Key Points:
- The Max Planck Institute Grand Ensemble in its CMIP6 version (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

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[7,8], Tatiana Ilyina[1] , Christopher Kadow[9] , Daniel Krieger[6,10] , Hongmei Li[1] , Jochem Marotzke[1,5] , Étienne Plésiat[9], Martin Schupfner[9], Fabian Wachsmann[9] , Lara Wallberg[1,6], Karl-Hermann Wieners[1] , and 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, Bergen, Norway, [8]Bjerknes Centre for Climate Research, Bergen, Norway, [9]German Climate Computing Centre (DKRZ), Hamburg, Germany, [10]Helmholtz-Zentrum Hereon, Geesthacht, Germany

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 currently 30 realizations 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 reanalyzes and compare it to MPI-GE. Then, we demonstrate with six 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 years 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 currently consists of 30 realizations 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 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 application examples how to use its power.

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 (Deser et al., 2012; Maher et al., 2019; Zelle et al., 2005), 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 investigate ocean ecosystem drivers (Rodgers et al., 2015), and 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 the risks associated with extreme values of multivariate extremes, which requires even larger ensemble sizes than conventional univariate extremes (Bevacqua et al., 2023; Burger et al., 2022). 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; Hawkins & Sutton, 2009, 2011; 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 is among the five modeling centers that first 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 (a) the analysis of climate extremes, (b) comparisons to model versions with higher horizontal resolution, (c) comparisons to other SMILEs with CMIP6 forcing, and (d) investigation of low-emission scenarios with high policy relevance.

Several SMILEs with CMIP6 forcing have been recently run by a number of modelling centers, including ensembles with high-frequency model output. Many of these SMILEs are listed in the Multi-Model Large Ensemble Archive (https://www.cesm.ucar.edu/community-projects/mmlea; Deser et al., 2020). Next to MPI-GE CMIP6, currently available SMILEs with CMIP6 forcing and at least 30 realizations 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 Shared Socioeconomic Pathways (SSPs) 1-1.9 and 1-2.6 that are both otherwise only provided by CanESM5 and SHMI-LENS. In contrast to most 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). This allows for an estimate of the climate response based on a single model with an average climate sensitivity, which can be contrasted with low- and high sensitivity models. 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.

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 reanalyzes, and compared to MPI-GE CMIP5. In Section 3, the power of MPI-GE CMIP6 is demonstrated with six application examples that

Table 1
Characteristics of MPI-GE CMIP6 and Other SMILEs With CMIP6 Forcing and At Least 30 Realizations

SMILE name	Model version	Horizontal resolution	High-frequency output	Realizations	Time period	Scenarios	ECS (°C)
MPI-GE CMIP6	MPI-ESM1.2-LR	1.8° atm.; 1.5° ocean	Daily for all para- meters provided, 3-hr, 6-hr for some (see Table 2 and Table S1 in Supporting Information S1)	30 (50 in 2024)	1850–2100	SSP1-1.9, 1-2.6, 2-4.5, 3-7.0, 5-8.5	2.80
ACCESS-ESM1.5	ACCESS-ESM1.5	1.88 × 1.25° 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
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
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
LENS2	CESM2	1.0° atm.; 1.0° ocean	Daily for all parameters provided, 3-hr, 6-hr for some	100	1850–2100	SSP3-7.0	5.16
SMHI-LENS	EC-Earth3.3.1	0.7° atm.; 1.0° ocean	Daily for many atm. parameters	50	1970–2100	SSP1-1.9, 1-2.6, 2-4.5, 3-7.0, 4-3.4, 4-6.0, 5-3.4 -OS, 5-8.5	4.31
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
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

Table 2
Available Simulations of MPI-ESM1.2 With Different Horizontal Resolution

Model version	Horizontal resolution	Realizations	Time period	Scenarios
MPI-ESM1.2-LR	T63, 1.8° atm.; GR15, 1.5° ocean	30 (50 in 2024)	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

Note. The MPI-ESM1.2-HR and -XR simulations were run as part of HighResMIP.

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 summarizes and concludes the paper.

2. MPI-GE CMIP6

2.1. Model Description

MPI-GE CMIP6 is currently 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) summarizes the updates that were introduced to MPI-ESM1.2, most importantly new radiation and aerosol parameterizations, 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).

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; Paulsen et al., 2017, HAMOCC6). 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 initialization and historical and future external forcing (i.e., atmospheric composition, solar cycle, volcanic eruptions, land use), and are concentration-driven. The 30-member ensemble of historical simulations covers the time period 1850–2014 and each member is initialized from a different state, approximately 25 years apart, of a quasi-stationary one-member 1,000-year long preindustrial simulation. This macro initialization from the preindustrial control state samples the full phase space of both the ocean and atmosphere states (Marotzke, 2019), and allows for investigating also the variability of climate quantities with long times of divergence (Hawkins & Sutton, 2016). Five scenario simulations (SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP5-8.5, 30 realizations each) cover the time period 2015–2100, and in each scenario the realizations are directly initialized from their corresponding realizations of the historical ensemble. By this, the new ensemble can play a key role in climate research. The number of ensemble members is currently extended to 50. All 50 ensemble members will be available in early 2024 for both the historical period and the five emission scenarios.

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, that is, for sea ice and land surface, are available for all simulations (Table S1 in Supporting Information 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 HAMOCC, 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 in Supporting Information S1). Model output can be accessed via DKRZ's ESGF server at https://esgf-data.dkrz.de/search/cmip6-dkrz/.

Table 3
Parameters With 3-Hourly and 6-Hourly Output on ESGF Available for All 30 Realizations

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 500 hPa	m	1

Note. The parameters with daily output are listed in Table S1 in Supporting Information S1. A full list of parameters subdivided for members r1–r10 and r11–r30 is given in Tables S2–S4 in Supporting Information S1.

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 reanalyzes (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 twenty-first century relative to the 1985–2014 reference period ranges from 0.4 K in SSP1-1.9 to 3.7 K in SSP5-8.5.

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 in Supporting Information 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 in Supporting Information S1). The different estimates from observational and reanalyzes 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
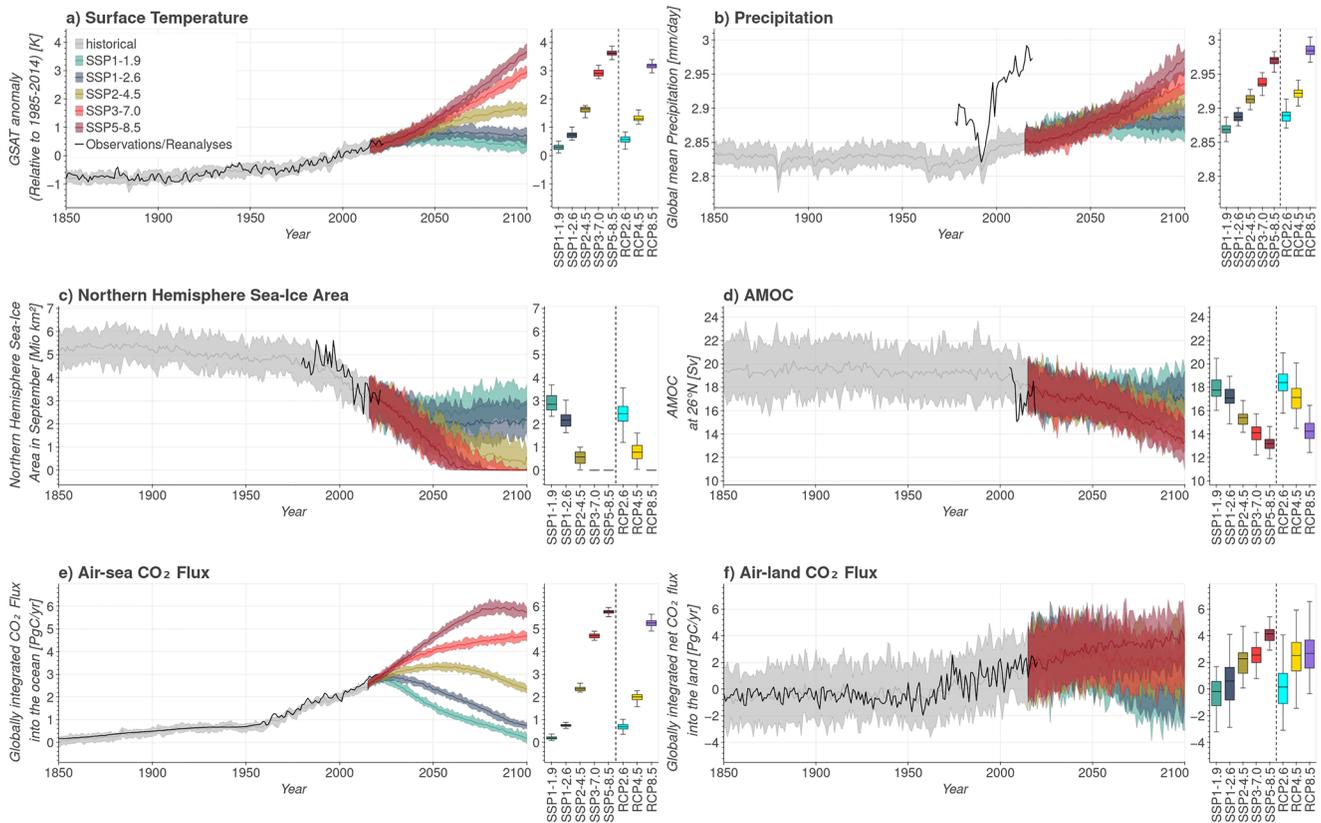
Figure 1. *Comparison of key climate quantities of Max Planck Institute Grand Ensemble CMIP6 (MPI-GE CMIP6) to observations or reanalyzes and MPI-GE CMIP5.* Annual mean ensemble spread (shading) and ensemble mean (thick lines) for the historical simulations (gray), and the five emission scenarios SSP1-1.9, 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 realizations) and MPI-GE CMIP5 (100 realizations). 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) at 26°N and 1,000 m depth, (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 reanalyzes, 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 (Friedlingstein et al., 2022; Global Carbon Project, 2021).

Northern Hemisphere sea-ice area is projected to shrink below the 1 million square kilometer threshold in the second half of the twenty-first century in SSP2-4.5, SSP3-7.0 and SSP5-8.5, but remains above that threshold in both SSP1-1.9 and SSP1-2.6 until the end of the twenty-first century, similar to previous findings on sea-ice decline in CMIP6 (Lee et al., 2021; Notz & Community, 2020). 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 estimated in the Global Carbon Project (Figures 1e and 1f). The evaluation of MPI-GE CMIP6 with observations and reanalyzes 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 twenty-first century. We find that MPI-GE CMIP6 shows slightly higher global-mean warming by the end of the twenty-first 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 Representative Concentration Pathway (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), likely caused by the stronger prescribed atmospheric CO_2 concentrations in the SSP than respective RCP scenarios (Meinshausen et al., 2011, 2020). The projected change in net CO_2 flux into the land is largely uncertain, but shows a similar response at the end of the twenty-first 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, that is, the same forcing, we find a very similar warming rate and variability (Figure S2 in Supporting Information S1). 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 (Fyfe et al., 2021; Wyser et al., 2020). 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 in Supporting Information S1). This high-frequency output comes at the expense of a smaller ensemble size of currently 30 realizations instead of 100 realizations, but makes MPI-GE CMIP6 specifically suited for the analysis of climate extremes including compound events. As such, the many realizations at LR provide benefits such as high sampling numbers that are not available with higher-resolution simulations.

Third, MPI-GE CMIP6 can be compared to higher-resolution simulations of the same model version (see Table 2), for instance 10 realizations of MPI-ESM1.2-HR (1.0° atm., 0.4° ocean, Müller et al., 2018) or a single realization 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 realizations 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 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. Analyzing 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, fully coupled climate models struggle to
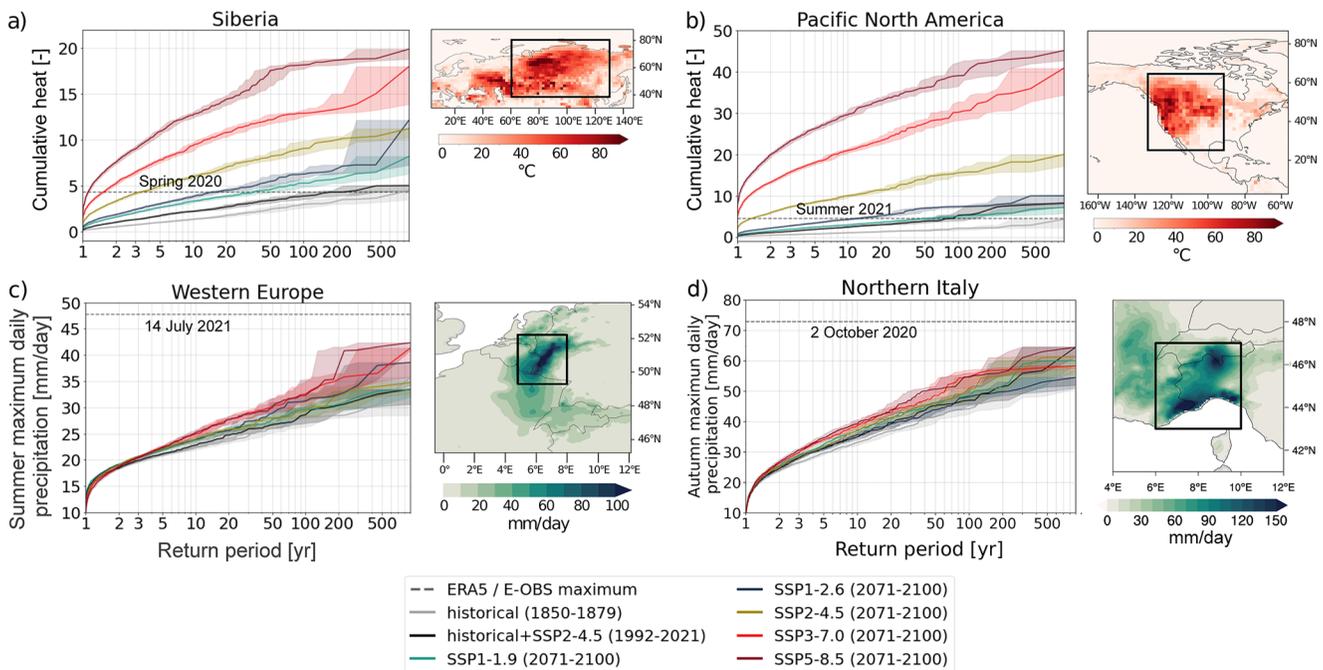
Figure 2. *Return periods from Max Planck Institute Grand Ensemble CMIP6 (MPI-GE CMIP6) for recently observed heat and precipitation extremes for different emission scenarios.* Return periods for (a) and (b) cumulative heat scaled with respect to climatology for (a) spring (MAM) 2020 Siberian heatwave 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, gray), the current climate (1992–2021, black), and the five SSP scenarios for the period 2071–2100 (colored). 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 14 July 2021 and 2 October 2020, respectively. Black boxes mark the regions of interest used for averaging.

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 simulating the magnitude of 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 90th percentile based on a centered 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 Figures 2a and 2b). We scale the cumulative heat with respect to climatology (1985–2014). We use return periods to quantify the time interval between two events of given magnitude to occur. 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 (Figures 2a and 2b). 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 4 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-year event by the end of the twenty-first 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 July 2021, and the one in northern Italy on 2 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 (Ibebuchi, 2022; Kreienkamp et al., 2021), 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 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 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 twenty-first 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., Myhre et al., 2019; Pendergrass et al., 2017). 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.
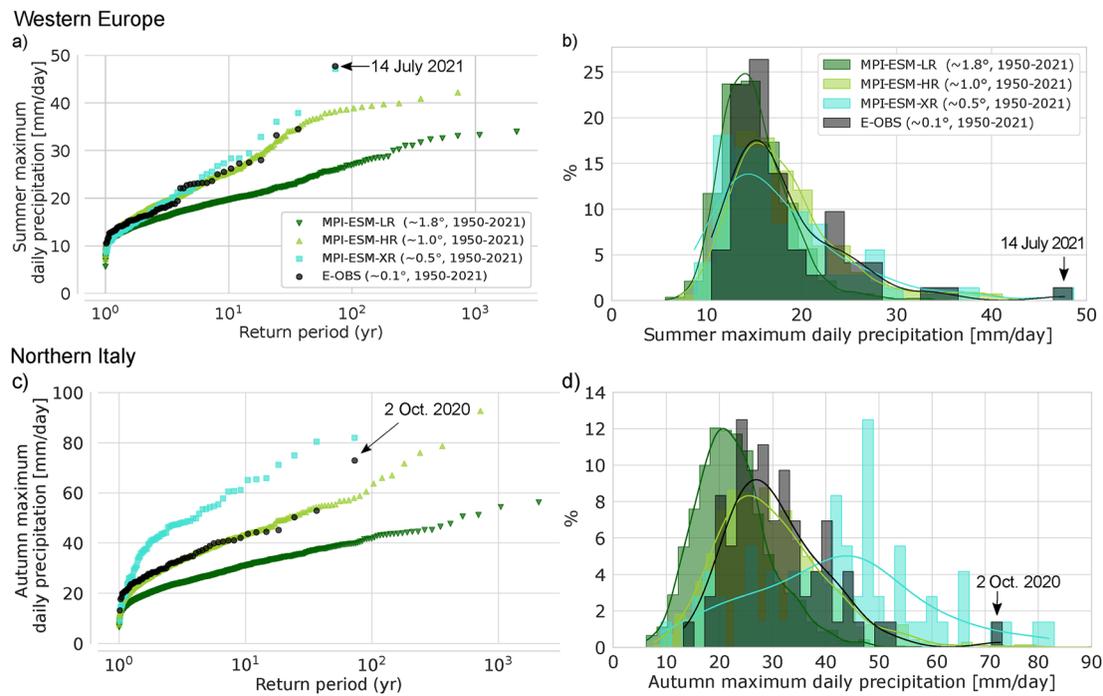
Figure 3. *Representation of precipitation extremes dependent on model resolution.* (a, b) Comparison of summer (JJA) maximum daily precipitation averaged across the western European box shown in Figure 2c from 1950 to 2021 in three model resolutions from MPI-ESM1.2 and in observations shown as (a) return periods and (b) probability density functions. (c, d) Comparison of autumn (SON) maximum daily precipitation averaged across the northern Italy box shown in Figure 2d from 1950 to 2021 in three model resolutions from MPI-ESM1.2 and in observations shown as (c) return periods and (d) probability density functions. Return periods are calculated empirically by ordering and ranking from the most extreme to the least extreme value. Values of all summers or autumns, respectively, and all realizations are merged for each ensemble. Further note that MPI-ESM-LR is based on 30 realizations, MPI-ESM-HR on 10 realizations and MPI-ESM-XR and the observed record on only a single realization. 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 July 2021 is 47.7 mm, and that of the event in northern Italy on 2 October 2020 is 72.9 mm.

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 parameterized (Iles et al., 2020; Kahraman et al., 2021; Kendon et al., 2021; Wehner et al., 2014). 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 realizations of MPI-ESM1.2-HR (Müller et al., 2018) with 1.0° atmospheric horizontal resolution, and a single realization 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 (Figures 3a and 3b). Strikingly, the single realization 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 in Supporting Information S1), 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. However, it remains unclear whether this ability is caused by an improved representation of the underlying processes. 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 (Figures 3c and 3d). 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

Marine heatwaves have been increasingly observed in the twentieth and twenty-first century (Hobday et al., 2016; Oliver et al., 2018) with implications for marine ecosystems on a global scale. The evolution of marine heatwaves in the past, present, and future has been linked with global warming (Frölicher et al., 2018; Oliver et al., 2019; Plecha & Soares, 2020). In recent times, more stress is put on marine ecosystems by ocean acidity extremes (Burger et al., 2020), partly even compounding with marine heatwaves (Burger et al., 2022). Here, we analyze daily mean sea surface temperature (SST) and hydrogen ion concentration ($[H^+]$) of MPI-GE CMIP6 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. Both SST and $[H^+]$ are defined as extreme, if they exceed the 90th percentile for five consecutive days (Burger et al., 2022; Hobday et al., 2016). 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 90th multi-year 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 characterize changes of both extremes with time and across scenarios.

Before the reference period 1985–2014, almost no marine heatwaves are detected globally. Between 1985 and 2014, 5–25 days per year are extreme with marine heatwaves being more frequent in the subpolar North Atlantic and the Southern Ocean (Figure 4a). Compared to observations (e.g., Oliver et al., 2018; Plecha & Soares, 2020) the ensemble-mean response of MPI-GE CMIP6 shows similar numbers of marine heatwaves for this time period. By 2030, between 40 and 120 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 marine heatwave days per year falls back to the 2030 numbers (Figure 4b, Figure S4 in Supporting Information S1). 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 twenty-first 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 average yearly number of days with extreme $[H^+]$ is about 6 days. However, there is a rapid increase in globally averaged extreme $[H^+]$ days after 2,000, to approximately 200 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 very slightly by the end of the twenty-first century (Figure 4f, Figure S4 in Supporting Information S1). Within the SSP1-1.9 scenario, ocean acidity extremes are projected to peak at approximately 340 days per year between 2025 and 2040 and decline thereafter to 300 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 and 2100 (Figures 4g and 4h). There is a notable difference in the global occurrence of ocean acidity extremes between SSP1-1.9 and SSP1-2.6 in the second half of the twenty-first century (Figure 4f), despite only small differences in terms of global mean temperature in both scenarios (Figure 1a).
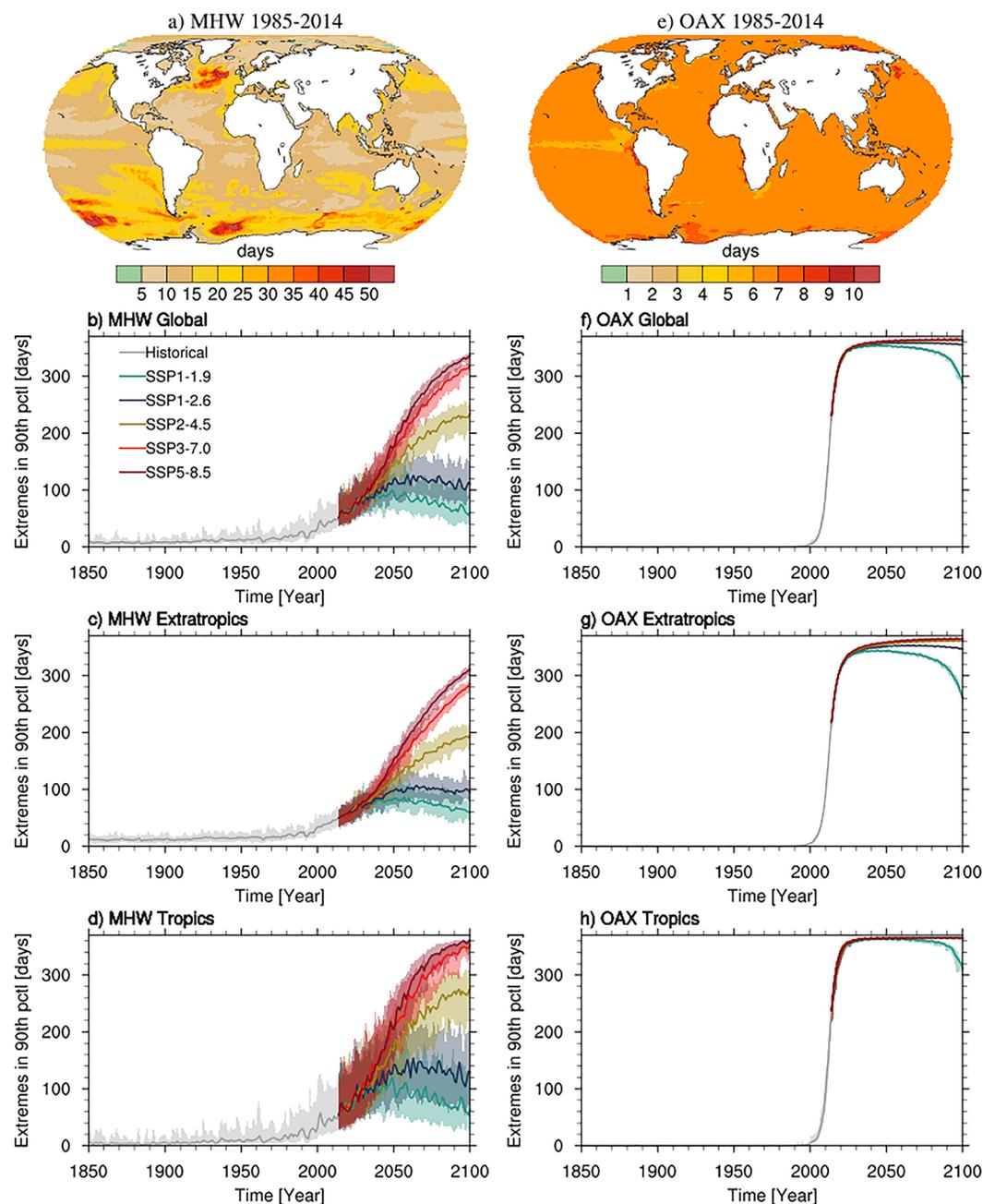
Figure 4. *Present and future frequency of marine heatwaves and ocean acidity extremes*. Maps of (a) the ensemble mean number of marine heatwave (MHW) days per year and (e) the number of ocean acidity extreme event (OAX) days per year, averaged over the reference period 1985–2014, based on the 90th percentile 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 (gray), 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).

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 is, based on a shifting baseline (Burger et al., 2022), 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 mean and extreme winds and their dependence on the emission scenario (Figures 5a and 5b and Figure S5 in Supporting Information S1). To detect projected global changes in wind speed, we first derive annual means and 95th 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 mean and extreme wind speeds over high-latitude oceans and decreasing mean and extreme wind speeds in most mid- and low-latitude ocean basins. Over land, increases 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 in Supporting Information S1).

We further analyze projected changes in storm activity in two regions that are known for the frequent passage of mature extratropical cyclones with often devastating impacts when they make landfall: the North Atlantic storm track (45–55°N, 10–40°W; Figure 5c) and the North Pacific storm track (30–40°N, 180°–150°W; Figure 5d). For both regions, we define storm activity as the standardized area-mean annual 95th percentiles of 3-hourly near-surface wind speeds. We therefore first calculate area means of grid-point-wise annual 95th percentiles of near-surface wind speeds for each ensemble member. We then standardize the area means by subtracting the 1985–2014 ensemble mean from each ensemble member, and divide by the 1985–2014 ensemble standard deviation.

For both the North Atlantic and North Pacific storm tracks, we find a decreasing storm activity in most future scenarios (Figures 5c and 5d and Figure S5 in Supporting Information S1). While the decrease in North Atlantic storm activity is almost scenario-independent, we find no notable change for North Pacific storm activity in scenarios SSP1-1.9 and SSP1-2.6, but strong decreases in the high-emission scenarios SSP3-7.0 and SSP5-8.5. This agrees with the projected change in surface wind speed, where the marine subtropics around 30°N show a strong signal of decreasing mean and extreme wind speeds in the SSP5-8.5 scenario (Figures 5a and 5b).

We further calculate the ensemble balance to characterize 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 0.5σ and -0.5σ, 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 Figures 5c and 5d). In the SSP1-1.9 and SSP1-2.6 scenarios, we find that the ensemble balance does not significantly deviate from 0 toward the end of the twenty-first century in the North Pacific, confirming the rather small projected change in storm activity. In the high-emission SSP3-7.0 and SSP5-8.5 scenarios, the ensemble balance falls to between -20 and -30 at the end of the twenty-first century, which indicates that most ensemble members agree on a decline in storm activity in the North Atlantic and North Pacific storm tracks.

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 extratropical storm tracks over both hemispheres (Priestley & Catto, 2022; 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 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
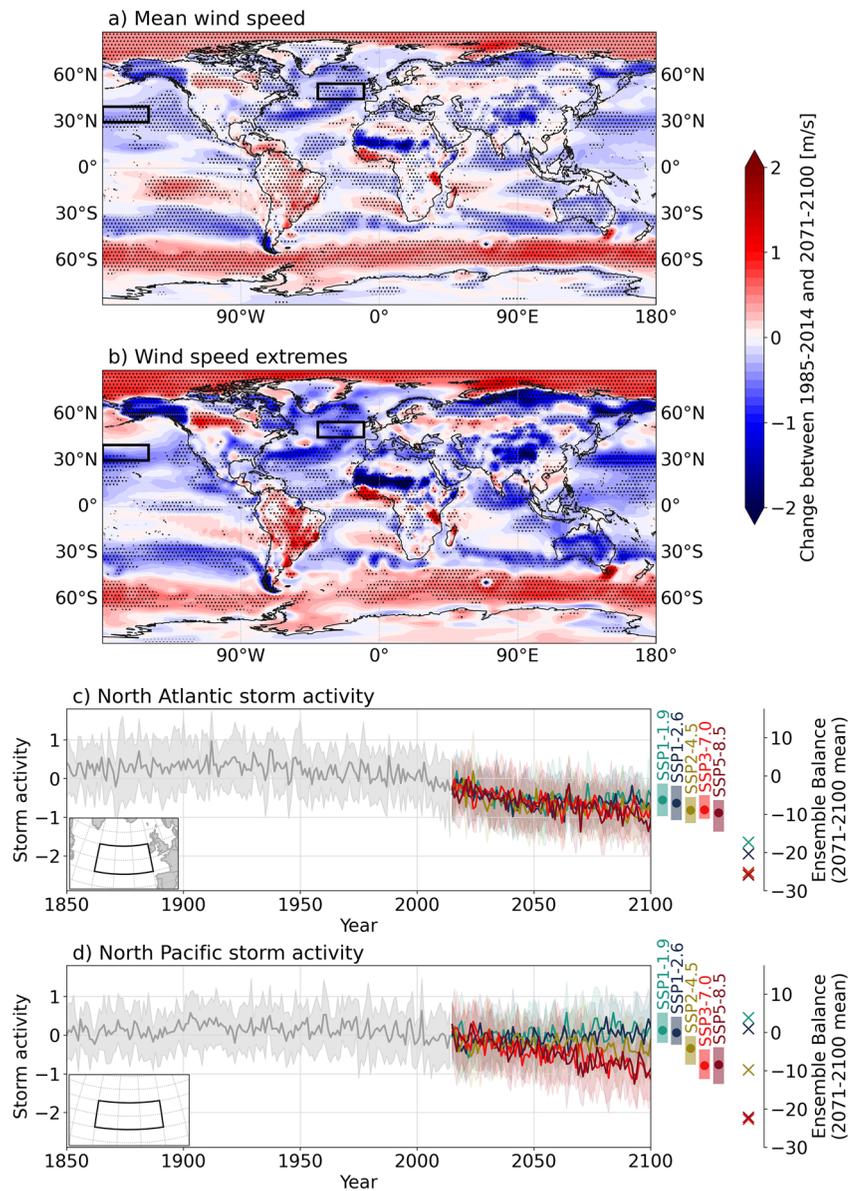
Figure 5. *Projected changes in near-surface wind speed and storm activity.* (a, b) Absolute change in (a) ensemble mean surface wind speed, and (b) ensemble mean 95th annual percentiles of surface wind speed between 1985–2014 and 2071–2100, based on SSP5-8.5 forcing. Black rectangles mark regions for which storm activity has been calculated. Stippling indicates significant changes for which the ensemble spreads in 1985–2014 and 2071–2100 show no overlap. Maps for the other four SSP scenarios are shown in Figure S5 in Supporting Information S1. (c, d) Ensemble mean storm activity (thick lines) and interquartile range (shading) for the historical simulations (gray) and the five scenarios (colored) over (c) the North Atlantic storm track and (d) the North Pacific storm track. Colored 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.

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 and 2°C, respectively.

To investigate the crossing probability of 1.5 and 2°C of global warming in MPI-GE CMIP6, we use annual mean, GSAT to compute for every year and each of the five scenarios the fraction of realizations (x/30 realizations) that crosses these temperature thresholds in a single year relative to the 1850–1900 reference period (Figures 6a and 6b). We find that in all emission scenarios, there is a non-zero chance of observing individual years above
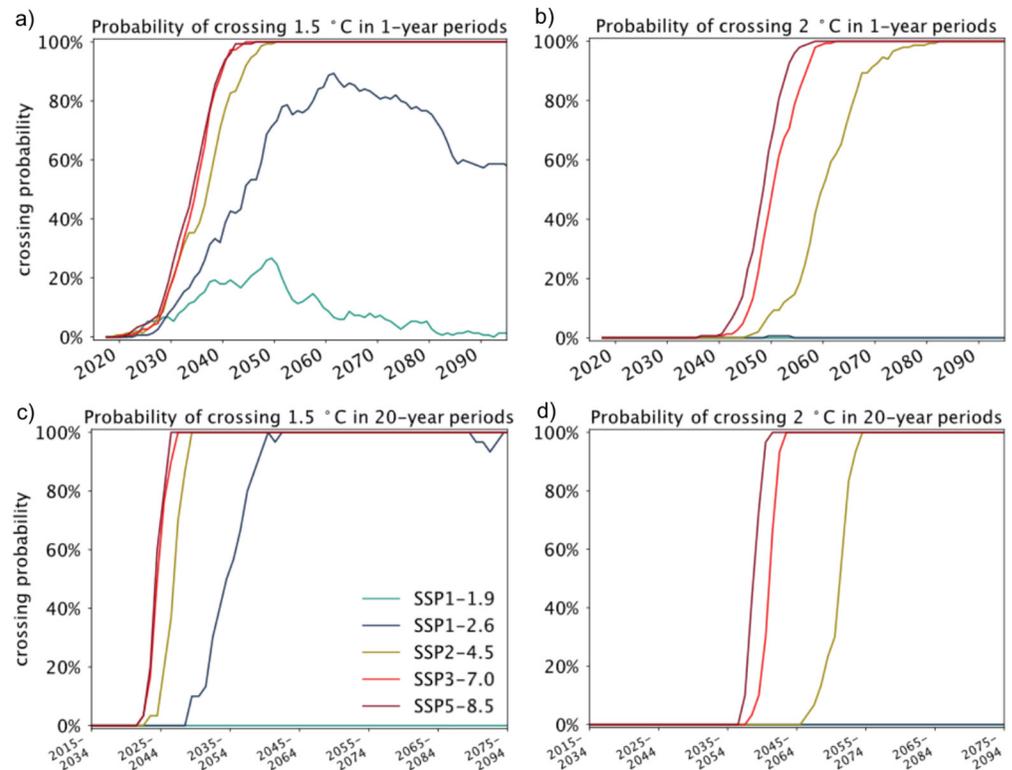
Figure 6. *Probability of crossing Paris Agreement global warming limits.* Probability of crossing (a) 1.5°C and (b) 2°C in a single year, and (c) 1.5°C and (d) 2°C in 20-year averages for the different emission scenarios until 2100. The crossing probability is defined as the fraction of the 30 realizations that cross the temperature threshold relative to the reference period 1850–1900. In (c, d), the 20-year mean global mean near-surface air temperature is plotted against the central year of that 20-year period.

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 realization and show for each 20-year window the fraction of realizations that crosses 1.5 or 2°C (Figures 6c and 6d). 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 realization. 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 2035–2050. 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 (Lee et al., 2021; Marotzke et al., 2022).

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 (ECS) 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 S5 in Supporting Information S1). 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 5 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 ECS 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 realizations 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 (Morice et al., 2021) have been trained using gridded global historical surface temperature anomalies from two large ensembles: (a) MPI-GE CMIP6, containing 30 realizations and spanning the 1850–2014 time period; and (b) MPI-GE CMIP5, containing 100 realizations (Maher et al., 2019) and spanning the 1850–2005 time period. Before the training, one ensemble member was excluded from each ensemble to create two testing data sets for the whole time series. Two validation data sets were created from the remaining ensemble members of each data set by pulling out randomly 1/8 of the data increments, here months. The remaining data were used to create the training data sets which contain 50,242 samples for MPI-GE CMIP6, and 162,162 samples for MPI-GE CMIP5. The training process for the neural networks sets the missing-value masks of the HadCRUT5 observations onto the two training data sets. Samples and masks of missing values are randomly selected for each iteration. This allows the neural network to learn interpolation and pattern-recognition strategies for all potential climate situations for all of the masks with the MPI-GE output. 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.

We compare the temperature climatologies of 1991–2020 and 1891–1920 of both the original HadCRUT5 data (Figure 7a) and the HadCRUT5 analysis (Figure 7b) with the artificial intelligence-infilled versions using the 100-member MPI-GE CMIP5 (Figure 7c) and the 30-member MPI-GE CMIP6 (Figure 7d). This comparison shows the substantial difference in the availability of information between the HadCRUT5 data sets and the artificial intelligence-infilled versions. Even the HadCRUT5 analysis, infilled using a method related to Kriging, does not cover the whole globe and provides a rather smoothed estimate of temperature change. With the artificial-intelligence models, we detect both strong warming and slight cooling in the regions where the HadCRUT5 data sets have missing data. In particular, the warming patterns reconstructed from the two ensembles show a strong warming signal over the century in northern high latitudes. Large areas in the Pacific also consistently show a warming between the two climatologies, despite the fact that the region is affected by strong El Niño-Southern Oscillation variability. The artificial intelligence-reconstructed data in the sparsely observed Southern Ocean and Antarctica reveal pronounced regional differences of warming and cooling. The annual global mean temperature time series reconstructed using the 100-member and the 30-member artificial-intelligence models are very similar to each other, but different when compared to the original HadCRUT5 data and its analysis (Figure S6 in Supporting Information S1). The early years of the artificial intelligence-infilled observations show in general a cooler temperature anomaly than the original HadCRUT5 data (Kadow et al., 2020). From the striking similarity in the reconstructed pattern and time series between MPI-GE CMIP5 and MPI-GE CMIP6, 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 SMILE 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:
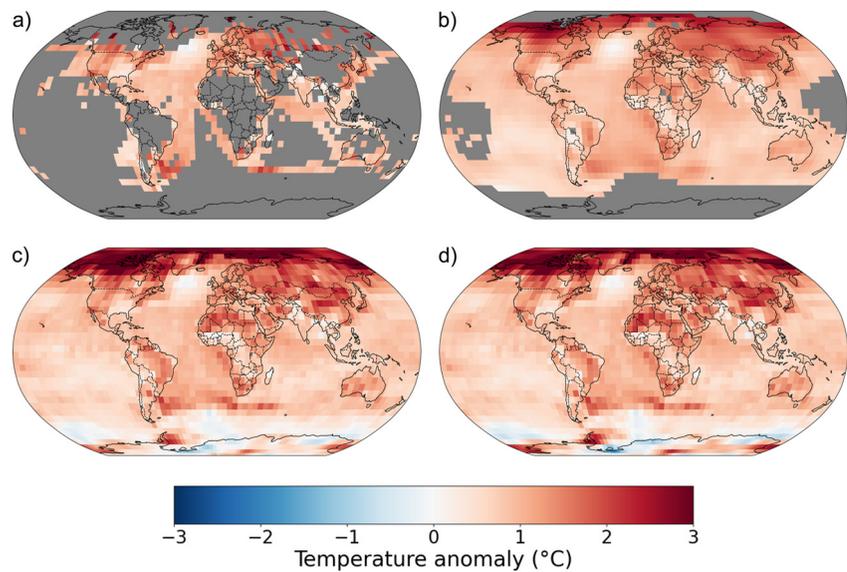
Figure 7. *Comparison of HadCRUT5 with Max Planck Institute Grand Ensemble CMIP6 (MPI-GE CMIP6) vs. MPI-GE CMIP5 for infilling observations of surface temperature with artificial intelligence.* 1991–2020 climatology referenced to the 1891–1920 climatology: (a) original "non-infilled" HadCRUT5 data set where gray pixels indicate missing values, (b) spatial reconstruction of the HadCRUT5 "analysis" data set from the UK Met Office (Morice et al., 2021), (c) spatial reconstruction of the HadCRUT5 data set using the artificial intelligence 100-members model (MPI-GE CMIP5, Maher et al., 2019), (d) spatial reconstruction of the HadCRUT5 data set using the artificial intelligence 30-members modiel (MPI-GE CMIP6). Mean values have been computed only for grid points containing at least 70% of valid values for the considered time period.

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., Landrum & Holland, 2020; Suarez-Gutierrez et al., 2020a, 2020b), 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 July 2021 and in northern Italy observed on 2 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 or 2°C. We find that the ensemble simulates a non-zero chance to observe individual years above 1.5°C even for the lowest emission scenario SSP1-1.9, which is consistent with the Paris Agreement pledges in this model. With its good representation of internal variability in GSAT and its ECS 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 (Fyfe et al., 2021; Wyser et al., 2020), primarily caused by the updated forcing in CMIP6. From combining MPI-GE CMIP6 with artificial intelligence, we find that 30 realizations 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.

Data Availability Statement

The MPI-ESM1.2-LR coupled climate model is distributed via http://www.mpimet.mpg.de/. The scripts and code for reproducing the plots are openly available through the publication repository of the Max Planck Society at https://hdl.handle.net/21.11116/0000-000D-B686-E. The simulations of MPI-GE CMIP6 can be accessed via DKRZ's ESGF server at https://esgf-data.dkrz.de/search/cmip6-dkrz/ by specifying the requested variable and frequency and the following IDs: Source: MPI-ESM1-2-LR; Institution: MPI-M (DKRZ for ssp119); Experiment: historical, ssp126, ssp245, ssp370, ssp585; Variant Label: r*i1p1f1. Please note that the ensemble size is currently extended to 50.

References

Adler, R., Sapiano, M., Huffman, G., Wang, J.-J., Gu, G., Bolvin, D., et al. (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. https://doi.org/10.3390/atmos9040138

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

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., et al. (2020). Quantifying progress across different CMIP phases with the ESMValTool. *Journal of Geophysical Research: Atmospheres*, *125*(21), e2019JD032321. https://doi.org/10.1029/2019JD032321

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

Burger, F. A., John, J. G., & Frölicher, T. L. (2020). Increase in ocean acidity variability and extremes under increasing atmospheric CO_2. *Biogeosciences*, *17*(18), 4633–4662. https://doi.org/10.5194/bg-17-4633-2020

Burger, F. A., Terhaar, J., & Frölicher, T. L. (2022). Compound marine heatwaves and ocean acidity extremes. *Nature Communications*, *13*(1), 4722. https://doi.org/10.1038/s41467-022-32120-7

Ciavarella, A., Cotterill, D., Stott, P., Kew, S., Philip, S., van Oldenborgh, G., et al. (2021). Prolonged Siberian heat of 2020 almost impossible without human influence. *Climatic Change*, *166*(9), 1689–1713. https://doi.org/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. https://doi.org/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. *Geoscientific Model Development*, *10*(9), 3297–3308. https://doi.org/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. https://doi.org/10.1016/j.wace.2022.100542

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., et al. (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. https://doi.org/10.1029/2019MS001895

Deser, C., Lehner, F., Rodgers, K. E. A., Ault, T., Delworth, T. L., DiNezio, P. N., et al. (2020). Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Climate Change*, *10*(4), 277–286. https://doi.org/10.1038/s41558-020-0731-2

Deser, C., Phillips, A., Bourdette, V. E. A., & Teng, H. (2012). Uncertainty in climate change projections: The role of internal variability. *Climate Dynamics*, *38*(3–4), 527–546. https://doi.org/10.1007/s00382-010-0977-x

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. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/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. https://doi.org/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

Fischer, E., Sippel, S., & Knutti, R. (2021). Increasing probability of record-shattering climate extremes. *Nature Climate Change*, *11*(8), 689–695. https://doi.org/10.1038/s41558-021-01092-9

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., et al. (2021). The Earth's energy budget, climate feedbacks, and climate sensitivity. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of Working Group I to the sixth assessment report of the Intergovernmental Panel on Climate Change* (pp. 923–1054). Cambridge University Press. https://doi.org/10.1017/9781009157896.009

Frajka-Williams, E., Moat, B., Smeed, D., Rayner, D., Johns, W., Baringer, M., et al. (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. https://doi.org/10.5285/cc1e34b3-3385-662b-e053-6c86abc03444

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Bakker, D. C., Hauck, J., et al. (2022). Global carbon budget 2021. *Earth System Science Data*, *14*(4), 1917–2005. https://doi.org/10.5194/essd-14-1917-2022

Frölicher, T. L., Fischer, E. M., & Gruber, N. (2018). Marine heatwaves under global warming. *Nature*, *560*(7718), 360–364. https://doi.org/10.1038/s41586-018-0383-9

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. https://doi.org/10.1073/pnas.2016549118

Global Carbon Project. (2021). Supplemental data of global carbon budget 2021 (Version 1.0) [Dataset]. Global Carbon Project. https://doi.org/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. https://doi.org/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. https://doi.org/10.1038/s41586-021-03981-7

Gutjahr, O., Putrasahan, D., Lohmann, K., Jungclaus, J. H., von Storch, J.-S., Brüggemann, N., et al. (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. https://doi.org/10.5194/gmd-12-3241-2019

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

Hawkins, E., & Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, *90*(8), 1095–1108. https://doi.org/10.1175/2009BAMS2607.1

Hawkins, E., & Sutton, R. (2011). The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, *37*(1–2), 407–418. https://doi.org/10.1007/s00382-010-0810-6

Hawkins, E., Sutton, R., Gregory, J. M., & Stainforth, D. A. (2016). Irreducible uncertainty in near-term climate projections. *Climate Dynamics*, *46*(11–12), 3807–3819. https://doi.org/10.1007/s00382-015-2806-8

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

Hewitt, H., Fox-Kemper, B., Pearson, B. E. A., Roberts, M., & Klocke, D. (2022). The small scales of the ocean may hold the key to surprises. *Nature Climate Change*, *12*(6), 496–499. https://doi.org/10.1038/s41558-022-01386-6

Hobday, A. J., Alexander, L. V., Perkins, S. E., Smale, D. A., Straub, S. C., Oliver, E. C., et al. (2016). A hierarchical approach to defining marine heatwaves. *Progress in Oceanography*, *141*, 227–238. https://doi.org/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. https://doi.org/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. https://doi.org/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. *Journal of Advances in Modeling Earth Systems*, *5*(2), 287–315. https://doi.org/10.1029/2012MS000178

Jungclaus, J. H., Fischer, N., Haak, H., Lohmann, K., Marotzke, J., Matei, D., et al. (2013). Characteristics of the ocean simulations in the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model. *Journal of Advances in Modeling Earth Systems*, *5*(2), 422–446. https://doi.org/10.1002/jame.20023

Kadow, C., Hall, D. M., & Ulbrich, U. (2020). Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, *13*(6), 408–413. 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. https://doi.org/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. https://doi.org/10.1098/rsta.2019.0547

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G., Gocheva, A., et al. (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. https://doi.org/10.1002/joc.773

Kreienkamp, F., Philip, S. Y., Tradowsky, J. S., Kew, S. F., Lorenz, P., Arrighi, J., et al. (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.

Landrum, L., & Holland, M. (2020). Extremes become routine in an emerging new Arctic. *Nature Climate Change*, *10*(10), 1108–1115. https://doi.org/10.1038/s41558-020-0892-z

Lee, J.-Y., Marotzke, J., Bala, G., Cao, L., Corti, S., Dunne, J., et al. (2021). Future global climate: Scenario-based projections and near-term information. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of Working Group I to the sixth assessment report of the Intergovernmental Panel on Climate Change* (pp. 553–672). Cambridge University Press. https://doi.org/10.1017/9781009157896.006

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

Lin, P., Zhao, B., Wei, J., Liu, H., Zhang, W., Chen, X., et al. (2022). The super-large ensemble experiments of CAS FGOALS-g3. *Advances in Atmospheric Sciences*, *39*(10), 1746–1765. 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). Springer International Publishing.

Loughran, T. F., Boysen, L., Bastos, A., Hartung, K., Havermann, F., Li, H., et al. (2021). Past and future climate variability uncertainties in the global carbon budget using the MPI grand ensemble. *Global Biogeochemical Cycles*, *35*(8), e2021GB007019. https://doi.org/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. https://doi.org/10.5194/esd-13-687-2022

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., et al. (2019). The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability. *Journal of Advances in Modeling Earth Systems*, *11*(7), 2050–2069. https://doi.org/10.1029/2019MS001639

Marotzke, J. (2019). Quantifying the irreducible uncertainty in near-term climate projections. *WIREs Climate Change*, *10*(1), e563. https://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. https://doi.org/10.1002/wea.4174

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. https://doi.org/10.1029/2018MS001400

Meinshausen, M., Nicholls, Z. R. J., Lewis, J., Gidden, M. J., Vogel, E., Freund, M., et al. (2020). The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geoscientific Model Development*, *13*(8), 3571–3605. https://doi.org/10.5194/gmd-13-3571-2020

Meinshausen, M., Smith, S., Calvin, K., Daniel, J. S., Kainuma, M. L. T., Lamarque, J. F., et al. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, *109*(1–2), 213–241. https://doi.org/10.1007/s10584-011-0156-z

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., et al. (2021). An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361. https://doi.org/10.1029/2019JD032361

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., et al. (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. https://doi.org/10.1029/2017MS001217

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

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

Oliver, E. C. J., Burrows, M. T., Donat, M. G., Sen Gupta, A., Alexander, L. V., Perkins-Kirkpatrick, S. E., et al. (2019). Projected marine heatwaves in the 21st century and the potential for ecological impact. *Frontiers in Marine Science*, *6*, 734. https://doi.org/10.3389/fmars.2019.00734

Oliver, E. C. J., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., et al. (2018). Longer and more frequent marine heatwaves over the past century. *Nature Communications*, *9*(1), 1324. https://doi.org/10.1038/s41467-018-03732-9

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. https://doi.org/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. *Nature Communications*, *12*(1), 7237. https://doi.org/10.1038/s41467-021-27515-x

Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. *Journal of Advances in Modeling Earth Systems*, *9*(1), 438–464. https://doi.org/10.1002/2016MS000737

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. https://doi.org/10.1038/s41598-017-17966-y

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

Philip, S. Y., Kew, S. F., van Oldenborgh, G. J., Anslow, F. S., Seneviratne, S. I., Vautard, R., et al. (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. https://doi.org/10.5194/esd-13-1689-2022

Pichelli, E., Coppola, E., Sobolowski, S. E. A., Ban, N., Giorgi, F., Stocchi, P., et al. (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*(11–12), 3581–3602. https://doi.org/10.1007/s00382-021-05657-4

Plecha, S. M., & Soares, P. M. M. (2020). Global marine heatwave events using the new CMIP6 multi-model ensemble: From shortcomings in present climate to future projections. *Environmental Research Letters*, *15*(12), 124058. https://doi.org/10.1088/1748-9326/abc847

Priestley, M. D. K., & Catto, J. L. (2022). Future changes in the extratropical storm tracks and cyclone intensity, wind speed, and structure. *Weather and Climate Dynamics*, *3*(1), 337–360. https://doi.org/10.5194/wcd-3-337-2022

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. https://doi.org/10.1088/1748-9326/ac5712

Reick, C. H., Raddatz, T., Brovkin, V., & Gayler, V. (2013). Representation of natural and anthropogenic land cover change in MPI-ESM. *Journal of Advances in Modeling Earth Systems*, *5*(3), 459–482. https://doi.org/10.1002/jame.20022

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

Rodgers, K. B., Lin, J., & Frölicher, T. L. (2015). Emergence of multiple ocean ecosystem drivers in a large ensemble suite with an Earth system model. *Biogeosciences*, *12*(11), 3301–3320. https://doi.org/10.5194/bg-12-3301-2015

Schulzweida, U. (2022). CDO user guide (2.1.0). *Zenodo*. https://doi.org/10.5281/zenodo.7112925

Seneviratne, S., Zhang, X., Badi, W., Dereczynski, C., Luca, A. D., Ghosh, S., et al. (2021). Weather and climate extreme events in a changing climate. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of Working Group I to the sixth assessment report of the Intergovernmental Panel on Climate Change* (pp. 1513–1766). Cambridge University Press. https://doi.org/10.1017/9781009157896.013

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

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., et al. (2013). Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, *5*(2), 146–172. https://doi.org/10.1002/jame.20015

Suarez-Gutierrez, L., Milinski, S., & Maher, N. (2021). Exploiting large ensembles for a better yet simpler climate model evaluation. *Climate Dynamics*, *57*(9–10), 2557–2580. https://doi.org/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. *Climate Dynamics*, *54*(9), 4351–4366. https://doi.org/10.1007/s00382-020-05233-2

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

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, *12*(11), 4823–4873. https://doi.org/10.5194/gmd-12-4823-2019

Tatebe, H., Ogura, T., Nitta, T., Komuro, Y., Ogochi, K., Takemura, T., et al. (2019). Description and basic evaluation of simulated mean state, internal variability, and climate sensitivity in MIROC6. *Geoscientific Model Development*, *12*(7), 2727–2765. https://doi.org/10.5194/gmd-12-2727-2019

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin America Meteorology Social*, *93*(4), 485–498. https://doi.org/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. https://doi.org/10.1029/2022GL099624

Wehner, M. F., Reed, K. A., Li, F., Prabhat, Bacmeister, J., Chen, C.-T., et al. (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. https://doi.org/10.1002/2013MS000276

White, R. H., Anderson, S., Booth, J. F., Braich, G., Draeger, C., Fei, C., et al. (2023). The unprecedented Pacific Northwest heatwave of June 2021. *Nature Communications*, *14*(727), 727. https://doi.org/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. https://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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/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. https://doi.org/10.1016/j.wace.2020.100266

Zelle, H., van Oldenborgh, G. J., Burgers, G., & Dijkstra, H. (2005). El Niño and greenhouse warming: Results from ensemble simulations with the NCAR CCSM. *Journal of Climate*, *18*(22), 4669–4683. https://doi.org/10.1175/JCLI3574.1

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., et al. (2020). The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 193–214. https://doi.org/10.1071/ES19035