The new Max Planck Institute Grand Ensemble with CMIP6 forcing and high-frequency model output

Dirk Olonscheck[1], Laura Suarez-Gutierrez[2], Sebastian Milinski[3], Goratz Beobide-Arsuaga[4], Johanna Baehr[4], Friederike Fröb[2], Lara Hellmich[2], Tatiana Ilyina[5], Christopher Kadow[6], Daniel Krieger[7], Hongmei Li[2], Jochem Marotzke[2], Étienne Plésiat[6], Martin Schupfner[6], Fabian Wachsmann[6], Karl-Hermann Wieners[2], and Sebastian Brune[8]

[1]MPI
[2]Max Planck Institute for Meteorology
[3]ECMWF
[4]Universität Hamburg, Center for Earth System Research and Sustainability
[5]Max Planck Institute of Meteorology
[6]DKRZ
[7]Helmholtz-Zentrum Geesthacht
[8]Universität Hamburg

May 4, 2023

Abstract

Single-model initial-condition large ensembles are powerful tools to quantify the forced response, internal climate variability, and their evolution under global warming. Here, we present the CMIP6 version of the Max Planck Institute Grand Ensemble (MPI-GE CMIP6) with 30 realisations for the historical period and five emission scenarios. The power of MPI-GE CMIP6 goes beyond its predecessor ensemble MPI-GE by providing high-frequency output, the full range of emission scenarios including the highly policy-relevant low emission scenarios SSP1-1.9 and SSP1-2.6, and the opportunity to compare the ensemble to complementary high-resolution simulations. First, we describe MPI-GE CMIP6, evaluate it with observations and reanalyses and compare it to MPI-GE. Then, we demonstrate with six novel application examples how to use the power of the ensemble to better quantify and understand present and future climate extremes, to inform about uncertainty in approaching Paris Agreement global warming limits, and to combine large ensembles and artificial intelligence. For instance, MPI-GE CMIP6 allows us to show that the recently observed Siberian and Pacific North American heatwaves would only avoid reaching 1-2 year return periods in 2071-2100 with low emission scenarios, that recently observed European precipitation extremes are captured only by complementary high-resolution simulations, and that 3-hourly output projects a decreasing activity of storms in mid-latitude oceans. Further, the ensemble is ideal for estimates of probabilities of crossing global warming limits and the irreducible uncertainty introduced by internal variability, and is sufficiently large to be used for infilling surface temperature observations with artificial intelligence.

The new Max Planck Institute Grand Ensemble with CMIP6 forcing and high-frequency model output

Dirk Olonscheck[1], Laura Suarez-Gutierrez[1,2,3], Sebastian Milinski[1,4], Goratz Beobide-Arsuaga[5,6], Johanna Baehr[5], Friederike Fröb[7], Lara Hellmich[1,6], Tatiana Ilyina[1], Christopher Kadow[8], Daniel Krieger[6,9], Hongmei Li[1], Jochem Marotzke[1,5], Étienne Plésiat[8], Martin Schupfner[8], Fabian Wachsmann[8], Karl-Hermann Wieners[1], Sebastian Brune[5]

[1]Max Planck Institute for Meteorology, Hamburg, Germany
[2]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[3]Institut Pierre-Simon Laplace, CNRS, Paris, France
[4]European Centre for Medium-Range Weather Forecasts, Bonn, Germany
[5]Center for Earth System Research and Sustainability, Universität Hamburg, Hamburg, Germany
[6]International Max Planck Research School on Earth System Modelling, Hamburg, Germany
[7]Geophysical Institute, University of Bergen, and Bjerknes Centre for Climate Research, Bergen, Norway
[8]German Climate Computing Centre (DKRZ), Hamburg, Germany
[9]Helmholtz-Zentrum Hereon, Geesthacht, Germany

Key Points:

- MPI-GE CMIP6 is a 30-member initial-condition large ensemble with up to 3-hourly model output and five emission scenarios
- The ensemble is specifically suited to investigate climate extremes and Paris Agreement global warming limits
- MPI-GE CMIP6 adequately represents heat extremes, while precipitation extremes are captured by complementary high-resolution simulations

Corresponding author: Dirk Olonscheck, dirk.olonscheck@mpimet.mpg.de

Abstract

Single-model initial-condition large ensembles are powerful tools to quantify the forced response, internal climate variability, and their evolution under global warming. Here, we present the CMIP6 version of the Max Planck Institute Grand Ensemble (MPI-GE CMIP6) with 30 realisations for the historical period and five emission scenarios. The power of MPI-GE CMIP6 goes beyond its predecessor ensemble MPI-GE by providing high-frequency output, the full range of emission scenarios including the highly policy-relevant low emission scenarios SSP1-1.9 and SSP1-2.6, and the opportunity to compare the ensemble to complementary high-resolution simulations. First, we describe MPI-GE CMIP6, evaluate it with observations and reanalyses and compare it to MPI-GE. Then, we demonstrate with six novel application examples how to use the power of the ensemble to better quantify and understand present and future climate extremes, to inform about uncertainty in approaching Paris Agreement global warming limits, and to combine large ensembles and artificial intelligence. For instance, MPI-GE CMIP6 allows us to show that the recently observed Siberian and Pacific North American heatwaves would only avoid reaching 1-2 year return periods in 2071-2100 with low emission scenarios, that recently observed European precipitation extremes are captured only by complementary high-resolution simulations, and that 3-hourly output projects a decreasing activity of storms in mid-latitude oceans. Further, the ensemble is ideal for estimates of probabilities of crossing global warming limits and the irreducible uncertainty introduced by internal variability, and is sufficiently large to be used for infilling surface temperature observations with artificial intelligence.

Plain Language Summary

Climate model simulations that start from different initial states and differ only due to the chaos in the climate system are used extensively to quantify the forced climate response, variability intrinsic to the climate system, and their change under global warming. Here, we present a new version of the Max Planck Institute Grand Ensemble (MPI-GE CMIP6) that is run as part of the latest generation of climate models. This single-model ensemble consists of 30 realisations for the historical period 1850-2014 and for five scenarios of possible future climates until 2100. The power of MPI-GE CMIP6 goes beyond its predecessor by not only providing monthly mean but also 3-hourly to daily model output, the full range of future scenarios including the two highly policy-relevant scenarios that were designed to match the Paris Agreement global warming limits of 1.5°C and 2°C, and the opportunity to compare the low-resolution ensemble to simulations of the same model version with higher horizontal resolution. In this paper, we describe the new ensemble and demonstrate with novel application examples how to use its power. For instance, the new ensemble allows us to show that recently observed heatwaves are projected to occur every year at the end of the 21st century if anthropogenic carbon emissions remain high, that recently observed precipitation extremes are captured only by simulations with higher horizontal resolution than that of MPI-GE CMIP6, and that the storminess in many ocean basins is projected to decrease. Further, the ensemble is ideal for estimates of crossing probabilities of Paris Agreement global warming limits, and is sufficiently large to be used to infill missing observations of surface temperature with artificial intelligence.

1 Introduction

Single-model initial-condition large ensembles (SMILEs) have become increasingly important to estimate the variability intrinsic to the climate system. A growing number of SMILEs are now available, reasonably sampling both model uncertainty and internal variability due to their ensemble size. SMILEs enabled substantial progress in understanding the Earth system. For instance, SMILEs were used to separate forced signals from internal variability to unprecedented precision (Maher et al., 2019), to quantify transient changes in the magnitude of climate variability (Olonscheck et al., 2021), and to evaluate how well climate models capture the variability and forced changes in the historical observational record (Suarez-Gutierrez et al., 2021). SMILEs are also used to identify systematic differences between simulated and observed patterns of sea-surface temperature and sea-level pressure change that are very unlikely to occur due to internal variability (Olonscheck et al., 2020; Wills et al., 2022). Furthermore, recent developments in compound event research highlight the importance of sufficiently sampling internal variability to robustly capture tail-risks in multivariate extremes, which requires even larger ensemble sizes than conventional univariate extremes (Bevacqua et al., 2023). The availability of SMILEs from multiple models further allows us to better quantify and differentiate sources of uncertainty in climate projections, especially uncertainties arising from internal variability and those from model differences (Deser et al., 2020; Lehner et al., 2020). These recent major advances in better understanding and quantifying climate variability and change show that SMILEs are increasingly useful tools for climate science.

The Max Planck Institute for Meteorology was one of the first modelling centres that produced a SMILE: the Max Planck Institute Grand Ensemble (MPI-GE, Maher et al. (2019)), which is still the largest SMILE available. MPI-GE – from here on called MPI-GE CMIP5 – is extremely successful and a powerful tool, but it is limited in various aspects: MPI-GE CMIP5 provides monthly model output with some daily output added later for one scenario only (e.g., Loughran et al., 2021; Raymond et al., 2022), it is run with CMIP5 forcing, and it provides three emission scenarios only. These limitations largely prevent the analysis of climate extremes across different emission scenarios because of the lack of high-frequency output, complicate direct comparisons of MPI-GE CMIP5 with SMILEs run with CMIP6 forcing, and restrict its usability for highly policy-relevant science. MPI-GE CMIP6 goes beyond these limitations by specifically enabling (1) the analysis of climate extremes, (2) comparisons to model versions with higher horizontal resolution, (3) comparisons to other SMILEs with CMIP6 forcing, and (4) investigation of low-emission scenarios with high policy relevance.

Several SMILEs with CMIP6 forcing have been recently run by a number of modelling centres, including ensembles with high-frequency model output. Next to MPI-GE CMIP6, currently available SMILEs with CMIP6 forcing and at least 30 realisations for both the historical and future period are ACCESS-ESM1.5 (Ziehn et al., 2020), CanESM5 (Swart et al., 2019), FGOALS (Lin et al., 2022), LENS2 (Rodgers et al., 2021), SMHI-LENS (Wyser et al., 2021), SPEAR-MED (Delworth et al., 2020), and MIROC6 (Tatebe et al., 2019). In comparison to the other CMIP6 SMILEs, MPI-GE CMIP6 provides the most extensive high-frequency output for the historical period and five different emission scenarios (Table 1). This includes the two highly policy-relevant scenarios SSP1-1.9 and SSP1-2.6 that are both otherwise only provided by CanESM5. In contrast to other SMILEs, MPI-GE CMIP6 has a climate sensitivity of 2.8°C which is close to the best estimate of 3°C of the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR6) (Forster et al., 2021). Furthermore, its predecessor MPI-GE CMIP5, based on a closely comparable model version, has shown to be one of the models that best represents the global and regional internal variability and forced response in annual observed temperatures (Suarez-Gutierrez et al., 2021) and precipitation (Wood et al., 2021). This good agreement with observations combined with the amount of high-frequency output for the full range of emission scenarios makes MPI-GE CMIP6 ideally suited for investigating future probabilities and magnitudes of climate extremes. The suitability of MPI-GE CMIP6 for studies on climate extremes is further enhanced by the possibility to compare the low-resolution ensemble to high-resolution ensembles or single simulations of the same model version that were run as part of the High Resolution Model Intercomparison Project (HighResMIP, Haarsma et al. (2016), compare Table 2). This unique combination of strengths makes MPI-GE CMIP6 a useful contribution to the CMIP6 multi-model ensemble and a powerful tool to investigate high-frequency climate variability and highly policy-relevant science questions.

Table 1: Characteristics of MPI-GE CMIP6 and other SMILEs with CMIP6 forcing and at least 30 realisations

SMILE name	Model version	Horizontal resolution	High-frequency output	Realisations	Time period	Scenarios	ECS
MPI-GE CMIP6	MPI-ESM1.2-LR	1.8°atm., 1.5°ocean	daily for all parameters, 3-hr, 6-hr for some (see Tables 2 and S1)	30	1850-2100	SSP1-1.9, 1-2.6, 2-4.5, 3-7.0, 5-8.5	2.80°C
ACCESS-ESM1.5	ACCESS-ESM1.5	1.88x1.25°atm.; 1.0°ocean	daily for many atm. parameters	40	1850-2100	SSP1-2.6, 2-4.5, 3-7.0, 5-8.5	3.87°C
CanESM5	CanESM5	2.8°atm., 1.0°ocean	daily for some atm. parameters	50	1850-2100	SSP1-1.9, 1-2.6, 2-4.5, 3-7.0, 5-8.5	5.62°C
FGOALS Super-large Ensemble	CAS FGOALS-g3	2.0°atm., 1.0°ocean	daily for many atm. parameters + tos, omldamax	110	1850-2100	SSP5-8.5	2.80°C
LENS2	CESM2	1.0°atm., 1.0°ocean	daily for all parameters, 3-hr, 6-hr for some	100	1850-2100	SSP3-7.0	5.16°C
SMHI-LENS	EC-Earth3.3.1	1.8°atm.; 1.0°ocean	daily for many atm. parameters	50	1970-2100	SSP1-1.9, 3-3.4, 5-3.4 -OS, 5-8.5	4.31°C
SPEAR-MED	GFDL AM4-LM4	0.5°atm., 1.0° (tropical refinement to 0.3°) ocean	daily for tas, tasmin, tasmax, pr, slp, uas, vas	30	1921-2100	SSP5-8.5	1.78°C
MIROC6	MIROC6	1.4°atm., 1.0°ocean	3-hr and daily for ta, tas, pr	50	1850-2100	SSP1-2.6, 2-4.5, 5-8.5	2.61°C

In this paper we present the new Max Planck Institute Grand Ensemble (MPI-GE CMIP6), and demonstrate its power beyond its predecessor ensemble MPI-GE CMIP5 (Maher et al., 2019) with six application examples. In section 2, MPI-GE CMIP6 is presented, evaluated with observations and reanalyses, and compared to MPI-GE CMIP5. In section 3, the power of MPI-GE CMIP6 is demonstrated with six application examples that specifically use the high-frequency model output for an improved understanding of climate extremes, the low-end emission scenarios for research on Paris Agreement global warming limits, and the medium ensemble size for an efficient combination of SMILEs with artificial intelligence. Section 4 summarises and concludes the paper.

2 MPI-GE CMIP6

2.1 Model description

MPI-GE CMIP6 is a 30-member ensemble simulated with the Max Planck Institute Earth System Model version 1.2 (MPI-ESM1.2, Mauritsen et al. (2019)), in the low resolution (LR) setup. In comparison to the MPI-GE CMIP5 simulations described in Maher et al. (2019), Mauritsen et al. (2019) summarises the updates that were introduced to MPI-ESM1.2, most importantly new radiation and aerosol parameterisations, and a nitrogen cycle for land biogeochemistry. Further, a major difference arises from the update of the external forcing from CMIP5 (Taylor et al., 2012) to CMIP6 (Eyring et al., 2016).

Table 2: Available simulations of MPI-ESM1.2 with different horizontal resolution. The MPI-ESM1.2-HR and -XR simulations were run as part of HighResMIP.

Model version	**Horizontal resolution**	**Realisations**	**Time period**	**Scenarios**
MPI-ESM1.2-LR	T63, 1.8°atm.; GR15, 1.5°ocean	30	1850-2100	SSP1-1.9, 1-2.6, 2-4.5, 3-7.0, 5-8.5
MPI-ESM1.2-HR	T127, 1.0°atm.; TP04, 0.4°ocean	10 (2)	1850-2100	SSP3-7.0 (SSP1-2.6, 2-4.5, 5-8.5)
MPI-ESM1.2-XR	T255, 0.5°atm.; TP04, 0.4°ocean	1	1950-2050	SSP5-8.5

MPI-GE CMIP6 is run with MPI-ESM version 1.2.01p7, with the atmosphere component ECHAM6 (Stevens et al. 2013, echam-6.3.05p2), which is directly coupled to the land component JSBACH (Reick et al. 2013, jsbach-3.20p1), and the ocean and sea-ice component MPIOM (Jungclaus et al. 2013, mpiom-1.6.3p4). MPIOM includes the ocean biogeochemistry module HAMOCC (Ilyina et al., 2013). The atmosphere/land and ocean components are coupled once a day by OASIS-MCT (Craig et al. (2017), oasis3mct-2.0). In MPI-ESM1.2-LR the atmosphere is resolved with spectral resolution T63 (equivalent to approx. 1.8° grid resolution) and 47 vertical levels, the ocean is resolved with a GR15 grid, nominal resolution 1.5°, at 40 vertical levels.

All simulations follow the CMIP6 protocol (Eyring et al., 2016) in terms of initialisation and historical and future external forcing (i.e. atmospheric composition, solar cycle, volcanic eruptions, land use). The 30-member ensemble of historical simulations covers the time period 1850-2014 and each member is initialised from a different state, approximately 25 years apart, of a quasi-stationary one-member 1000-year long preindustrial simulation. This macro initialisation from the preindustrial control state samples the full phase space of both the ocean and atmosphere states (Marotzke, 2019). Five scenario simulations (SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5, 30 realisations each) cover the time period 2015-2100, and in each scenario the realisations are directly initialised from their corresponding realisations of the historical ensemble.

2.2 Availability of high-frequency model output

In addition to standard CMIP6 monthly mean output, daily mean 3D fields of the state of atmosphere and ocean as well as selected daily mean 2D fields, i.e. for sea ice and land surface, are available for all simulations (Table S1 for details). Additionally, a number of atmospheric and land surface parameters are available on the 3-hourly time scale as listed in Table 3. Standard ocean biogeochemistry output from HAMMOC, 3D and 2D, is available on a monthly mean basis, with additional daily means for selected surface 2D or integrated 2D fields (see Table S1). Model output can be accessed via DKRZ's ESGF server at https://esgf-data.dkrz.de/search/cmip6-dkrz/.

2.3 Model evaluation and comparison to MPI-GE CMIP5

MPI-GE CMIP6 performs well in representing key climate quantities as derived from observations and reanalyses (Figure 1). The simulated range of global mean near-surface air temperature (GSAT) anomaly captures the interannual variability and the warming rate of HadCRUT5 well (Morice et al. (2021), Figure 1a). The projected ensemble mean GSAT warming at the end of the 21^{st} century relative to the 1985-2014 reference period ranges from 0.4K in SSP1-1.9 to 3.7K in SSP5-8.5.

Table 3: Parameters with 3-hourly and 6-hourly output on ESGF available for all 30 realisations. The parameters with daily output are listed in Table S1. A full list of parameters subdivided for members r1-r10 and r11-r30 is given in Tables S2-S4.

name	**parameter long name**	**unit**	**level**
3-hourly atmosphere / land			
mrro	Total Runoff	kg m-2 s-1	1
psl	Sea Level Pressure	Pa	1
sfcWind	Near-Surface Wind Speed	m s-1	1
tas	Near-Surface Air Temperature	K	1
uas	Eastward Near-Surface Wind	m s-1	1
vas	Northward Near-Surface Wind	m s-1	1
6-hourly atmosphere / land			
hurs	Near-Surface Relative Humidity	%	1
hus	Specific Humidity	1	47
huss	Near-Surface Specific Humidity	1	1
mrsol	Total Water Content of Soil Layer	kg m-2	5
mrsos	Moisture in Upper Portion of Soil Column	kg m-2	1
pr	Precipitation	kg m-2 s-1	1
ps	Surface Air Pressure	Pa	1
psl	Sea Level Pressure	Pa	1
ta	Air Temperature	K	47
tas	Near-Surface Air Temperature	K	1
tsl	Temperature of Soil	K	1
ua	Eastward Wind	m s-1	47
uas	Eastward Near-Surface Wind	m s-1	1
va	Northward Wind	m s-1	47
vas	Northward Near-Surface Wind	m s-1	1
wap	Omega (=dp/dt)	Pa s-1	4
zg	Geopotential Height	m	28
zg500	Geopotential Height at 500hPa	m	1

For global mean precipitation, MPI-GE CMIP6 underestimates both the magnitude and the interannual variability estimated from the ERA5 reanalysis (Figure 1b), as well as that of ERA-Interim (Figure S1). However, when comparing global mean precipitation in MPI-GE CMIP6 to the observational product of the Global Precipitation Climatology Project (GPCP, Adler et al. (2018)), we find that MPI-GE CMIP6 overestimates the observed global mean precipitation, but still shows too little interannual variability (Figure S1). The different estimates from observational and reanalyses products confirm previous findings that global mean precipitation products have large uncertainty of up to 40% (Bosilovich et al., 2016; Bock et al., 2020). Thus, MPI-GE CMIP6 is well within the range of observational uncertainty, but underestimates interannual variability. For the September Northern Hemisphere sea-ice area, the simulated range captures the observed evolution as derived from the sea-ice index (Fetterer et al. (2017), Figure 1c). September Northern Hemisphere sea-ice area is projected to shrink below the 1 million square kilometre threshold in the second half of the 21st century in SSP2-4.5, SSP3-7.0 and SSP5-8.5, but remains in both SSP1-1.9 and SSP1-2.6 until the end of the 21st century, similar to previous findings on sea-ice decline in CMIP6 (Notz & Community, 2020; Lee et al., 2021). The simulated range of the Atlantic meridional overturning circulation (AMOC) at 26° N is similar to the observed strength and interannual variability of the RAPID observations (Frajka-Williams et al. (2021), Figure 1d). However, the observations suggest that MPI-GE CMIP6 slightly overestimates the AMOC strength. The simulated range of the globally integrated CO_2 flux into the ocean and the net CO_2 flux into the land agrees well with the magnitude as reconstructed in the Global Carbon Project (Friedlingstein et al. (2022)), with simulated estimates of the globally integrated net CO_2 flux into the land exhibiting larger deviations from the mean state than those observed (Figure 1e-f). The evaluation of MPI-GE CMIP6 with observations and reanalyses shows that the ensemble realistically simulates both the long-term evolution and – except for precipitation – also the interannual variability of key climate quantities.

We further compare MPI-GE CMIP6 to MPI-GE CMIP5 with respect to the response of the key climate quantities to the various emission scenarios at the end of the 21st century. We find that MPI-GE CMIP6 shows slightly higher global-mean warming by the end of the 21st century than MPI-GE CMIP5 especially for the respective highest-emission scenarios (Figure 1a). In line with this, September Northern Hemisphere sea-ice area is projected to decline more in the respective SSP than RCP scenarios in the ensemble mean (Figure 1c). Similarly, the ensemble-mean decline in AMOC is substantially stronger in all SSP scenarios than in their respective RCP scenarios (Figure 1d). The globally integrated CO_2 flux into the ocean is larger in the mid and high-end SSP than in the respective RCP scenarios (Figure 1e). The projected change in net CO_2 flux into the land is largely uncertain, but shows a similar response at the end of the 21st century, except for SSP5-8.5 which shows a substantially stronger ensemble-mean increase than RCP8.5 (Figure 1f). In contrast to the stronger changes in MPI-GE CMIP6 compared to MPI-GE CMIP5, global mean precipitation is projected to increase less in the respective SSP than RCP scenarios (Figure 1b). From comparing the global mean temperature response of both model versions to a 1%CO_2 increase per year, i.e. the same forcing, we find a very similar warming rate and variability (Figure S2). This implies that the stronger changes in most quantities can be largely explained by the slightly stronger radiative forcing in the SSP compared to RCP scenarios, as has been shown for other models too (Wyser et al., 2020; Fyfe et al., 2021). We conclude that differences between MPI-GE CMIP6 and MPI-GE CMIP5 largely stem from the updated forcing in CMIP6 compared to CMIP5 rather than from differences in the model formulation.

3 Power of MPI-GE CMIP6 beyond MPI-GE CMIP5

MPI-GE CMIP5 (Maher et al., 2019) is extremely successful and a powerful tool to quantify climate variability and its change under global warming. However, the applicability of MPI-GE CMIP6 goes beyond MPI-GE CMIP5 in at least four critical aspects:

First, MPI-GE CMIP5 is run with CMIP5 forcing which limits direct comparisons to the large number of SMILEs that were run with CMIP6 forcing. MPI-GE CMIP6 provides the opportunity to compare MPI-ESM with other SMILEs run with CMIP6 forcing, and to investigate the impact of different forcings between MPI-GE CMIP5 and MPI-GE CMIP6.

Second, MPI-GE CMIP5 does not provide high-frequency model output across different emission scenarios, but only monthly mean output in most cases which strongly limits the usefulness for investigating short-lived climate extremes and their drivers (Suarez-Gutierrez et al., 2020a). In contrast, MPI-GE CMIP6 provides high-frequency output with 3-hourly and 6-hourly output for some variables (see Table 3) and daily output for all variables (see Table S1). This high-frequency output comes at the expense of a smaller ensemble size of 30 realisations instead of 100 realisations, but makes MPI-GE CMIP6 specifically suited for the analysis of climate extremes.

Third, MPI-GE CMIP6 can be compared to higher-resolution simulations of the same model version (see Table 2), for instance 10 realisations of MPI-ESM1.2-HR (1.0° atm., 0.4° ocean, Müller et al. (2018)) or a single realisation of MPI-ESM1.2-XR which provides also higher horizontal resolution in the atmosphere (0.5° atm., 0.4° ocean, Gutjahr et al. (2019)). This allows for the combination of high-frequency output in relatively low horizontal resolution of MPI-GE CMIP6 with high-resolution simulations, which is not possible with MPI-GE CMIP5.

Fourth, MPI-GE CMIP6 provides five instead of three emission scenarios. The five scenarios with 30 realisations each span the full range of IPCC scenarios from the low-emission scenario SSP1-1.9 to the high-emission scenario SSP5-8.5. With the scenarios SSP1-1.9 and SSP1-2.6, MPI-GE CMIP6 provides ensembles of two scenarios that were designed for projections of the Paris Agreement global warming limits of a 1.5°C and 2°C warmer world by the end of this century. This makes MPI-GE CMIP6 one of the few models that provide large ensembles for the two scenarios aligned with the Paris Agreement pledges, which allows for timely and highly policy-relevant science.

In the following, we exemplify the power of MPI-GE CMIP6 with six application examples. These examples include the analysis of heat, precipitation, wind, and ocean acidity extremes (Section 3.1), the probability of crossing Paris Agreement global warming limits (Section 3.2), and the potential of combining SMILEs with artificial intelligence methods for infilling observations (Section 3.3).

3.1 Analysing climate extremes

Climate extremes are among the most devastating and costly events, and their frequency and intensity is projected to increase with global warming (Seneviratne et al., 2021). However, climate models struggle to represent observed extremes because of large internal climate variability and their limited horizontal and temporal resolution (e.g., Slingo et al., 2022). Given the ensemble size and high-frequency output of MPI-GE CMIP6, we first investigate projected changes in heat and precipitation extremes and evaluate whether the new ensemble is capable of realistically simulating recently observed heat and precipitation extremes (Section 3.1.1). We then test whether observed precipitation extremes are better captured by model versions with higher horizontal resolution (Section 3.1.2). Finally, we investigate projected changes in marine heatwaves and ocean acidity extremes (Section 3.1.3) as well as in wind extremes (Section 3.1.4). For these analyses we choose a fixed baseline climatology over the time period 1985-2014.

3.1.1 Continental heat and precipitation extremes

We first evaluate whether MPI-GE CMIP6 is capable of simulating heat and precipitation extremes that were recently observed (Figure 2). We focus on the Siberian heatwave in spring 2020 (Ciavarella et al., 2021), the Pacific North American heatwave in summer 2021 (Philip et al., 2022), the extreme precipitation event in western Europe in summer 2021 (Ibebuchi, 2022; Tuel et al., 2022), and the extreme precipitation event in northern Italy in autumn 2020 (Davolio et al., 2023). To do so, we use daily surface maximum temperature and daily precipitation from MPI-GE CMIP6, and use ERA5 (Hersbach et al., 2020) and E-OBS (Klein Tank et al., 2002) as observational reference.

For continental heat extremes, we use the metric heat excess, which takes into account both heatwave intensity and persistence into one single metric (Perkins-Kirkpatrick & Lewis, 2020). To calculate heat excess, we identify heatwaves on a grid-point level when daily maximum near-surface air temperature exceeds the 90^{th} percentile based on a centred 15-day running window of the historical period 1985-2014 for at least three consecutive days. The cumulative heat is then calculated by seasonal integration of the exceeding heat above the threshold during heatwave days. In addition, we weight the cumulative heat of each grid point by the cosine of the latitude and spatially integrate it. For the 2020 Siberian heatwave we integrate the cumulative heat over boreal spring (MAM) and 40° N-80° N and 60° E-130° E. For the 2021 Pacific North American heatwave we integrate the cumulative heat over boreal summer (JJA) and 25° N-65° N and 90° W-130° W (see maps in Figure 2a,b). We scale the cumulative heat with respect to climatology (1985-2014). We compute the return periods for historical climate (1850-1879), the current climate (1992-2021) and the five SSP scenarios (SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-5.8; 2071-2100), and compare them to the two recent heatwaves in ERA5 (Figure 2a,b). The cumulative heat estimated by ERA5 in spring 2020 and summer 2021 integrated over the respective domains is 4.3 and 4.5.

These two record-shattering heat extremes led to devastating impacts. The Siberian heatwave was linked to large wildfires that causes a release of 56 megatons of CO_2 in June 2020, and to the melting of large permafrost areas which led to widespread infrastructure and environmental damages (Ciavarella et al., 2021). The Pacific North American heatwave also led to hundreds of attributable deaths, marine life mass-mortality events, reduced crop and fruit yields, river flooding from rapid snow and glacier melt, and a substantial increase in wildfires (White et al., 2023). In line with previous attribution studies (Ciavarella et al., 2021; Philip et al., 2022), we find that both heatwaves were virtually impossible in the preindustrial MPI-GE CMIP6 world, and have over 100-year return periods in current climate conditions. However, under the moderate emission scenario SSP2-4.5, heat excess levels as high as those during the 2020 Siberian heatwave could occur every four years (Figure 2a), and more than every other year for the 2021 Pacific North American heatwave (Figure 2b). In SSP5-8.5, MPI-GE CMIP6 projections show that a comparable 1-in-100-years event by the end of the 21$^{\text{st}}$ century reaches heat excess levels 5 to 8 times higher than the 2020 and 2021 levels, respectively. Only in the low emission scenarios SSP1-1.9 or SSP1-2.6 return periods below 10 years for such heat extremes can be avoided.

For precipitation extremes, we focus on two recently observed record-shattering events: the extreme precipitation event in western Europe on the 14$^{\text{th}}$ of July 2021, and the one in northern Italy on 2$^{\text{nd}}$ of October 2020. The extreme precipitation event in western Europe caused unprecedented flooding of the rivers Ahr and Erft. A rapid attribution study shows that observations over a larger region and different regional climate models give high confidence that human-induced climate change has increased the likelihood and intensity of events like the western European precipitation extreme (Kreienkamp et al., 2021; Ibebuchi, 2022), in line with the intensification of observed extreme precipitation in central Europe during the last century related to Northern Hemispheric warming (Zeder & Fischer, 2020). When integrated over 49° N-52° N and 5° E-8° E, the daily precipitation as observed by the E-OBS data set (Klein Tank et al., 2002) on 14$^{\text{th}}$ of July 2021 is 47.7 mm which represents the maximum daily precipitation in summer in the 72-year long observed record (see map in Figure 2c). The extreme precipitation event in northern Italy caused devastating large-scale flooding and represents an unprecedented strong event in a region that shows a high frequency of precipitation extremes (Davolio et al., 2023; Grazzini et al., 2021). The event was caused by a superposition of an upper-level trough over the western Mediterranean basin and moisture transport from the tropics by an atmospheric river (Davolio et al., 2023). When integrated over 43° N-47° N and 6° E-10° E, the daily precipitation observed by E-OBS on 2$^{\text{nd}}$ of October 2020 is 72.9 mm.

We use daily precipitation from MPI-GE CMIP6 and E-OBS, and compare the observed extreme precipitation events to the seasonal maximum daily precipitation simulated for the historical climate (1850-1879), the current climate (1992-2021), and the five SSP scenarios for the period 2071-2100. We find that MPI-GE CMIP6 does not simulate a summer and autumn daily precipitation event as intense as observed, not even until the end of the 21$^{\text{st}}$ century (Figure 2c). This implies that in any of the climate conditions simulated by MPI-GE CMIP6 an event as intense as the ones observed in 2020 and 2021 is virtually impossible, with return periods exceeding 900 years for all scenarios. We further find that simulated summer and autumn maximum daily precipitation is larger for higher emission scenarios than for lower scenarios in 2071-2100 and for the historical and current climate, in line with the fact that warmer air can hold more water leading to increased precipitation (e.g., Pendergrass et al., 2017; Myhre et al., 2019). However, the spread from the emission scenarios largely overlaps, suggesting that the uncertainty due to internal variability dominates scenario uncertainty and thus events typical for higher emission scenarios could also occur in a lower warming world due to internal variability. The results show that precipitation extremes as intense as the ones observed are not captured by MPI-GE CMIP6 possibly because the horizontal resolution of MPI-GE CMIP6 is too low to simulate real-world mechanisms leading to such small-scale precipitation extremes (Slingo et al., 2022). Given the increased probability of extremes that are unprecedented in the observed record and the often substantial impacts (Fischer et al., 2021), a realistic representation of such extreme events by climate models is highly needed.

3.1.2 Resolution dependence of representing precipitation extremes

Higher horizontal resolution of climate models improves the simulation of extreme precipitation because higher-resolution models reflect smaller spatial scales of extreme precipitation and key processes such as deep convection do not need to be parameterised (Wehner et al., 2014; Iles et al., 2020; Kendon et al., 2021; Kahraman et al., 2021). To test whether the inability of MPI-GE CMIP6 to represent the two observed precipitation extremes is caused by the model's coarse horizontal resolution, we investigate whether these events are better captured in higher-resolution versions of the same model, namely 10 realisations of MPI-ESM1.2-HR (Müller et al., 2018) with 1.0° atmospheric horizontal resolution, and a single realisation of MPI-ESM1.2-XR (Gutjahr et al., 2019) with 0.5° atmospheric horizontal resolution (see Table 2).

For the western European event, we find that MPI-ESM1.2-HR and MPI-ESM1.2-XR show higher agreement with the observed distribution of summer maximum daily precipitation over the period 1950-2021 than MPI-ESM1.2-LR, the low-resolution model version used for MPI-GE CMIP6 (Figure 3a,b). Strikingly, the single realisation of MPI-ESM1.2-XR simulates a single daily precipitation as intense as the one observed with a more widespread but still similar pattern (compare Figure S3), while MPI-ESM1.2-LR and MPI-ESM1.2-HR do not simulate such high daily precipitation amounts. Although the horizontal resolution of MPI-ESM1.2-XR is still not sufficient to resolve important processes such as moist convection (Hewitt et al., 2022; Slingo et al., 2022), our finding suggests that its resolution is sufficient to represent the recently observed regional precipitation extreme. Alternatively, MPI-ESM1.2-XR might overestimate the real-world precipitation intensity, which could also explain why the single simulation captures an event as intense as observed.

For autumn precipitation in northern Italy, we find that MPI-ESM1.2-HR much better represents the observed frequency of autumn maximum daily precipitation than MPI-ESM1.2-LR (Figure 3c,d). MPI-ESM1.2-XR shows generally too high autumn maximum precipitation, simulating precipitation amounts as large as observed with higher frequency. This is in line with previous findings that in the Mediterranean coastal region autumn precipitation intensity is larger at convection-permitting resolution than at coarse resolution because realistically representing deep convection is central for such events (Luu et al., 2020; Pichelli et al., 2021). The comparison between the western European and northern Italian events suggests that the model is able to simulate larger-scale autumn precipitation at coarser horizontal resolution than convective summer precipitation (Feldmann et al., 2008; Luu et al., 2020; Williams & O'Gorman, 2022). We conclude that while MPI-GE CMIP6 fails to simulate the observed precipitation extremes in western Europe and northern Italy, high-resolution simulations of the same model version are able to capture these extreme events, highlighting the potential for investigating regional precipitation extremes from comparing high-frequency model output of MPI-GE CMIP6 with simulations of higher horizontal resolution.

3.1.3 Marine heatwaves and ocean acidity extremes

We analyse daily mean sea surface temperature (SST) and hydrogen ion concentration ($[H^+]$) to identify marine heatwaves and ocean acidity extremes between 1850 and 2100 (Figure 4).

We use a percentile-based threshold and the reference period 1985-2014 for both extremes such that the probability of the occurrence of marine heatwaves and ocean acidity extremes in a year is the same. SST and $[H^+]$ are defined as extreme, if they exceed the 99th percentile for five consecutive days (Hobday et al., 2016; Burger et al., 2020). Although applying a duration criterion for ocean acidity extremes is not common, here it ensures comparability with marine heatwaves. The percentiles are calculated as the 20-member ensemble mean (only members 11 to 30 contain daily mean output for $[H+]$) over the 99th multiyear daily running percentile with a 5-day window length at every grid cell between 1985 and 2014. Finally, we calculate the number of extreme days per year to characterise changes of both extremes with time and across scenarios.

Before the reference period 1985-2014, almost no marine heatwaves are detected. Between 1985 and 2014, less than ten days per year are extreme with marine heatwaves being more frequent in the subpolar North Atlantic and the Southern Ocean (Figure 4a). By 2030, between five and 70 days per year are extreme with substantial overlap among different scenarios. By 2100, the SSP5-8.5 scenario projects the most marine heatwaves, with the entire ocean being in almost a constant state of extreme; while in the SSP1-1.9 scenario the number of extreme days per year does not exceed 15 by 2100 (Figure 4b, Figure S4). There is a much larger difference between the SSP1-1.9 and SSP5-8.5 scenarios in terms of global marine heatwave days at the end of the 21st century when compared to the difference in terms of global mean temperature between these scenarios (compare Figures 1a and 4b), indicating an amplified impact of global warming on marine heatwaves.

Over the historical period, globally, no ocean acidity extreme is detectable prior to the reference period. Within the reference period 1985-2014 (Figure 4e), the number of days with extreme $[H^+]$ increases to approximately five days per year in 2010 and continues to increase substantially to nearly 40 days per year in 2014. Locally, within the reference period, only very weak spatial gradients in the ensemble-mean number of ocean acidity extremes exist (Figure 4e). Until 2030, the entire ocean area moves rapidly to a near-permanent extreme state with more than 300 extreme days per year for all five future scenarios. By 2100, almost all days of a year show ocean acidity extremes in the SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios, while in the SSP1-2.6 scenario, the number of ocean acidity extreme days is projected to decline slightly by the end of the 21st century (Figure 4f, Figure S4). Within the SSP1-1.9 scenario, ocean acidity extremes are projected to peak at approximately 330 days per year between 2025-2040 and decline thereafter to 140 days per year by 2100. In this scenario, ocean acidity extremes occur less frequently in the Arctic Ocean and in the Southern Ocean compared to the Tropics between 2071-2100 (Figure 4g,h). There is a striking difference in the global occurrence of ocean acidity extremes between SSP1-1.9 and SSP1-2.6 in the second half of the 21st century (Figure 4f), despite only small differences in terms of global mean temperature in both scenarios (Figure 1a).

The CO_2 system in seawater and the mixing ratio of atmospheric CO_2 are tightly related, which leads to the smooth response in the mean surface ocean $[H^+]$. Sea surface temperature on the other hand is more variable across space and time than $[H^+]$, therefore the number of marine heatwaves varies more than the number of ocean acidity extremes across ensemble members. The number of detected extremes is sensitive to the definition, affected by the choice of threshold and reference period (Gruber et al., 2021). While using the same definition for both marine heatwaves and ocean acidity extremes is helpful to illustrate the different internal variability structure of the underlying parameters, understanding the governing processes may require a different extreme event definition that would ultimately lead to a different number of detected events.

3.1.4 Wind extremes

Future changes in wind extremes are among the most uncertain impacts of anthropogenic climate change (Seneviratne et al., 2021). We use the 3-hourly output of MPI-GE CMIP6 to project global changes in wind extremes and their dependence on the emission scenario (Figure 5a and Figure S5). To detect projected global changes in wind speed, we first derive 95$^{\text{th}}$ annual percentiles of near-surface wind speeds for each grid point from the entire 30-member ensemble and then calculate the absolute difference between the 2071-2100 mean and the 1985-2014 reference mean. Here, we focus on SSP5-8.5 because the projected changes are most distinct: Over the ocean, we find a latitudinal contrasting pattern with increasing wind extremes over high-latitude oceans and decreasing wind extremes in most mid- and low-latitude ocean basins. Over land, increases in wind extremes are projected for South America, Western and Eastern Africa and parts of the Northern mid- to high-latitudes, whereas substantial decreases are projected for Alaska, Siberia, Central Asia and the Western Sahara. Weaker changes but with the same pattern are found for lower-emission scenarios (Figure S5).

We further analyse projected changes in storm activity in two regions that are known for the frequent passage of mature hurricanes and typhoons with often devastating impacts when they make landfall: north-west of Bermuda in the North Atlantic (Figure 5b) and south-east of Japan in the North Pacific (Figure 5c). For both regions, we select three grid points that form a triangle spanning the area of interest (Table S5). We then use 3-hourly mean sea-level pressure data from MPI-GE CMIP6 at the selected grid points and derive geostrophic winds v_g from the horizontal mean sea-level pressure gradients $\partial p/\partial x$ and $\partial p/\partial y$ according to Krieger et al. (2020) via

$$v_g = \left(v_x^2 + v_y^2\right)^{1/2}, \tag{1}$$

with

$$v_x = -\frac{1}{\rho f}\frac{\partial p}{\partial y} \qquad \text{and} \qquad v_y = \frac{1}{\rho f}\frac{\partial p}{\partial x}, \tag{2}$$

where ρ is the density of air (set at 1.25 kg m^{-3}) and f the average of the Coriolis parameter at the three corners of the triangle. We chose the grid points so that the resulting triangle is sufficiently close to an equilateral triangle. This requirement is necessary to avoid a large error propagation of pressure uncertainties, which would cause a shift of the wind direction towards the main axis of the triangle (Krieger et al., 2020). We then define storm activity as the standardised annual 95$^{\text{th}}$ percentiles of 3-hourly geostrophic wind speeds. We therefore first calculate annual 95$^{\text{th}}$ percentiles of geostrophic winds for each ensemble member. We then standardise by subtracting the 1985-2014 ensemble mean from each ensemble member, and divide by the 1985-2014 ensemble standard deviation.

For both north-west of Bermuda and south-east of Japan, we find a decreasing storm activity with strongest decreases for high-emission scenarios, while we find no notable change in scenario SSP1-1.9 (Figure 5b,c and Figure S5). This agrees with the projected change in surface wind speed, where the marine subtropics around 30° N show a strong signal of decreasing wind speeds in the SSP5-8.5 scenario (Figure 5a).

We further calculate the ensemble balance to characterise whether changes in the ensemble mean are caused by a shift in the majority of the ensemble members or by a few strong outliers. To do so, we first apply a moving Gaussian low-pass filter to the storm activity time series of each ensemble member. We then define thresholds for high and low activity periods at $0.5\,\sigma$ and $-0.5\,\sigma$, and count for how many members the low-pass filtered curve exceeds these thresholds in a certain year. The difference in the number of high-activity and low-activity members is then regarded as the ensemble balance (crosses on the secondary y-axis in Figure 5b,c). In the SSP1-1.9 and SSP1-2.6 scenarios, we find that the ensemble balance does not significantly deviate from 0 towards the end of the 21$^{\text{st}}$ century in both focus regions, confirming the rather small projected change in storm activity. In the high-emission SSP5-8.5 scenario, the ensemble balance falls to near -30 at the end of the 21$^{\text{st}}$ century, which indicates that nearly all ensemble members agree on a decline in storm activity both north-west of Bermuda and south-east of Japan.

The proxy for storm activity is based on the hypothetical geostrophic wind and its long-term statistics, as proposed originally by Schmidt and von Storch (1993). For high latitudes, where the synoptic-scale wind in higher altitudes is close to geostrophic, it has been shown that the statistics of the geostrophic wind closely resemble the statistics of the near-surface wind (Krueger & von Storch, 2011). In latitudes closer to the equator this assumption does not hold, as most of the wind extremes occur in or near tropical cyclones, which are not fully in geostrophic balance. The proxy should therefore not be used as a single tool to make conclusions about future changes in the intensity or frequency of tropical cyclones. However, the decreasing storm activity for mid-latitude hurricanes and typhoons is in line with recent findings of a decreasing frequency of tropical cyclones (Chand et al., 2022). As the proxy only describes storm activity with one quantity, it cannot distinguish between changes in the frequency and changes in the intensity of storms. A change in storm activity can thus be interpreted as a change in either number or intensity of cyclones, or a combined change thereof. Also, changes connected to smaller-scale features such as fronts or convective wind gusts within cyclones cannot be detected by the proxy, as the derived geostrophic wind acts as an area mean over the entire triangle.

Overall, MPI-GE CMIP6 projects increasing wind extremes over high-latitude oceans and decreasing wind extremes in most mid- and low-latitude oceans, in line with current understanding of observed changes in wind extremes caused by a poleward shift of extra-tropical storm tracks over both hemispheres (Seneviratne et al., 2021). We conclude that MPI-GE CMIP6 with its 3-hourly model output is a powerful tool to understand changes in the frequency and intensity of wind extremes for different emission scenarios.

3.2 Investigating crossing probabilities of 1.5°C and 2°C global warming

The Paris Agreement in 2015 states the goal to keep global warming well below 2°C, and to pursue efforts to limit global warming to 1.5°C above preindustrial levels to avoid devastating and unmanageable consequences of climate change. MPI-GE CMIP6 is suited to investigate the uncertainty in crossing these global warming limits because one can account for internal climate variability with ensemble simulations for five different emission scenarios, including the scenarios SSP1-1.9 and SSP1-2.6 that project a global warming of 1.5°C and 2°C, respectively.

To investigate the crossing probability of 1.5°C and 2°C of global warming in MPI-GE CMIP6, we use annual mean, global mean near-surface air temperature (GSAT) to compute for every year and each of the five scenarios the fraction of realisations (x / 30 realisations) that crosses these temperature thresholds in a single year relative to the 1850-1900 reference period (Figure 6a,b). We find that in all emission scenarios, there is a non-zero chance of observing individual years above 1.5°C within the next decades, including the SSP1-1.9 scenario that represents the strongest mitigation efforts. However, this finding does not imply that every scenario crosses the Paris agreement 1.5°C global warming limit because whether a temperature threshold will be crossed or not is commonly evaluated for 20-year mean temperatures (Lee et al., 2021). To account for this definition, we also compute the 20-year running mean GSAT time series for each realisation and show for each 20-year window the fraction of realisations that crosses 1.5°C or 2°C (Figure 6c,d). We find that MPI-GE CMIP6 with the SSP1-1.9 scenario is consistent with the 1.5°C warming limit, whereas all other scenarios cross this threshold. We stress that when 1.5°C are crossed for 20-year means is still affected by internal variability: for SSP1-2.6, 1.5°C may be crossed around the 20-year mean of the period starting in 2030, but only 10 years later it is virtually certain that 1.5°C is crossed in the 20-year mean of any realisation. Further, the SSP1-1.9 and SSP1-2.6 scenarios will not cross 2°C neither in single years nor for 20-year means while all other scenarios will cross this threshold between 20-year means starting in 2035 to 2050. These estimates are at the upper range of the IPCC AR6 central estimate of crossing the 1.5°C threshold which lies in the early 2030s for all scenarios except SSP5-8.5 (Marotzke et al., 2022; Lee et al., 2021).

We note that the IPCC AR6 uncertainty range includes uncertainties in historical warming, climate sensitivity and internal variability (Lee et al., 2021), whereas MPI-GE CMIP6 has a fixed climate sensitivity and the uncertainty range is only due to internal variability. However, the observed internal variability in GSAT is well simulated by the model (Suarez-Gutierrez et al., 2021) and its equilibrium climate sensitivity of 2.8°C is close to the central estimate of the IPCC AR6 assessment of 3°C. Comparing the central estimates of crossing times for 1.5°C between MPI-GE CMIP6 and the IPCC AR6 assessment shows that the MPI-GE CMIP6 estimates are systematically later than in AR6 (Table S6). Most notably, SSP1-1.9 does not cross 1.5°C in the model, the crossing in SSP1-2.6 occurs a decade later, and the crossing in all other scenarios about five years later than in IPCC AR6. This shows that the MPI-GE CMIP6 estimates are broadly consistent with but slightly more conservative than the IPCC AR6 assessment.

We conclude that with its good representation of internal variability in GSAT and its equilibrium climate sensitivity close to the central estimate of the IPCC AR6 assessment, MPI-GE CMIP6 offers a unique framework to investigate timing and local impacts of crossing temperature thresholds such as 1.5°C.

3.3 Combining SMILEs and artificial intelligence

SMILEs and artificial intelligence can be combined powerfully because the multiple realisations of a same model provide testing, validation and training data sets to infill gaps in observational data. We provide one example by using a method that is based on an inpainting technique developed by Liu et al. (2018) to repair corrupted images. It makes use of a U-Net neural network made of partial convolutional layers and a state-of-the-art loss function designed to produce semantically meaningful predictions. As shown in Kadow et al. (2020), the method can infill large and irregular regions of missing climate data and is able to reconstruct specific climate patterns that are not captured by standard interpolation techniques such as the Kriging method (Cowtan & Way, 2014).

We here test whether the ensemble size of MPI-GE CMIP6 is sufficiently large to be used for infilling the HadCRUT5 data set with similar capability than the 100-member MPI-GE CMIP5. The models used to infill the HadCRUT5 data set (Dunn et al., 2020) have been trained using gridded global historical surface temperature anomalies from three large ensembles: 1) MPI-GE CMIP6, containing 30 realisations and spanning the 1850-2014 time period; 2) MPI-GE CMIP5, containing 100 realisations and spanning the 1850-2005 time period; and 3) a subset of MPI-GE CMIP5 containing the first 30 ensemble members, here called MPI-GE CMIP5(30). Before the training, one ensemble member was excluded from each ensemble to create three testing data sets. Three validation data sets were created from the remaining ensemble members of each data set by pulling out the data every 8 timesteps for MPI-GE CMIP6 and MPI-GE CMIP5(30), and every 7 timesteps for MPI-GE CMIP5. The remaining data were used to create the training data sets which contain 50.242 samples for MPI-GE CMIP6, 47.502 samples for MPI-GE CMIP5(30) and 162.162 samples for MPI-GE CMIP5. For this work, additional features have been implemented to the original version of the code (Kadow et al., 2020) to improve the computational performance and the quality of the reconstruction. In particular, a custom padding operation accounting for the boundary conditions of the global data is now applied before each partial convolution, to account for the sphere of the Earth.

The annual global mean temperature time series reconstructed using the 100 member and the 30 member models are very similar, especially when compared to the original HadCRUT5 data (Figure 7). For all three ensembles, we detect an overall warming signal also on a regional scale around the globe by comparing the climatologies 2020-1991 and 1920-1891 with a century apart (insets in Figure 7 and Figure S6). In particular, the warming patterns reconstructed from the three ensembles show a strong century warming signal in northern polar regions, where the original HadCRUT5 data set has missing data. Large areas in the Pacific also consistently show a warming between the two climatologies, despite the fact that the region is affected by strong ENSO variability. The infilled data in the sparsely observed Antarctica show a less strong, but more mixed warming signal as observed when reconstructed with the different ensembles. From the striking similarity in the reconstructed pattern, we conclude that MPI-GE CMIP6 allowed us to train a model with equivalent capabilities to MPI-GE CMIP5 but at a lower computational cost.

4 Summary and Conclusions

MPI-GE CMIP6 is a new 30-member single-model initial-condition large ensemble which power goes beyond its predecessor MPI-GE CMIP5 (Maher et al., 2019) in several aspects and allows for novel analyses with broad societal relevance:

First, MPI-GE CMIP6 provides 3-hourly, 6-hourly and daily model output that is together with its ensemble size well suited to investigate present and future changes in climate extremes, their drivers, and their changing characteristics across different emission scenarios. While several studies used MPI-GE CMIP5 to study present and future changes in climate extremes (e.g., Suarez-Gutierrez et al., 2020a, 2020b; Landrum & Holland, 2020), the high-frequency output of MPI-GE CMIP6 now allows one to also investigate the drivers and causal links of these changes which can be compared across different emission scenarios. For instance, we find from daily output that the recently observed Siberian and Pacific North American heatwaves will occur every year in 2071-2100 in high-emission scenarios but substantially less frequent in the low-emission scenarios. We further find from the 3-hourly output that the frequency of wind extremes is projected to decrease in tropical to mid-latitude oceans in all five emission scenarios. These findings illustrate that MPI-GE CMIP6 is specifically suited to investigate climate extremes and can be used to study high-impact events.

Second, MPI-GE CMIP6 provides the opportunity to compare the ensemble to high-resolution simulations of the same model version, including a 10-member ensemble of MPI-ESM-HR (1.0° atmosphere, 0.4° ocean), and a single member of MPI-ESM-XR (0.5° atmosphere, 0.4° ocean). While MPI-GE CMIP6 is not able to represent the unprecedented precipitation extreme in western Europe observed on 14$^{\text{th}}$ of July 2021 and in northern Italy observed on 2$^{\text{nd}}$ of October 2020, we find that these events are captured by high-resolution simulations of the same model version. This finding illustrates the benefit of comparing low-resolution SMILEs with high-frequency output to high-resolution simulations of the same model version for investigating regional climate extremes.

Third, MPI-GE CMIP6 provides historical simulations and the five emission scenarios SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5 which enable the investigation of different climate futures and the quantification of uncertainty from internal variability. We find that the frequencies of marine heatwaves and ocean acidity extremes are projected to substantially increase in all emissions scenarios, with substantial recovery by 2100 only under SSP1-1.9. Moreover, the ensemble simulations of the scenarios SSP1-1.9 and SSP1-2.6 specifically allow for quantifying irreducible uncertainty when aiming to limit global mean warming to 1.5°C or 2°C. We find that in MPI-GE CMIP6, even for the lowest emission scenario SSP1-1.9, which is consistent with the Paris Agreement pledges in this model, there is a non-zero chance to observe individual years above 1.5°C. With its good representation of internal variability in GSAT and its equilibrium climate sensitivity close to the central estimate of the AR6 assessment, MPI-GE CMIP6 as a single-model ensemble provides new opportunities to quantify uncertainty in when global warming thresholds might be crossed. Such analyses on irreducible uncertainty from internal variability are highly relevant for investigating transition pathways to carbon-neutral economies to meet the Paris Agreement pledges.

Fourth, MPI-GE CMIP6 is run with CMIP6 forcing and provides the opportunity to compare the ensemble to other SMILEs with CMIP6 forcing. This facilitates comparisons to the growing number of SMILEs. From comparing the respective scenarios from MPI-GE CMIP6 to the ones from its predecessor MPI-GE CMIP5, we find that the change from CMIP5 to CMIP6 forcing causes a slightly stronger climate response, in line with findings from other SMILEs (Wyser et al., 2020; Fyfe et al., 2021), primarily caused by the updated forcing in CMIP6. From combining MPI-GE CMIP6 with artificial intelligence, we find that 30 realisations have equivalent capabilities as the 100-member MPI-GE CMIP5 when training a model to infill surface temperature observations.

Overall, MPI-GE CMIP6 beneficially complements the number of available SMILEs by a unique combination of a moderate ensemble size, high-frequency model output, the full range of emission scenarios including the lower end, and the availability of high-resolution simulations of the same model version. Consequently, MPI-GE CMIP6 allows a better understanding of changes in climate variability and extremes, and to quantify related uncertainties. This improved quantification will help to better inform society on the likelihood of plausible changes in the climate system to occur, including climate extremes.

Acknowledgments

This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project IDs uo1075 and mh0469. We thank Nicola Maher for conducting a pre-submission review. This research was funded by the Max Planck Society (MPG) and the German Research Foundation (DFG) under Germany's Excellence Strategy—EXC 2037 'CLICCS—Climate, Climatic Change, and Society'—Project No.: 390683824, contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg, and the Cluster of Excellence CliSAP (EXC177), Universität Hamburg. D.O. received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 820829 (CONSTRAIN). S.B. and J.B. were supported by Copernicus Climate Change Service, funded by the EU, under Contract C3S2-370. G.B.-A. is supported by the German Ministry of Education and Research (BMBF) under the ClimXtreme project NA2EE (grant no. 01LP1902F). F.F., T.I., and H.L. received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 820989 (COMFORT). D.K. received funding from the project WAKOS – Wasser an den Küsten Ostfrieslands, which is funded by the German Federal Ministry of Education and Research (BMBF; grant no. 01LR2003A). L.S.-G. received funding from the German Ministry of Education and Research (BMBF) under the ClimXtreme project DecHeat (grant no. 01LP1901F) and from the European Union's Horizon Europe Framework Programme under the Marie Skłodowska-Curie grant agreement no. 101064940. The applied data standardisation tools and workflow have been developed within the BMBF funded project DICAD (grant no. 01LP1605A). E.P. received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 101003876 (CLINT). Post-processing of data was performed with CDO (Schulzweida, 2022).

Open Research

The MPI-ESM1.2-LR coupled climate model is distributed via http://www.mpimet.mpg.de/. The simulation run scripts and code for reproducing the plots will be openly available through the publication repository of the Max Planck Society.

References

Adler, R., Sapiano, M., Huffman, G., Wang, J.-J., Gu, G., Bolvin, D., … Shin, D.-B. (2018). The Global Precipitation Climatology Project (GPCP) Monthly

Analysis (New Version 2.3) and a Review of 2017 Global Precipitation. *Atmosphere*, *9*(4), 2073-4433. doi: 10.3390/atmos9040138

Bevacqua, L., E.and Suarez-Gutierrez, Jézéquel, A., Lehner, F., Vrac, M., Yiou, P., & J., Z. (2023). Advancing our understanding of compound weather and climate events via large ensemble model simulations. *Nature Communications*, *14*(2145). doi: 10.1038/s41467-023-37847-5

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., ... Eyring, V. (2020). Quantifying progress across different cmip phases with the esmvaltool. *Journal of Geophysical Research: Atmospheres*, *125*(21), e2019JD032321. doi: 10.1029/2019JD032321

Bosilovich, M. G., Lucchesi, R., & Suarez, M. (2016). MERRA-2: File Specification. GMAO Office Note No. 9 (Version 1.1). , 73 pp.

Burger, F., John, J., & Frölicher, T. (2020). Increase in ocean acidity variability and extremes under increasing atmospheric CO2. *Biogeosciences*, *17*(18), 4633-4662. doi: 10.5194/bg-17-4633-2020

Chand, S. S., Walsh, K. J. E., Camargo, S. J., Kossin, J. P., Tory, K. J., Wehner, M. F., ... Murakami, H. (2022). Declining tropical cyclone frequency under global warming. *Nature Climate Change*, *12*(7), 655–661. doi: 10.1038/s41558-022-01388-4

Ciavarella, A., Cotterill, D., Stott, P., Kew, S., Philip, S., van Oldenborgh, G., ... Zolina, O. (2021). Prolonged siberian heat of 2020 almost impossible without human influence. *Climatic Change*, *166*(9), 1689–1713. doi: 10.1007/s10584-021-03052-w

Cowtan, K., & Way, R. G. (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1935–1944. doi: 10.1002/qj.2297

Craig, A., Valcke, S., & Coquart, L. (2017). Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0. *Geosci Model Dev*, *10*(9), 3297–3308. doi: 10.5194/gmd-10-3297-2017

Davolio, S., Vercellino, M., Miglietta, M.-M., Drago Pitura, L., Laviola, S., & Levizzani, V. (2023). The influence of an atmospheric river on a heavy precipitation event over the western Alps. *Weather and Climate Extremes*, *39*, 100542. doi: 10.1016/j.wace.2022.100542

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., ... Zhao, M. (2020). SPEAR: The Next Generation GFDL Modeling System for Seasonal to Multidecadal Prediction and Projection. *Journal of Advances in Modeling Earth Systems*, *12*(3), e2019MS001895. doi: 10.1029/2019MS001895

Deser, C., Lehner, F., & Rodgers, K. e. a. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Chang.*, *10*, 277–286. doi: 10.1038/s41558-020-0731-2

Dunn, R. J. H., Alexander, L. V., Donat, M. G., Zhang, X., Bador, M., Herold, N., ... Bin Hj Yussof, M. N. (2020). Development of an Updated Global Land In Situ-Based Data Set of Temperature and Precipitation Extremes: HadEX3. *Journal of Geophysical Research: Atmospheres*, *125*(16), e2019JD032263. doi: 10.1029/2019JD032263

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

Feldmann, H., Früh, B., Schädler, G., Panitz, H.-J., Keuler, K., Jacob, D., & Lorenz, P. (2008). Evaluation of the precipitation for South-western Germany from high resolution simulations with regional climate models. *Meteorologische Zeitschrift*, *17*(4), 455–465. doi: 10.1127/0941-2948/2008/0295

Fetterer, F., Knowles, K., Meier, W., Savoie, M., & Windnagel, A. (2017). *Sea* *ice index, version 3.* National Snow and Ice Data Center. Retrieved from https://nsidc.org/data/G02135/versions/3 doi: 10.7265/N5K072F8

Fischer, E., Sippel, S., & Knutti, R. (2021). Increasing probability of record-shattering climate extremes. *Nat. Clim. Chang.*, *11*, 689–695. doi: 10.1038/s41558-021-01092-9

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., … Zhang, H. (2021). The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. , 923–1054. doi: 10.1017/9781009157896.009

Frajka-Williams, E., Moat, B., Smeed, D., Rayner, D., Johns, W., Baringer, M., … Collins, J. (2021). Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2020 (v2020.1). NERC EDS British Oceanographic Data Centre NOC. doi: 10.5285/cc1e34b3-3385-662b-e053-6c86abc03444

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C., Hauck, J., … others (2022). Global carbon budget 2021. *Earth System Science Data*, *14*(4), 1917–2005.

Fyfe, J. C., Kharin, V. V., Santer, B. D., Cole, J. N. S., & Gillett, N. P. (2021). Significant impact of forcing uncertainty in a large ensemble of climate model simulations. *Proceedings of the National Academy of Sciences*, *118*(23), e2016549118. doi: 10.1073/pnas.2016549118

Global Carbon Project. (2021). *Supplemental data of Global Carbon Budget 2021 (Version 1.0) [Data set]. Global Carbon Project.* doi: 10.18160/gcp-2021

Grazzini, F., Fragkoulidis, G., Teubler, F., Wirth, V., & Craig, G. C. (2021). Extreme precipitation events over northern Italy. Part II: Dynamical precursors. *Quarterly Journal of the Royal Meteorological Society*, *147*(735), 1237-1257. doi: 10.1002/qj.3969

Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. *Nature*, *600*(7889), 395–407. doi: 10.1038/s41586-021-03981-7

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., … Stössel, A. (2019). Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP). *Geoscientific Model Development*, *12*(7), 3241–3281. doi: 10.5194/gmd-12-3241-2019

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., … von Storch, J.-S. (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, *9*(11), 4185–4208. doi: 10.5194/gmd-9-4185-2016

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: 10.1002/qj.3803

Hewitt, H., Fox-Kemper, B., & Pearson, B. e. a. (2022). The small scales of the ocean may hold the key to surprises. *Nat. Clim. Chang.*, *12*, 496–499. doi: 10.1038/s41558-022-01386-6

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C., … Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. *Prog Oceanogr*, *141*, 227-238. doi: 10.1016/j.pocean.2015.12.014

Ibebuchi, C. C. (2022). Patterns of atmospheric circulation in Western Europe linked to heavy rainfall in Germany: preliminary analysis into the 2021 heavy rainfall episode. *Theoretical and Applied Climatology*, *148*(1), 269–283. doi: 10.1007/s00704-022-03945-5

Iles, C. E., Vautard, R., Strachan, J., Joussaume, S., Eggen, B. R., & Hewitt, C. D. (2020). The benefits of increasing resolution in global and regional climate simulations for European climate extremes. *Geoscientific Model Development*, *13*(11), 5583–5607. doi: 10.5194/gmd-13-5583-2020

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations. *J. Adv. Mod. Earth Sys.*, *5*(2), 287-315. doi: 10.1029/2012MS000178

Kadow, C., Hall, D. M., & Ulbrich, U. (2020). Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, *13*(6), 408–413. doi: https://doi.org/10.1038/s41561-020-0582-5

Kahraman, A., Kendon, E. J., Chan, S. C., & Fowler, H. J. (2021). Quasi-Stationary Intense Rainstorms Spread Across Europe Under Climate Change. *Geophysical Research Letters*, *48*(13), e2020GL092361. doi: 10.1029/2020GL092361

Kendon, E. J., Prein, A. F., Senior, C. A., & Stirling, A. (2021). Challenges and outlook for convection-permitting climate modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2195), 20190547. doi: 10.1098/rsta.2019.0547

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G., Gocheva, A., ... Petrovic, P. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology*, *22*(12), 1441-1453. doi: 10.1002/joc.773

Kreienkamp, F., Philip, S. Y., Tradowsky, J. S., Kew, S. F., et al., & Wanders, N. (2021). *Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021* (Tech. Rep.). World Weather Attribution. (Published online August 23, 2021)

Krieger, D., Krueger, O., Feser, F., Weisse, R., Tinz, B., & Storch, H. (2020). German Bight storm activity, 1897–2018. *International Journal of Climatology*. doi: 10.1002/joc.6837

Krueger, O., & von Storch, H. (2011). Evaluation of an air pressure–based proxy for storm activity. *Journal of Climate*, *24*(10), 2612 - 2619. doi: 10.1175/2011JCLI3913.1

Landrum, L., & Holland, M. (2020). Extremes become routine in an emerging new Arctic. *Nat. Clim. Chang.*(10), 1108—1115. doi: 10.1038/s41558-020-0892-z

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., ... Zhou, T. (2021). Future Global Climate: Scenario-Based Projections and Near-Term Information. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. , 553–672. doi: 10.1017/9781009157896.006

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., ... Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics*, *11*(2), 491–508. doi: 10.5194/esd-11-491-2020

Lin, P., Zhao, B., Wei, J., Liu, H., Zhang, W., Chen, X., ... Zhou, T. (2022). The Super-large Ensemble Experiments of CAS FGOALS-g3. *Advances in Atmo-*

spheric Sciences. doi: https://doi.org/10.1007/s00376-022-1439-1

Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., & Catanzaro, B. (2018). Image Inpainting for Irregular Holes Using Partial Convolutions. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (pp. 89–105). Cham: Springer International Publishing.

Loughran, T. F., Boysen, L., Bastos, A., Hartung, K., Havermann, F., Li, H., ... Pongratz, J. (2021). Past and future climate variability uncertainties in the global carbon budget using the mpi grand ensemble. *Global Biogeochemical Cycles*, *35*(8), e2021GB007019. doi: 10.1029/2021GB007019

Luu, L. N., Vautard, R., Yiou, P., & Soubeyroux, J.-M. (2020). Evaluation of convection-permitting extreme precipitation simulations for the south of France. *Earth System Dynamics*, *13*(1), 687–702. doi: 10.5194/esd-13-687-2022

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., ... Marotzke, J. (2019). The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability. *J. Adv. Model. Earth Syst.*, *11*(7), 2050-2069. doi: 10.1029/2019MS001639

Marotzke, J. (2019). Quantifying the irreducible uncertainty in near-term climate projections. *WIREs Climate Change*, *10*(1), e563. doi: doi.org/10.1002/wcc.563

Marotzke, J., Milinski, S., & Jones, C. (2022). How close are we to $1.5°C$ or $2°C$ of global warming? *Weather*, *77*(4), 147-148. doi: 10.1002/wea.4174

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., ... Roeckner, E. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998-1038. doi: 10.1029/2018MS001400

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., ... Simpson, I. R. (2021). An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. *Journal of Geophysical Research: Atmospheres*, *126*(3). doi: 10.1029/2019JD032361

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., ... Marotzke, J. (2018). A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383–1413. doi: 10.1029/2017MS001217

Myhre, G., Alterskjær, K., Stjern, C. W., Hodnebrog, Ø., Marelle, L., Samset, B. H., ... Stohl, A. (2019). Frequency of extreme precipitation increases extensively with event rareness under global warming. *Scientific Reports*, *9*(1), 16063. doi: 10.1038/s41598-019-52277-4

Notz, D., & Community, S. (2020). Arctic Sea Ice in CMIP6. *Geophysical Research Letters*, *47*(10), e2019GL086749. doi: 10.1029/2019GL086749

Olonscheck, D., Rugenstein, M., & Marotzke, J. (2020). Broad consistency between observed and simulated trends in sea surface temperature patterns. *Geophysical Research Letters*, *47*(10), e2019GL086773. doi: 10.1029/2019GL086773

Olonscheck, D., Schurer, A., Lücke, L., & Hegerl, G. (2021). Large-scale emergence of regional changes in year-to-year temperature variability by the end of the 21st century. *Nat. Commun.*, *12*, 7237. doi: 10.1038/s41467-021-27515-x

Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C., & Sanderson, B. M. (2017). Precipitation variability increases in a warmer climate. *Scientific Reports*, *7*(1), 17966. doi: 10.1038/s41598-017-17966-y

Perkins-Kirkpatrick, S., & Lewis, S. (2020). Increasing trends in regional heatwaves. *Nature Communications*, *11*(1), 3357. doi: 10.1038/s41467-020-16970-7

Philip, S. Y., Kew, S. F., van Oldenborgh, G. J., Anslow, F. S., Seneviratne, S. I., Vautard, R., ... Otto, F. E. L. (2022). Rapid attribution analysis of the extraordinary heat wave on the pacific coast of the us and canada in june 2021. *Earth System Dynamics*, *13*(4), 1689–1713. doi: 10.5194/esd-13-1689-2022

Pichelli, E., Coppola, E., & Sobolowski, S. e. a. (2021). The first multi-model ensemble of regional climate simulations at kilometer-scale resolution part 2: historical and future simulations of precipitation. *Climate Dynamics*, *56*, 3581—3602. doi: 10.1007/s00382-021-05657-4

Raymond, C., Suarez-Gutierrez, S., Kornhuber, K., Pascolini-Campbell, M., Sillmann, J., & Waliser, D. (2022). Increasing spatiotemporal proximity of heat and precipitation extremes in a warming world quantified by a large model ensemble. *Environmental Research Letters*, *17*(3), 035005. doi: 10.1088/1748-9326/ac5712

Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., ... Yeager, S. G. (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics*, *12*(4), 1393–1411. doi: 10.5194/esd-12-1393-2021

Schmidt, H., & von Storch, H. (1993). German Bight storms analysed. *Nature*, *365*(6449), 791. doi: 10.1038/365791a0

Schulzweida, U. (2022). CDO User Guide (2.1.0). *Zenodo*. doi: 10.5281/zenodo.7112925

Seneviratne, S., X. Zhang, M. A., Badi, W., Dereczynski, C., Luca, A. D., Ghosh, S., ... Zhou, B. (2021). Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. , 1513–1766. doi: 10.1017/9781009157896.013

Slingo, J., Bates, P., Bauer, P., Belcher, S., Palmer, T., Stephens, G., ... Teutsch, G. (2022). Ambitious partnership needed for reliable climate prediction. *Nature Climate Change*, *12*(6), 499–503. doi: 10.1038/s41558-022-01384-8

Suarez-Gutierrez, L., Milinski, S., & Maher, N. (2021). Exploiting large ensembles for a better yet simpler climate model evaluation. *Clim Dyn*(57), 2557–2580. doi: 10.1007/s00382-021-05821-w

Suarez-Gutierrez, L., Müller, W., Li, C., & Marotzke, J. (2020a). Dynamical and thermodynamical drivers of variability in european summer heat extremes. *Clim Dyn*, *54*(9), 4351–4366. doi: 10.1007/s00382-020-05233-2

Suarez-Gutierrez, L., Müller, W., Li, C., & Marotzke, J. (2020b). Hotspots of extreme heat under global warming. *Clim Dyn*(55), 429–447. doi: 10.1007/s00382-020-05263-w

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., ... Winter, B. (2019). The canadian earth system model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. Retrieved 2020-08-22, from `10.5194/gmd-12-4823-2019`

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., ... Kimoto, M. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. Retrieved 2019-12-10, from `10.5194/gmd-12-2727-2019`

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteor. Soc.*, *93*(4), 485–498. doi: 10.1175/BAMS-D-11-00094.1

Tuel, A., Steinfeld, D., Ali, S. M., Sprenger, M., & Martius, O. (2022). Large-scale drivers of persistent extreme weather during early summer 2021 in europe. *Geophysical Research Letters*, *49*(18), e2022GL099624. doi: 10.1029/2022GL099624

Wehner, M. F., Reed, K. A., Li, F., Prabhat, Bacmeister, J., Chen, C.-T., ... Jablonowski, C. (2014). The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. *Journal of Advances in Modeling Earth Systems*, *6*(4), 980-997. doi: 10.1002/2013MS000276

White, R. H., Anderson, S., Booth, J. F., Braich, G., Draeger, C., Fei, C., ... West, G. (2023). The unprecedented Pacific Northwest heatwave of June 2021. *Nature Communications*, *14*(727), 2041-1723. doi: 10.1038/s41467-023-36289-3

Williams, A. I. L., & O'Gorman, P. A. (2022). Summer-Winter Contrast in the Response of Precipitation Extremes to Climate Change Over Northern Hemisphere Land. *Geophysical Research Letters*, *49*(10), e2021GL096531. doi: doi.org/10.1029/2021GL096531

Wills, R. C. J., Dong, Y., Proistosecu, C., Armour, K. C., & Battisti, D. S. (2022). Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change. *Geophysical Research Letters*, *49*(17), e2022GL100011. doi: 10.1029/2022GL100011

Wood, R. R., Lehner, F., Pendergrass, A. G., & Schlunegger, S. (2021). Changes in precipitation variability across time scales in multiple global climate model large ensembles. *Environmental Research Letters*, *16*(8), 084022. doi: 10.1088/1748-9326/ac10dd

Wyser, K., Koenigk, T., Fladrich, U., Fuentes-Franco, R., Karami, M. P., & Kruschke, T. (2021). The SMHI large ensemble (SMHI-LENS) with EC-Earth3.3.1. *Geoscientific Model Development*, *14*(7), 4781–4796. doi: 10.5194/gmd-14-4781-2021

Wyser, K., van Noije, T., Yang, S., von Hardenberg, J., O'Donnell, D., & Döscher, R. (2020). On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6. *Geoscientific Model Development*, *13*(8), 3465–3474. doi: 10.5194/gmd-13-3465-2020

Zeder, J., & Fischer, E. M. (2020). Observed extreme precipitation trends and scaling in Central Europe. *Weather and Climate Extremes*, *29*, 100266. doi: 10.1016/j.wace.2020.100266

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., ... Srbinovsky, J. (2020). The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, *70*, 193–214. doi: 10.1071/ES19035

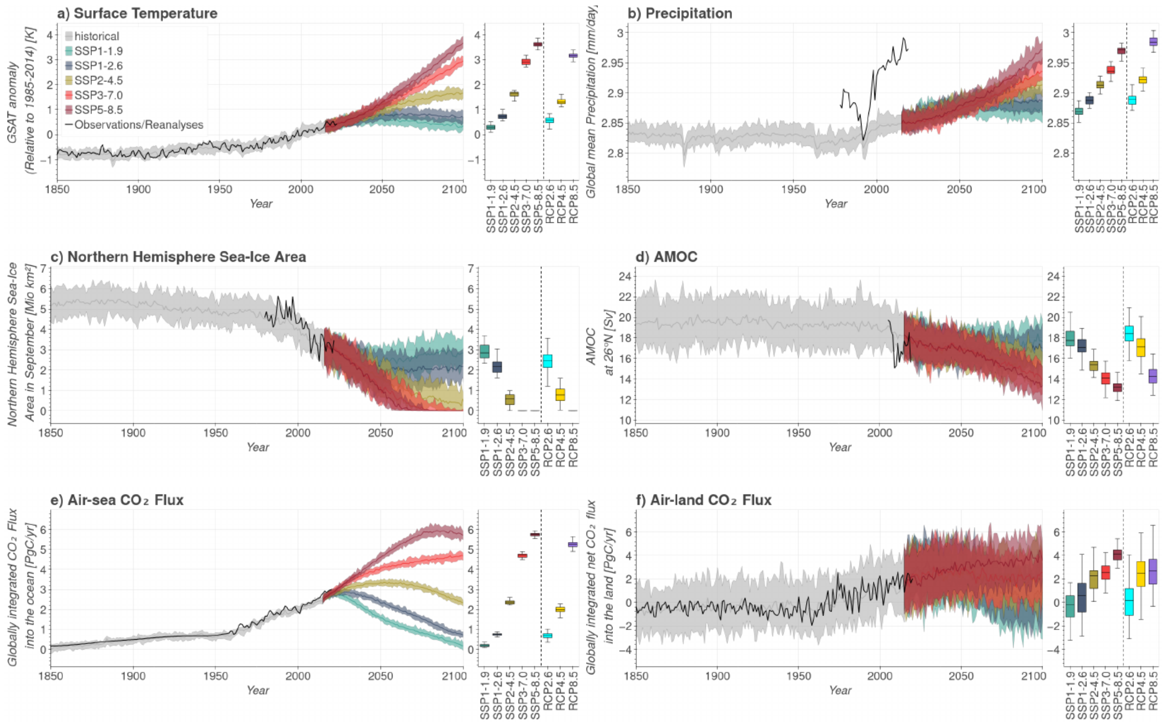

Figure 1: **Comparison of key climate quantities of MPI-GE CMIP6 to observations or reanalyses and MPI-GE CMIP5.** Ensemble spread (shading) and ensemble mean (thick lines) for the historical simulations (grey), and the five emission scenarios SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5. Right hand-side panels show the projected mean and range in year 2099 for the different scenarios of MPI-GE CMIP6 (30 realisations) and MPI-GE CMIP5 (100 realisations). Shown for **a)** global mean near-surface air temperature (GSAT) anomalies (relative to 1985–2014), **b)** global mean precipitation, **c)** Northern Hemisphere sea-ice area in September, **d)** Atlantic Meridional Overturning Circulation (AMOC), **e)** globally integrated CO_2 flux into the ocean and **f)** globally integrated net CO_2 flux into the land. Thick black lines show observations or reanalyses, specifically in **a)** HadCRUT5 (Morice et al., 2021), **b)** ERA5 (Hersbach et al., 2020), **c)** Sea-Ice Index (Fetterer et al., 2017), **d)** RAPID (Frajka-Williams et al., 2021), **e,f)** Global Carbon Project (Global Carbon Project, 2021; Friedlingstein et al., 2022).

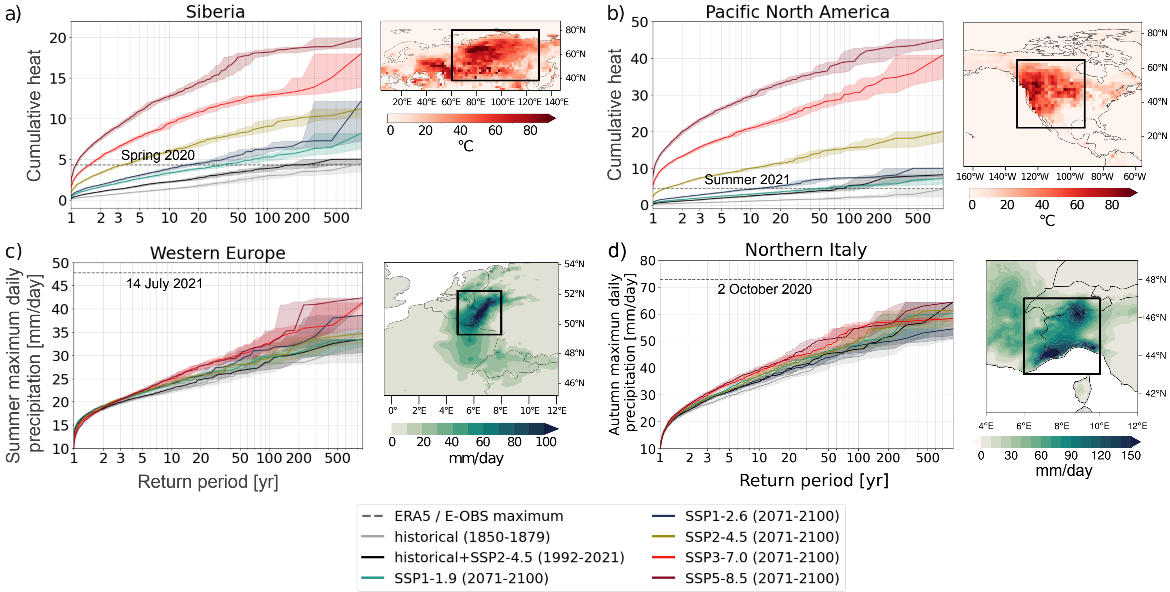

Figure 2: **Return periods from MPI-GE CMIP6 for recently observed heat and precipitation extremes for different emission scenarios.** Return periods for **a-b)** cumulative heat scaled with respect to climatology for **a)** spring (MAM) 2020 Siberian heatwave and **b)** summer (JJA) 2021 Pacific North American heatwave, and **c-d)** seasonal maximum daily precipitation for **c)** western Europe in summer (JJA) and **d)** northern Italy in autumn for the historical climate (1850-1879, grey), the current climate (1992-2021, black), and the five SSP scenarios for the period 2071-2100 (coloured). Shading denotes 95% confidence intervals calculated by bootstrapping with re-sampling. The horizontal dashed line in **a)** and **b)** marks the maximum cumulative heat as calculated from ERA5, and in **c)** and **d)** the observed maximum daily precipitation of the respective season from E-OBS (Klein Tank et al., 2002). The observed spatial pattern of these events is shown as maps in **a)** and **b)** for cumulative heat for spring 2020 and summer 2021, respectively, and in **c)** and **d)** for precipitation on 14th of July 2021 and 2nd of October 2020, respectively. Black boxes mark the regions of interest used for averaging.

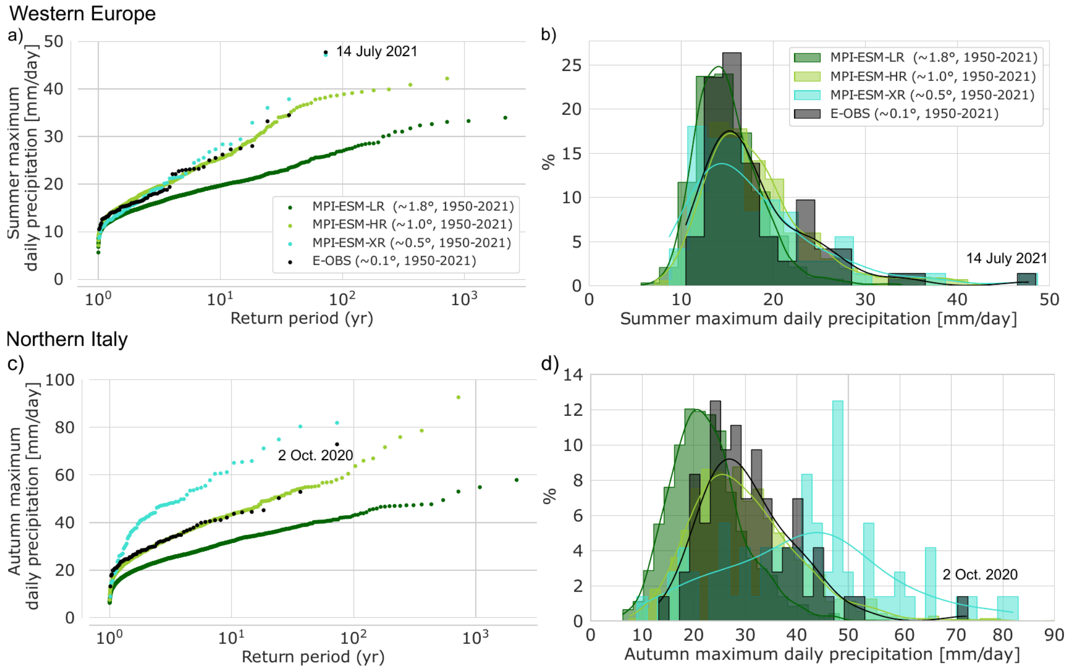

Figure 3: **Representation of precipitation extremes dependent on model resolution. a-b)** Comparison of summer (JJA) maximum daily precipitation averaged across the western European box shown in Fig. 2c from 1950-2021 in three model resolutions from MPI-ESM1.2 and in observations shown as **a)** return periods and **b)** probability density functions. **c-d)** Comparison of autumn (SON) maximum daily precipitation averaged across the northern Italy box shown in Fig. 2d from 1950-2021 in three model resolutions from MPI-ESM1.2 and in observations shown as **c)** return periods and **d)** probability density functions. Note that the return periods are calculated empirically. Values of all summers or autumns, respectively, and all realisations are merged for each ensemble. Further note that MPI-ESM-LR is based on 30 realisations, MPI-ESM-HR on 10 realisations and MPI-ESM-XR and the observed record on only a single realisation. The sample size of MPI-ESM-HR and MPI-ESM-XR might be insufficient to determine return levels above a few years robustly. The domain-averaged maximum daily precipitation of the western European extreme event on 14th of July 2021 is 47.7 mm, and that of the event in northern Italy on 2nd of October 2020 is 72.9 mm.

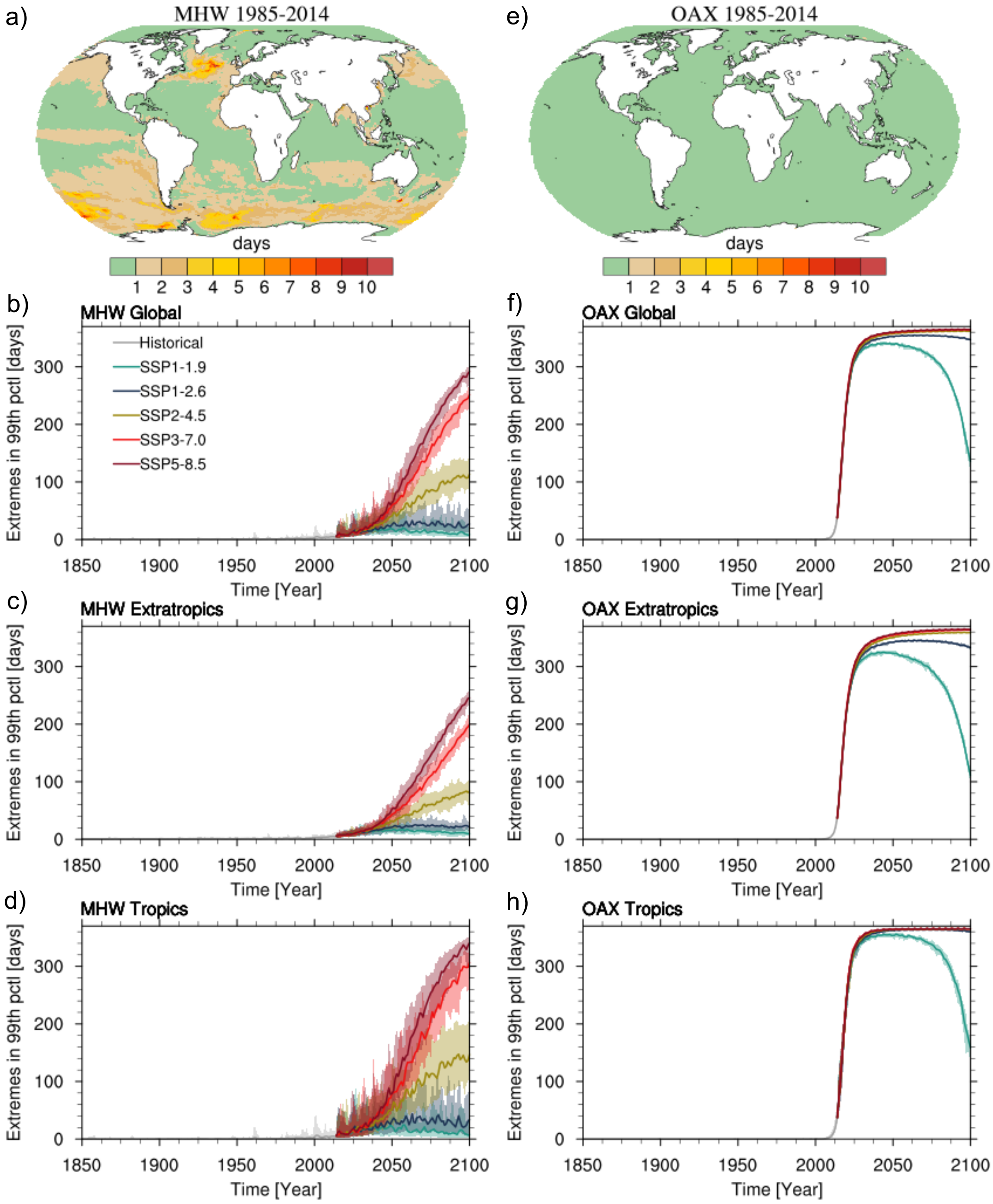

Figure 4: **Present and future frequency of marine heatwaves and ocean acidity extremes.** Maps of **a)** the ensemble mean number of marine heatwave (MHW) days per year and **e)** the number of ocean acidity extreme event (OAX) days per year in the reference period 1985-2014, based on the 99th percentile of daily mean sea surface temperature, and of daily mean surface hydrogen ion concentration, respectively. **b-d)** Globally and regionally averaged number of MHW days per year (global, extratropics: outside of 30°N/30°S, tropics: within 30°N/30°S) for the historical period 1850-2014 (grey), and scenarios SSP1-1.9 (green), SSP1-2.6 (blue), SSP2-4.5 (yellow), SSP3-7.0 (red), SSP5-8.5 (purple) for the period 2015-2100. The shadings cover the ensemble spread, thick lines show the 20-member ensemble mean. **f-h)** Globally and regionally averaged number of OAX days per year and region, similar to **b-d)**.

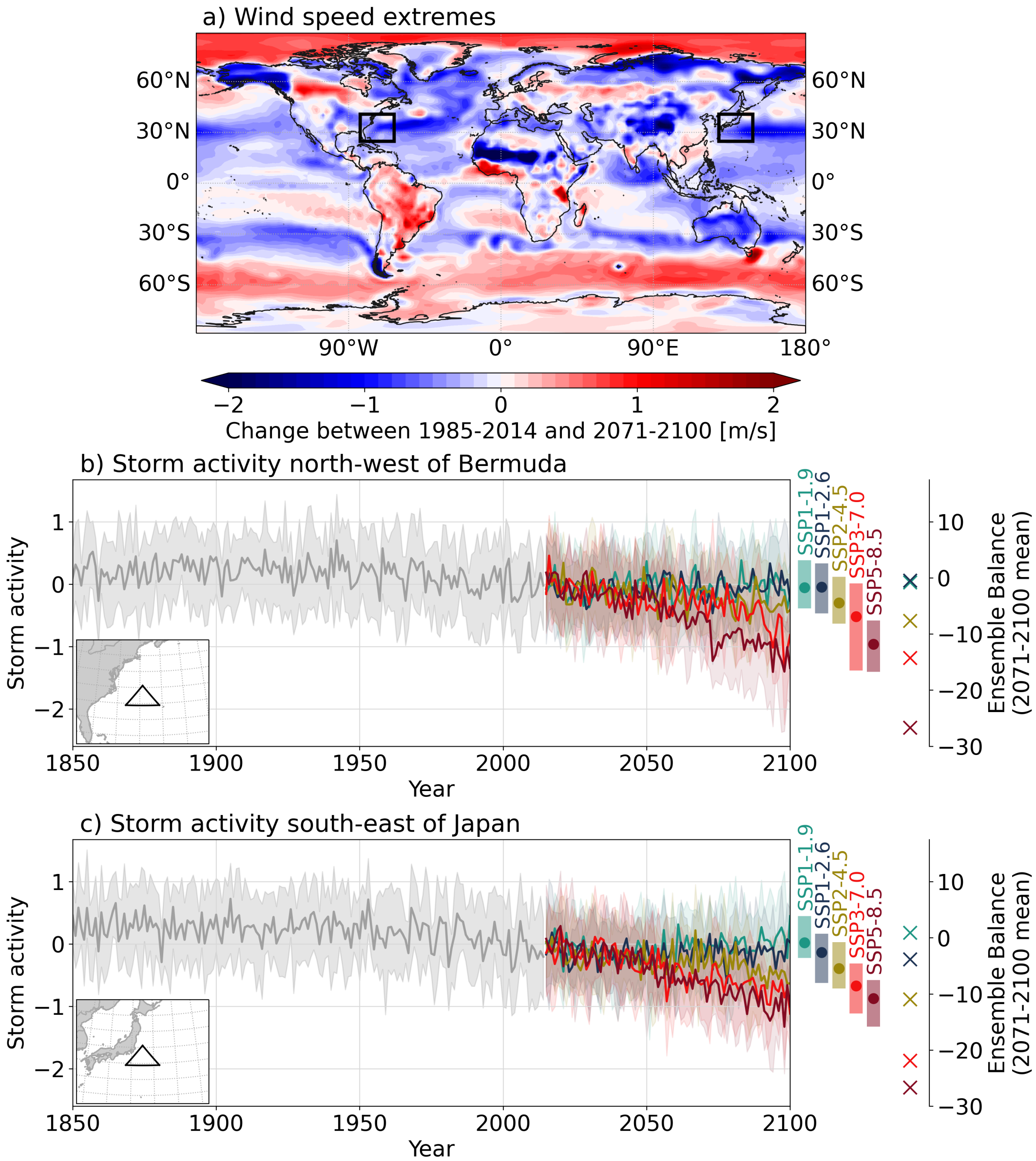

Figure 5: **Projected changes in near-surface wind speed and storm activity. a)** Absolute change in ensemble mean 95th annual percentiles of surface wind speed between 1985-2014 and 2071-2100, based on SSP5-8.5 forcing. Black circles mark regions for which storm activity has been calculated. Maps for the other four SSP scenarios are shown in Figure S5. **b-c)** Ensemble mean storm activity (thick lines) and interquartile range (shading) for the historical simulations (grey) and the five scenarios (coloured) over **b)** the Atlantic Ocean north-west of Bermuda and **c)** the Pacific Ocean south-east of Japan. Coloured dots and bars indicate the 2071-2100 average and range of the ensemble mean for each scenario, and crosses show the 2071-2100 mean ensemble balance.

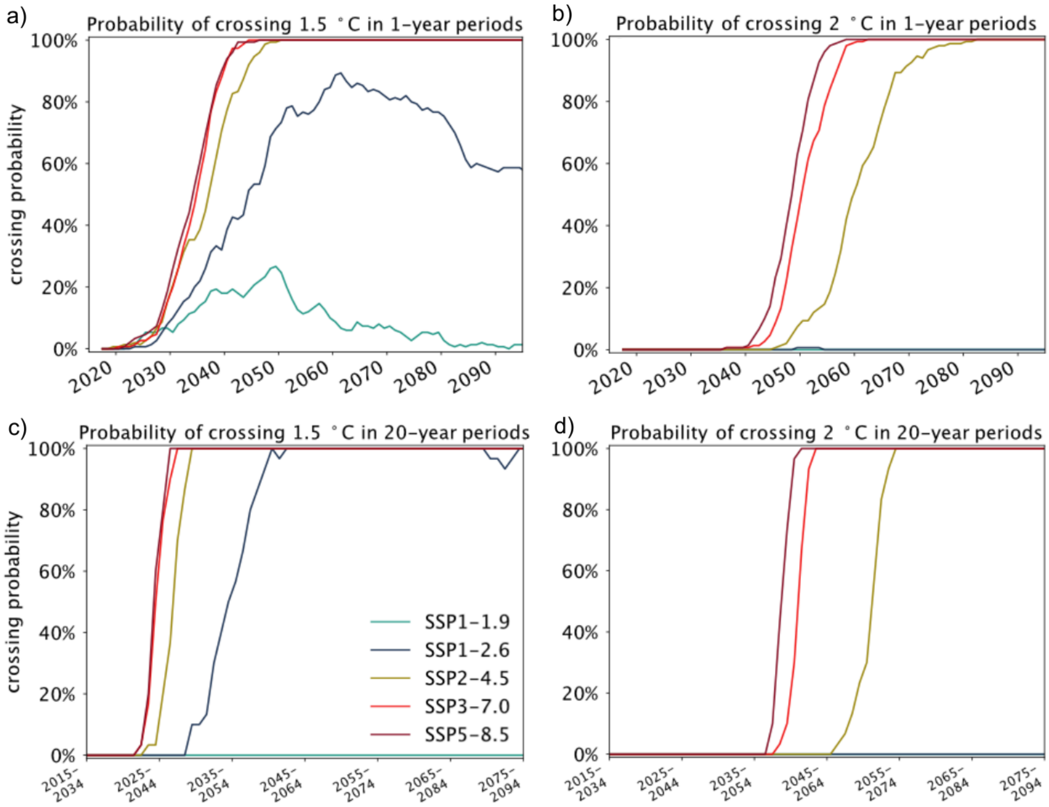

Figure 6: **Probability of crossing Paris Agreement global warming limits.** Probability of crossing **a)** 1.5°C and **b)** 2°C in a single year, and **c)** 1.5°C and **d)** 2°C in 20-year averages for the different emission scenarios until 2100. The crossing probability is defined as the fraction of the 30 realisations that cross the temperature threshold relative to the reference period 1850-1900. In **c,d)**, the 20-year mean GSAT is plotted against the central year of that 20-year period.

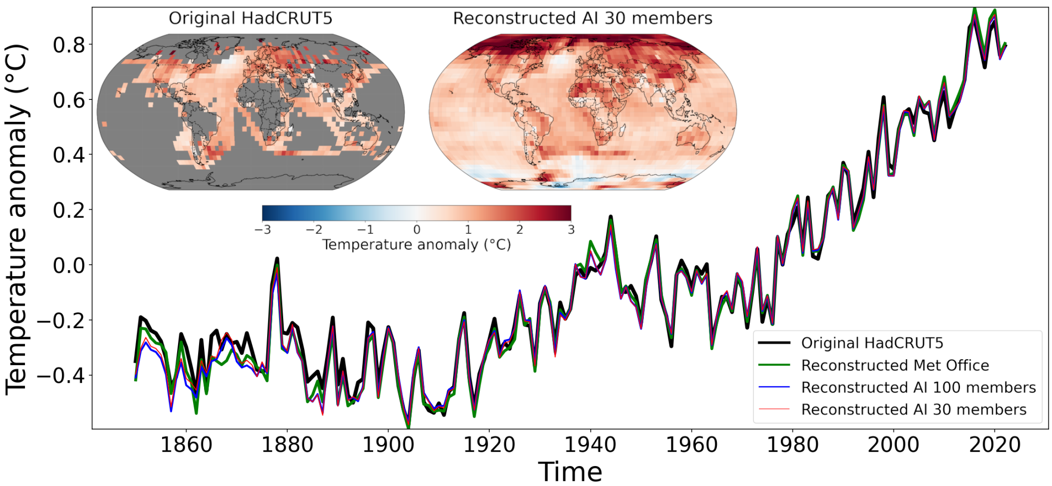

Figure 7: **Comparison of MPI-GE CMIP6 vs MPI-GE CMIP5 for infilling observations of surface temperature with artificial intelligence.** Annual global mean anomaly temperature with respect to the 1961–1990 climatology obtained by using: the gridded original "non-infilled" HadCRUT5 data set (black curve), the partially reconstructed HadCRUT5 data set from the Met Office (Morice et al., 2021), the fully reconstructed HadCRUT5 data set obtained with the AI 100 members model (blue curve, using MPI-GE CMIP5 (Maher et al., 2019)), the fully reconstructed HadCRUT5 obtained with our AI 30 members model (red curve, using MPI-GE CMIP6). Insets: 2020-1991 climatology referenced to the 1920-1891 climatology. Left inset: Original HadCRUT5 data set where gray pixels indicate missing values. Mean values have been computed only for grid points containing at least 70% of valid values for the considered time period. Right inset: Spatial reconstruction of the HadCRUT5 data set using the AI 30 members model.

The new Max Planck Institute Grand Ensemble with CMIP6 forcing and high-frequency model output

Dirk Olonscheck[1], Laura Suarez-Gutierrez[1,2,3], Sebastian Milinski[1,4], Goratz Beobide-Arsuaga[5,6], Johanna Baehr[5], Friederike Fröb[7], Lara Hellmich[1,6], Tatiana Ilyina[1], Christopher Kadow[8], Daniel Krieger[6,9], Hongmei Li[1], Jochem Marotzke[1,5], Étienne Plésiat[8], Martin Schupfner[8], Fabian Wachsmann[8], Karl-Hermann Wieners[1], Sebastian Brune[5]

[1]Max Planck Institute for Meteorology, Hamburg, Germany
[2]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland
[3]Institut Pierre-Simon Laplace, CNRS, Paris, France
[4]European Centre for Medium-Range Weather Forecasts, Bonn, Germany
[5]Center for Earth System Research and Sustainability, Universität Hamburg, Hamburg, Germany
[6]International Max Planck Research School on Earth System Modelling, Hamburg, Germany
[7]Geophysical Institute, University of Bergen, and Bjerknes Centre for Climate Research, Bergen, Norway
[8]German Climate Computing Centre (DKRZ), Hamburg, Germany
[9]Helmholtz-Zentrum Hereon, Geesthacht, Germany

Key Points:

- MPI-GE CMIP6 is a 30-member initial-condition large ensemble with up to 3-hourly model output and five emission scenarios
- The ensemble is specifically suited to investigate climate extremes and Paris Agreement global warming limits
- MPI-GE CMIP6 adequately represents heat extremes, while precipitation extremes are captured by complementary high-resolution simulations

Corresponding author: Dirk Olonscheck, `dirk.olonscheck@mpimet.mpg.de`

Abstract

Single-model initial-condition large ensembles are powerful tools to quantify the forced response, internal climate variability, and their evolution under global warming. Here, we present the CMIP6 version of the Max Planck Institute Grand Ensemble (MPI-GE CMIP6) with 30 realisations for the historical period and five emission scenarios. The power of MPI-GE CMIP6 goes beyond its predecessor ensemble MPI-GE by providing high-frequency output, the full range of emission scenarios including the highly policy-relevant low emission scenarios SSP1-1.9 and SSP1-2.6, and the opportunity to compare the ensemble to complementary high-resolution simulations. First, we describe MPI-GE CMIP6, evaluate it with observations and reanalyses and compare it to MPI-GE. Then, we demonstrate with six novel application examples how to use the power of the ensemble to better quantify and understand present and future climate extremes, to inform about uncertainty in approaching Paris Agreement global warming limits, and to combine large ensembles and artificial intelligence. For instance, MPI-GE CMIP6 allows us to show that the recently observed Siberian and Pacific North American heatwaves would only avoid reaching 1-2 year return periods in 2071-2100 with low emission scenarios, that recently observed European precipitation extremes are captured only by complementary high-resolution simulations, and that 3-hourly output projects a decreasing activity of storms in mid-latitude oceans. Further, the ensemble is ideal for estimates of probabilities of crossing global warming limits and the irreducible uncertainty introduced by internal variability, and is sufficiently large to be used for infilling surface temperature observations with artificial intelligence.

Plain Language Summary

Climate model simulations that start from different initial states and differ only due to the chaos in the climate system are used extensively to quantify the forced climate response, variability intrinsic to the climate system, and their change under global warming. Here, we present a new version of the Max Planck Institute Grand Ensemble (MPI-GE CMIP6) that is run as part of the latest generation of climate models. This single-model ensemble consists of 30 realisations for the historical period 1850-2014 and for five scenarios of possible future climates until 2100. The power of MPI-GE CMIP6 goes beyond its predecessor by not only providing monthly mean but also 3-hourly to daily model output, the full range of future scenarios including the two highly policy-relevant scenarios that were designed to match the Paris Agreement global warming limits of 1.5°C and 2°C, and the opportunity to compare the low-resolution ensemble to simulations of the same model version with higher horizontal resolution. In this paper, we describe the new ensemble and demonstrate with novel application examples how to use its power. For instance, the new ensemble allows us to show that recently observed heatwaves are projected to occur every year at the end of the 21st century if anthropogenic carbon emissions remain high, that recently observed precipitation extremes are captured only by simulations with higher horizontal resolution than that of MPI-GE CMIP6, and that the storminess in many ocean basins is projected to decrease. Further, the ensemble is ideal for estimates of crossing probabilities of Paris Agreement global warming limits, and is sufficiently large to be used to infill missing observations of surface temperature with artificial intelligence.

1 Introduction

Single-model initial-condition large ensembles (SMILEs) have become increasingly important to estimate the variability intrinsic to the climate system. A growing number of SMILEs are now available, reasonably sampling both model uncertainty and internal variability due to their ensemble size. SMILEs enabled substantial progress in understanding the Earth system. For instance, SMILEs were used to separate forced signals from internal variability to unprecedented precision (Maher et al., 2019), to quantify transient changes in the magnitude of climate variability (Olonscheck et al., 2021), and to evaluate how well climate models capture the variability and forced changes in the historical observational record (Suarez-Gutierrez et al., 2021). SMILEs are also used to identify systematic differences between simulated and observed patterns of sea-surface temperature and sea-level pressure change that are very unlikely to occur due to internal variability (Olonscheck et al., 2020; Wills et al., 2022). Furthermore, recent developments in compound event research highlight the importance of sufficiently sampling internal variability to robustly capture tail-risks in multivariate extremes, which requires even larger ensemble sizes than conventional univariate extremes (Bevacqua et al., 2023). The availability of SMILEs from multiple models further allows us to better quantify and differentiate sources of uncertainty in climate projections, especially uncertainties arising from internal variability and those from model differences (Deser et al., 2020; Lehner et al., 2020). These recent major advances in better understanding and quantifying climate variability and change show that SMILEs are increasingly useful tools for climate science.

The Max Planck Institute for Meteorology was one of the first modelling centres that produced a SMILE: the Max Planck Institute Grand Ensemble (MPI-GE, Maher et al. (2019)), which is still the largest SMILE available. MPI-GE – from here on called MPI-GE CMIP5 – is extremely successful and a powerful tool, but it is limited in various aspects: MPI-GE CMIP5 provides monthly model output with some daily output added later for one scenario only (e.g., Loughran et al., 2021; Raymond et al., 2022), it is run with CMIP5 forcing, and it provides three emission scenarios only. These limitations largely prevent the analysis of climate extremes across different emission scenarios because of the lack of high-frequency output, complicate direct comparisons of MPI-GE CMIP5 with SMILEs run with CMIP6 forcing, and restrict its usability for highly policy-relevant science. MPI-GE CMIP6 goes beyond these limitations by specifically enabling (1) the analysis of climate extremes, (2) comparisons to model versions with higher horizontal resolution, (3) comparisons to other SMILEs with CMIP6 forcing, and (4) investigation of low-emission scenarios with high policy relevance.

Several SMILEs with CMIP6 forcing have been recently run by a number of modelling centres, including ensembles with high-frequency model output. Next to MPI-GE CMIP6, currently available SMILEs with CMIP6 forcing and at least 30 realisations for both the historical and future period are ACCESS-ESM1.5 (Ziehn et al., 2020), CanESM5 (Swart et al., 2019), FGOALS (Lin et al., 2022), LENS2 (Rodgers et al., 2021), SMHI-LENS (Wyser et al., 2021), SPEAR-MED (Delworth et al., 2020), and MIROC6 (Tatebe et al., 2019). In comparison to the other CMIP6 SMILEs, MPI-GE CMIP6 provides the most extensive high-frequency output for the historical period and five different emission scenarios (Table 1). This includes the two highly policy-relevant scenarios SSP1-1.9 and SSP1-2.6 that are both otherwise only provided by CanESM5. In contrast to other SMILEs, MPI-GE CMIP6 has a climate sensitivity of 2.8°C which is close to the best estimate of 3°C of the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC AR6) (Forster et al., 2021). Furthermore, its predecessor MPI-GE CMIP5, based on a closely comparable model version, has shown to be one of the models that best represents the global and regional internal variability and forced response in annual observed temperatures (Suarez-Gutierrez et al., 2021) and precipitation (Wood et al., 2021). This good agreement with observations combined with the amount of high-frequency output for the full range of emission scenarios makes MPI-GE CMIP6 ideally suited for investigating future probabilities and magnitudes of climate extremes. The suitability of MPI-GE CMIP6 for studies on climate extremes is further enhanced by the possibility to compare the low-resolution ensemble to high-resolution ensembles or single simulations of the same model version that were run as part of the High Resolution Model Intercomparison Project (HighResMIP, Haarsma et al. (2016), compare Table 2). This unique combination of strengths makes MPI-GE CMIP6 a useful contribution to the CMIP6 multi-model ensemble and a powerful tool to investigate high-frequency climate variability and highly policy-relevant science questions.

Table 1: Characteristics of MPI-GE CMIP6 and other SMILEs with CMIP6 forcing and at least 30 realisations

SMILE name	Model version	Horizontal resolution	High-frequency output	Realisations	Time period	Scenarios	ECS
MPI-GE CMIP6	MPI-ESM1.2-LR	1.8°atm., 1.5°ocean	daily for all parameters, 3-hr, 6-hr for some (see Tables 2 and S1)	30	1850-2100	SSP1-1.9, 1-2.6, 2-4.5, 3-7.0, 5-8.5	2.80°C
ACCESS-ESM1.5	ACCESS-ESM1.5	1.88x1.25°atm.; 1.0°ocean	daily for many atm. parameters	40	1850-2100	SSP1-2.6, 2-4.5, 3-7.0, 5-8.5	3.87°C
CanESM5	CanESM5	2.8°atm., 1.0°ocean	daily for some atm. parameters	50	1850-2100	SSP1-1.9, 1-2.6, 2-4.5, 3-7.0, 5-8.5	5.62°C
FGOALS Super-large Ensemble	CAS FGOALS-g3	2.0°atm., 1.0°ocean	daily for many atm. parameters + tos, omldamax	110	1850-2100	SSP5-8.5	2.80°C
LENS2	CESM2	1.0°atm., 1.0°ocean	daily for all parameters, 3-hr, 6-hr for some	100	1850-2100	SSP3-7.0	5.16°C
SMHI-LENS	EC-Earth3.3.1	1.8°atm.; 1.0°ocean	daily for many atm. parameters	50	1970-2100	SSP1-1.9, 3-3.4, 5-3.4 -OS, 5-8.5	4.31°C
SPEAR-MED	GFDL AM4-LM4	0.5°atm., 1.0° (tropical refinement to 0.3°) ocean	daily for tas, tasmin, tasmax, pr, slp, uas, vas	30	1921-2100	SSP5-8.5	1.78°C
MIROC6	MIROC6	1.4°atm., 1.0°ocean	3-hr and daily for ta, tas, pr	50	1850-2100	SSP1-2.6, 2-4.5, 5-8.5	2.61°C

In this paper we present the new Max Planck Institute Grand Ensemble (MPI-GE CMIP6), and demonstrate its power beyond its predecessor ensemble MPI-GE CMIP5 (Maher et al., 2019) with six application examples. In section 2, MPI-GE CMIP6 is presented, evaluated with observations and reanalyses, and compared to MPI-GE CMIP5. In section 3, the power of MPI-GE CMIP6 is demonstrated with six application examples that specifically use the high-frequency model output for an improved understanding of climate extremes, the low-end emission scenarios for research on Paris Agreement global warming limits, and the medium ensemble size for an efficient combination of SMILEs with artificial intelligence. Section 4 summarises and concludes the paper.

2 MPI-GE CMIP6

2.1 Model description

MPI-GE CMIP6 is a 30-member ensemble simulated with the Max Planck Institute Earth System Model version 1.2 (MPI-ESM1.2, Mauritsen et al. (2019)), in the low resolution (LR) setup. In comparison to the MPI-GE CMIP5 simulations described in Maher et al. (2019), Mauritsen et al. (2019) summarises the updates that were introduced to MPI-ESM1.2, most importantly new radiation and aerosol parameterisations, and a nitrogen cycle for land biogeochemistry. Further, a major difference arises from the update of the external forcing from CMIP5 (Taylor et al., 2012) to CMIP6 (Eyring et al., 2016).

Table 2: Available simulations of MPI-ESM1.2 with different horizontal resolution. The MPI-ESM1.2-HR and -XR simulations were run as part of HighResMIP.

Model version	**Horizontal resolution**	**Realisations**	**Time period**	**Scenarios**
MPI-ESM1.2-LR	T63, 1.8°atm.; GR15, 1.5°ocean	30	1850-2100	SSP1-1.9, 1-2.6, 2-4.5, 3-7.0, 5-8.5
MPI-ESM1.2-HR	T127, 1.0°atm.; TP04, 0.4°ocean	10 (2)	1850-2100	SSP3-7.0 (SSP1-2.6, 2-4.5, 5-8.5)
MPI-ESM1.2-XR	T255, 0.5°atm.; TP04, 0.4°ocean	1	1950-2050	SSP5-8.5

MPI-GE CMIP6 is run with MPI-ESM version 1.2.01p7, with the atmosphere component ECHAM6 (Stevens et al. 2013, echam-6.3.05p2), which is directly coupled to the land component JSBACH (Reick et al. 2013, jsbach-3.20p1), and the ocean and sea-ice component MPIOM (Jungclaus et al. 2013, mpiom-1.6.3p4). MPIOM includes the ocean biogeochemistry module HAMOCC (Ilyina et al., 2013). The atmosphere/land and ocean components are coupled once a day by OASIS-MCT (Craig et al. (2017), oasis3mct-2.0). In MPI-ESM1.2-LR the atmosphere is resolved with spectral resolution T63 (equivalent to approx. 1.8° grid resolution) and 47 vertical levels, the ocean is resolved with a GR15 grid, nominal resolution 1.5°, at 40 vertical levels.

All simulations follow the CMIP6 protocol (Eyring et al., 2016) in terms of initialisation and historical and future external forcing (i.e. atmospheric composition, solar cycle, volcanic eruptions, land use). The 30-member ensemble of historical simulations covers the time period 1850-2014 and each member is initialised from a different state, approximately 25 years apart, of a quasi-stationary one-member 1000-year long preindustrial simulation. This macro initialisation from the preindustrial control state samples the full phase space of both the ocean and atmosphere states (Marotzke, 2019). Five scenario simulations (SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5, 30 realisations each) cover the time period 2015-2100, and in each scenario the realisations are directly initialised from their corresponding realisations of the historical ensemble.

2.2 Availability of high-frequency model output

In addition to standard CMIP6 monthly mean output, daily mean 3D fields of the state of atmosphere and ocean as well as selected daily mean 2D fields, i.e. for sea ice and land surface, are available for all simulations (Table S1 for details). Additionally, a number of atmospheric and land surface parameters are available on the 3-hourly time scale as listed in Table 3. Standard ocean biogeochemistry output from HAMMOC, 3D and 2D, is available on a monthly mean basis, with additional daily means for selected surface 2D or integrated 2D fields (see Table S1). Model output can be accessed via DKRZ's ESGF server at https://esgf-data.dkrz.de/search/cmip6-dkrz/.

2.3 Model evaluation and comparison to MPI-GE CMIP5

MPI-GE CMIP6 performs well in representing key climate quantities as derived from observations and reanalyses (Figure 1). The simulated range of global mean near-surface air temperature (GSAT) anomaly captures the interannual variability and the warming rate of HadCRUT5 well (Morice et al. (2021), Figure 1a). The projected ensemble mean GSAT warming at the end of the 21st century relative to the 1985-2014 reference period ranges from 0.4K in SSP1-1.9 to 3.7K in SSP5-8.5.

Table 3: Parameters with 3-hourly and 6-hourly output on ESGF available for all 30 realisations. The parameters with daily output are listed in Table S1. A full list of parameters subdivided for members r1-r10 and r11-r30 is given in Tables S2-S4.

name	parameter long name	unit	level
3-hourly atmosphere / land			
mrro	Total Runoff	kg m-2 s-1	1
psl	Sea Level Pressure	Pa	1
sfcWind	Near-Surface Wind Speed	m s-1	1
tas	Near-Surface Air Temperature	K	1
uas	Eastward Near-Surface Wind	m s-1	1
vas	Northward Near-Surface Wind	m s-1	1
6-hourly atmosphere / land			
hurs	Near-Surface Relative Humidity	%	1
hus	Specific Humidity	1	47
huss	Near-Surface Specific Humidity	1	1
mrsol	Total Water Content of Soil Layer	kg m-2	5
mrsos	Moisture in Upper Portion of Soil Column	kg m-2	1
pr	Precipitation	kg m-2 s-1	1
ps	Surface Air Pressure	Pa	1
psl	Sea Level Pressure	Pa	1
ta	Air Temperature	K	47
tas	Near-Surface Air Temperature	K	1
tsl	Temperature of Soil	K	1
ua	Eastward Wind	m s-1	47
uas	Eastward Near-Surface Wind	m s-1	1
va	Northward Wind	m s-1	47
vas	Northward Near-Surface Wind	m s-1	1
wap	Omega (=dp/dt)	Pa s-1	4
zg	Geopotential Height	m	28
zg500	Geopotential Height at 500hPa	m	1

For global mean precipitation, MPI-GE CMIP6 underestimates both the magnitude and the interannual variability estimated from the ERA5 reanalysis (Figure 1b), as well as that of ERA-Interim (Figure S1). However, when comparing global mean precipitation in MPI-GE CMIP6 to the observational product of the Global Precipitation Climatology Project (GPCP, Adler et al. (2018)), we find that MPI-GE CMIP6 overestimates the observed global mean precipitation, but still shows too little interannual variability (Figure S1). The different estimates from observational and reanalyses products confirm previous findings that global mean precipitation products have large uncertainty of up to 40% (Bosilovich et al., 2016; Bock et al., 2020). Thus, MPI-GE CMIP6 is well within the range of observational uncertainty, but underestimates interannual variability. For the September Northern Hemisphere sea-ice area, the simulated range captures the observed evolution as derived from the sea-ice index (Fetterer et al. (2017), Figure 1c). September Northern Hemisphere sea-ice area is projected to shrink below the 1 million square kilometre threshold in the second half of the 21st century in SSP2-4.5, SSP3-7.0 and SSP5-8.5, but remains in both SSP1-1.9 and SSP1-2.6 until the end of the 21st century, similar to previous findings on sea-ice decline in CMIP6 (Notz & Community, 2020; Lee et al., 2021). The simulated range of the Atlantic meridional overturning circulation (AMOC) at 26° N is similar to the observed strength and interannual variability of the RAPID observations (Frajka-Williams et al. (2021), Figure 1d). However, the observations suggest that MPI-GE CMIP6 slightly overestimates the AMOC strength. The simulated range of the globally integrated CO_2 flux into the ocean and the net CO_2 flux into the land agrees well with the magnitude as reconstructed in the Global Carbon Project (Friedlingstein et al. (2022)), with simulated estimates of the globally integrated net CO_2 flux into the land exhibiting larger deviations from the mean state than those observed (Figure 1e-f). The evaluation of MPI-GE CMIP6 with observations and reanalyses shows that the ensemble realistically simulates both the long-term evolution and – except for precipitation – also the interannual variability of key climate quantities.

We further compare MPI-GE CMIP6 to MPI-GE CMIP5 with respect to the response of the key climate quantities to the various emission scenarios at the end of the 21^{st} century. We find that MPI-GE CMIP6 shows slightly higher global-mean warming by the end of the 21^{st} century than MPI-GE CMIP5 especially for the respective highest-emission scenarios (Figure 1a). In line with this, September Northern Hemisphere sea-ice area is projected to decline more in the respective SSP than RCP scenarios in the ensemble mean (Figure 1c). Similarly, the ensemble-mean decline in AMOC is substantially stronger in all SSP scenarios than in their respective RCP scenarios (Figure 1d). The globally integrated CO_2 flux into the ocean is larger in the mid and high-end SSP than in the respective RCP scenarios (Figure 1e). The projected change in net CO_2 flux into the land is largely uncertain, but shows a similar response at the end of the 21^{st} century, except for SSP5-8.5 which shows a substantially stronger ensemble-mean increase than RCP8.5 (Figure 1f). In contrast to the stronger changes in MPI-GE CMIP6 compared to MPI-GE CMIP5, global mean precipitation is projected to increase less in the respective SSP than RCP scenarios (Figure 1b). From comparing the global mean temperature response of both model versions to a $1\%CO_2$ increase per year, i.e. the same forcing, we find a very similar warming rate and variability (Figure S2). This implies that the stronger changes in most quantities can be largely explained by the slightly stronger radiative forcing in the SSP compared to RCP scenarios, as has been shown for other models too (Wyser et al., 2020; Fyfe et al., 2021). We conclude that differences between MPI-GE CMIP6 and MPI-GE CMIP5 largely stem from the updated forcing in CMIP6 compared to CMIP5 rather than from differences in the model formulation.

3 Power of MPI-GE CMIP6 beyond MPI-GE CMIP5

MPI-GE CMIP5 (Maher et al., 2019) is extremely successful and a powerful tool to quantify climate variability and its change under global warming. However, the applicability of MPI-GE CMIP6 goes beyond MPI-GE CMIP5 in at least four critical aspects:

First, MPI-GE CMIP5 is run with CMIP5 forcing which limits direct comparisons to the large number of SMILEs that were run with CMIP6 forcing. MPI-GE CMIP6 provides the opportunity to compare MPI-ESM with other SMILEs run with CMIP6 forcing, and to investigate the impact of different forcings between MPI-GE CMIP5 and MPI-GE CMIP6.

Second, MPI-GE CMIP5 does not provide high-frequency model output across different emission scenarios, but only monthly mean output in most cases which strongly limits the usefulness for investigating short-lived climate extremes and their drivers (Suarez-Gutierrez et al., 2020a). In contrast, MPI-GE CMIP6 provides high-frequency output with 3-hourly and 6-hourly output for some variables (see Table 3) and daily output for all variables (see Table S1). This high-frequency output comes at the expense of a smaller ensemble size of 30 realisations instead of 100 realisations, but makes MPI-GE CMIP6 specifically suited for the analysis of climate extremes.

Third, MPI-GE CMIP6 can be compared to higher-resolution simulations of the same model version (see Table 2), for instance 10 realisations of MPI-ESM1.2-HR (1.0° atm., 0.4° ocean, Müller et al. (2018)) or a single realisation of MPI-ESM1.2-XR which provides also higher horizontal resolution in the atmosphere (0.5° atm., 0.4° ocean, Gutjahr et al. (2019)). This allows for the combination of high-frequency output in relatively low horizontal resolution of MPI-GE CMIP6 with high-resolution simulations, which is not possible with MPI-GE CMIP5.

Fourth, MPI-GE CMIP6 provides five instead of three emission scenarios. The five scenarios with 30 realisations each span the full range of IPCC scenarios from the low-emission scenario SSP1-1.9 to the high-emission scenario SSP5-8.5. With the scenarios SSP1-1.9 and SSP1-2.6, MPI-GE CMIP6 provides ensembles of two scenarios that were designed for projections of the Paris Agreement global warming limits of a 1.5°C and 2°C warmer world by the end of this century. This makes MPI-GE CMIP6 one of the few models that provide large ensembles for the two scenarios aligned with the Paris Agreement pledges, which allows for timely and highly policy-relevant science.

In the following, we exemplify the power of MPI-GE CMIP6 with six application examples. These examples include the analysis of heat, precipitation, wind, and ocean acidity extremes (Section 3.1), the probability of crossing Paris Agreement global warming limits (Section 3.2), and the potential of combining SMILEs with artificial intelligence methods for infilling observations (Section 3.3).

3.1 Analysing climate extremes

Climate extremes are among the most devastating and costly events, and their frequency and intensity is projected to increase with global warming (Seneviratne et al., 2021). However, climate models struggle to represent observed extremes because of large internal climate variability and their limited horizontal and temporal resolution (e.g., Slingo et al., 2022). Given the ensemble size and high-frequency output of MPI-GE CMIP6, we first investigate projected changes in heat and precipitation extremes and evaluate whether the new ensemble is capable of realistically simulating recently observed heat and precipitation extremes (Section 3.1.1). We then test whether observed precipitation extremes are better captured by model versions with higher horizontal resolution (Section 3.1.2). Finally, we investigate projected changes in marine heatwaves and ocean acidity extremes (Section 3.1.3) as well as in wind extremes (Section 3.1.4). For these analyses we choose a fixed baseline climatology over the time period 1985-2014.

3.1.1 Continental heat and precipitation extremes

We first evaluate whether MPI-GE CMIP6 is capable of simulating heat and precipitation extremes that were recently observed (Figure 2). We focus on the Siberian heatwave in spring 2020 (Ciavarella et al., 2021), the Pacific North American heatwave in summer 2021 (Philip et al., 2022), the extreme precipitation event in western Europe in summer 2021 (Ibebuchi, 2022; Tuel et al., 2022), and the extreme precipitation event in northern Italy in autumn 2020 (Davolio et al., 2023). To do so, we use daily surface maximum temperature and daily precipitation from MPI-GE CMIP6, and use ERA5 (Hersbach et al., 2020) and E-OBS (Klein Tank et al., 2002) as observational reference.

For continental heat extremes, we use the metric heat excess, which takes into account both heatwave intensity and persistence into one single metric (Perkins-Kirkpatrick & Lewis, 2020). To calculate heat excess, we identify heatwaves on a grid-point level when daily maximum near-surface air temperature exceeds the 90^{th} percentile based on a centred 15-day running window of the historical period 1985-2014 for at least three consecutive days. The cumulative heat is then calculated by seasonal integration of the exceeding heat above the threshold during heatwave days. In addition, we weight the cumulative heat of each grid point by the cosine of the latitude and spatially integrate it. For the 2020 Siberian heatwave we integrate the cumulative heat over boreal spring (MAM) and 40° N-80° N and 60° E-130° E. For the 2021 Pacific North American heatwave we integrate the cumulative heat over boreal summer (JJA) and 25° N-65° N and 90° W-130° W (see maps in Figure 2a,b). We scale the cumulative heat with respect to climatology (1985-2014). We compute the return periods for historical climate (1850-1879), the current climate (1992-2021) and the five SSP scenarios (SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, SSP5-5.8; 2071-2100), and compare them to the two recent heatwaves in ERA5 (Figure 2a,b). The cumulative heat estimated by ERA5 in spring 2020 and summer 2021 integrated over the respective domains is 4.3 and 4.5.

These two record-shattering heat extremes led to devastating impacts. The Siberian heatwave was linked to large wildfires that causes a release of 56 megatons of CO_2 in June 2020, and to the melting of large permafrost areas which led to widespread infrastructure and environmental damages (Ciavarella et al., 2021). The Pacific North American heatwave also led to hundreds of attributable deaths, marine life mass-mortality events, reduced crop and fruit yields, river flooding from rapid snow and glacier melt, and a substantial increase in wildfires (White et al., 2023). In line with previous attribution studies (Ciavarella et al., 2021; Philip et al., 2022), we find that both heatwaves were virtually impossible in the preindustrial MPI-GE CMIP6 world, and have over 100-year return periods in current climate conditions. However, under the moderate emission scenario SSP2-4.5, heat excess levels as high as those during the 2020 Siberian heatwave could occur every four years (Figure 2a), and more than every other year for the 2021 Pacific North American heatwave (Figure 2b). In SSP5-8.5, MPI-GE CMIP6 projections show that a comparable 1-in-100-years event by the end of the 21st century reaches heat excess levels 5 to 8 times higher than the 2020 and 2021 levels, respectively. Only in the low emission scenarios SSP1-1.9 or SSP1-2.6 return periods below 10 years for such heat extremes can be avoided.

For precipitation extremes, we focus on two recently observed record-shattering events: the extreme precipitation event in western Europe on the 14th of July 2021, and the one in northern Italy on 2nd of October 2020. The extreme precipitation event in western Europe caused unprecedented flooding of the rivers Ahr and Erft. A rapid attribution study shows that observations over a larger region and different regional climate models give high confidence that human-induced climate change has increased the likelihood and intensity of events like the western European precipitation extreme (Kreienkamp et al., 2021; Ibebuchi, 2022), in line with the intensification of observed extreme precipitation in central Europe during the last century related to Northern Hemispheric warming (Zeder & Fischer, 2020). When integrated over 49° N-52° N and 5° E-8° E, the daily precipitation as observed by the E-OBS data set (Klein Tank et al., 2002) on 14th of July 2021 is 47.7 mm which represents the maximum daily precipitation in summer in the 72-year long observed record (see map in Figure 2c). The extreme precipitation event in northern Italy caused devastating large-scale flooding and represents an unprecedented strong event in a region that shows a high frequency of precipitation extremes (Davolio et al., 2023; Grazzini et al., 2021). The event was caused by a superposition of an upper-level trough over the western Mediterranean basin and moisture transport from the tropics by an atmospheric river (Davolio et al., 2023). When integrated over 43° N-47° N and 6° E-10° E, the daily precipitation observed by E-OBS on 2nd of October 2020 is 72.9 mm.

We use daily precipitation from MPI-GE CMIP6 and E-OBS, and compare the observed extreme precipitation events to the seasonal maximum daily precipitation simulated for the historical climate (1850-1879), the current climate (1992-2021), and the five SSP scenarios for the period 2071-2100. We find that MPI-GE CMIP6 does not simulate a summer and autumn daily precipitation event as intense as observed, not even until the end of the 21st century (Figure 2c). This implies that in any of the climate conditions simulated by MPI-GE CMIP6 an event as intense as the ones observed in 2020 and 2021 is virtually impossible, with return periods exceeding 900 years for all scenarios. We further find that simulated summer and autumn maximum daily precipitation is larger for higher emission scenarios than for lower scenarios in 2071-2100 and for the historical and current climate, in line with the fact that warmer air can hold more water leading to increased precipitation (e.g., Pendergrass et al., 2017; Myhre et al., 2019). However, the spread from the emission scenarios largely overlaps, suggesting that the uncertainty due to internal variability dominates scenario uncertainty and thus events typical for higher emission scenarios could also occur in a lower warming world due to internal variability. The results show that precipitation extremes as intense as the ones observed are not captured by MPI-GE CMIP6 possibly because the horizontal resolution of MPI-GE CMIP6 is too low to simulate real-world mechanisms leading to such small-scale precipitation extremes (Slingo et al., 2022). Given the increased probability of extremes that are unprecedented in the observed record and the often substantial impacts (Fischer et al., 2021), a realistic representation of such extreme events by climate models is highly needed.

3.1.2 Resolution dependence of representing precipitation extremes

Higher horizontal resolution of climate models improves the simulation of extreme precipitation because higher-resolution models reflect smaller spatial scales of extreme precipitation and key processes such as deep convection do not need to be parameterised (Wehner et al., 2014; Iles et al., 2020; Kendon et al., 2021; Kahraman et al., 2021). To test whether the inability of MPI-GE CMIP6 to represent the two observed precipitation extremes is caused by the model's coarse horizontal resolution, we investigate whether these events are better captured in higher-resolution versions of the same model, namely 10 realisations of MPI-ESM1.2-HR (Müller et al., 2018) with 1.0° atmospheric horizontal resolution, and a single realisation of MPI-ESM1.2-XR (Gutjahr et al., 2019) with 0.5° atmospheric horizontal resolution (see Table 2).

For the western European event, we find that MPI-ESM1.2-HR and MPI-ESM1.2-XR show higher agreement with the observed distribution of summer maximum daily precipitation over the period 1950-2021 than MPI-ESM1.2-LR, the low-resolution model version used for MPI-GE CMIP6 (Figure 3a,b). Strikingly, the single realisation of MPI-ESM1.2-XR simulates a single daily precipitation as intense as the one observed with a more widespread but still similar pattern (compare Figure S3), while MPI-ESM1.2-LR and MPI-ESM1.2-HR do not simulate such high daily precipitation amounts. Although the horizontal resolution of MPI-ESM1.2-XR is still not sufficient to resolve important processes such as moist convection (Hewitt et al., 2022; Slingo et al., 2022), our finding suggests that its resolution is sufficient to represent the recently observed regional precipitation extreme. Alternatively, MPI-ESM1.2-XR might overestimate the real-world precipitation intensity, which could also explain why the single simulation captures an event as intense as observed.

For autumn precipitation in northern Italy, we find that MPI-ESM1.2-HR much better represents the observed frequency of autumn maximum daily precipitation than MPI-ESM1.2-LR (Figure 3c,d). MPI-ESM1.2-XR shows generally too high autumn maximum precipitation, simulating precipitation amounts as large as observed with higher frequency. This is in line with previous findings that in the Mediterranean coastal region autumn precipitation intensity is larger at convection-permitting resolution than at coarse resolution because realistically representing deep convection is central for such events (Luu et al., 2020; Pichelli et al., 2021). The comparison between the western European and northern Italian events suggests that the model is able to simulate larger-scale autumn precipitation at coarser horizontal resolution than convective summer precipitation (Feldmann et al., 2008; Luu et al., 2020; Williams & O'Gorman, 2022). We conclude that while MPI-GE CMIP6 fails to simulate the observed precipitation extremes in western Europe and northern Italy, high-resolution simulations of the same model version are able to capture these extreme events, highlighting the potential for investigating regional precipitation extremes from comparing high-frequency model output of MPI-GE CMIP6 with simulations of higher horizontal resolution.

3.1.3 Marine heatwaves and ocean acidity extremes

We analyse daily mean sea surface temperature (SST) and hydrogen ion concentration ([H^+]) to identify marine heatwaves and ocean acidity extremes between 1850 and 2100 (Figure 4).

We use a percentile-based threshold and the reference period 1985-2014 for both extremes such that the probability of the occurrence of marine heatwaves and ocean acidity extremes in a year is the same. SST and $[H^+]$ are defined as extreme, if they exceed the 99^{th} percentile for five consecutive days (Hobday et al., 2016; Burger et al., 2020). Although applying a duration criterion for ocean acidity extremes is not common, here it ensures comparability with marine heatwaves. The percentiles are calculated as the 20-member ensemble mean (only members 11 to 30 contain daily mean output for $[H+]$) over the 99^{th} multiyear daily running percentile with a 5-day window length at every grid cell between 1985 and 2014. Finally, we calculate the number of extreme days per year to characterise changes of both extremes with time and across scenarios.

Before the reference period 1985-2014, almost no marine heatwaves are detected. Between 1985 and 2014, less than ten days per year are extreme with marine heatwaves being more frequent in the subpolar North Atlantic and the Southern Ocean (Figure 4a). By 2030, between five and 70 days per year are extreme with substantial overlap among different scenarios. By 2100, the SSP5-8.5 scenario projects the most marine heatwaves, with the entire ocean being in almost a constant state of extreme; while in the SSP1-1.9 scenario the number of extreme days per year does not exceed 15 by 2100 (Figure 4b, Figure S4). There is a much larger difference between the SSP1-1.9 and SSP5-8.5 scenarios in terms of global marine heatwave days at the end of the 21st century when compared to the difference in terms of global mean temperature between these scenarios (compare Figures 1a and 4b), indicating an amplified impact of global warming on marine heatwaves.

Over the historical period, globally, no ocean acidity extreme is detectable prior to the reference period. Within the reference period 1985-2014 (Figure 4e), the number of days with extreme $[H^+]$ increases to approximately five days per year in 2010 and continues to increase substantially to nearly 40 days per year in 2014. Locally, within the reference period, only very weak spatial gradients in the ensemble-mean number of ocean acidity extremes exist (Figure 4e). Until 2030, the entire ocean area moves rapidly to a near-permanent extreme state with more than 300 extreme days per year for all five future scenarios. By 2100, almost all days of a year show ocean acidity extremes in the SSP2-4.5, SSP3-7.0, and SSP5-8.5 scenarios, while in the SSP1-2.6 scenario, the number of ocean acidity extreme days is projected to decline slightly by the end of the 21^{st} century (Figure 4f, Figure S4). Within the SSP1-1.9 scenario, ocean acidity extremes are projected to peak at approximately 330 days per year between 2025-2040 and decline thereafter to 140 days per year by 2100. In this scenario, ocean acidity extremes occur less frequently in the Arctic Ocean and in the Southern Ocean compared to the Tropics between 2071-2100 (Figure 4g,h). There is a striking difference in the global occurrence of ocean acidity extremes between SSP1-1.9 and SSP1-2.6 in the second half of the 21st century (Figure 4f), despite only small differences in terms of global mean temperature in both scenarios (Figure 1a).

The CO_2 system in seawater and the mixing ratio of atmospheric CO_2 are tightly related, which leads to the smooth response in the mean surface ocean $[H^+]$. Sea surface temperature on the other hand is more variable across space and time than $[H^+]$, therefore the number of marine heatwaves varies more than the number of ocean acidity extremes across ensemble members. The number of detected extremes is sensitive to the definition, affected by the choice of threshold and reference period (Gruber et al., 2021). While using the same definition for both marine heatwaves and ocean acidity extremes is helpful to illustrate the different internal variability structure of the underlying parameters, understanding the governing processes may require a different extreme event definition that would ultimately lead to a different number of detected events.

3.1.4 Wind extremes

Future changes in wind extremes are among the most uncertain impacts of anthropogenic climate change (Seneviratne et al., 2021). We use the 3-hourly output of MPI-GE CMIP6 to project global changes in wind extremes and their dependence on the emission scenario (Figure 5a and Figure S5). To detect projected global changes in wind speed, we first derive 95^{th} annual percentiles of near-surface wind speeds for each grid point from the entire 30-member ensemble and then calculate the absolute difference between the 2071-2100 mean and the 1985-2014 reference mean. Here, we focus on SSP5-8.5 because the projected changes are most distinct: Over the ocean, we find a latitudinal contrasting pattern with increasing wind extremes over high-latitude oceans and decreasing wind extremes in most mid- and low-latitude ocean basins. Over land, increases in wind extremes are projected for South America, Western and Eastern Africa and parts of the Northern mid- to high-latitudes, whereas substantial decreases are projected for Alaska, Siberia, Central Asia and the Western Sahara. Weaker changes but with the same pattern are found for lower-emission scenarios (Figure S5).

We further analyse projected changes in storm activity in two regions that are known for the frequent passage of mature hurricanes and typhoons with often devastating impacts when they make landfall: north-west of Bermuda in the North Atlantic (Figure 5b) and south-east of Japan in the North Pacific (Figure 5c). For both regions, we select three grid points that form a triangle spanning the area of interest (Table S5). We then use 3-hourly mean sea-level pressure data from MPI-GE CMIP6 at the selected grid points and derive geostrophic winds v_g from the horizontal mean sea-level pressure gradients $\partial p/\partial x$ and $\partial p/\partial y$ according to Krieger et al. (2020) via

$$v_g = (v_x^2 + v_y^2)^{1/2}, \tag{1}$$

with

$$v_x = -\frac{1}{\rho f}\frac{\partial p}{\partial y} \qquad \text{and} \qquad v_y = \frac{1}{\rho f}\frac{\partial p}{\partial x}, \tag{2}$$

where ρ is the density of air (set at 1.25 kg m^{-3}) and f the average of the Coriolis parameter at the three corners of the triangle. We chose the grid points so that the resulting triangle is sufficiently close to an equilateral triangle. This requirement is necessary to avoid a large error propagation of pressure uncertainties, which would cause a shift of the wind direction towards the main axis of the triangle (Krieger et al., 2020). We then define storm activity as the standardised annual 95^{th} percentiles of 3-hourly geostrophic wind speeds. We therefore first calculate annual 95^{th} percentiles of geostrophic winds for each ensemble member. We then standardise by subtracting the 1985-2014 ensemble mean from each ensemble member, and divide by the 1985-2014 ensemble standard deviation.

For both north-west of Bermuda and south-east of Japan, we find a decreasing storm activity with strongest decreases for high-emission scenarios, while we find no notable change in scenario SSP1-1.9 (Figure 5b,c and Figure S5). This agrees with the projected change in surface wind speed, where the marine subtropics around 30° N show a strong signal of decreasing wind speeds in the SSP5-8.5 scenario (Figure 5a).

We further calculate the ensemble balance to characterise whether changes in the ensemble mean are caused by a shift in the majority of the ensemble members or by a few strong outliers. To do so, we first apply a moving Gaussian low-pass filter to the storm activity time series of each ensemble member. We then define thresholds for high and low activity periods at $0.5\,\sigma$ and $-0.5\,\sigma$, and count for how many members the low-pass filtered curve exceeds these thresholds in a certain year. The difference in the number of high-activity and low-activity members is then regarded as the ensemble balance (crosses on the secondary y-axis in Figure 5b,c). In the SSP1-1.9 and SSP1-2.6 scenarios, we find that the ensemble balance does not significantly deviate from 0 towards the end of the 21^{st} century in both focus regions, confirming the rather small projected change in storm activity. In the high-emission SSP5-8.5 scenario, the ensemble balance falls to near -30 at the end of the 21^{st} century, which indicates that nearly all ensemble members agree on a decline in storm activity both north-west of Bermuda and south-east of Japan.

The proxy for storm activity is based on the hypothetical geostrophic wind and its long-term statistics, as proposed originally by Schmidt and von Storch (1993). For high latitudes, where the synoptic-scale wind in higher altitudes is close to geostrophic, it has been shown that the statistics of the geostrophic wind closely resemble the statistics of the near-surface wind (Krueger & von Storch, 2011). In latitudes closer to the equator this assumption does not hold, as most of the wind extremes occur in or near tropical cyclones, which are not fully in geostrophic balance. The proxy should therefore not be used as a single tool to make conclusions about future changes in the intensity or frequency of tropical cyclones. However, the decreasing storm activity for mid-latitude hurricanes and typhoons is in line with recent findings of a decreasing frequency of tropical cyclones (Chand et al., 2022). As the proxy only describes storm activity with one quantity, it cannot distinguish between changes in the frequency and changes in the intensity of storms. A change in storm activity can thus be interpreted as a change in either number or intensity of cyclones, or a combined change thereof. Also, changes connected to smaller-scale features such as fronts or convective wind gusts within cyclones cannot be detected by the proxy, as the derived geostrophic wind acts as an area mean over the entire triangle.

Overall, MPI-GE CMIP6 projects increasing wind extremes over high-latitude oceans and decreasing wind extremes in most mid- and low-latitude oceans, in line with current understanding of observed changes in wind extremes caused by a poleward shift of extra-tropical storm tracks over both hemispheres (Seneviratne et al., 2021). We conclude that MPI-GE CMIP6 with its 3-hourly model output is a powerful tool to understand changes in the frequency and intensity of wind extremes for different emission scenarios.

3.2 Investigating crossing probabilities of 1.5°C and 2°C global warming

The Paris Agreement in 2015 states the goal to keep global warming well below 2°C, and to pursue efforts to limit global warming to 1.5°C above preindustrial levels to avoid devastating and unmanageable consequences of climate change. MPI-GE CMIP6 is suited to investigate the uncertainty in crossing these global warming limits because one can account for internal climate variability with ensemble simulations for five different emission scenarios, including the scenarios SSP1-1.9 and SSP1-2.6 that project a global warming of 1.5°C and 2°C, respectively.

To investigate the crossing probability of 1.5°C and 2°C of global warming in MPI-GE CMIP6, we use annual mean, global mean near-surface air temperature (GSAT) to compute for every year and each of the five scenarios the fraction of realisations (x / 30 realisations) that crosses these temperature thresholds in a single year relative to the 1850-1900 reference period (Figure 6a,b). We find that in all emission scenarios, there is a non-zero chance of observing individual years above 1.5°C within the next decades, including the SSP1-1.9 scenario that represents the strongest mitigation efforts. However, this finding does not imply that every scenario crosses the Paris agreement 1.5°C global warming limit because whether a temperature threshold will be crossed or not is commonly evaluated for 20-year mean temperatures (Lee et al., 2021). To account for this definition, we also compute the 20-year running mean GSAT time series for each realisation and show for each 20-year window the fraction of realisations that crosses 1.5°C or 2°C (Figure 6c,d). We find that MPI-GE CMIP6 with the SSP1-1.9 scenario is consistent with the 1.5°C warming limit, whereas all other scenarios cross this threshold. We stress that when 1.5°C are crossed for 20-year means is still affected by internal variability: for SSP1-2.6, 1.5°C may be crossed around the 20-year mean of the period starting in 2030, but only 10 years later it is virtually certain that 1.5°C is crossed in the 20-year mean of any realisation. Further, the SSP1-1.9 and SSP1-2.6 scenarios will not cross 2°C neither in single years nor for 20-year means while all other scenarios will cross this threshold between 20-year means starting in 2035 to 2050. These estimates are at the upper range of the IPCC AR6 central estimate of crossing the 1.5°C threshold which lies in the early 2030s for all scenarios except SSP5-8.5 (Marotzke et al., 2022; Lee et al., 2021).

We note that the IPCC AR6 uncertainty range includes uncertainties in historical warming, climate sensitivity and internal variability (Lee et al., 2021), whereas MPI-GE CMIP6 has a fixed climate sensitivity and the uncertainty range is only due to internal variability. However, the observed internal variability in GSAT is well simulated by the model (Suarez-Gutierrez et al., 2021) and its equilibrium climate sensitivity of 2.8°C is close to the central estimate of the IPCC AR6 assessment of 3°C. Comparing the central estimates of crossing times for 1.5°C between MPI-GE CMIP6 and the IPCC AR6 assessment shows that the MPI-GE CMIP6 estimates are systematically later than in AR6 (Table S6). Most notably, SSP1-1.9 does not cross 1.5°C in the model, the crossing in SSP1-2.6 occurs a decade later, and the crossing in all other scenarios about five years later than in IPCC AR6. This shows that the MPI-GE CMIP6 estimates are broadly consistent with but slightly more conservative than the IPCC AR6 assessment.

We conclude that with its good representation of internal variability in GSAT and its equilibrium climate sensitivity close to the central estimate of the IPCC AR6 assessment, MPI-GE CMIP6 offers a unique framework to investigate timing and local impacts of crossing temperature thresholds such as 1.5°C.

3.3 Combining SMILEs and artificial intelligence

SMILEs and artificial intelligence can be combined powerfully because the multiple realisations of a same model provide testing, validation and training data sets to infill gaps in observational data. We provide one example by using a method that is based on an inpainting technique developed by Liu et al. (2018) to repair corrupted images. It makes use of a U-Net neural network made of partial convolutional layers and a state-of-the-art loss function designed to produce semantically meaningful predictions. As shown in Kadow et al. (2020), the method can infill large and irregular regions of missing climate data and is able to reconstruct specific climate patterns that are not captured by standard interpolation techniques such as the Kriging method (Cowtan & Way, 2014).

We here test whether the ensemble size of MPI-GE CMIP6 is sufficiently large to be used for infilling the HadCRUT5 data set with similar capability than the 100-member MPI-GE CMIP5. The models used to infill the HadCRUT5 data set (Dunn et al., 2020) have been trained using gridded global historical surface temperature anomalies from three large ensembles: 1) MPI-GE CMIP6, containing 30 realisations and spanning the 1850-2014 time period; 2) MPI-GE CMIP5, containing 100 realisations and spanning the 1850-2005 time period; and 3) a subset of MPI-GE CMIP5 containing the first 30 ensemble members, here called MPI-GE CMIP5(30). Before the training, one ensemble member was excluded from each ensemble to create three testing data sets. Three validation data sets were created from the remaining ensemble members of each data set by pulling out the data every 8 timesteps for MPI-GE CMIP6 and MPI-GE CMIP5(30), and every 7 timesteps for MPI-GE CMIP5. The remaining data were used to create the training data sets which contain 50.242 samples for MPI-GE CMIP6, 47.502 samples for MPI-GE CMIP5(30) and 162.162 samples for MPI-GE CMIP5. For this work, additional features have been implemented to the original version of the code (Kadow et al., 2020) to improve the computational performance and the quality of the reconstruction. In particular, a custom padding operation accounting for the boundary conditions of the global data is now applied before each partial convolution, to account for the sphere of the Earth.

The annual global mean temperature time series reconstructed using the 100 member and the 30 member models are very similar, especially when compared to the original HadCRUT5 data (Figure 7). For all three ensembles, we detect an overall warming signal also on a regional scale around the globe by comparing the climatologies 2020-1991 and 1920-1891 with a century apart (insets in Figure 7 and Figure S6). In particular, the warming patterns reconstructed from the three ensembles show a strong century warming signal in northern polar regions, where the original HadCRUT5 data set has missing data. Large areas in the Pacific also consistently show a warming between the two climatologies, despite the fact that the region is affected by strong ENSO variability. The infilled data in the sparsely observed Antarctica show a less strong, but more mixed warming signal as observed when reconstructed with the different ensembles. From the striking similarity in the reconstructed pattern, we conclude that MPI-GE CMIP6 allowed us to train a model with equivalent capabilities to MPI-GE CMIP5 but at a lower computational cost.

4 Summary and Conclusions

MPI-GE CMIP6 is a new 30-member single-model initial-condition large ensemble which power goes beyond its predecessor MPI-GE CMIP5 (Maher et al., 2019) in several aspects and allows for novel analyses with broad societal relevance:

First, MPI-GE CMIP6 provides 3-hourly, 6-hourly and daily model output that is together with its ensemble size well suited to investigate present and future changes in climate extremes, their drivers, and their changing characteristics across different emission scenarios. While several studies used MPI-GE CMIP5 to study present and future changes in climate extremes (e.g., Suarez-Gutierrez et al., 2020a, 2020b; Landrum & Holland, 2020), the high-frequency output of MPI-GE CMIP6 now allows one to also investigate the drivers and causal links of these changes which can be compared across different emission scenarios. For instance, we find from daily output that the recently observed Siberian and Pacific North American heatwaves will occur every year in 2071-2100 in high-emission scenarios but substantially less frequent in the low-emission scenarios. We further find from the 3-hourly output that the frequency of wind extremes is projected to decrease in tropical to mid-latitude oceans in all five emission scenarios. These findings illustrate that MPI-GE CMIP6 is specifically suited to investigate climate extremes and can be used to study high-impact events.

Second, MPI-GE CMIP6 provides the opportunity to compare the ensemble to high-resolution simulations of the same model version, including a 10-member ensemble of MPI-ESM-HR (1.0° atmosphere, 0.4° ocean), and a single member of MPI-ESM-XR (0.5° atmosphere, 0.4° ocean). While MPI-GE CMIP6 is not able to represent the unprecedented precipitation extreme in western Europe observed on 14^{th} of July 2021 and in northern Italy observed on 2^{nd} of October 2020, we find that these events are captured by high-resolution simulations of the same model version. This finding illustrates the benefit of comparing low-resolution SMILEs with high-frequency output to high-resolution simulations of the same model version for investigating regional climate extremes.

Third, MPI-GE CMIP6 provides historical simulations and the five emission scenarios SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5 which enable the investigation of different climate futures and the quantification of uncertainty from internal variability. We find that the frequencies of marine heatwaves and ocean acidity extremes are projected to substantially increase in all emissions scenarios, with substantial recovery by 2100 only under SSP1-1.9. Moreover, the ensemble simulations of the scenarios SSP1-1.9 and SSP1-2.6 specifically allow for quantifying irreducible uncertainty when aiming to limit global mean warming to 1.5°C or 2°C. We find that in MPI-GE CMIP6, even for the lowest emission scenario SSP1-1.9, which is consistent with the Paris Agreement pledges in this model, there is a non-zero chance to observe individual years above 1.5°C. With its good representation of internal variability in GSAT and its equilibrium climate sensitivity close to the central estimate of the AR6 assessment, MPI-GE CMIP6 as a single-model ensemble provides new opportunities to quantify uncertainty in when global warming thresholds might be crossed. Such analyses on irreducible uncertainty from internal variability are highly relevant for investigating transition pathways to carbon-neutral economies to meet the Paris Agreement pledges.

Fourth, MPI-GE CMIP6 is run with CMIP6 forcing and provides the opportunity to compare the ensemble to other SMILEs with CMIP6 forcing. This facilitates comparisons to the growing number of SMILEs. From comparing the respective scenarios from MPI-GE CMIP6 to the ones from its predecessor MPI-GE CMIP5, we find that the change from CMIP5 to CMIP6 forcing causes a slightly stronger climate response, in line with findings from other SMILEs (Wyser et al., 2020; Fyfe et al., 2021), primarily caused by the updated forcing in CMIP6. From combining MPI-GE CMIP6 with artificial intelligence, we find that 30 realisations have equivalent capabilities as the 100-member MPI-GE CMIP5 when training a model to infill surface temperature observations.

Overall, MPI-GE CMIP6 beneficially complements the number of available SMILEs by a unique combination of a moderate ensemble size, high-frequency model output, the full range of emission scenarios including the lower end, and the availability of high-resolution simulations of the same model version. Consequently, MPI-GE CMIP6 allows a better understanding of changes in climate variability and extremes, and to quantify related uncertainties. This improved quantification will help to better inform society on the likelihood of plausible changes in the climate system to occur, including climate extremes.

Acknowledgments

This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project IDs uo1075 and mh0469. We thank Nicola Maher for conducting a pre-submission review. This research was funded by the Max Planck Society (MPG) and the German Research Foundation (DFG) under Germany's Excellence Strategy—EXC 2037 'CLICCS—Climate, Climatic Change, and Society'—Project No.: 390683824, contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg, and the Cluster of Excellence CliSAP (EXC177), Universität Hamburg. D.O. received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 820829 (CONSTRAIN). S.B. and J.B. were supported by Copernicus Climate Change Service, funded by the EU, under Contract C3S2-370. G.B.-A. is supported by the German Ministry of Education and Research (BMBF) under the ClimXtreme project NA2EE (grant no. 01LP1902F). F.F., T.I., and H.L. received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 820989 (COMFORT). D.K. received funding from the project WAKOS – Wasser an den Küsten Ostfrieslands, which is funded by the German Federal Ministry of Education and Research (BMBF; grant no. 01LR2003A). L.S.-G. received funding from the German Ministry of Education and Research (BMBF) under the ClimXtreme project DecHeat (grant no. 01LP1901F) and from the European Union's Horizon Europe Framework Programme under the Marie Skłodowska-Curie grant agreement no. 101064940. The applied data standardisation tools and workflow have been developed within the BMBF funded project DICAD (grant no. 01LP1605A). E.P. received funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 101003876 (CLINT). Post-processing of data was performed with CDO (Schulzweida, 2022).

Open Research

The MPI-ESM1.2-LR coupled climate model is distributed via http://www.mpimet.mpg.de/. The simulation run scripts and code for reproducing the plots will be openly available through the publication repository of the Max Planck Society.

References

Adler, R., Sapiano, M., Huffman, G., Wang, J.-J., Gu, G., Bolvin, D., … Shin, D.-B. (2018). The Global Precipitation Climatology Project (GPCP) Monthly

Analysis (New Version 2.3) and a Review of 2017 Global Precipitation. *Atmosphere*, *9*(4), 2073-4433. doi: 10.3390/atmos9040138

Bevacqua, L., E.and Suarez-Gutierrez, Jézéquel, A., Lehner, F., Vrac, M., Yiou, P., & J., Z. (2023). Advancing our understanding of compound weather and climate events via large ensemble model simulations. *Nature Communications*, *14*(2145). doi: 10.1038/s41467-023-37847-5

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., ... Eyring, V. (2020). Quantifying progress across different cmip phases with the esmvaltool. *Journal of Geophysical Research: Atmospheres*, *125*(21), e2019JD032321. doi: 10.1029/2019JD032321

Bosilovich, M. G., Lucchesi, R., & Suarez, M. (2016). MERRA-2: File Specification. GMAO Office Note No. 9 (Version 1.1). , 73 pp.

Burger, F., John, J., & Frölicher, T. (2020). Increase in ocean acidity variability and extremes under increasing atmospheric CO2. *Biogeosciences*, *17*(18), 4633-4662. doi: 10.5194/bg-17-4633-2020

Chand, S. S., Walsh, K. J. E., Camargo, S. J., Kossin, J. P., Tory, K. J., Wehner, M. F., ... Murakami, H. (2022). Declining tropical cyclone frequency under global warming. *Nature Climate Change*, *12*(7), 655–661. doi: 10.1038/s41558-022-01388-4

Ciavarella, A., Cotterill, D., Stott, P., Kew, S., Philip, S., van Oldenborgh, G., ... Zolina, O. (2021). Prolonged siberian heat of 2020 almost impossible without human influence. *Climatic Change*, *166*(9), 1689–1713. doi: 10.1007/s10584-021-03052-w

Cowtan, K., & Way, R. G. (2014). Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1935–1944. doi: 10.1002/qj.2297

Craig, A., Valcke, S., & Coquart, L. (2017). Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0. *Geosci Model Dev*, *10*(9), 3297–3308. doi: 10.5194/gmd-10-3297-2017

Davolio, S., Vercellino, M., Miglietta, M.-M., Drago Pitura, L., Laviola, S., & Levizzani, V. (2023). The influence of an atmospheric river on a heavy precipitation event over the western Alps. *Weather and Climate Extremes*, *39*, 100542. doi: 10.1016/j.wace.2022.100542

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., ... Zhao, M. (2020). SPEAR: The Next Generation GFDL Modeling System for Seasonal to Multidecadal Prediction and Projection. *Journal of Advances in Modeling Earth Systems*, *12*(3), e2019MS001895. doi: 10.1029/2019MS001895

Deser, C., Lehner, F., & Rodgers, K. e. a. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nat. Clim. Chang.*, *10*, 277–286. doi: 10.1038/s41558-020-0731-2

Dunn, R. J. H., Alexander, L. V., Donat, M. G., Zhang, X., Bador, M., Herold, N., ... Bin Hj Yussof, M. N. (2020). Development of an Updated Global Land In Situ-Based Data Set of Temperature and Precipitation Extremes: HadEX3. *Journal of Geophysical Research: Atmospheres*, *125*(16), e2019JD032263. doi: 10.1029/2019JD032263

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, *9*(5), 1937–1958. doi: 10.5194/gmd-9-1937-2016

Feldmann, H., Früh, B., Schädler, G., Panitz, H.-J., Keuler, K., Jacob, D., & Lorenz, P. (2008). Evaluation of the precipitation for South-western Germany from high resolution simulations with regional climate models. *Meteorologische Zeitschrift*, *17*(4), 455–465. doi: 10.1127/0941-2948/2008/0295

Fetterer, F., Knowles, K., Meier, W., Savoie, M., & Windnagel, A. (2017). *Sea* *ice index, version 3.* National Snow and Ice Data Center. Retrieved from `https://nsidc.org/data/G02135/versions/3` doi: 10.7265/N5K072F8

Fischer, E., Sippel, S., & Knutti, R. (2021). Increasing probability of record-shattering climate extremes. *Nat. Clim. Chang.*, *11*, 689–695. doi: 10.1038/s41558-021-01092-9

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., ... Zhang, H. (2021). The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. , 923–1054. doi: 10.1017/9781009157896.009

Frajka-Williams, E., Moat, B., Smeed, D., Rayner, D., Johns, W., Baringer, M., ... Collins, J. (2021). Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2020 (v2020.1). NERC EDS British Oceanographic Data Centre NOC. doi: 10.5285/cc1e34b3-3385-662b-e053-6c86abc03444

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C., Hauck, J., ... others (2022). Global carbon budget 2021. *Earth System Science Data*, *14*(4), 1917–2005.

Fyfe, J. C., Kharin, V. V., Santer, B. D., Cole, J. N. S., & Gillett, N. P. (2021). Significant impact of forcing uncertainty in a large ensemble of climate model simulations. *Proceedings of the National Academy of Sciences*, *118*(23), e2016549118. doi: 10.1073/pnas.2016549118

Global Carbon Project. (2021). *Supplemental data of Global Carbon Budget 2021 (Version 1.0) [Data set]. Global Carbon Project.* doi: 10.18160/gcp-2021

Grazzini, F., Fragkoulidis, G., Teubler, F., Wirth, V., & Craig, G. C. (2021). Extreme precipitation events over northern Italy. Part II: Dynamical precursors. *Quarterly Journal of the Royal Meteorological Society*, *147*(735), 1237-1257. doi: 10.1002/qj.3969

Gruber, N., Boyd, P. W., Frölicher, T. L., & Vogt, M. (2021). Biogeochemical extremes and compound events in the ocean. *Nature*, *600*(7889), 395–407. doi: 10.1038/s41586-021-03981-7

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., ... Stössel, A. (2019). Max Planck Institute Earth System Model (MPI-ESM1.2) for the High-Resolution Model Intercomparison Project (HighResMIP). *Geoscientific Model Development*, *12*(7), 3241–3281. doi: 10.5194/gmd-12-3241-2019

Haarsma, R. J., Roberts, M. J., Vidale, P. L., Senior, C. A., Bellucci, A., Bao, Q., ... von Storch, J.-S. (2016). High Resolution Model Intercomparison Project (HighResMIP v1.0) for CMIP6. *Geoscientific Model Development*, *9*(11), 4185–4208. doi: 10.5194/gmd-9-4185-2016

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. doi: 10.1002/qj.3803

Hewitt, H., Fox-Kemper, B., & Pearson, B. e. a. (2022). The small scales of the ocean may hold the key to surprises. *Nat. Clim. Chang.*, *12*, 496–499. doi: 10.1038/s41558-022-01386-6

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C., ... Wernberg, T. (2016). A hierarchical approach to defining marine heatwaves. *Prog Oceanogr*, *141*, 227-238. doi: 10.1016/j.pocean.2015.12.014

Ibebuchi, C. C. (2022). Patterns of atmospheric circulation in Western Europe linked to heavy rainfall in Germany: preliminary analysis into the 2021 heavy rainfall episode. *Theoretical and Applied Climatology*, *148*(1), 269–283. doi: 10.1007/s00704-022-03945-5

Iles, C. E., Vautard, R., Strachan, J., Joussaume, S., Eggen, B. R., & Hewitt, C. D. (2020). The benefits of increasing resolution in global and regional climate simulations for European climate extremes. *Geoscientific Model Development*, *13*(11), 5583–5607. doi: 10.5194/gmd-13-5583-2020

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations. *J. Adv. Mod. Earth Sys.*, *5*(2), 287-315. doi: 10.1029/2012MS000178

Kadow, C., Hall, D. M., & Ulbrich, U. (2020). Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, *13*(6), 408–413. doi: https://doi.org/10.1038/s41561-020-0582-5

Kahraman, A., Kendon, E. J., Chan, S. C., & Fowler, H. J. (2021). Quasi-Stationary Intense Rainstorms Spread Across Europe Under Climate Change. *Geophysical Research Letters*, *48*(13), e2020GL092361. doi: 10.1029/2020GL092361

Kendon, E. J., Prein, A. F., Senior, C. A., & Stirling, A. (2021). Challenges and outlook for convection-permitting climate modelling. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2195), 20190547. doi: 10.1098/rsta.2019.0547

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G., Gocheva, A., … Petrovic, P. (2002). Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology*, *22*(12), 1441-1453. doi: 10.1002/joc.773

Kreienkamp, F., Philip, S. Y., Tradowsky, J. S., Kew, S. F., et al., & Wanders, N. (2021). *Rapid attribution of heavy rainfall events leading to the severe flooding in Western Europe during July 2021* (Tech. Rep.). World Weather Attribution. (Published online August 23, 2021)

Krieger, D., Krueger, O., Feser, F., Weisse, R., Tinz, B., & Storch, H. (2020). German Bight storm activity, 1897–2018. *International Journal of Climatology*. doi: 10.1002/joc.6837

Krueger, O., & von Storch, H. (2011). Evaluation of an air pressure–based proxy for storm activity. *Journal of Climate*, *24*(10), 2612 - 2619. doi: 10.1175/2011JCLI3913.1

Landrum, L., & Holland, M. (2020). Extremes become routine in an emerging new Arctic. *Nat. Clim. Chang.*(10), 1108—1115. doi: 10.1038/s41558-020-0892-z

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., … Zhou, T. (2021). Future Global Climate: Scenario-Based Projections and Near-Term Information. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. , 553–672. doi: 10.1017/9781009157896.006

Lehner, F., Deser, C., Maher, N., Marotzke, J., Fischer, E. M., Brunner, L., … Hawkins, E. (2020). Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6. *Earth System Dynamics*, *11*(2), 491–508. doi: 10.5194/esd-11-491-2020

Lin, P., Zhao, B., Wei, J., Liu, H., Zhang, W., Chen, X., … Zhou, T. (2022). The Super-large Ensemble Experiments of CAS FGOALS-g3. *Advances in Atmo-*

spheric Sciences. doi: https://doi.org/10.1007/s00376-022-1439-1

Liu, G., Reda, F. A., Shih, K. J., Wang, T.-C., Tao, A., & Catanzaro, B. (2018). Image Inpainting for Irregular Holes Using Partial Convolutions. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (pp. 89–105). Cham: Springer International Publishing.

Loughran, T. F., Boysen, L., Bastos, A., Hartung, K., Havermann, F., Li, H., ... Pongratz, J. (2021). Past and future climate variability uncertainties in the global carbon budget using the mpi grand ensemble. *Global Biogeochemical Cycles*, *35*(8), e2021GB007019. doi: 10.1029/2021GB007019

Luu, L. N., Vautard, R., Yiou, P., & Soubeyroux, J.-M. (2020). Evaluation of convection-permitting extreme precipitation simulations for the south of France. *Earth System Dynamics*, *13*(1), 687–702. doi: 10.5194/esd-13-687-2022

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., ... Marotzke, J. (2019). The Max Planck Institute Grand Ensemble: Enabling the Exploration of Climate System Variability. *J. Adv. Model. Earth Syst.*, *11*(7), 2050-2069. doi: 10.1029/2019MS001639

Marotzke, J. (2019). Quantifying the irreducible uncertainty in near-term climate projections. *WIREs Climate Change*, *10*(1), e563. doi: doi.org/10.1002/wcc.563

Marotzke, J., Milinski, S., & Jones, C. (2022). How close are we to 1.5°C or 2°C of global warming? *Weather*, *77*(4), 147-148. doi: 10.1002/wea.4174

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., ... Roeckner, E. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998-1038. doi: 10.1029/2018MS001400

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., ... Simpson, I. R. (2021). An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. *Journal of Geophysical Research: Atmospheres*, *126*(3). doi: 10.1029/2019JD032361

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., ... Marotzke, J. (2018). A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383–1413. doi: 10.1029/2017MS001217

Myhre, G., Alterskjær, K., Stjern, C. W., Hodnebrog, Ø., Marelle, L., Samset, B. H., ... Stohl, A. (2019). Frequency of extreme precipitation increases extensively with event rareness under global warming. *Scientific Reports*, *9*(1), 16063. doi: 10.1038/s41598-019-52277-4

Notz, D., & Community, S. (2020). Arctic Sea Ice in CMIP6. *Geophysical Research Letters*, *47*(10), e2019GL086749. doi: 10.1029/2019GL086749

Olonscheck, D., Rugenstein, M., & Marotzke, J. (2020). Broad consistency between observed and simulated trends in sea surface temperature patterns. *Geophysical Research Letters*, *47*(10), e2019GL086773. doi: 10.1029/2019GL086773

Olonscheck, D., Schurer, A., Lücke, L., & Hegerl, G. (2021). Large-scale emergence of regional changes in year-to-year temperature variability by the end of the 21st century. *Nat. Commun.*, *12*, 7237. doi: 10.1038/s41467-021-27515-x

Pendergrass, A. G., Knutti, R., Lehner, F., Deser, C., & Sanderson, B. M. (2017). Precipitation variability increases in a warmer climate. *Scientific Reports*, *7*(1), 17966. doi: 10.1038/s41598-017-17966-y

Perkins-Kirkpatrick, S., & Lewis, S. (2020). Increasing trends in regional heatwaves. *Nature Communications*, *11*(1), 3357. doi: 10.1038/s41467-020-16970-7

Philip, S. Y., Kew, S. F., van Oldenborgh, G. J., Anslow, F. S., Seneviratne, S. I., Vautard, R., ... Otto, F. E. L. (2022). Rapid attribution analysis of the extraordinary heat wave on the pacific coast of the us and canada in june 2021. *Earth System Dynamics*, *13*(4), 1689–1713. doi: 10.5194/esd-13-1689-2022

Pichelli, E., Coppola, E., & Sobolowski, S. e. a. (2021). The first multi-model ensemble of regional climate simulations at kilometer-scale resolution part 2: historical and future simulations of precipitation. *Climate Dynamics*, *56*, 3581—3602. doi: 10.1007/s00382-021-05657-4

Raymond, C., Suarez-Gutierrez, S., Kornhuber, K., Pascolini-Campbell, M., Sillmann, J., & Waliser, D. (2022). Increasing spatiotemporal proximity of heat and precipitation extremes in a warming world quantified by a large model ensemble. *Environmental Research Letters*, *17*(3), 035005. doi: 10.1088/1748-9326/ac5712

Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., ... Yeager, S. G. (2021). Ubiquity of human-induced changes in climate variability. *Earth System Dynamics*, *12*(4), 1393–1411. doi: 10.5194/esd-12-1393-2021

Schmidt, H., & von Storch, H. (1993). German Bight storms analysed. *Nature*, *365*(6449), 791. doi: 10.1038/365791a0

Schulzweida, U. (2022). CDO User Guide (2.1.0). *Zenodo*. doi: 10.5281/zenodo.7112925

Seneviratne, S., X. Zhang, M. A., Badi, W., Dereczynski, C., Luca, A. D., Ghosh, S., ... Zhou, B. (2021). Weather and Climate Extreme Events in a Changing Climate. In Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. , 1513–1766. doi: 10.1017/9781009157896.013

Slingo, J., Bates, P., Bauer, P., Belcher, S., Palmer, T., Stephens, G., ... Teutsch, G. (2022). Ambitious partnership needed for reliable climate prediction. *Nature Climate Change*, *12*(6), 499–503. doi: 10.1038/s41558-022-01384-8

Suarez-Gutierrez, L., Milinski, S., & Maher, N. (2021). Exploiting large ensembles for a better yet simpler climate model evaluation. *Clim Dyn*(57), 2557–2580. doi: 10.1007/s00382-021-05821-w

Suarez-Gutierrez, L., Müller, W., Li, C., & Marotzke, J. (2020a). Dynamical and thermodynamical drivers of variability in european summer heat extremes. *Clim Dyn*, *54*(9), 4351–4366. doi: 10.1007/s00382-020-05233-2

Suarez-Gutierrez, L., Müller, W., Li, C., & Marotzke, J. (2020b). Hotspots of extreme heat under global warming. *Clim Dyn*(55), 429–447. doi: 10.1007/s00382-020-05263-w

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., ... Winter, B. (2019). The canadian earth system model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. Retrieved 2020-08-22, from `10.5194/gmd-12-4823-2019`

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., ... Kimoto, M. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. Retrieved 2019-12-10, from `10.5194/gmd-12-2727-2019`

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An Overview of CMIP5 and the Experiment Design. *Bull. Amer. Meteor. Soc.*, *93*(4), 485–498. doi: 10.1175/BAMS-D-11-00094.1

Tuel, A., Steinfeld, D., Ali, S. M., Sprenger, M., & Martius, O. (2022). Large-scale drivers of persistent extreme weather during early summer 2021 in europe. *Geophysical Research Letters*, *49*(18), e2022GL099624. doi: 10.1029/2022GL099624

Wehner, M. F., Reed, K. A., Li, F., Prabhat, Bacmeister, J., Chen, C.-T., ... Jablonowski, C. (2014). The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1. *Journal of Advances in Modeling Earth Systems*, *6*(4), 980-997. doi: 10.1002/2013MS000276

White, R. H., Anderson, S., Booth, J. F., Braich, G., Draeger, C., Fei, C., ... West, G. (2023). The unprecedented Pacific Northwest heatwave of June 2021. *Nature Communications*, *14*(727), 2041-1723. doi: 10.1038/s41467-023-36289-3

Williams, A. I. L., & O'Gorman, P. A. (2022). Summer-Winter Contrast in the Response of Precipitation Extremes to Climate Change Over Northern Hemisphere Land. *Geophysical Research Letters*, *49*(10), e2021GL096531. doi: doi.org/10.1029/2021GL096531

Wills, R. C. J., Dong, Y., Proistosecu, C., Armour, K. C., & Battisti, D. S. (2022). Systematic climate model biases in the large-scale patterns of recent sea-surface temperature and sea-level pressure change. *Geophysical Research Letters*, *49*(17), e2022GL100011. doi: 10.1029/2022GL100011

Wood, R. R., Lehner, F., Pendergrass, A. G., & Schlunegger, S. (2021). Changes in precipitation variability across time scales in multiple global climate model large ensembles. *Environmental Research Letters*, *16*(8), 084022. doi: 10.1088/1748-9326/ac10dd

Wyser, K., Koenigk, T., Fladrich, U., Fuentes-Franco, R., Karami, M. P., & Kruschke, T. (2021). The SMHI large ensemble (SMHI-LENS) with EC-Earth3.3.1. *Geoscientific Model Development*, *14*(7), 4781–4796. doi: 10.5194/gmd-14-4781-2021

Wyser, K., van Noije, T., Yang, S., von Hardenberg, J., O'Donnell, D., & Döscher, R. (2020). On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6. *Geoscientific Model Development*, *13*(8), 3465–3474. doi: 10.5194/gmd-13-3465-2020

Zeder, J., & Fischer, E. M. (2020). Observed extreme precipitation trends and scaling in Central Europe. *Weather and Climate Extremes*, *29*, 100266. doi: 10.1016/j.wace.2020.100266

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., ... Srbinovsky, J. (2020). The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, *70*, 193–214. doi: 10.1071/ES19035

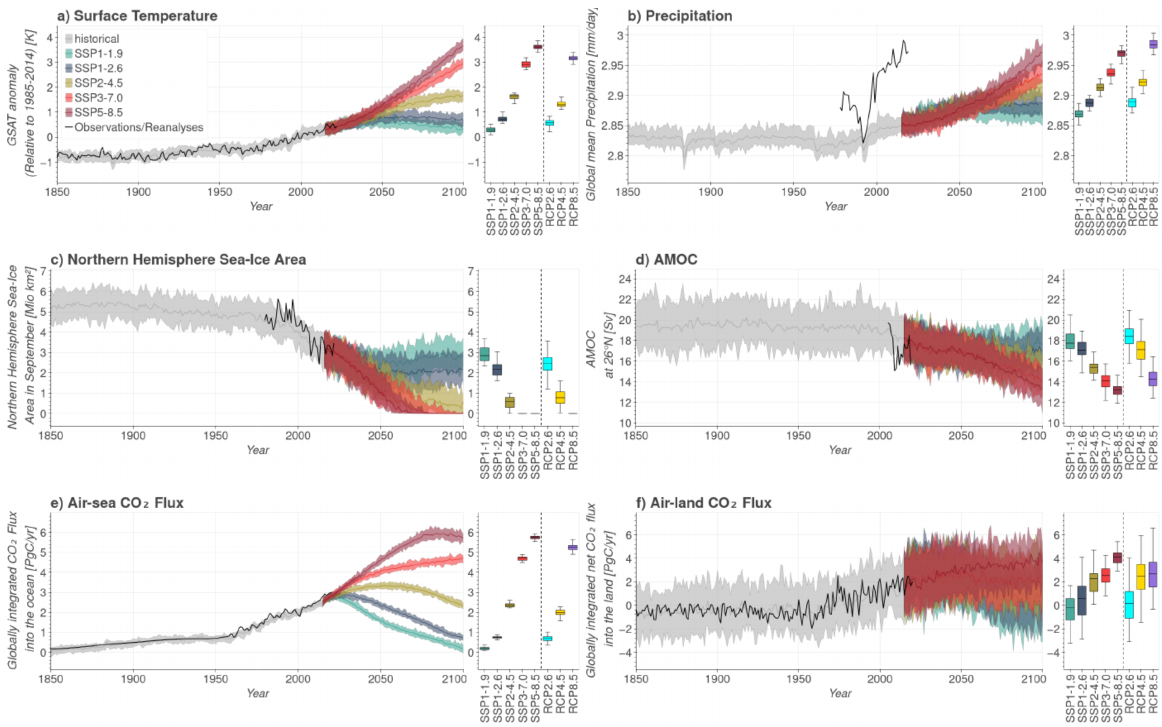

Figure 1: **Comparison of key climate quantities of MPI-GE CMIP6 to observations or reanalyses and MPI-GE CMIP5.** Ensemble spread (shading) and ensemble mean (thick lines) for the historical simulations (grey), and the five emission scenarios SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5. Right hand-side panels show the projected mean and range in year 2099 for the different scenarios of MPI-GE CMIP6 (30 realisations) and MPI-GE CMIP5 (100 realisations). Shown for **a)** global mean near-surface air temperature (GSAT) anomalies (relative to 1985–2014), **b)** global mean precipitation, **c)** Northern Hemisphere sea-ice area in September, **d)** Atlantic Meridional Overturning Circulation (AMOC), **e)** globally integrated CO_2 flux into the ocean and **f)** globally integrated net CO_2 flux into the land. Thick black lines show observations or reanalyses, specifically in **a)** HadCRUT5 (Morice et al., 2021), **b)** ERA5 (Hersbach et al., 2020), **c)** Sea-Ice Index (Fetterer et al., 2017), **d)** RAPID (Frajka-Williams et al., 2021), **e,f)** Global Carbon Project (Global Carbon Project, 2021; Friedlingstein et al., 2022).

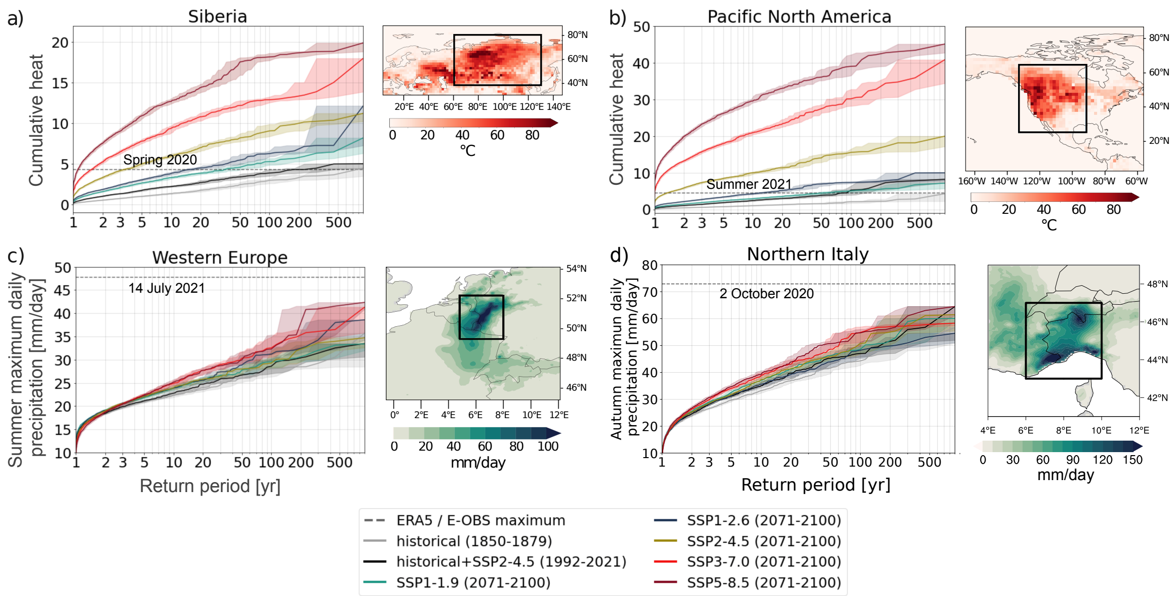

Figure 2: **Return periods from MPI-GE CMIP6 for recently observed heat and precipitation extremes for different emission scenarios.** Return periods for **a-b)** cumulative heat scaled with respect to climatology for **a)** spring (MAM) 2020 Siberian heatwave and **b)** summer (JJA) 2021 Pacific North American heatwave, and **c-d)** seasonal maximum daily precipitation for **c)** western Europe in summer (JJA) and **d)** northern Italy in autumn for the historical climate (1850-1879, grey), the current climate (1992-2021, black), and the five SSP scenarios for the period 2071-2100 (coloured). Shading denotes 95% confidence intervals calculated by bootstrapping with re-sampling. The horizontal dashed line in **a)** and **b)** marks the maximum cumulative heat as calculated from ERA5, and in **c)** and **d)** the observed maximum daily precipitation of the respective season from E-OBS (Klein Tank et al., 2002). The observed spatial pattern of these events is shown as maps in **a)** and **b)** for cumulative heat for spring 2020 and summer 2021, respectively, and in **c)** and **d)** for precipitation on 14th of July 2021 and 2nd of October 2020, respectively. Black boxes mark the regions of interest used for averaging.

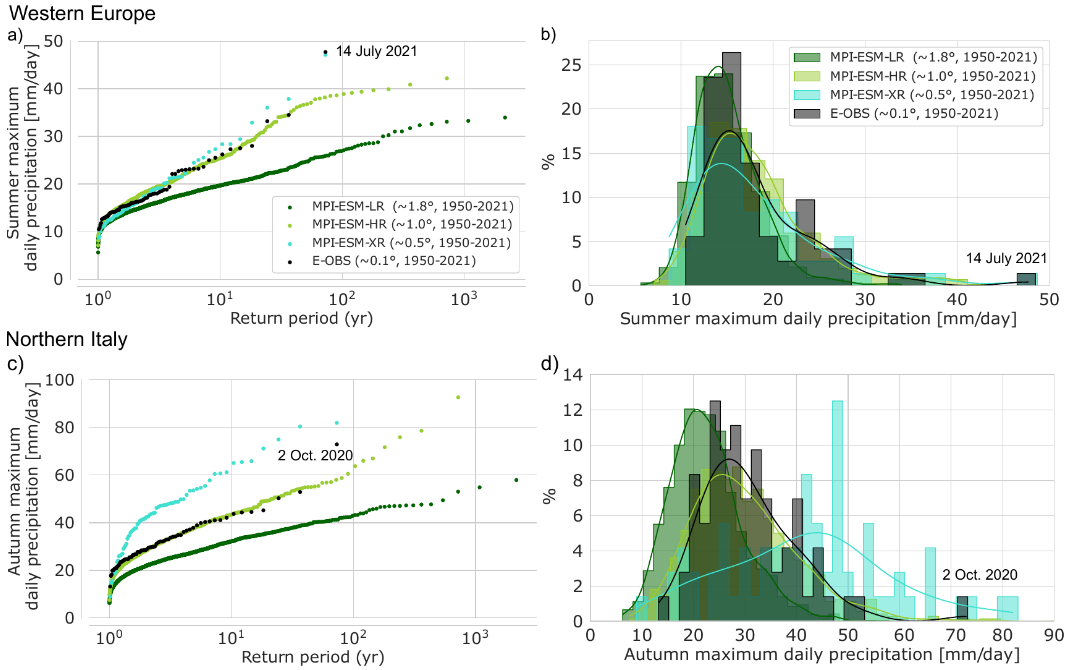

Figure 3: **Representation of precipitation extremes dependent on model resolution. a-b)** Comparison of summer (JJA) maximum daily precipitation averaged across the western European box shown in Fig. 2c from 1950-2021 in three model resolutions from MPI-ESM1.2 and in observations shown as **a)** return periods and **b)** probability density functions. **c-d)** Comparison of autumn (SON) maximum daily precipitation averaged across the northern Italy box shown in Fig. 2d from 1950-2021 in three model resolutions from MPI-ESM1.2 and in observations shown as **c)** return periods and **d)** probability density functions. Note that the return periods are calculated empirically. Values of all summers or autumns, respectively, and all realisations are merged for each ensemble. Further note that MPI-ESM-LR is based on 30 realisations, MPI-ESM-HR on 10 realisations and MPI-ESM-XR and the observed record on only a single realisation. The sample size of MPI-ESM-HR and MPI-ESM-XR might be insufficient to determine return levels above a few years robustly. The domain-averaged maximum daily precipitation of the western European extreme event on 14$^{\text{th}}$ of July 2021 is 47.7 mm, and that of the event in northern Italy on 2$^{\text{nd}}$ of October 2020 is 72.9 mm.

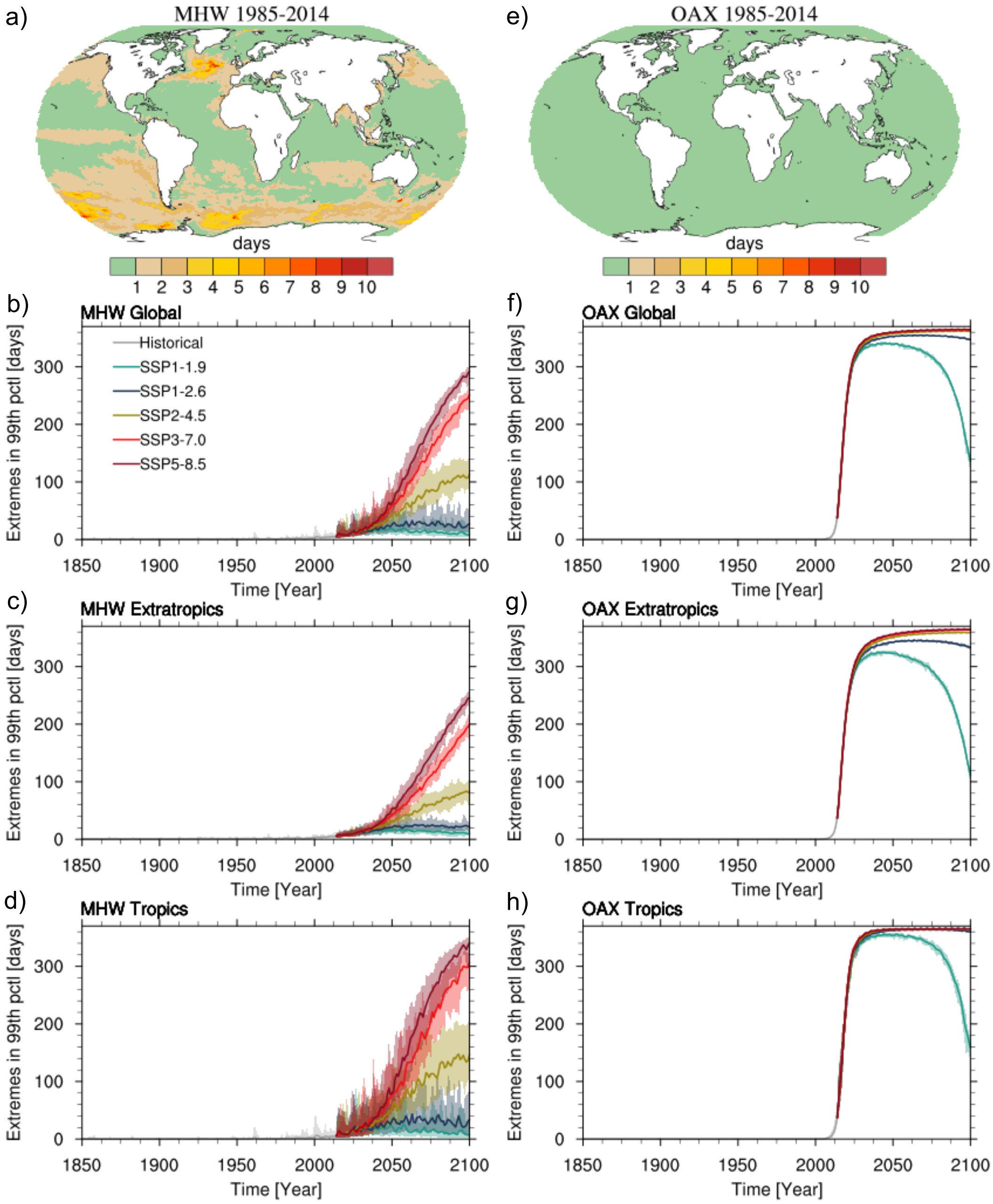

Figure 4: **Present and future frequency of marine heatwaves and ocean acidity extremes.** Maps of **a)** the ensemble mean number of marine heatwave (MHW) days per year and **e)** the number of ocean acidity extreme event (OAX) days per year in the reference period 1985-2014, based on the 99th percentile of daily mean sea surface temperature, and of daily mean surface hydrogen ion concentration, respectively. **b-d)** Globally and regionally averaged number of MHW days per year (global, extratropics: outside of 30°N/30°S, tropics: within 30°N/30°S) for the historical period 1850-2014 (grey), and scenarios SSP1-1.9 (green), SSP1-2.6 (blue), SSP2-4.5 (yellow), SSP3-7.0 (red), SSP5-8.5 (purple) for the period 2015-2100. The shadings cover the ensemble spread, thick lines show the 20-member ensemble mean. **f-h)** Globally and regionally averaged number of OAX days per year and region, similar to **b-d)**.

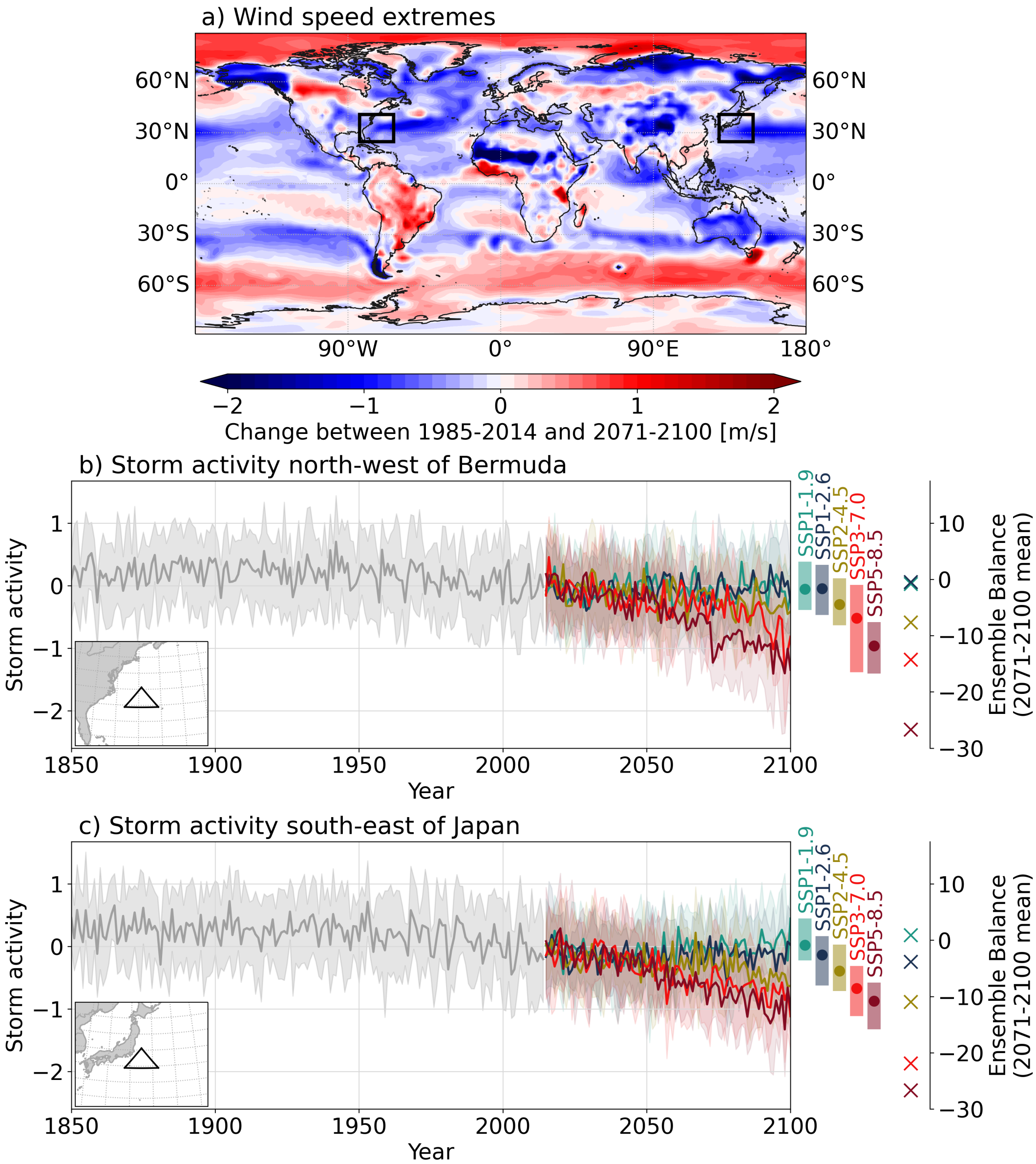

Figure 5: **Projected changes in near-surface wind speed and storm activity. a)** Absolute change in ensemble mean 95th annual percentiles of surface wind speed between 1985-2014 and 2071-2100, based on SSP5-8.5 forcing. Black circles mark regions for which storm activity has been calculated. Maps for the other four SSP scenarios are shown in Figure S5. **b-c)** Ensemble mean storm activity (thick lines) and interquartile range (shading) for the historical simulations (grey) and the five scenarios (coloured) over **b)** the Atlantic Ocean north-west of Bermuda and **c)** the Pacific Ocean south-east of Japan. Coloured dots and bars indicate the 2071-2100 average and range of the ensemble mean for each scenario, and crosses show the 2071-2100 mean ensemble balance.

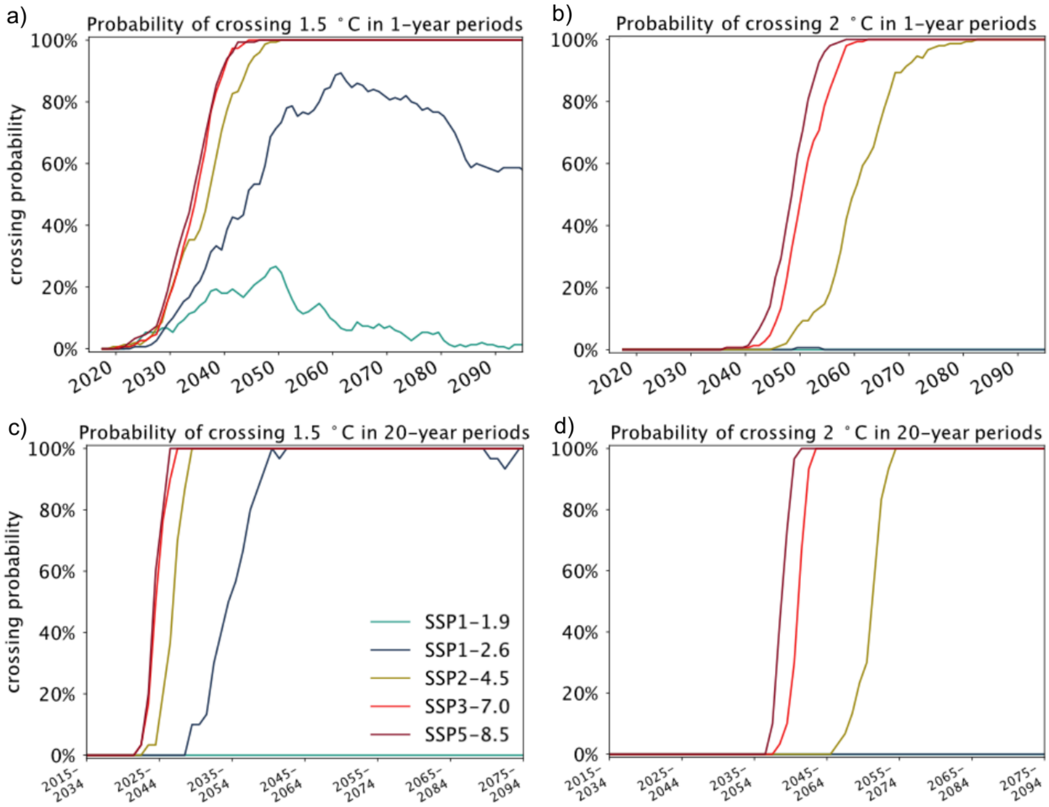

Figure 6: **Probability of crossing Paris Agreement global warming limits.** Probability of crossing **a)** 1.5°C and **b)** 2°C in a single year, and **c)** 1.5°C and **d)** 2°C in 20-year averages for the different emission scenarios until 2100. The crossing probability is defined as the fraction of the 30 realisations that cross the temperature threshold relative to the reference period 1850-1900. In **c,d)**, the 20-year mean GSAT is plotted against the central year of that 20-year period.

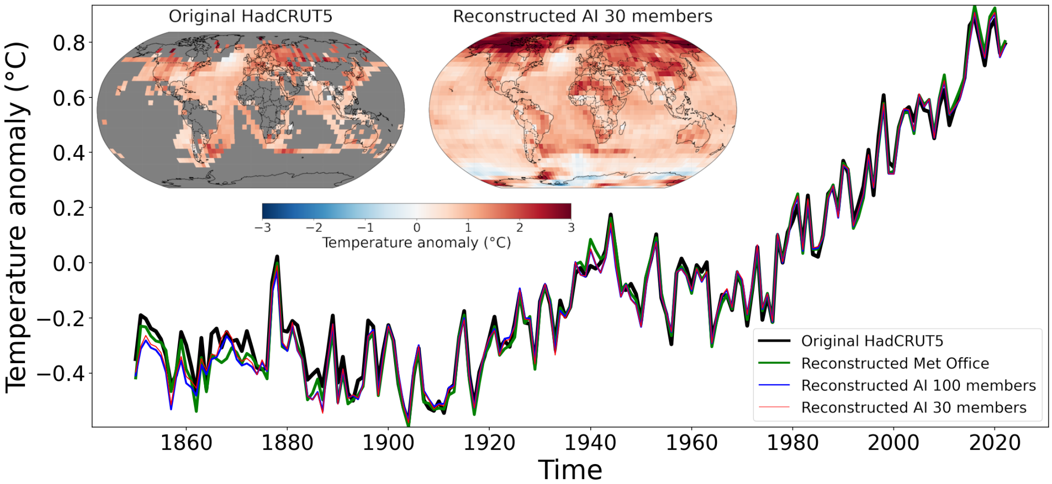

Figure 7: **Comparison of MPI-GE CMIP6 vs MPI-GE CMIP5 for infilling observations of surface temperature with artificial intelligence.** Annual global mean anomaly temperature with respect to the 1961–1990 climatology obtained by using: the gridded original "non-infilled" HadCRUT5 data set (black curve), the partially reconstructed HadCRUT5 data set from the Met Office (Morice et al., 2021), the fully reconstructed HadCRUT5 data set obtained with the AI 100 members model (blue curve, using MPI-GE CMIP5 (Maher et al., 2019)), the fully reconstructed HadCRUT5 obtained with our AI 30 members model (red curve, using MPI-GE CMIP6). Insets: 2020-1991 climatology referenced to the 1920-1891 climatology. Left inset: Original HadCRUT5 data set where gray pixels indicate missing values. Mean values have been computed only for grid points containing at least 70% of valid values for the considered time period. Right inset: Spatial reconstruction of the HadCRUT5 data set using the AI 30 members model.

Supporting Information for "The new Max Planck Institute Grand Ensemble with CMIP6 forcing and high-frequency model output"

Dirk Olonscheck[1], Laura Suarez-Gutierrez[1,2,3], Sebastian Milinski[1,4], Goratz Beobide-Arsuaga[5,6], Johanna Baehr[5], Friederike Fröb[7], Lara Hellmich[1,6], Tatiana Ilyina[1], Christopher Kadow[8], Daniel Krieger[6,9], Hongmei Li[1], Jochem Marotzke[1,5], Étienne Plésiat[8], Martin Schupfner[8], Fabian Wachsmann[8], Karl-Hermann Wieners[1], Sebastian Brune[5]

[1]Max Planck Institute for Meteorology, Hamburg, Germany

[2]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

[3]Institut Pierre-Simon Laplace, CNRS, Paris, France

[4]European Centre for Medium-Range Weather Forecasts, Bonn, Germany

[5]Center for Earth System Research and Sustainability, Universität Hamburg, Hamburg, Germany

[6]International Max Planck Research School on Earth System Modelling, Hamburg, Germany

[7]Geophysical Institute, University of Bergen, and Bjerknes Centre for Climate Research, Bergen, Norway

[8]German Climate Computing Centre (DKRZ), Hamburg, Germany

[9]Helmholtz-Zentrum Hereon, Geesthacht, Germany

Contents of this file

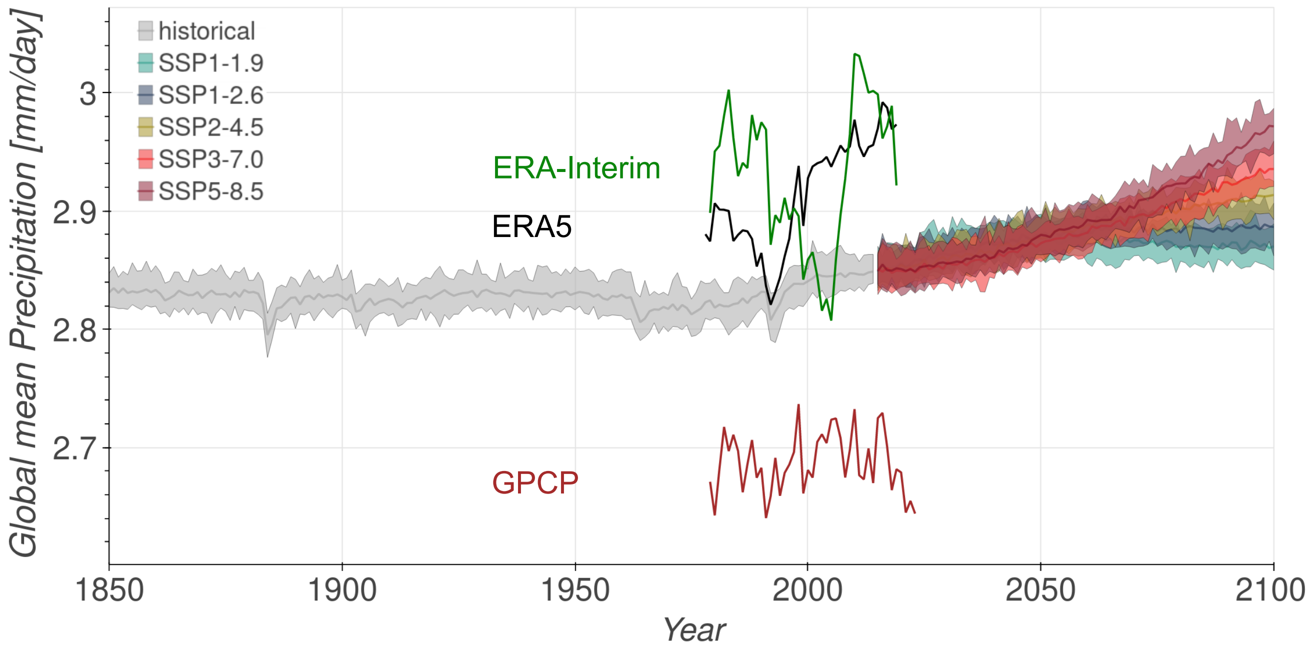

Figure S1. Global mean precipitation in MPI-GE CMIP6 compared to different reanalyses and observations. Same as Figure 1b) but showing both ERA5, ERA-Interim and the observational product of the Global Precipitation Climatology Project (GPCP) version 2.3.

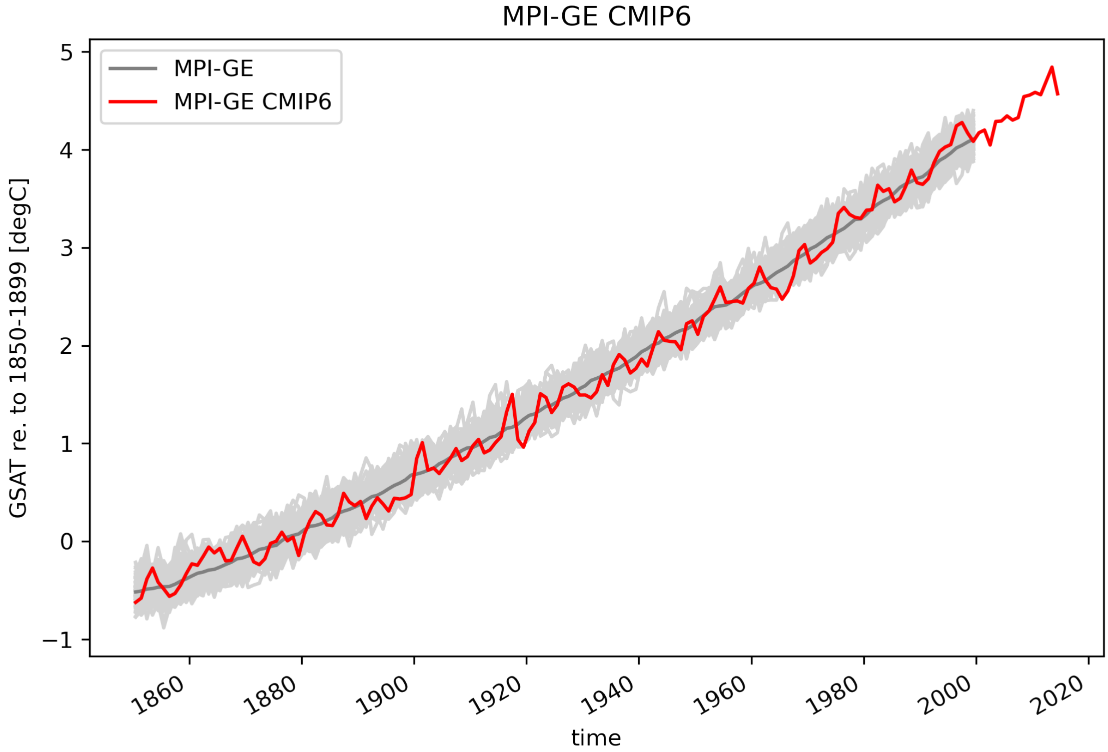

Figure S2. Comparison of the global mean temperature response of MPI-GE CMIP5 and MPI-GE CMIP6 to a 1%CO_2 increase per year relative to 1850-1899. The 100 realisations of MPI-GE CMIP5 are shown in light grey and the ensemble mean in dark grey. A single realisation of MPI-GE CMIP6 is shown in red. Note that the 100 realisations for the historical period of MPI-GE CMIP5 end in year 2005.

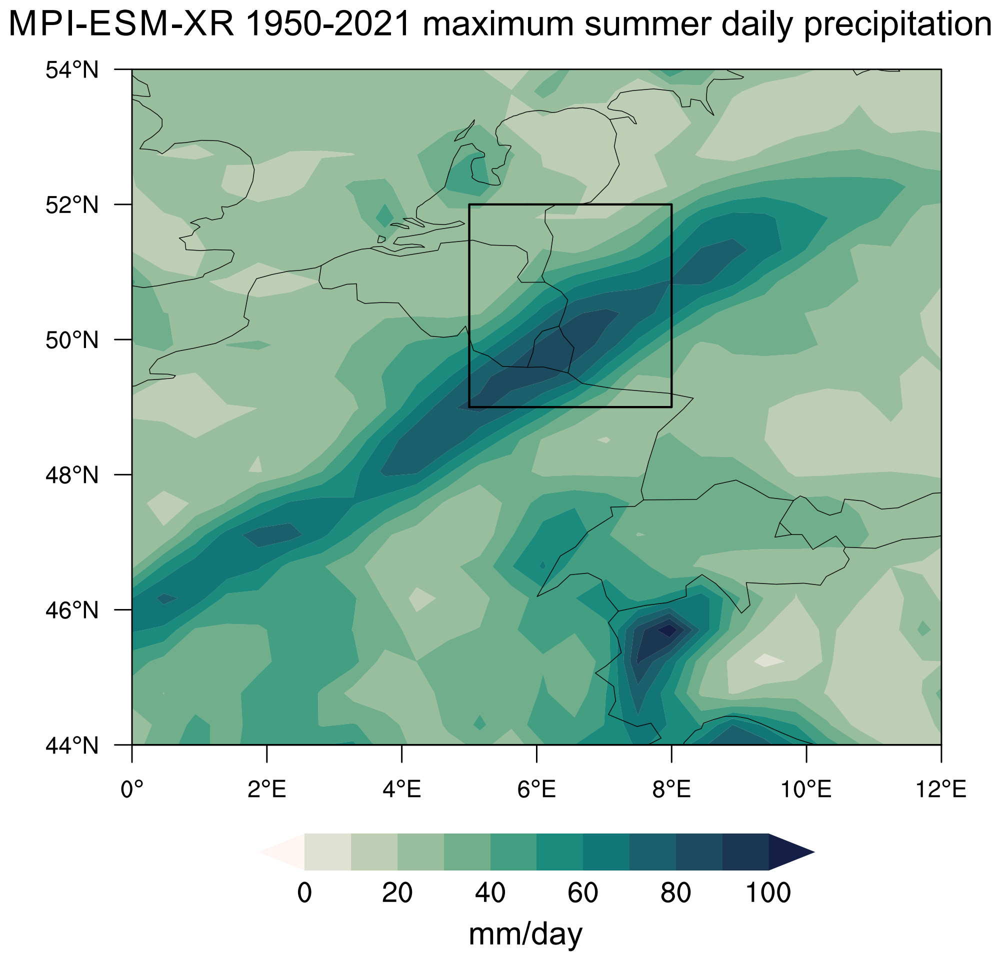

Figure S3. Spatial pattern of the maximum daily summer precipitation in western Europe between 1950-2021 as simulated by MPI-ESM-XR. The black box marks the region of interest averaged for Figure 2 and 3.

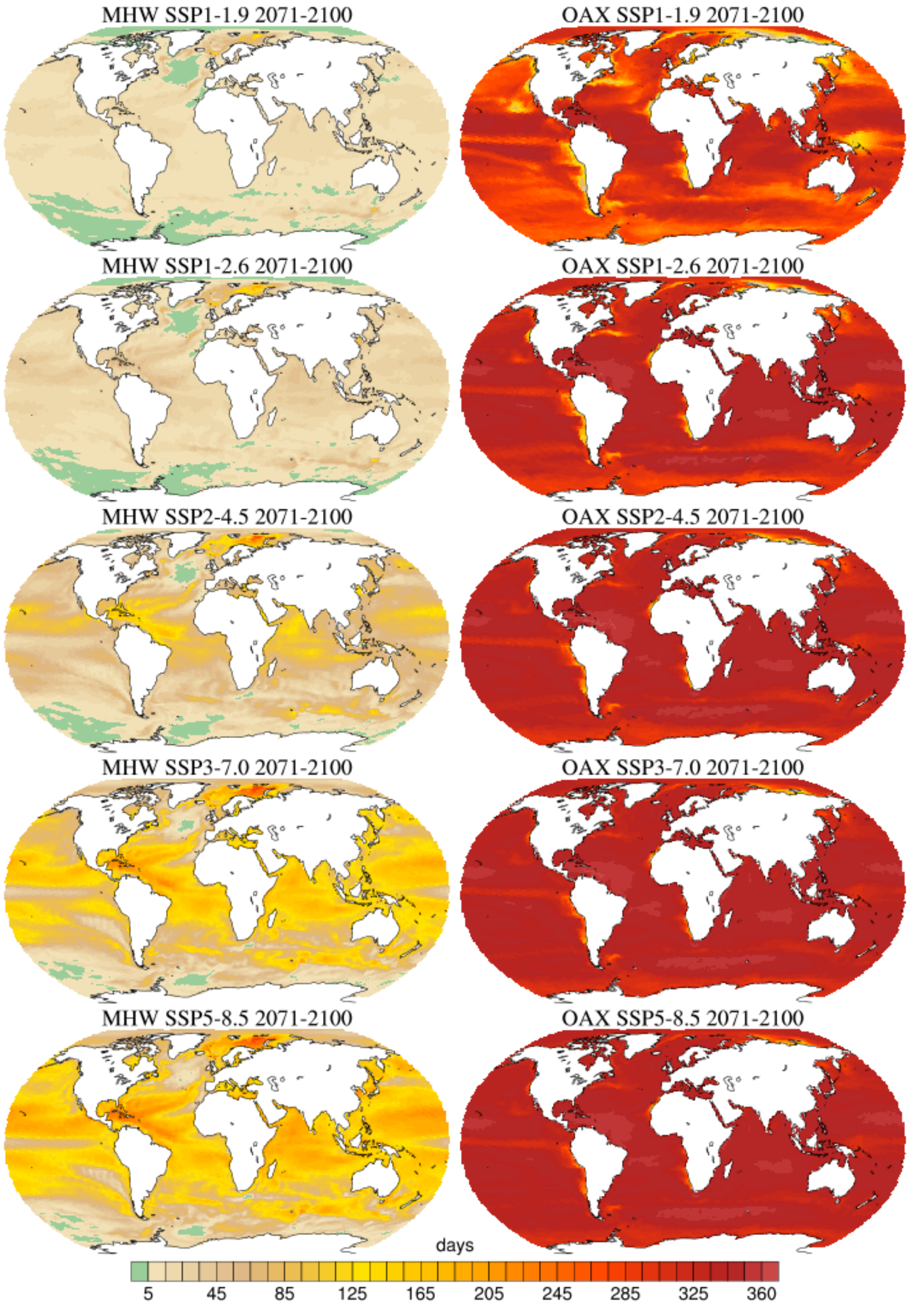

Figure S4. Spatial distribution of marine heat waves (MHW) and ocean acidity extremes (OAX) for different emission scenarios. Ensemble mean number of MHW days per year (left panels) and number of OAX days per year (right panels) during 2071-2100 under the emission scenarios SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0, and SSP5-8.5. The MHW and OAX are defined based on the 99$^{\text{th}}$ percentile of daily mean sea surface temperature and of daily mean surface hydrogen ion concentration, respectively.

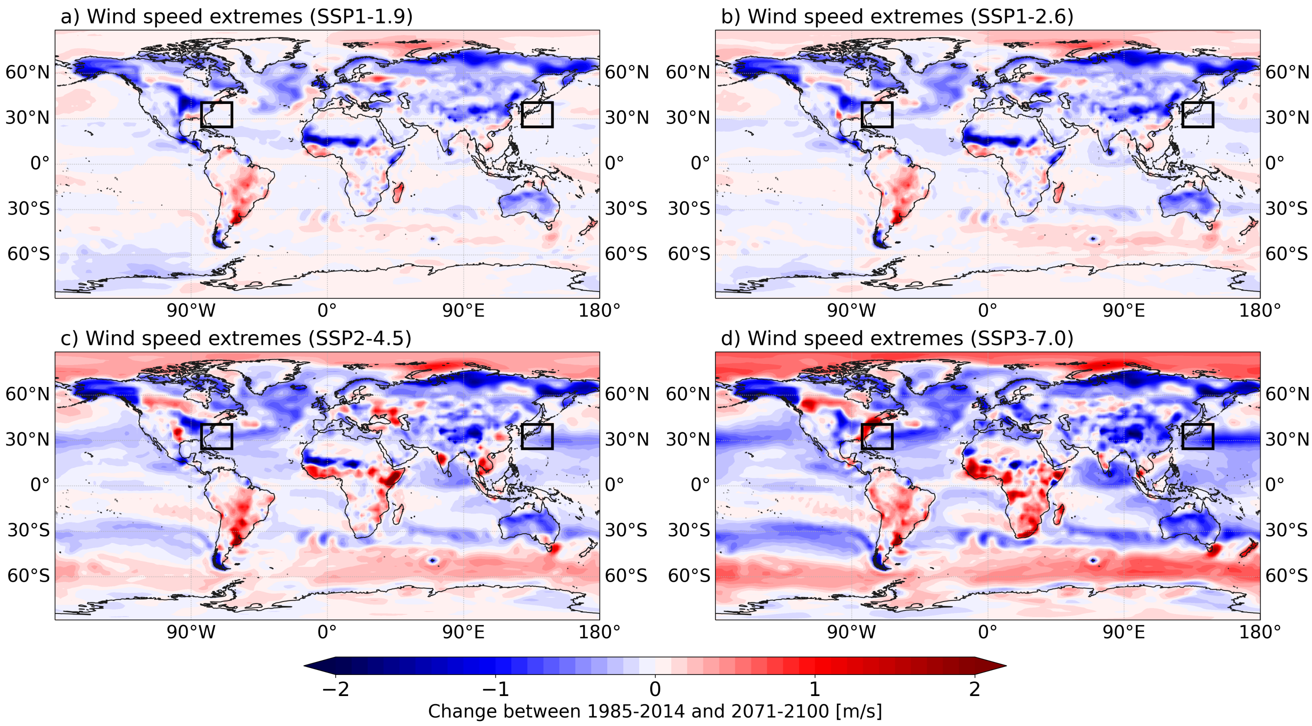

Figure S5. Projected changes in near-surface wind speed for lower-emission scenarios. Absolute change in ensemble mean 95$^{\text{th}}$ annual percentiles of surface wind speed between 1985-2014 and 2071-2100, based on **a)** SSP1-1.9, **b)** SSP1-2.6, **c)** SSP2-4.5, **d)** SSP3-7.0 forcing. Black rectangles mark regions for which storm activity has been calculated.

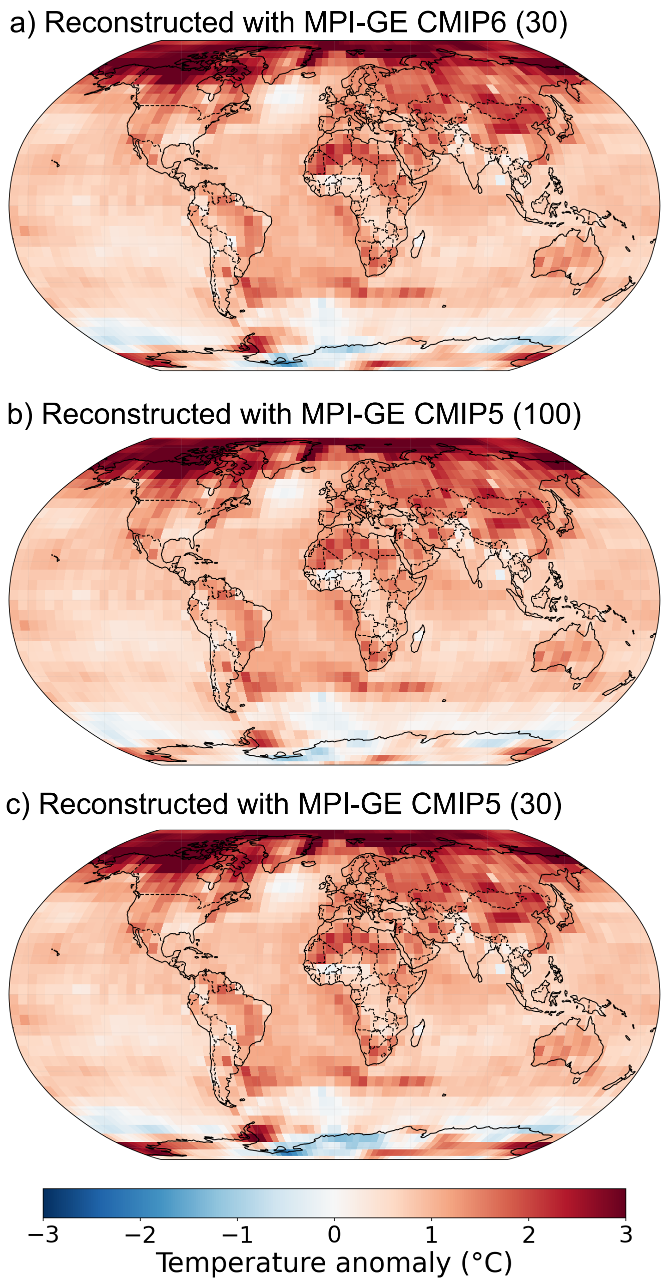

Figure S6. Comparison of using MPI-GE CMIP6 and MPI-GE CMIP5 to infill observations of surface temperature with artificial intelligence. Spatial reconstruction of the HadCRUT5 data set using **a)** the AI 30 members model based on MPI-GE CMIP6, **b)** the AI 100 members model based on MPI-GE CMIP5, and **c)** the AI 30 members model based on a first 30 members of MPI-GE CMIP5.

Data as listed in the following tables can be accessed either via DKRZ ESGF server or DKRZ WDCC long term archive (DKRZ LTA):

- ESGF: https://esgf-data.dkrz.de/search/cmip6-dkrz/
- DKRZ LTA 3hourly: http://hdl.handle.net/21.14106/5bb56765ffe486031cd6600a3d34ba3ad99c7f20
- DKRZ LTA 6hourly: http://hdl.handle.net/21.14106/b61690b4d0080648815e2ceba91f5a764a3addc3
- DKRZ LTA daily: http://hdl.handle.net/21.14106/1ce9699e340e6c46f4b34626bae2b65714696c56

Table S1: Parameters with daily output on ESGF available for all 30 realisations.

name	parameter long name	unit	level
daily atmosphere / land			
clt	Total Cloud Cover Percentage	%	1
cnc	Canopy Covered Area Percentage	%	1
es	Bare Soil Evaporation	kg m-2 s-1	1
hfls	Surface Upward Latent Heat Flux	W m-2	1
hfss	Surface Upward Sensible Heat Flux	W m-2	1
hur	Relative Humidity	%	47
hurs	Near-Surface Relative Humidity	%	1
hursmax	Daily Maximum Near-Surface Relative Humidity	%	1
hursmin	Daily Minimum Near-Surface Relative Humidity	%	1
hus	Specific Humidity	1	47
hus850	Specific Humidity at 850hPa	1	1
huss	Near-Surface Specific Humidity	1	1
lai	Leaf Area Index	1	1
mlotst	Ocean Mixed Layer Thickness Defined by Sigma T	m	1
mrro	Total Runoff	kg m-2 s-1	1
mrso	Total Soil Moisture Content	kg m-2	1
mrsol	Total Water Content of Soil Layer	kg m-2	1
mrsos	Moisture in Upper Portion of Soil Column	kg m-2	1
od550aer	Ambient Aerosol Optical Thickness at 550nm	1	1
pr	Precipitation	kg m-2 s-1	1
prc	Convective Precipitation	kg m-2 s-1	1
prsn	Snowfall Flux	kg m-2 s-1	1
ps	Surface Air Pressure	Pa	1
psl	Sea Level Pressure	Pa	1
rlds	Surface Downwelling Longwave Radiation	W m-2	1
rldscs	Surface Downwelling Clear-Sky Longwave Radiation	W m-2	1
rlus	Surface Upwelling Longwave Radiation	W m-2	1
rlut	TOA Outgoing Longwave Radiation	W m-2	1
rlutcs	TOA Outgoing Clear-Sky Longwave Radiation	W m-2	1
rsds	Surface Downwelling Shortwave Radiation	W m-2	1
rsdscs	Surface Downwelling Clear-Sky Shortwave Radiation	W m-2	1
rsdt	TOA Incident Shortwave Radiation	W m-2	1
rsus	Surface Upwelling Shortwave Radiation	W m-2	1
rsuscs	Surface Upwelling Clear-Sky Shortwave Radiation	W m-2	1
rsut	TOA Outgoing Shortwave Radiation	W m-2	1
rsutcs	TOA Outgoing Clear-Sky Shortwave Radiation	W m-2	1
rzwc	Root Zone Soil Moisture	kg m-2	1
sbl	Surface Snow and Ice Sublimation Flux	kg m-2 s-1	1
sfcWind	Daily-Mean Near-Surface Wind Speed	m s-1	1
sfcWindmax	Daily Maximum Near-Surface Wind Speed	m s-1	1
snc	Snow Area Percentage	%	1
snw	Surface Snow Amount	kg m-2	1

Continued on next page

Table S1 – continued from previous page

name	parameter long name	unit	level
snwc	Snow water equivalent intercepted by the vegetation	kg m-2	1
ta	Air Temperature	K	47
ta500	Air Temperature at 500hPa	K	1
ta700	Air Temperature at 700hPa	K	1
ta850	Air Temperature at 850hPa	K	1
tas	Near-Surface Air Temperature	K	1
tasmax	Daily Maximum Near-Surface Air Temperature	K	1
tasmin	Daily Minimum Near-Surface Air Temperature	K	1
tauu	Surface Downward Eastward Wind Stress	Pa	1
tauv	Surface Downward Northward Wind Stress	Pa	1
tdps	2m Dewpoint Temperature	K	1
tr	Surface Radiative Temperature	K	1
ts	Surface Temperature	K	1
tsl	Temperature of Soil	K	1
ua	Eastward Wind	m s-1	47
ua10	Eastward Wind at 10hPa	m s-1	1
uas	Eastward Near-Surface Wind	m s-1	1
va	Northward Wind	m s-1	47
vas	Northward Near-Surface Wind	m s-1	1
wap	Omega (=dp/dt)	Pa s-1	47
wap500	Pressure Tendency	Pa s-1	1
zg	Geopotential Height	m	47
zg10	Geopotential Height at 10hPa	m	1
zg100	Geopotential Height at 100hPa	m	1
zg1000	Geopotential Height at 1000hPa	m	1
zg500	Geopotential Height at 500hPa	m	1
daily ocean / sea ice / biogeochem			
chlos	Surface Mass Concentration of Total Phytoplankton Expressed as Chlorophyll in Sea Water	kg m-3	1
omldamax	Mean Daily Maximum Ocean Mixed Layer Thickness Defined by Mixing Scheme	m	1
phycos	Sea Surface Phytoplankton Carbon Concentration	mol m-3	1
siconc	Sea-Ice Area Percentage (Ocean Grid)	%	1
sisnthick	Snow Thickness	m	1
sispeed	Sea-Ice Speed	m s-1	1
sithick	Sea Ice Thickness	m	1
sitimefrac	Fraction of Time Steps with Sea Ice	1	1
siu	X-Component of Sea-Ice Velocity	m s-1	1
siv	Y-Component of Sea-Ice Velocity	m s-1	1
sos	Sea Surface Salinity	0.001	1
sossq	Square of Sea Surface Salinity	1.00E-06	1
t20d	Depth of 20 degree Celsius Isotherm	m	1
tos	Sea Surface Temperature	degC	1
tossq	Square of Sea Surface Temperature	degC2	1

Table S2: Parameters with 3-hourly output on either ESGF or *DKRZ LTA (*)* for any of the 30 realisations.

name	**parameter long name**	**unit**	**level**	**r1-r10**	**r11-r30**
atmosphere / land					
clt	Total Cloud Cover Percentage	%	1	x	
hfls	Surface Upward Latent Heat Flux	W m-2	1	x	
hfss	Surface Upward Sensible Heat Flux	W m-2	1	x	
hus	Specific Humidity	1	47	x	
huss	Near-Surface Specific Humidity	1	1	x	
mrro	Total Runoff	kg m-2 s-1	1	x	x
mrsos	Moisture in Upper Portion of Soil Column	kg m-2	1	x	
pr	Precipitation	kg m-2 s-1	1	x	
prc	Convective Precipitation	kg m-2 s-1	1	x	
prra	Rainfall Flux	kg m-2 s-1	1	x	
prsn	Snowfall Flux	kg m-2 s-1	1	x	
ps	Surface Air Pressure	Pa	1	x	
psl	Sea Level Pressure	Pa	1	x	x
rlds	Surface Downwelling Longwave Radiation	W m-2	1	x	
rldscs	Surface Downwelling Clear-Sky Longwave Radiation	W m-2	1	x	
rlus	Surface Upwelling Longwave Radiation	W m-2	1	x	
rlut	TOA Outgoing Longwave Radiation	W m-2	1	x	
rlutcs	TOA Outgoing Clear-Sky Longwave Radiation	W m-2	1	x	
rsds	Surface Downwelling Shortwave Radiation	W m-2	1	x	
rsdscs	Surface Downwelling Clear-Sky Shortwave Radiation	W m-2	1	x	
rsdt	TOA Incident Shortwave Radiation	W m-2	1	x	
rsucs	Upwelling Clear-Sky Shortwave Radiation	W m-2	48	x	
rsus	Surface Upwelling Shortwave Radiation	W m-2	1	x	
rsuscs	Surface Upwelling Clear-Sky Shortwave Radiation	W m-2	1	x	
rsut	TOA Outgoing Shortwave Radiation	W m-2	1	x	
rsutcs	TOA Outgoing Clear-Sky Shortwave Radiation	W m-2	1	x	
sfcWind	Near-Surface Wind Speed	m s-1	1	x	x
ta	Air Temperature	K	47	x	
tas	Near-Surface Air Temperature	K	1	x	*x**
ua	Eastward Wind	m s-1	7	x	
uas	Eastward Near-Surface Wind	m s-1	1	x	x
va	Northward Wind	m s-1	7	x	
vas	Northward Near-Surface Wind	m s-1	1	x	x
wap	Omega (=dp/dt)	Pa s-1	7	x	
ocean / sea ice / biogeochem					
tos	Sea Surface Temperature	degC	1	x	

Table S3: Parameters with 6-hourly output on either ESGF or *DKRZ LTA (*)* for any of the 30 realisations.

name	**parameter long name**	**unit**	**level**	**r1-r10**	**r11-r30**
atmosphere / land					
*hur**	*Relative Humidity**	*1**	*47**		*r11**
hurs	Near-Surface Relative Humidity	%	1	x	x
hus	Specific Humidity	1	47	x	x
huss	Near-Surface Specific Humidity	1	1	x	x
mrsol	Total Water Content of Soil Layer	kg m-2	5	x	x
mrsos	Moisture in Upper Portion of Soil Column	kg m-2	1	x	x
pr	Precipitation	kg m-2 s-1	1	x	x
ps	Surface Air Pressure	Pa	1	x	x
psl	Sea Level Pressure	Pa	1	x	x
sfcWind	Near-Surface Wind Speed	m s-1	1		x
snw	Surface Snow Amount	kg m-2	1		x
ta	Air Temperature	K	47	x	x
tas	Near-Surface Air Temperature	K	1	x	x
ts	Surface Temperature	K	1		x
tsl	Temperature of Soil	K	1	x	x
ua	Eastward Wind	m s-1	47	x	x
uas	Eastward Near-Surface Wind	m s-1	1	x	x
va	Northward Wind	m s-1	47	x	x
vas	Northward Near-Surface Wind	m s-1	1	x	x
wap	Omega (=dp/dt)	Pa s-1	4	x	x
zg	Geopotential Height	m	28	x	x
zg500	Geopotential Height at 500hPa	m	1	x	x

Table S4: Parameters with daily output on either ESGF or *DKRZ LTA (*)* for any of the 30 realisations.

name	parameter long name	unit	level	r1-r10	r11-r30
atmosphere / land					
ares	Aerodynamic Resistance	s m-1	1	x	
cct	Air Pressure at Convective Cloud Top	Pa	1		x
cl	Percentage Cloud Cover	%	47		x
cli	Mass Fraction of Cloud Ice	kg kg-1	47		x
clivi	Ice Water Path	kg m-2	1		x
clt	Total Cloud Cover Percentage	%	1	x	x
clw	Mass Fraction of Cloud Liquid Water	kg kg-1	47		x
clwvi	Condensed Water Path	kg m-2	1		x
cnc	Canopy Covered Area Percentage	%	1	x	x
es	Bare Soil Evaporation	kg m-2 s-1	1	x	x
hfls	Surface Upward Latent Heat Flux	W m-2	1	x	x
hfss	Surface Upward Sensible Heat Flux	W m-2	1	x	x
hur	Relative Humidity	%	47	x	x
hurs	Near-Surface Relative Humidity	%	1	x	x
hursmax	Daily Maximum Near-Surface Relative Humidity	%	1	x	x
hursmin	Daily Minimum Near-Surface Relative Humidity	%	1	x	x
hus	Specific Humidity	1	47	x	x
hus850	Specific Humidity at 850hPa	1	1	x	x
huss	Near-Surface Specific Humidity	1	1	x	x
lai	Leaf Area Index	1	1	x	x
mc	Convective Mass Flux	kg m-2 s-1	48		x
mlotst	Ocean Mixed Layer Thickness Defined by Sigma T	m	1	x	x
mrro	Total Runoff	kg m-2 s-1	1	x	x
mrrob	Subsurface Runoff	kg m-2 s-1	1	x	
mrros	Surface Runoff	kg m-2 s-1	1	x	
mrso	Total Soil Moisture Content	kg m-2	1	x	x
mrsol	Total Water Content of Soil Layer	kg m-2	1	x	x
mrsos	Moisture in Upper Portion of Soil Column	kg m-2	1	x	x
od550aer	Ambient Aerosol Optical Thickness at 550nm	1	1	x	x
pr	Precipitation	kg m-2 s-1	1	x	x
prc	Convective Precipitation	kg m-2 s-1	1	x	x
prra	Rainfall Flux over Land	kg m-2 s-1	1	x	
prsn	Snowfall Flux	kg m-2 s-1	1	x	x
prw	Water Vapor Path	kg m-2	1		x
ps	Surface Air Pressure	Pa	1	x	x
psl	Sea Level Pressure	Pa	1	x	x
rlds	Surface Downwelling Longwave Radiation	W m-2	1	x	x
rldscs	Surface Downwelling Clear-Sky Longwave Radiation	W m-2	1	x	x
rlus	Surface Upwelling Longwave Radiation	W m-2	1	x	x

Continued on next page

Table S4 – continued from previous page

name	parameter long name	unit	level	r1-r10	r11-r30
rlut	TOA Outgoing Longwave Radiation	W m-2	1	x	x
rlutcs	TOA Outgoing Clear-Sky Longwave Radiation	W m-2	1	x	x
rsds	Surface Downwelling Shortwave Radiation	W m-2	1	x	x
rsdscs	Surface Downwelling Clear-Sky Shortwave Radiation	W m-2	1	x	x
rsdt	TOA Incident Shortwave Radiation	W m-2	1	x	x
rsus	Surface Upwelling Shortwave Radiation	W m-2	1	x	x
rsuscs	Surface Upwelling Clear-Sky Shortwave Radiation	W m-2	1	x	x
rsut	TOA Outgoing Shortwave Radiation	W m-2	1	x	x
rsutcs	TOA Outgoing Clear-Sky Shortwave Radiation	W m-2	1	x	x
rzwc	Root Zone Soil Moisture	kg m-2	1	x	x
sbl	Surface Snow and Ice Sublimation Flux	kg m-2 s-1	1	x	x
sfcWind	Daily-Mean Near-Surface Wind Speed	m s-1	1	x	x
sfcWindmax	Daily Maximum Near-Surface Wind Speed	m s-1	1	x	x
snc	Snow Area Percentage	%	1	x	x
snm	Surface Snow Melt	kg m-2 s-1	1	x	
snw	Surface Snow Amount	kg m-2	1	x	x
snwc	snow water equivalent intercepted by the vegetation	kg m-2	1	x	x
ta	Air Temperature	K	47	x	x
ta500	Air Temperature at 500hPa	K	1	x	x
ta700	Air Temperature at 700hPa	K	1	x	x
ta850	Air Temperature at 850hPa	K	1	x	x
tas	Near-Surface Air Temperature	K	1	x	x
tasmax	Daily Maximum Near-Surface Air Temperature	K	1	x	x
tasmin	Daily Minimum Near-Surface Air Temperature	K	1	x	x
tauu	Surface Downward Eastward Wind Stress	Pa	1	x	x
tauv	Surface Downward Northward Wind Stress	Pa	1	x	x
tdps	2m Dewpoint Temperature	K	1	x	x
tr	Surface Radiative Temperature	K	1	x	x
tran	Transpiration	kg m-2 s-1	1		x
ts	Surface Temperature	K	1	x	x
tsl	Temperature of Soil	K	1	x	x
ua	Eastward Wind	m s-1	47	x	x
ua10	Eastward Wind at 10hPa	m s-1	1	x	x
uas	Eastward Near-Surface Wind	m s-1	1	x	x
utendnogw	Eastward Acceleration Due to Non-Orographic Gravity Wave Drag	m s-2	39		x
utendogw	Eastward Acceleration Due to Orographic Gravity Wave Drag	m s-2	39		x
va	Northward Wind	m s-1	47	x	x
vas	Northward Near-Surface Wind	m s-1	1	x	x
wap	Omega (=dp/dt)	Pa s-1	47	x	x
wap500	Pressure Tendency	Pa s-1	1	x	x

Continued on next page

Table S4 – continued from previous page

name	**parameter long name**	**unit**	**level**	**r1-r10**	**r11-r30**
zg	Geopotential Height	m	47	x	x
zg10	Geopotential Height at 10hPa	m	1	x	x
zg100	Geopotential Height at 100hPa	m	1	x	x
zg1000	Geopotential Height at 1000hPa	m	1	x	x
zg500	Geopotential Height at 500hPa	m	1	x	x
ocean / sea ice / biogeochem					
chlos	Surface Mass Concentration of Total Phytoplankton Expressed as Chlorophyll in Sea Water	kg m-3	1	x	x
*fgco2**	*Surface Downward Mass Flux of Carbon Dioxide Expressed as Carbon**	*kg m-2 s-1**	*1**		*x**
*intpp**	*Integrated Primary Production**	*mol C m-2 s-1**	*1**		*x**
omldamax	Mean Daily Maximum Ocean Mixed Layer Thickness Defined by Mixing Scheme	m	1	x	x
*mlotst**	*Ocean Mixed Layer Thickness Defined by Sigma T**	*m**	*1**		*x**
*ph**	*Surface Hydrogen Ion Concentration**	*kmol m-3**	*1**		*x**
phycos	Sea Surface Phytoplankton Carbon Concentration	mol m-3	1	x	x
siconc	Sea-Ice Area Percentage (Ocean Grid)	%	1	x	x
sisnthick	Snow Thickness	m	1	x	x
sispeed	Sea-Ice Speed	m s-1	1	x	x
sitemptop	Surface Temperature of Sea Ice	K	1		x
sithick	Sea Ice Thickness	m	1	x	x
sitimefrac	Fraction of Time Steps with Sea Ice	1	1	x	x
siu	X-Component of Sea-Ice Velocity	m s-1	1	x	x
siv	Y-Component of Sea-Ice Velocity	m s-1	1	x	x
sos	Sea Surface Salinity	0.001	1	x	x
sossq	Square of Sea Surface Salinity	1.00E-06	1	x	x
*spco2**	*Surface Partial Pressure of Carbon Dioxide in Sea Water**	*Pa**	*1**		*x**
t20d	Depth of 20 degree Celsius Isotherm	m	1	x	x
tos	Sea Surface Temperature	degC	1	x	x
tossq	Square of Sea Surface Temperature	degC2	1	x	x
*zos**	*Sea Surface Height above Geoid**	*m**	*1**		*x**

Table S5. Coordinates of the grid points used for calculating storm activity in the model.

Grid point	**Latitude**	**Longitude**
NW of Bermuda - North	36.372° N	69.375° W
NW of Bermuda - West	32.642° N	73.125° W
NW of Bermuda - East	32.642° N	65.625° W
SE of Japan - North	36.372° N	142.500° E
SE of Japan - West	32.642° N	138.750° E
SE of Japan - East	32.642° N	146.250° E

Table S6. Comparison of central estimates of 20-year mean crossing times of the 1.5° C global warming threshold for MPI-GE CMIP6, IPCC AR6, and MPI-GE CMIP6 when using the historical warming of IPCC AR6 instead of the model's own historical warming. The time ranges for MPI-GE CMIP6 only stem from internal variability whereas those for AR6 include uncertainties in historical warming, climate sensitivity and internal variability.

Scenario	**MPI-GE CMIP6**	**AR6**	**Difference**	**With AR6 historical warming**
SSP1-1.9	NA	2025-2044	NA	NA
SSP1-2.6	2034-2053	2023-2042	11	2042-2061
SSP2-4.5	2027-2046	2021-2040	6	2030-2049
SSP3-7.0	2025-2044	2021-2040	4	2027-2046
SSP5-8.5	2024-2043	2018-2037	6	2027-2046