions that regionally exceeds 1K. The most

prominent exception from this cooling response can be found over northern Eurasia in boreal winter. The underlying mecha-

nisms 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

170   lead to an equatorward deflection of planetary waves. This reduces higher latitude wave breaking and hence disturbances of the

polar vortex leading to enhanced zonal warm air advection. In ModE-Sim we can find anonamously negative SLP response in

the northern polar latitudes in connection with a band of positive SLP anomaly that spans over the North Atlantic and Northern

Eurasia. The most prominent feature of the precipitation response is a shift of the innertroppical convergence zone towards the summer hemisphere that is strongest over the central and West Pacific.

175   Exemplary for atmospheric response to the ocean boundary conditions we analyzed the response to 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 epoche 1a, averaged these anomalies for the Nino 3.4 region (5°N to 5°S, 170 °W to 120°W) and applyied a 5-month running mean. Afterwards, we computed regressions of atmospheric surface quantities in boreal winter on this index for all sets of epoche 1a. ModE-Sim is able to reproduce the main features of the known ENSO

180   teleconnection (Fig. 5): The SLP response shows good agreement to other models and reanalysis data (e.g., Döscher et al., 2022, fig 18 therein) with a streghening of the Aleutian low, a weekening of the NAO, negative SLP anomalies in the eastern tropical Pacific and a band in the subpolar Southern Ocean, moderate positive SLP anomalies over the Arctic, the Antarctic and a strongly positive SLP anomaly west over the Amundsen Sea. The observed 2m 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

185   over Kanada and Alaska connected to the weakened Aleutian Low and 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 over Australia and Malaysia.

3.3   Differences between the individual sets in epoch 1

For epoch 1, we computed 3 sets that differ in the ocean boundary conditions and the radiative forcing (see section 2.2.2). A

190   resulting question is whether these 3 sets show substantial differences in mean state and ensemble spread. Fig. 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.

195   The main differences in the mean state are found in the high latitudes and are the plausible response to the different sea ice forcing. In the lower latitudes, the main features are warm anomalies over subtropical North Africa and along the Brazilian North East 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 2m-temperature (suppl. fig. S3). Also, the differences in terms of the SST forcing are relatively small compared to the variations caused by different

200   states of the ocean circulation.

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). The method proposed by the authors has the strength that it evaluates internal variability without making a priori assumptions but by using the overestimation of the occurrence of

205   observations lying outside the ensemble spread as a criterion for model bias in the representation of internal variability. We

apply the same method as Suarez-Gutierrez et al. (2021) 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.

For 2m-temperature ModE-Sim performs best over Eurasia and tropical South America, as well as over parts of North

210   America (Fig. 6a - 6d). 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 observations fall within the 12.5 to 87.5 percentile range in an over-proportional number of time steps. On seasonal timescales, it shows that the performance in boreal winter is better than in the summer season (suppl. fig. S4). The results for temperature hold even when analyzing each set separately, indicating that only 20 ensemble

215   members already give a reasonably good estimate of internal variability (suppl. 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 ensemble spread is found over parts of Europa and western Siberia and parts of North

220   America. The ensemble spread is also large in Northern Africa, over Eastern Siberia and the Tibetian plateau. On monthly time scales, ModE-Sim shows the tendency to be too wet in many timesteps in the subtropical desert regions (Fig. 6f).

3.5   Ability of simulating extreme events

Beyond its ability to separate internal variability and externally forced events, a large ensemble can also be a useful tool for

225   studying climate extremes. By definition, extreme events are rare and therefore usually only a limited number 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

230   underlying processes triggering climate extremes.

Fig. 7 shows that ModE-Sim is able to produce a reasonable number of heatwaves 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

235   between the number of observed heat waves and the ModE-Sim ensemble means indicates that a large part of the heat waves can be linked to external forcing. This accounts particularly for the southern hemisphere. 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 AGCM simulations covering the period from 1420 to
240 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 analyz-
ing epoch1 and epoch 2 separately when using the dataset. Except for localized effects in the high latitudes when switching
245 between climatological sea ice forcing and analogs, 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 is 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 shar-
250 ing 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 the 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
255 data assimilation approach. The according reanalysis product, ModE-RA (Valler et al., 2023, submitted, data available at
https://www.wdc-climate.de/ui/entry?acronym=ModE-RA), is published in a consistent data structure to easily allow direct
comparision between the AGCM ensemble and the climate reconstruction based on it.

Code availability. ECHAM6 was published by the Max Plack 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
260 modelling department of MPIMET: https://mpimet.mpg.de/en/research/modelling. 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. Code used to create EVA inputs and generate perturbed
volcanic forcing can be found here: https://doi.org/10.5281/zenodo.7669569.

Data availability. A subset of ModE-Sim variables (including, but not limited to, these used in this manuscript), forcings and boundary
265 conditions (unless standard PMIP4/HadISST2) and example run scripts are made available through the World Data Center for Climate:
https://www.wdc-climate.de/ui/entry?acronym=ModE-Sim. Individual ModE-Sim sets also can be accessed through the following DOIs:
https://doi.org/10.26050/WDCC/ModE-Sim_s14201 (Set 1420-1), https://doi.org/10.26050/WDCC/ModE-Sim_s14202 (Set 1420-2),

https://doi.org/10.26050/WDCC/ModE-Sim_s14203 (Set 1420-3), https://doi.org/10.26050/WDCC/ModE-Sim_s18501 (Set 1850-1), https://doi.org/10.26050/WDCC/ModE-Sim_s18502 (Set 1850-2). Further variables are avaiable upon request by contacting the authors.

270 *Author contributions.* Ralf Hand was leading the writing of the manuscript, performed the AGCM experiments and did most of the presented evaluation. Eric Samakinwa provided the ocean boundary coonditions and generated the perturbed volcanic forcing. Laura Lipfert did the analysis on the heat waves part. All authors contributed to the discussion on the experimental setup, the analysis and the manuscript.

Competing interests. None

Acknowledgements. We would like to thank Doris Folini, Sylvaine Ferrachat and the climate modeling group at ETH Zurich and Johann
275 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 HPC team. Furthermore we would like to thank Eileen Hertwig, 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. RH, ES and LL were funded by the European Research Council through H2020 (ERC Grant PALAEO-RA 787574).

References

280 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 System Science Data, 5, 71–99, https://doi.org/10.5194/essd-5-71-2013, publisher: Copernicus GmbH, 2013.

Bhend, J., Franke, J., Folini, D., Wild, M., and Brönnimann, S.: An ensemble-based approach to climate reconstructions, Climate of the Past,
285 8, 963–976, https://doi.org/10.5194/cp-8-963-2012, publisher: Copernicus GmbH, 2012.

Bittner, M., Timmreck, C., Schmidt, H., Toohey, M., and Krüger, K.: The impact of wave-mean flow interaction on the North-
ern Hemisphere polar vortex after tropical volcanic eruptions, Journal of Geophysical Research: Atmospheres, 121, 5281–5297,
https://doi.org/10.1002/2015JD024603, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2015JD024603, 2016.

Brönnimann, S.: Impact of El Niño–Southern Oscillation on European climate, Reviews of Geophysics, 45,
290 https://doi.org/10.1029/2006RG000199, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2006RG000199, 2007.

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, number: 4 Publisher: Nature Publishing Group, 2020.

295 Driscoll, S., Bozzo, A., Gray, L. J., Robock, A., and Stenchikov, G.: Coupled Model Intercomparison Project 5 (CMIP5) simulations of cli-
mate following volcanic eruptions, Journal of Geophysical Research: Atmospheres, 117, https://doi.org/10.1029/2012JD017607, _eprint:
https://onlinelibrary.wiley.com/doi/pdf/10.1029/2012JD017607, 2012.

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-
300 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., Nieradzik, L., van Noije, T., Nolan, P., O'Donnell, D., Ollinaho, 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., Wårlind, 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,
305 Geoscientific Model Development, 15, 2973–3020, https://doi.org/10.5194/gmd-15-2973-2022, publisher: Copernicus GmbH, 2022.

Fischer, E. M., Luterbacher, J., Zorita, E., Tett, S. F. B., Casty, C., and Wanner, H.: European climate response to tropical vol-
canic eruptions over the last half millennium, Geophysical Research Letters, 34, https://doi.org/10.1029/2006GL027992, _eprint:
https://onlinelibrary.wiley.com/doi/pdf/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
310 past climatic variations, Scientific Data, 4, 170 076, https://doi.org/10.1038/sdata.2017.76, number: 1 Publisher: Nature Publishing Group,
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., Timm-
315 reck, 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, Journal of Advances in Modeling Earth Systems, 5, 572–597, https://doi.org/10.1002/jame.20038, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jame.20038, 2013.

Graft, H.-F., Kirchner, I., Robock, A., and Schult, I.: Pinatubo eruption winter climate effects: model versus observations, Climate Dynamics, 9, 81–93, https://doi.org/10.1007/BF00210011, 1993.

320 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, Climate of the Past, 13, 629–648, https://doi.org/10.5194/cp-13-629-2017, publisher: Copernicus GmbH, 2017.

Jungclaus, J. H., Bard, E., Baroni, M., Braconnot, P., Cao, J., Chini, L. P., Egorova, T., Evans, M., González-Rouco, J. F., Goosse, H., Hurtt, G. C., Joos, F., Kaplan, J. O., Khodri, M., Klein Goldewijk, K., Krivova, N., LeGrande, A. N., Lorenz, S. J., Luterbacher, J., Man,
325 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, Geoscientific Model Development, 10, 4005–4033, https://doi.org/https://doi.org/10.5194/gmd-10-4005-2017, 2017.

330 Landrum, L. and Holland, M. M.: Extremes become routine in an emerging new Arctic, Nature Climate Change, 10, 1108–1115, https://doi.org/10.1038/s41558-020-0892-z, number: 12 Publisher: Nature Publishing Group, 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, Nature Communications, 11, 3827, https://doi.org/10.1038/s41467-020-17683-7, number: 1 Publisher: Nature Publishing Group, 2020.

335 Maher, N., McGregor, S., England, M. H., and Gupta, A. S.: Effects of volcanism on tropical variability, Geophysical Research Letters, 42, 6024–6033, https://doi.org/10.1002/2015GL064751, _eprint: https://onlinelibrary.wiley.com/doi/pdf/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, Journal of Advances in Modeling
340 Earth Systems (JAMES), 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 System Dynamics, 12, 401–418, https://doi.org/10.5194/esd-12-401-2021, 2021.

Milinski, S., Maher, N., and Olonscheck, D.: How large does a large ensemble need to be?, Earth System Dynamics, 11, 885–901, https://doi.org/10.5194/esd-11-885-2020, publisher: Copernicus GmbH, 2020.

345 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), Journal of Advances in Modeling Earth Systems, 10, 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
350 over the preindustrial Common Era, Nature, 571, 550–554, https://doi.org/10.1038/s41586-019-1401-2, number: 7766 Publisher: Nature Publishing Group, 2019.

Olonscheck, D., Rugenstein, M., and Marotzke, J.: Broad Consistency Between Observed and Simulated Trends in Sea Surface Temperature Patterns, Geophysical Research Letters, 47, e2019GL086 773, https://doi.org/10.1029/2019GL086773, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2019GL086773, 2020.

355 Robock, A.: Volcanic eruptions and climate, Reviews of Geophysics, 38, 191–219, https://doi.org/10.1029/1998RG000054, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/1998RG000054, 2000.

Rohde, R. A. 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, publisher: Copernicus GmbH, 2020.

Samakinwa, E., Valler, V., Hand, R., Neukom, R., Gómez-Navarro, J. J., Kennedy, J., Rayner, N. A., and Brönnimann, S.: An en-
360 semble reconstruction of global monthly sea surface temperature and sea ice concentration 1000–1849, Scientific Data, 8, 261, https://doi.org/10.1038/s41597-021-01043-1, number: 1 Publisher: Nature Publishing Group, 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 System Science Data, 14, 3167–3196, https://doi.org/10.5194/essd-14-3167-2022, publisher: Copernicus GmbH, 2022.

365 Sjolte, J., Adolphi, F., Guðlaugsdòttir, H., and Muscheler, R.: Major Differences in Regional Climate Impact Between High- and Low-Latitude Volcanic Eruptions, Geophysical Research Letters, 48, e2020GL092 017, https://doi.org/10.1029/2020GL092017, _eprint: https://onlinelibrary.wiley.com/doi/pdf/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
370 System Model: ECHAM6, Journal of Advances in Modeling Earth Systems, 5, 146–172, https://doi.org/10.1002/jame.20015, 2013.

Suarez-Gutierrez, L., , S., and Maher, N.: Exploiting large ensembles for a better yet simpler climate model evaluation, Climate Dynamics, 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, Journal of Geophysical Research: Atmospheres, 119, 2864–2889, https://doi.org/10.1002/2013JD020316, _eprint:
375 https://onlinelibrary.wiley.com/doi/pdf/10.1002/2013JD020316, 2014.

Toohey, M. and Sigl, M.: Volcanic stratospheric sulfur injections and aerosol optical depth from 500 BCE to 1900 CE, Earth System Science Data, 9, 809–831, https://doi.org/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, Geoscientific Model Development, 9, 4049–4070, https://doi.org/10.5194/gmd-9-4049-2016, publisher: Copernicus GmbH,
380 2016.

Trenberth, K. E.: The Definition of El Niño, Bulletin of the American Meteorological Society, 78, 2771–2777, https://doi.org/http://dx.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, Geoscience Data Journal, 9, 89–107, https://doi.org/10.1002/gdj3.121, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/gdj3.121,
385 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), submitted to Nature Scientific Data, 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 System Dynamics, 11, 1013–1031, https://doi.org/10.5194/esd-11-1013-2020, publisher: Copernicus GmbH, 2020.

390

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, Journal of Geophysical Research: Atmospheres, 122, 7971–7989, https://doi.org/10.1002/2017JD026728, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017JD026728, 2017.
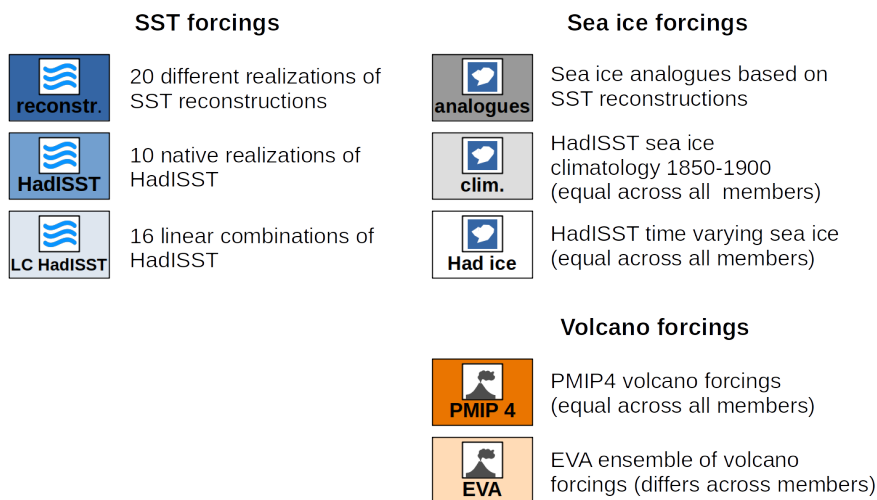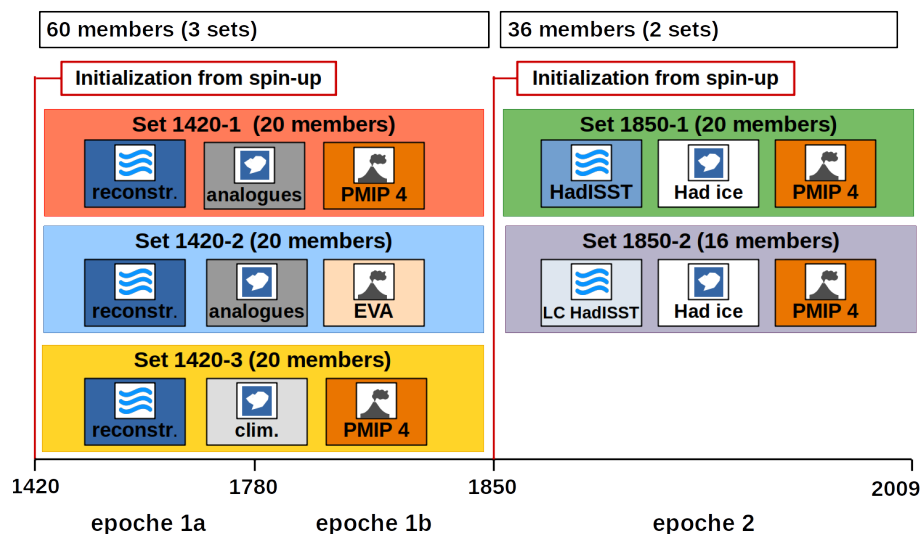
Figure 1. Overview on the setup of the experiment sets in ModE-Sim and their forcings.
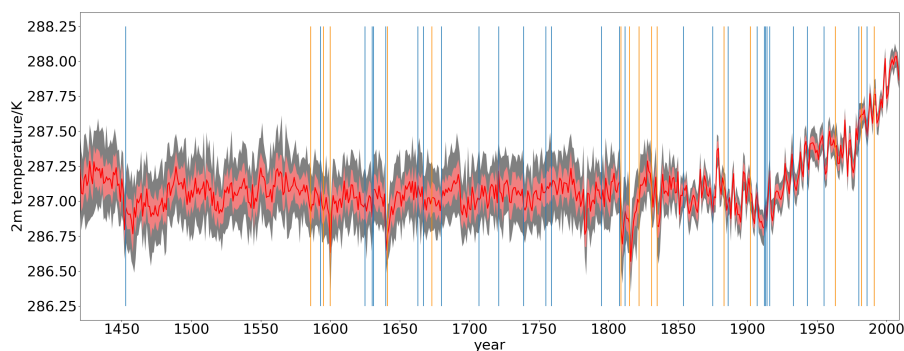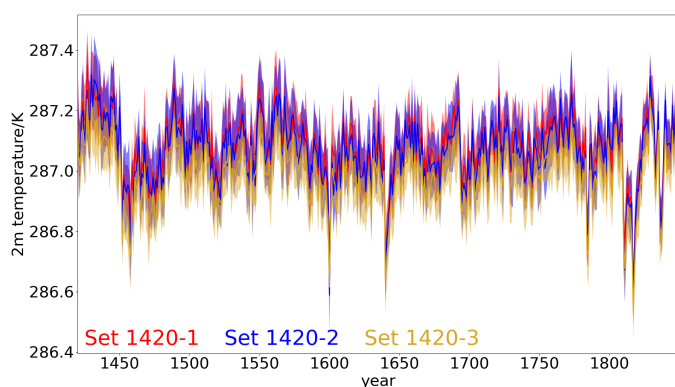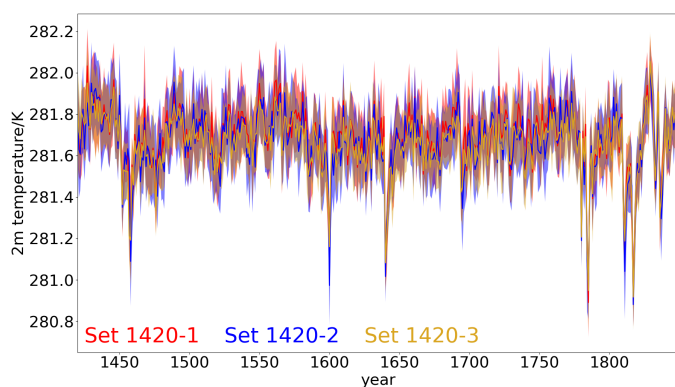
Figure 2. Time series of globally averaged 2m temperature. (a) Ensemble mean (red line), +/- 1 ensemble standard deviation (light red shading), and ensemble minimum/maximum (grey shading) of all sets in epoch 1. Vertical orange lines indicate volcanoes used for the composites in section 3.2, vertical blue lines indicate additional volcanic eruptions with a volcanic explosivity index ≥ 5. (b) Sets 1420-1 (red), 1420-2 (blue) and 1420-3 (dark yellow) separated. The coloured lines indicate the ensemble mean of each set, the shadings the ensemble minimum-maximum range. (c) same as (b), but limited to grid points over land.
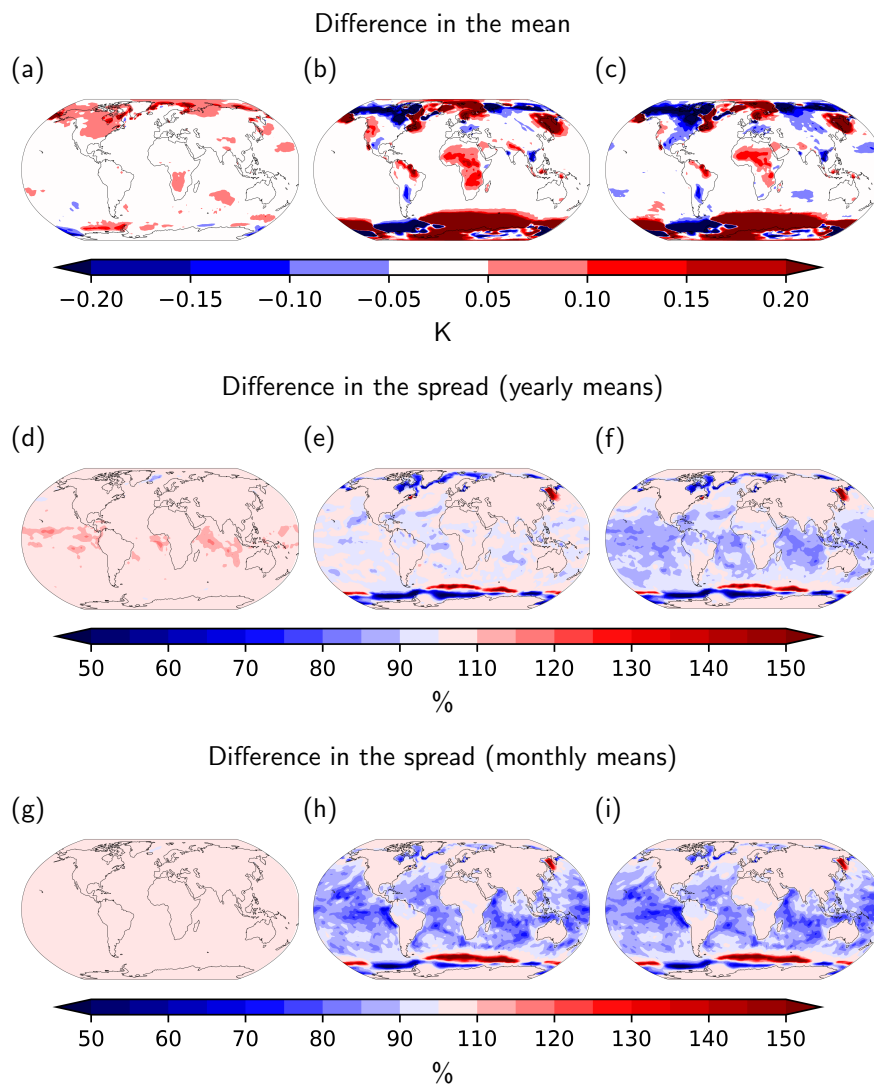
Figure 3. Difference in the temporally averaged ensemble mean of 2m temperature (in K, upper row), the ratio between the temporally averaged ensemble standard deviations on annual (in %, middle row) and monthly (in %, lower row) timescales between the different sets in epoch 1a.
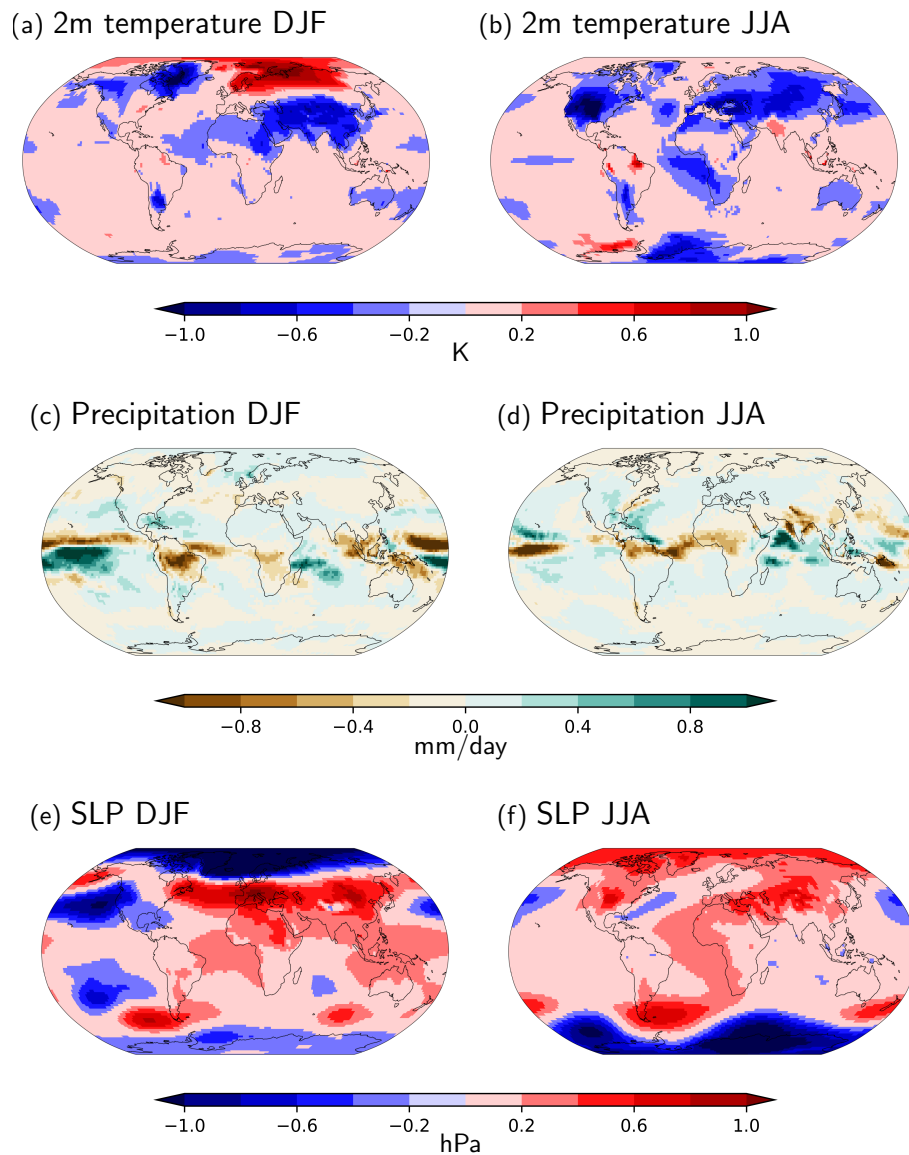
Figure 4. Response to volcanic forcing averaged over 15 maior eruptions (same as in Fischer et al. (2007), table 1 therein.). Difference between the ensemble mean of the first winter (DJF, left column) and summer (JJA, right column) after an eruption for 2m temperature (top), precipitation (middle) and sea level pressure (bottom) and the same quantity avaraged over the 5 previous summers/winters.
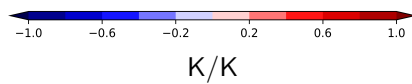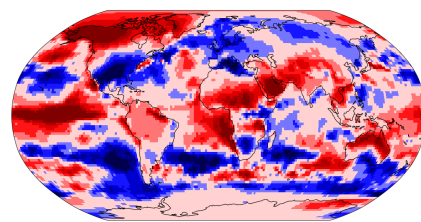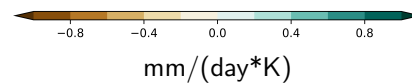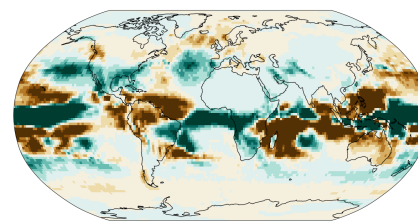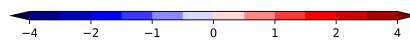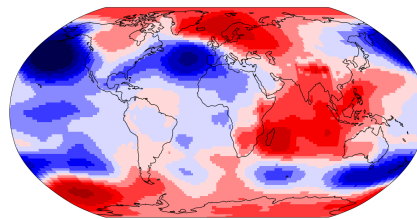
Figure 5. Atmosperic response to ENSO. Regression of winter (DJF) means of (a) 2m temperature, (b) precipitation, and (c) sea level pressure on the Nino 3.4 index for all sets of epoche 1a.
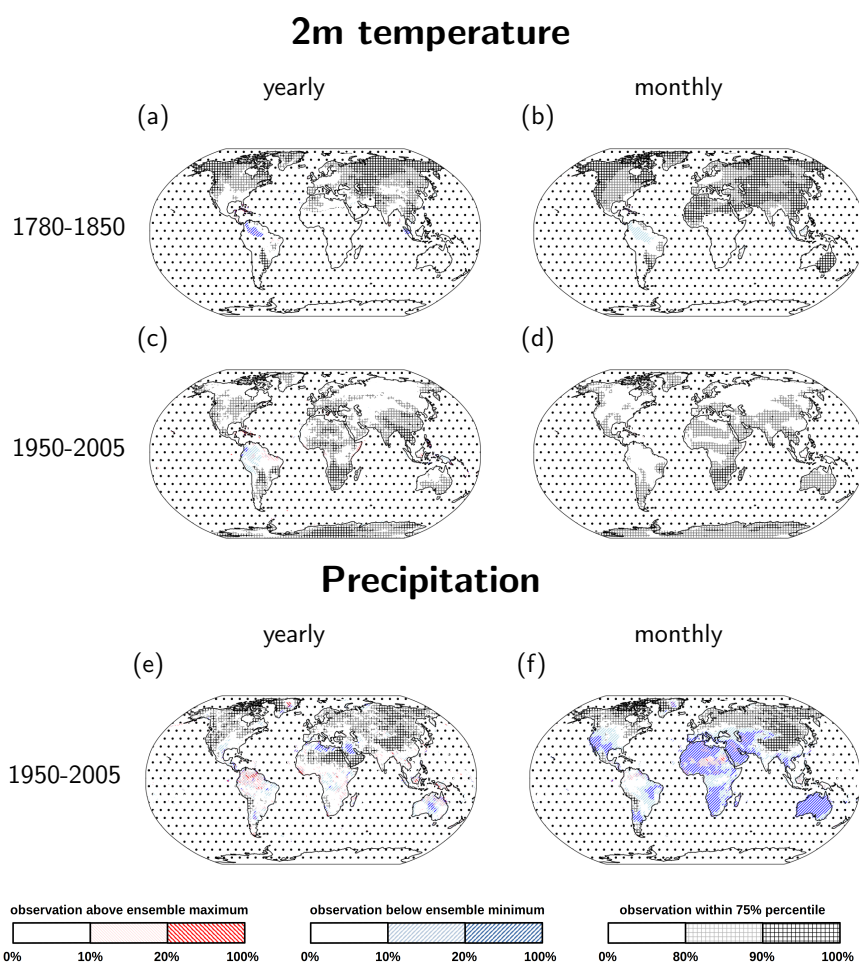
Figure 6. Ability of ModE-Sim to capture internal variability for 2m temperature in the period 1780 to 1850 (upper row) and 1950 to 2005 (middle row) and precipitation (lower row) on yearly (left) and monthly (right) timescales. Light red (Dark red) shadings indicate regions where Berkeley Earth (for temperature) respectively GPPC (for precipitation) observations lie below the ensemble maximum of the 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) respectively 90 (black hatching) % of the time steps fall within the 12.5 to 87.5 percentile range. Stippling indicates regions where observations are available for less than 10% of the time steps.

(a)   (b)
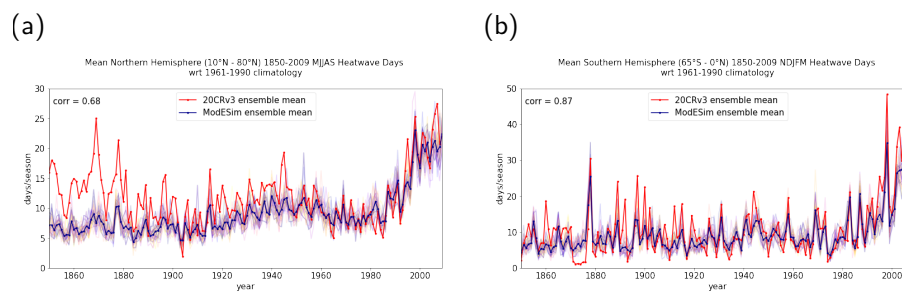
Figure 7. Ability of ModE-Sim to capture late 19th and 20th heat waves for (a) northern hemisphere (10°N to 80°N) in boreal summer (May to September) and (b) southern hemisphere (65°S-0°) in austral summer (November to March). Shown is the number of heatwave days per season defined as days when 2m temperature exceeds the 95th percentile of the 1961 to 1990 reference periode. Light colours show the individual ModE-Sim ensemble members, the blue line the ModE-Sim ensemble mean and the red line the 20CRv3 ensemble mean.