Supplement for: Three decades of simulating global temperature patterns with coupled global climate models

Lukas Brunner,[1,2*] Rohit Ghosh,[3] Leopold Haimberger,[1] Cathy Hohenegger,[4] Dian Putrasahan,[4] Thomas Rackow,[5] Reto Knutti,[6] and Aiko Voigt[1]

[1] Department of Meteorology and Geophysics, University of Vienna, Vienna, Austria

[2] Research Unit Sustainability and Climate Risk, Center for Earth System Research and Sustainability (CEN), Universität Hamburg, Hamburg, Germany

[3] Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research (AWI), Bremerhaven, Germany

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

[5] European Centre for Medium-Range Weather Forecasts (ECMWF), Bonn, Germany

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

*To whom correspondence should be addressed; E-mail: lukas.brunner@uni-hamburg.de

The PDF file includes:

- Supplementary Methods
- Supplementary figures S1 to S12
- Supplementary tables S1 to S10
- Supplementary References

The additional online supplement includes:

- Additional figures, source code, and data: Brunner [2]

Supplementary Methods

Reference data quality assessment

It seems clear that the 10 reference datasets used in this study are not fully equal estimates of the (unknown) true state of the climate system. It has, for example, been shown that more recent references represent the climate system better than their predecessors [e.g., 11, 16] and that individual reference datasets suffer from bugs or other issues in some regions and/or time periods, deteriorating their reliability (e.g., [26]; outliers in figure S10).

Notwithstanding, we draw on all 10 reference datasets to (1) account for the (historical) uncertainty in our knowledge of the true state of the climate system, (2) avoid the potential of underestimating model error due to dependencies between individual models and references, and (3) allow a comparison of model performance against reference performance in a leave-one-out cross-validation where one reference is compared against the others.

We provide several "sanity checks" for the references in the following. Each of them emerges as the coldest or warmest for 2 % to 22 % of Earth's surface (figure S9) and is considered an outlier (below the 25th percentile minus 1.5× the interquartile range or about the 75th percentile plus 1.5× the interquartile range) for 1 % to 11 % of the surface area (figure S10). Based on this we argue that at least none of the used datasets is clearly worse than the others and that newer datasets are not necessarily better everywhere. As an example: ERA5, as one of the latest reanalysis datasets, has a relatively large percentage of cold regions and is considered an outlier across much of the tropical warm pool region. Somewhat surprisingly, the relatively old 2nd generation reanalysis ERA-40, which was used already extensively for intercomparison with models and optimal fingerprinting at its time [e.g., 20], is the best performing reference dataset for the outlier metric.

Finally, some studies have used models to evaluate observation-based datasets [12, 21] and we show that using this approach also all 10 references perform reasonably (figure S8). We also note that we have excluded JRA25 [17] from the analysis based on the

considerations outlined above.

Relation between the reference range and the inside fraction

We relate the spatial pattern of reference range width to the number of models found inside of this range to examine whether the reference range or model performance is the main driver of this metric. We start with showing the width of the reference range (i.e., $\tau_{\text{ref\_range}} = \tau_{\text{ref\_max}} - \tau_{\text{ref\_min}}$) at each grid cell in figure S1a. The colorbar is chosen to highlight three rough domains of magnitudes in this metric: ocean (mostly: $\tau_{\text{ref\_range}} < 1\,\text{K}$), land ($1\,\text{K} \leq \tau_{\text{ref\_range}} < 3\,\text{K}$), and high latitudes ($3\,\text{K} \leq \tau_{\text{ref\_range}}$).

The fraction of all 165 CMIP models falling within the reference range at each grid cell is shown in figure S1b. Areas, where the fraction is large, indicate regions where (1) our knowledge of the true climate is low (i.e., where the width of the reference range is high) and/or (2) it is easy for models to correctly simulate surface temperature. To investigate this in more detail we also show how much of Earth's surface is simulated within the bounds of the references (figure S1c) and how this inside fraction is determined by the reference range width (figure S1d).

There is not a single grid cell where no model manages to fall between the 10 reference datasets. In addition, almost the entire Earth (95 %) has at least 10 % of the 165 models falling within the reference range. This means that there are no regions, where observed mean temperature can not be simulated by at least some models within observational uncertainty. For half of the Earth's surface, mean temperature is simulated correctly by at least a quarter of models and about 15 % even by more than half of the models (see figure S1c).

To disentangle what is mainly driving the models' ability to simulate a given grid cell, we show the correlation between the reference range width and the fraction of models inside the reference range in figure S1d. Evaluated over all grid cells, about half of the variance (coefficient of determination $r^2 = 0.48$) in the inside fraction can be explained

by the reference uncertainty. This correlation strongly depends on the domain, with the reference range strongly determining how many models fall within the reference range at high latitudes (r^2 : 0.80 pole-wards of ±60°N). Over the land and ocean outside the high latitudes, in turn, about 40 % and 30 % of the inside fraction is determined by the reference range.

All these findings are robust when using available historical model runs instead and when calculating the reference range based on a Gaussian fit and only based on ERA5, MERRA2, and JRA55 (table S4).

Supplementary figures

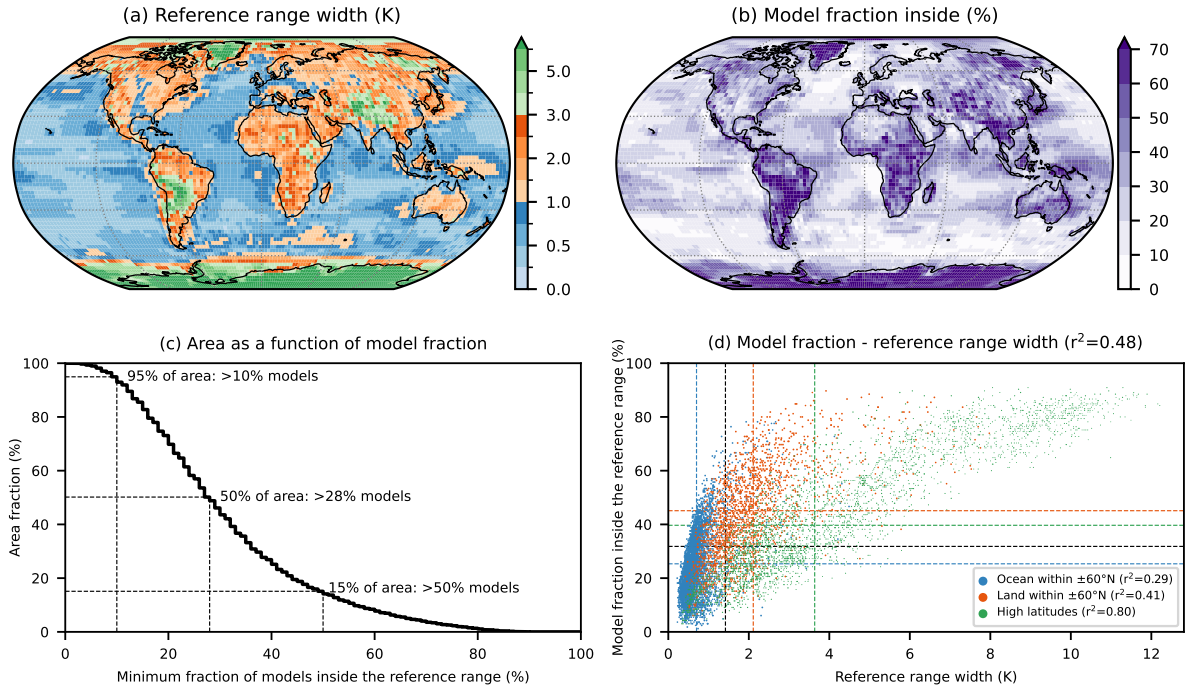

Figure S1: Relation between the reference range width and fraction of models inside the reference range. (a) Width of the range for all 10 reference datasets. Note that different step sizes are used for each of the three color regimes. (b) Fraction of 165 CMIP models falling within the reference range. (c) Fraction of surface area as a function of the fraction of models inside the reference range. (d) Relation between the reference range width and fraction of models inside the range for each of the 10'368 grid cells. The grid cells are separated into low- and mid-latitude ocean (blue) and land (orange) as well as high latitudes (green). The dot sizes scale with the cosine of the latitude of each grid cell and the coefficients of determination (r^2) are area-weighted. The dashed lines indicate the respective mean values.

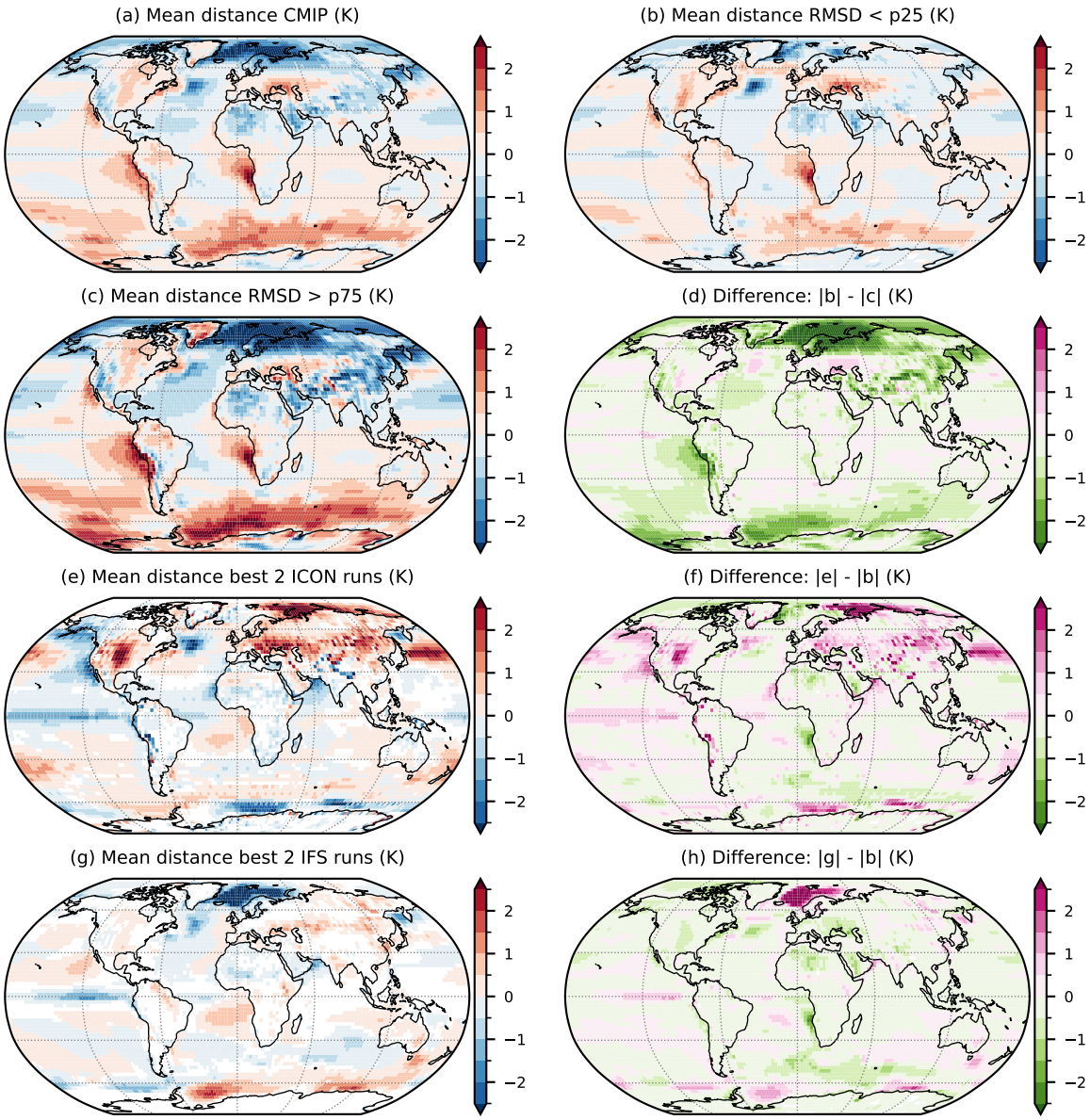

Figure S2: Multi-model mean bias for (a) all 165 CMIP models, (b) the 25 % models with the lowest RMSD (41 models), (c) the 25 % models with the highest RMSD, and (d) the difference of the absolute values of (b) and (c). Mean of the two best km-scale runs for the (e) ICON and (g) IFS model, as well as the absolute value of differences to the case shown in (c).

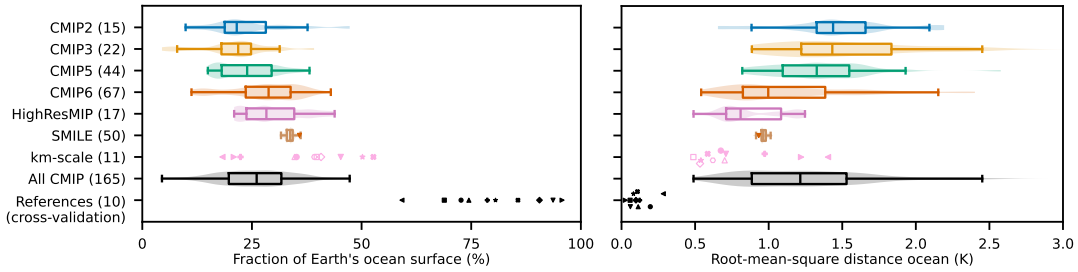

Figure S3: Same as figuere 1a,b in the main manuscript but land gird cells masked.

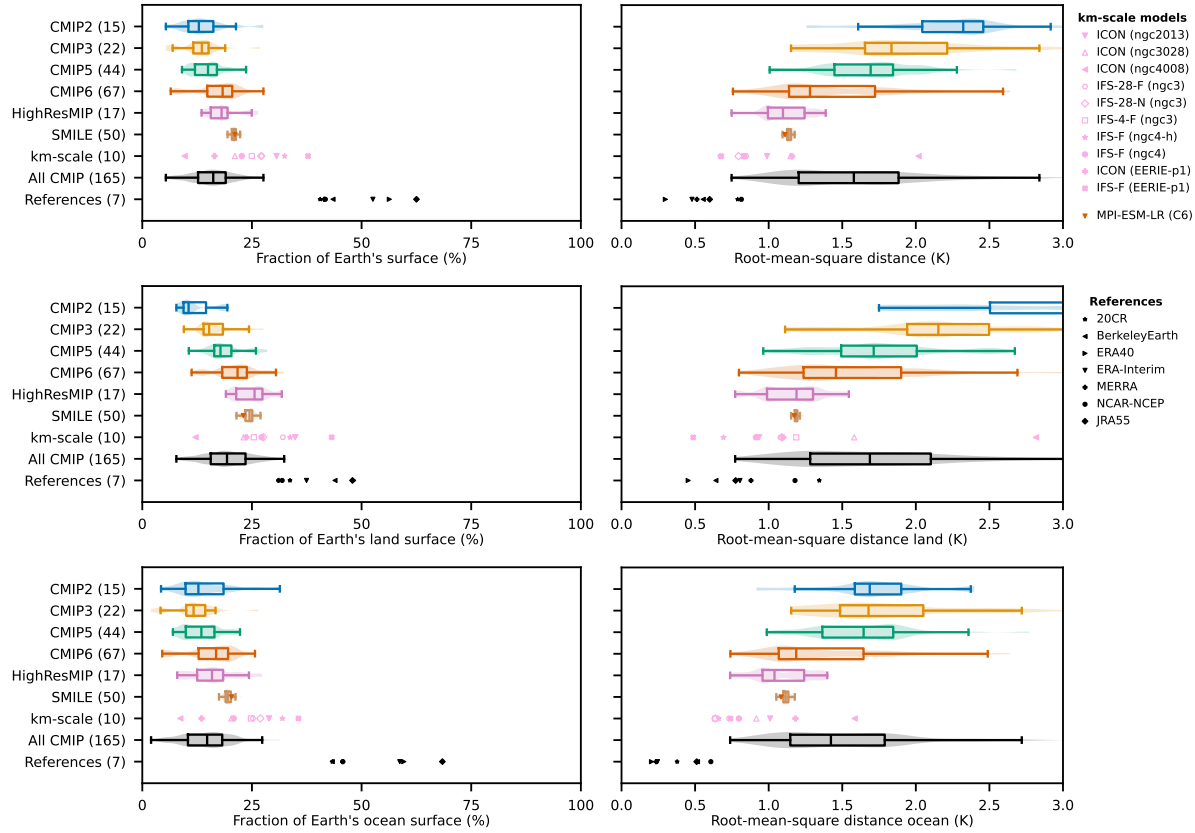

Figure S4: Same as figuere 1 in the main manuscript but using only the three modern reanalysis datasets ERA5, JRA3Q, and MERRA2. Note that for this case the Reference category simply uses the remaining seven references and no leave-one-out cross-validation.

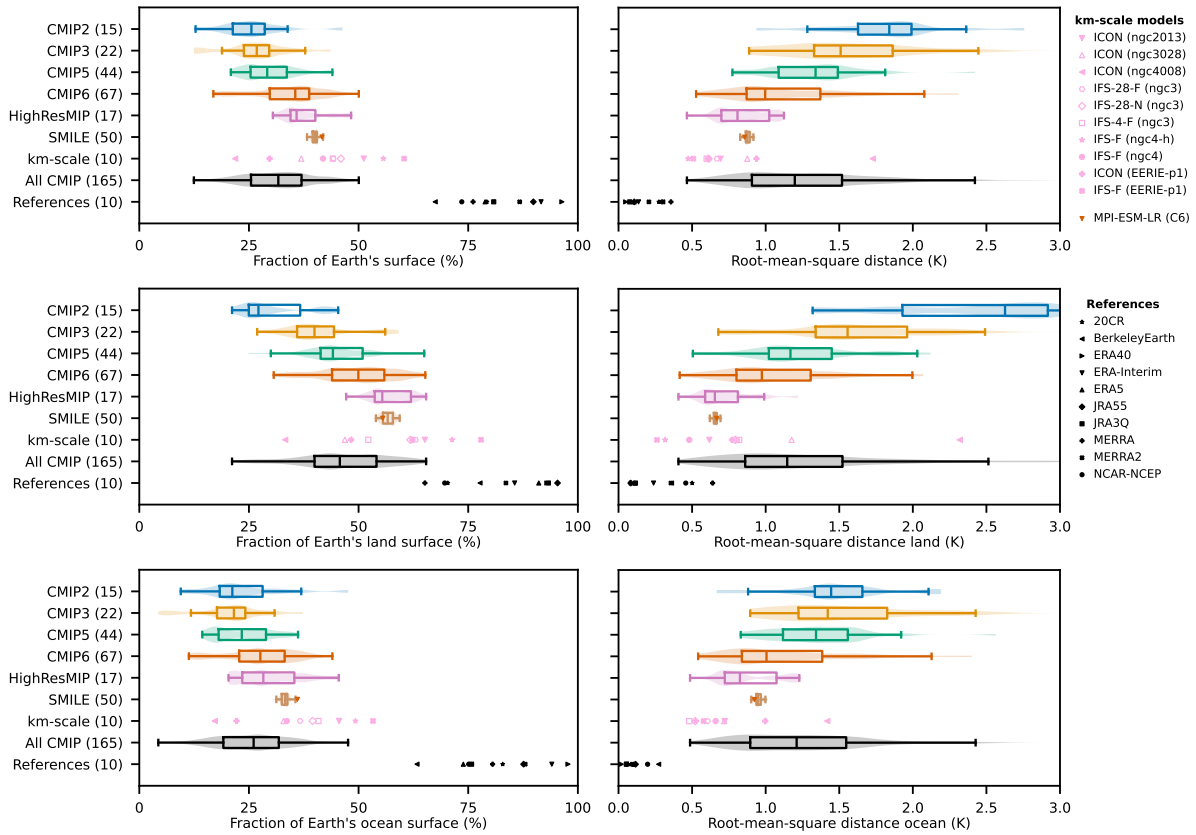

Figure S5: Same as figuere 1 in the main manuscript but using the 90 % range of a Gaussian fit as reference. Note that this case each reference is evaluated against a Gaussian fitted to all 10 references.

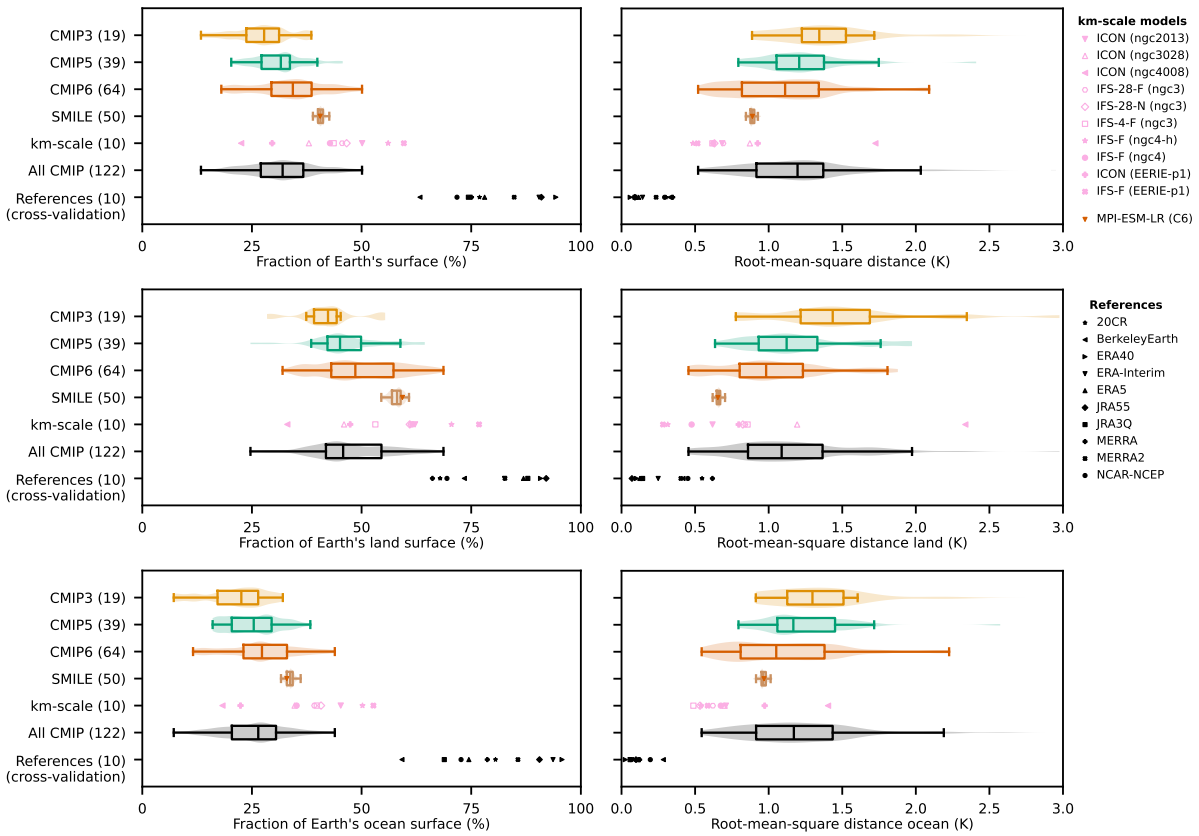

Figure S6: Same as figuere 1 in the main manuscript but using historical simulations for the CMIP models if available.

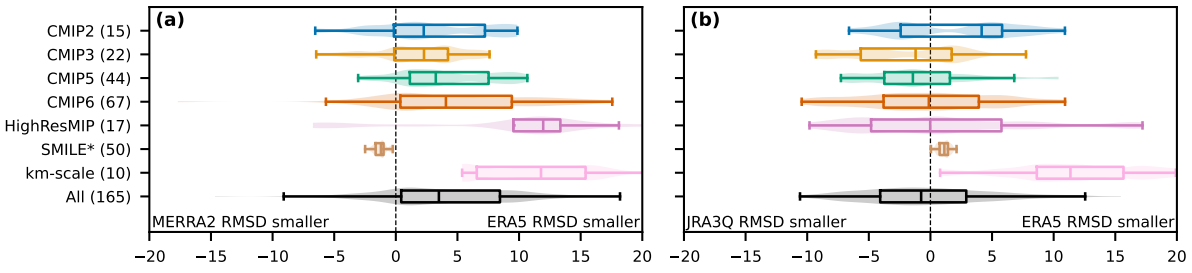

Figure S7: Same as figuere 2b in the main manuscript but comparing MERRA2-ERA5 and JRA3Q-ERA5.

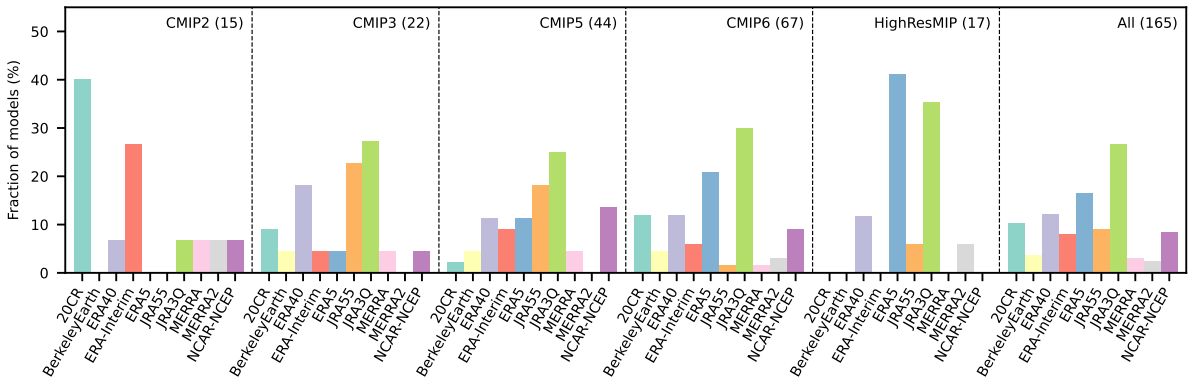

Figure S8: Fraction of models closest to each reference. Shown is the fraction of models for which each reference dataset (shown on the x-axis) leads to the lowest RMSD. km-scale models are not included as they are all closest to ERA5. See also table S5 for a full list.

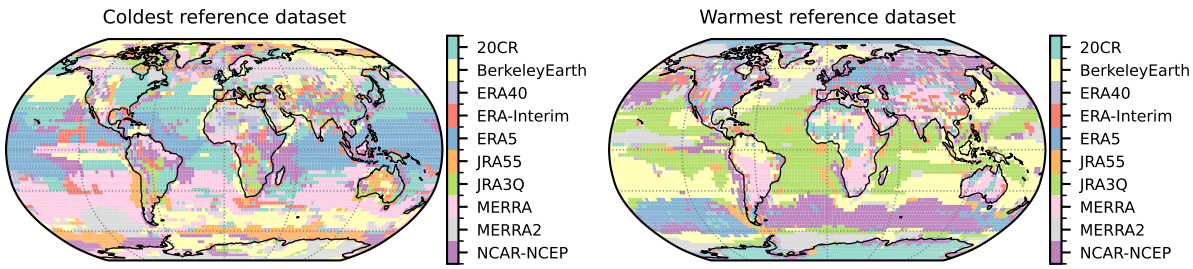

Figure S9: Coldest and warmest reference dataset at each grid cell.

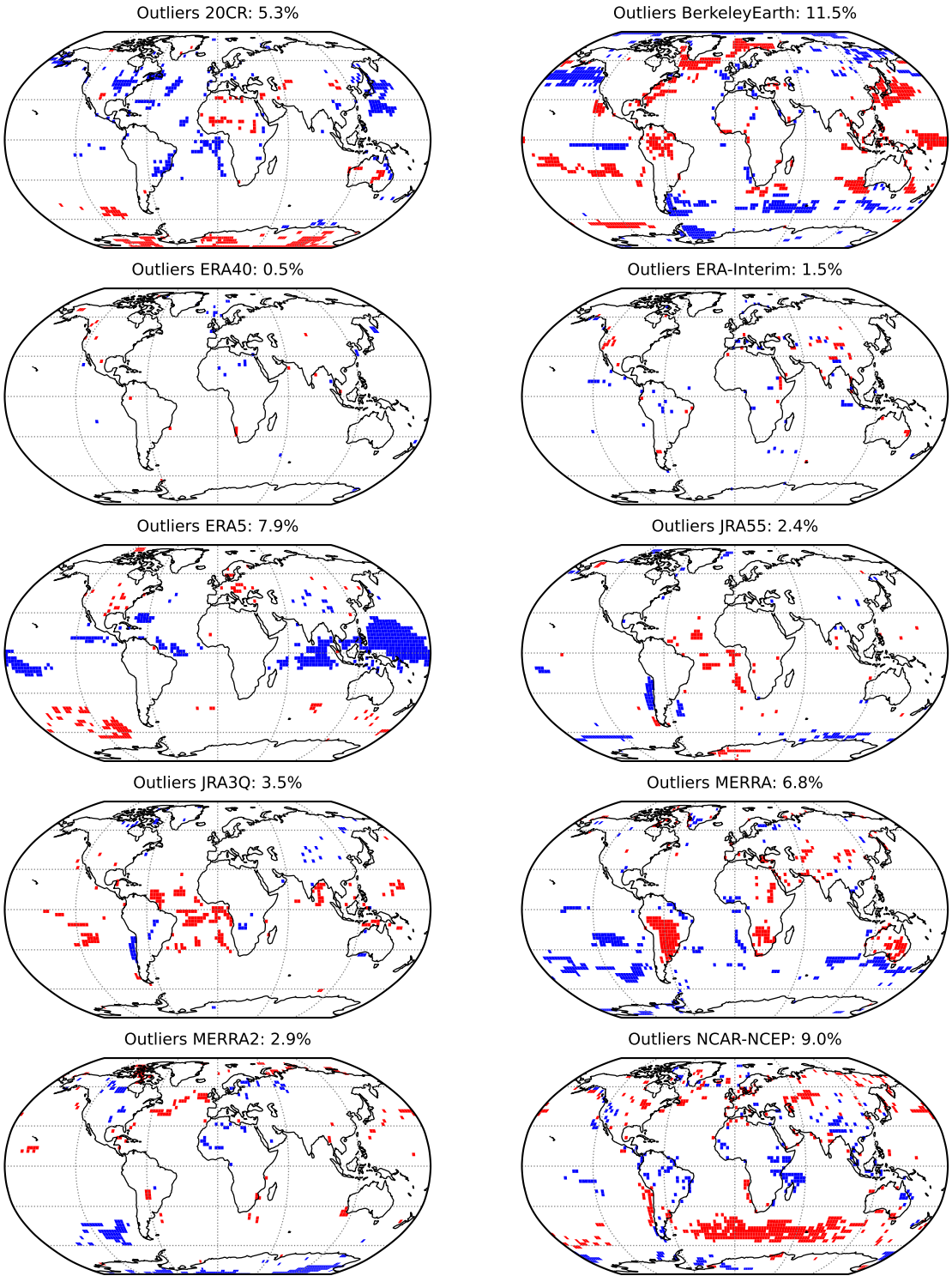

Figure S10: Grid cells where each reference is considered an outlier: below the 25th percentile minus 1.5× the interquartile range (blue) or above the 75th percentile plus 1.5× the interquartile range (red) of all 10 references.

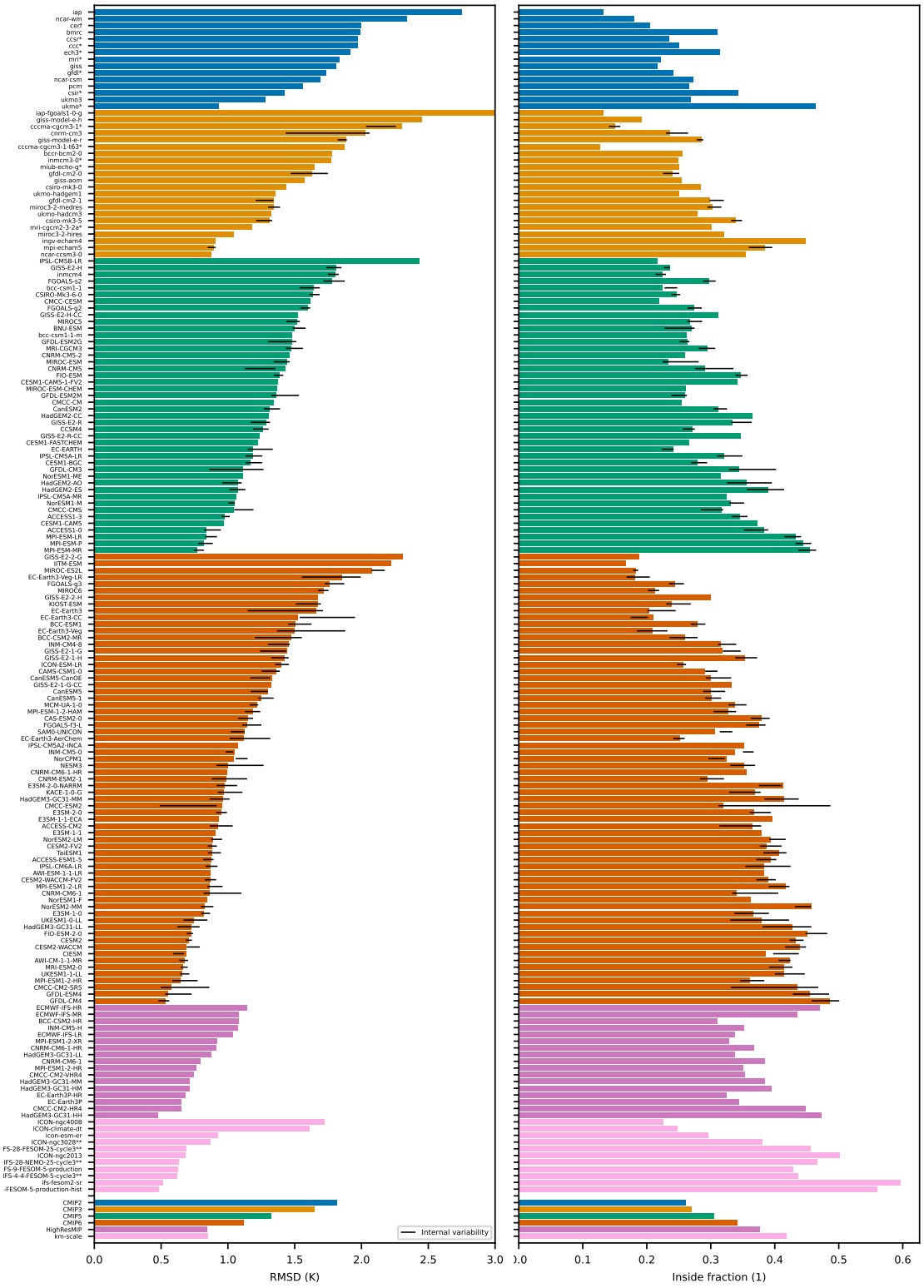

Figure S11: Root-mean-square distance (RMSD) and inside area fraction for all 176 models. Black horizontal lines: for models with more than 221 years of pre-industrial control run available, an additional 20 20-year climatologies are used to estimate the effect of internal variability.

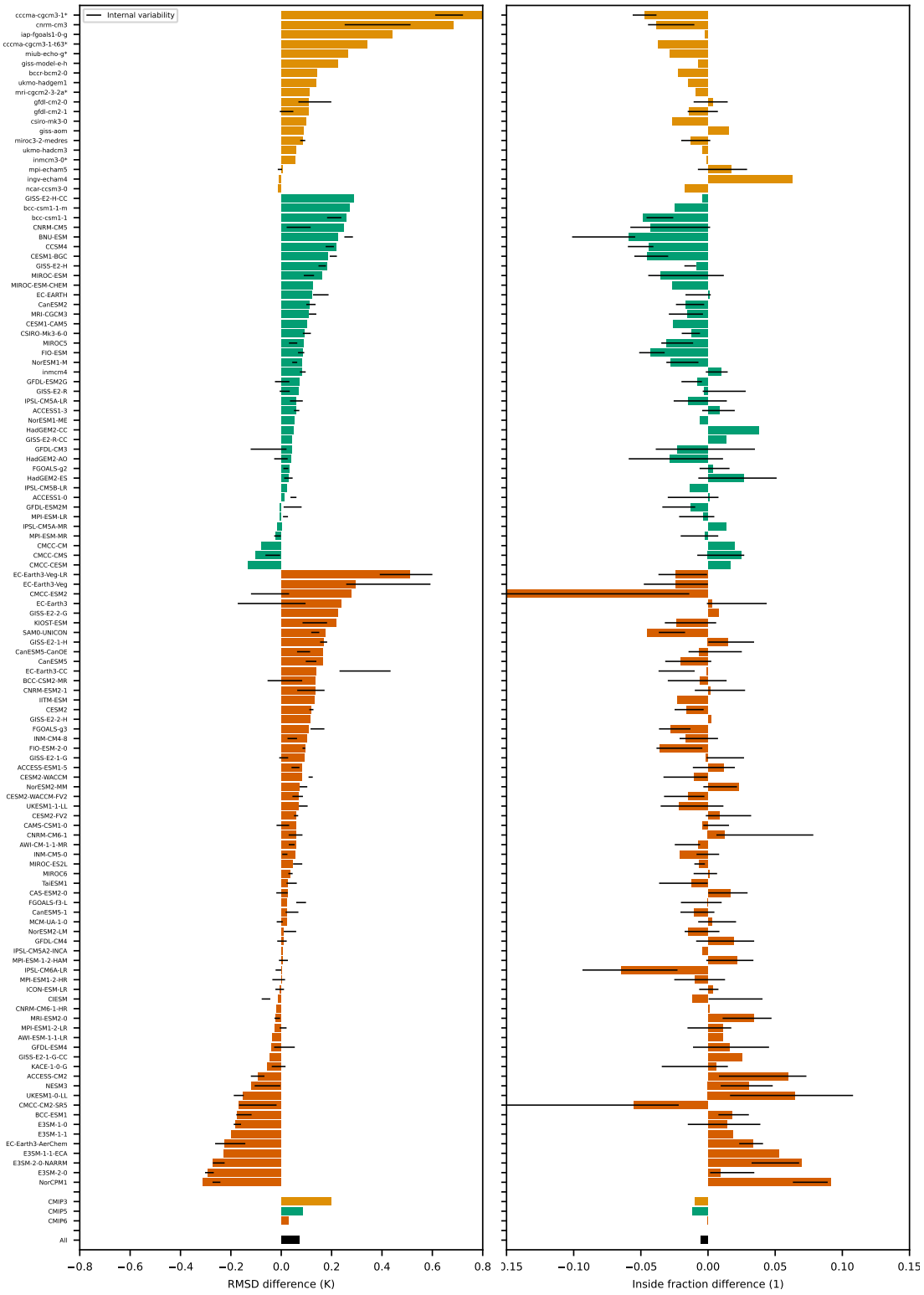

Figure S12: Root-mean-square distance (RMSD) and inside area fraction difference between pre-industrial control and historical simulations for all models which provide both. Black horizontal lines: same as in figure S11.

Supplementary tables

Table S1: Summary of root-mean-square distance (RMSD) changes with model resolution calculated over all models in a generation and over high-low resolution pairs of the same model.

Base	Pairs /Count	Res. (km)	Slope	Further information
Theil-Sen CMIP2	15		−0.23 K/100 km	CI 95 %: −0.45 to 0.00 K/100 km
Theil-Sen CMIP3	22		−0.38 K/100 km	CI 95 %: −0.54 to −0.21 K/100 km
cccma-cgcm3-1	t63/-	250/333	−0.50 K/100 km	[22]
miroc3-2	hires/medres	100/250	−0.18 K/100 km	[22]
Theil-Sen CMIP5	44		−0.16 K/100 km	CI 95 %: −0.33 to −0.01 K/100 km
bcc-csm1-1	m/-	100/250	−0.10 K/100 km	[27]
CMCC	CM/CMS	67/166	−0.30 K/100 km	[23]
CESM1-CAM5	-/FV2	96/192	−0.38 K/100 km	[14]
IPSL-CM5A	MR/LR	157/235	−0.15 K/100 km	[7]
Theil-Sen CMIP6	67		−0.24 K/100 km	CI 95 %: −0.39 to −0.08 K/100 km
CESM2	-/FV2	96/192	−0.18 K/100 km	[5]
CESM2-WACCM	-/FV2	96/192	−0.18 K/100 km	[8]
CNRM-CM6-1	HR/-	44/125	0.17 K/100 km	[24]
EC-Earth3-Veg	-/LR	62/100	−0.96 K/100 km	[6]
HadGEM3-GC31	MM/LL	60/136	0.31 K/100 km	[1]
MPI-ESM1-2	HR/LR	83/166	−0.26 K/100 km	[13]
NorESM2	MM/LM	96/192	−0.06 K/100 km	[25]
Theil-Sen CMIP	148		−0.30 K/100 km	CI 95 %: −0.36 to −0.23 K/100 km
Mean CMIP slopes	4		−0.25 K/100 km	
Median pairs	13		−0.18 K/100 km	
Theil-Sen HRMIP	17		−0.08 K/100 km	CI 95 %: −0.50 to 0.24 K/100 km
CMCC-CM2	VHR4/-HR4	24/96	0.14 K/100 km	[4]
CNRM-CM6-1	HR/-	44/125	0.15 K/100 km	[24]
EC-Earth3P	HR/-	31/62	0.10 K/100 km	[10]
ECMWF-IFS	HR/MR	44/88	0.13 K/100 km	[18]
HadGEM3-GC31	HH/LL	60/136	−0.36 K/100 km	[19]
MPI-ESM1-2	XR/HR	42/83	0.38 K/100 km	[13]

Table S2: **Summary of km-scale models used in this study.** The "Years" column gives the number of years used, for runs with less than 20 years, all available years are used. The "Res." column gives the atmospheric and ocean resolution, respectively. The "DCP" column indicates if deep convection parametrization was used. The "Scenario" column gives the forcing used.

*on with 1/5 cloud-base mass flux

Acronym	Years	Res. (km)	DCP	Scenario
ICON (ngc2013)	20	10/10	off	fixed 2020
ICON (ngc3028)	5	5/5	off	fixed 2020
ICON (ngc4008)	20	10/5	off	SSP3-7.0 from 2020
ICON (ClimateDT)	20	10/10	off	historical/SSP3-7.0
ICON (EERIE-p1)	20	10/5	off	1950-control
IFS-4-F (ngc3)	5	4.4/5	slightly on*	fixed 2020
IFS-28-N (ngc3)	5	28/25	on	fixed 2020
IFS-28-F (ngc3)	5	28/25	on	fixed 2020
IFS-F (ngc4)	20	9/5	slightly on*	SSP3-7.0 from 2020
IFS-F (ngc4-h)	20	9/5	slightly on*	historical/SSP3-7.0
IFS-F (EERIE-p1)	20	9/5	on	1950-control

Table S3: The default regridding strategy in this study is conservative remapping. However, not all km-scale model grids allow this approach. In these, cases we have use the next best regridding strategy available from the grid that best matched our target grid of 2.5° as listed below. Healpix zoom level 5 corresponds to 10'958 grid cells compared to 10'368 in the target resolution.

Acronym	Spatial remapping strategy
ICON (ngc2013)	conservative remapping from native resolution
ICON (ngc3028)	nearest-neighbour remapping from healpix zoom level 5
IFS-4-F (ngc3)	nearest-neighbour remapping from 1° unstructured grid
IFS-28-F (ngc3)	nearest-neighbour remapping from 1° unstructured grid
IFS-28-N (ngc3)	nearest-neighbour remapping from 1° unstructured grid
ICON (ngc4008)	nearest-neighbour remapping from healpix zoom level 5
IFS-F (ngc4)	bilinear remapping from healpix zoom level 5
IFS-F (ngc4-h)	bilinear remapping from healpix zoom level 5
ICON (EERIE-p1)	conservative remapping from 0.25°
IFS-F (EERIE-p1)	conservative remapping from 0.25°

Table S4: Reference range and corresponding inside area fraction statistics for different domains: ocean within ±60°N, land within ±60°N, high latitude areas outside of ±60°N, and global. The "Width" columns give the area-mean reference range width and the fraction of models within the reference range is given in the "Frac." column. The r^2 columns give the area-weighted coefficient of determination. The "All references" category corresponds to the case in the main paper, "Gauss" corresponds to the Gaussian fit, and "Strict" to the case using only ERA5, JRA55, and MERRA2. The "Historical" category uses the same references as "All references" but only models with provide a historical simulation (122/165).

	All references			**Gauss**			**Strict**			**Historical**	
Domain	Width	Frac.	r^2	Range	Frac.	r^2	Range	Frac.	r^2	Frac.	r^2
Ocean	0.7 K	25 %	0.29	0.7 K	25 %	0.27	0.4 K	14 %	0.43	25 %	0.26
Land	2.1 K	45 %	0.41	2.1 K	45 %	0.41	0.9 K	19 %	0.51	46 %	0.43
High lat.	3.6 K	40 %	0.80	3.5 K	39 %	0.81	2.3 K	25 %	0.80	42 %	0.79
Global	1.4 K	32 %	0.48	1.4 K	32 %	0.48	0.7 K	17 %	0.49	32 %	0.50

Table S5: Summary of all datasets included in the study and their properties. "Δd": mean horizontal resolution of the provided atmospheric grid; "RMSD": root-mean-square distance to the reference range; "IF": inside fraction; "T": global mean of the 20-year mean surface temperature using the pre-industrial control run (note that this value is removed from each model before any analysis); "Closest ref.": reference dataset leading to the lowest RMSD. If a historical run is available its values are in brackets, for "T" the values in brackets give the differences historical minus piControl. CMIP2 and CMIP3 models using flux adjustment are marked with an * based on Gupta et al. [9] and Meehl et al. [15]. km-scale models with less than 20 years of data available are marked with an *.

CMIP2	Δd (km)	RMSD (K)	IF (%)	T (K)	Closest ref.
bmrc	377	2.0	31 %	288.8	ERA-Interim
ccc*	333	2.0	25 %	287.4	ERA-Interim
ccsr*	499	2.0	23 %	287.7	20CR
cerf	333	2.0	21 %	289.9	NCAR-NCEP
csir*	377	1.4	34 %	286.9	20CR
ech3*	499	1.9	31 %	288.1	ERA-Interim
gfdl*	515	1.7	24 %	284.8	20CR
giss	392	1.8	22 %	287.5	MERRA
iap	515	2.8	13 %	287.0	20CR
mri*	392	1.8	22 %	288.7	20CR
ncar-csm	250	1.7	27 %	285.8	ERA-Interim
ncar-wm	515	2.3	18 %	290.4	20CR
pcm	250	1.6	27 %	285.4	ERA40
ukmo3	270	1.3	27 %	286.7	JRA3Q
ukmo*	270	0.9	46 %	286.7	MERRA2
CMIP3	Δd (km)	RMSD (K)	IF (%)	T (K)	Closest ref.
bccr-bcm2-0	250	1.8 (1.6)	26 % (28 %)	285.4 (+0.5)	JRA3Q
cccma-cgcm3-1-t63*	250	1.9 (1.5)	13 % (16 %)	285.2 (+1.1)	JRA55
cccma-cgcm3-1*	333	2.3 (1.5)	15 % (20 %)	285.4 (+1.5)	JRA55
cnrm-cm3	250	2.0 (1.3)	24 % (27 %)	285.4 (+1.4)	ERA40
csiro-mk3-0	166	1.4 (1.3)	28 % (31 %)	285.2 (+1.0)	JRA3Q
csiro-mk3-5	166	1.3	34 %	288.1	JRA55
gfdl-cm2-0	198	1.6 (1.5)	24 % (24 %)	285.7 (+0.1)	JRA3Q
gfdl-cm2-1	198	1.3 (1.2)	30 % (31 %)	286.4 (+0.4)	ERA40
giss-aom	307	1.6 (1.5)	25 % (24 %)	286.9 (+0.6)	MERRA
giss-model-e-h	392	2.4 (2.2)	19 % (20 %)	288.1 (+0.4)	20CR
giss-model-e-r	392	1.9	29 %	287.2	20CR
iap-fgoals1-0-g	258	4.4 (3.9)	13 % (13 %)	284.8 (+0.9)	JRA55

ingv-echam4	100	0.9 (0.9)	45 % (39 %)	287.9 (+0.7)	ERA40
inmcm3-0*	397	1.8 (1.7)	25 % (25 %)	285.9 (+0.3)	NCAR-NCEP
miroc3-2-hires	100	1.0	32 %	287.1	ERA5
miroc3-2-medres	250	1.3 (1.3)	30 % (31 %)	286.6 (+0.1)	ERA-Interim
miub-echo-g*	333	1.6 (1.4)	25 % (28 %)	286.3 (+0.6)	JRA3Q
mpi-echam5	166	0.9 (0.9)	38 % (37 %)	287.4 (+0.2)	JRA3Q
mri-cgcm2-3-2a*	250	1.2 (1.1)	30 % (31 %)	285.7 (+0.7)	BerkeleyEarth
ncar-ccsm3-0	125	0.9 (0.9)	35 % (37 %)	286.3 (+0.5)	ERA40
ukmo-hadcm3	270	1.3 (1.3)	28 % (28 %)	286.7 (+0.5)	JRA3Q
ukmo-hadgem1	135	1.4 (1.2)	25 % (27 %)	285.7 (+0.4)	JRA55

CMIP5	Δd (km)	RMSD (K)	IF (%)	T (K)	Closest ref.
ACCESS1-0	135	0.8 (0.8)	38 % (38 %)	287.1 (+0.0)	ERA40
ACCESS1-3	135	1.0 (0.9)	35 % (34 %)	287.3 (+0.1)	JRA55
BNU-ESM	250	1.5 (1.3)	27 % (33 %)	286.0 (+0.9)	JRA55
CCSM4	96	1.3 (1.0)	27 % (32 %)	286.4 (+0.8)	JRA55
CESM1-BGC	96	1.2 (1.0)	28 % (32 %)	286.5 (+0.8)	JRA55
CESM1-CAM5-1-FV2	192	1.4	34 %	286.6	BerkeleyEarth
CESM1-CAM5	96	1.0 (0.9)	37 % (40 %)	286.3 (+0.4)	JRA3Q
CESM1-FASTCHEM	96	1.2	27 %	286.5	JRA55
CMCC-CESM	333	1.6 (1.7)	22 % (20 %)	286.9 (+0.3)	ERA-Interim
CMCC-CMS	166	1.0 (1.1)	32 % (29 %)	287.0 (+0.3)	JRA3Q
CMCC-CM	67	1.3 (1.4)	25 % (23 %)	286.6 (+0.3)	JRA3Q
CNRM-CM5-2	125	1.5	26 %	286.1	JRA3Q
CNRM-CM5	125	1.4 (1.2)	29 % (33 %)	286.4 (+0.5)	JRA3Q
CSIRO-Mk3-6-0	166	1.6 (1.5)	25 % (26 %)	285.9 (+0.3)	BerkeleyEarth
CanESM2	250	1.3 (1.2)	31 % (33 %)	286.8 (+0.6)	NCAR-NCEP
EC-EARTH	100	1.2 (1.1)	24 % (24 %)	285.5 (+0.8)	ERA5
FGOALS-g2	258	1.6 (1.6)	27 % (27 %)	285.5 (+0.4)	JRA55
FGOALS-s2	192	1.8	30 %	286.7	NCAR-NCEP
FIO-ESM	250	1.4 (1.3)	35 % (39 %)	286.3 (+0.9)	NCAR-NCEP
GFDL-CM3	198	1.1 (1.1)	34 % (37 %)	287.2 (-0.4)	ERA5
GFDL-ESM2G	198	1.5 (1.4)	26 % (27 %)	286.4 (+0.4)	ERA40
GFDL-ESM2M	198	1.4 (1.4)	26 % (27 %)	287.0 (+0.4)	ERA40
GISS-E2-H-CC	198	1.5 (1.2)	31 % (31 %)	287.4 (+0.9)	JRA3Q
GISS-E2-H	198	1.8 (1.6)	24 % (24 %)	288.2 (+0.5)	ERA5
GISS-E2-R-CC	198	1.2 (1.2)	35 % (33 %)	287.4 (+0.6)	MERRA
GISS-E2-R	198	1.3 (1.2)	33 % (34 %)	287.5 (+0.5)	MERRA
HadGEM2-AO	135	1.1 (1.0)	36 % (38 %)	287.0 (+0.4)	ERA40
HadGEM2-CC	135	1.3 (1.3)	37 % (33 %)	286.5 (+0.2)	ERA-Interim
HadGEM2-ES	135	1.1 (1.0)	39 % (36 %)	286.9 (+0.0)	ERA-Interim
IPSL-CM5A-LR	235	1.2 (1.1)	32 % (34 %)	285.3 (+0.8)	ERA40

IPSL-CM5A-MR	157	1.1 (1.1)	32 % (31 %)	286.1 (+0.8)	ERA5
IPSL-CM5B-LR	235	2.4 (2.4)	22 % (23 %)	286.2 (+0.5)	JRA3Q
MIROC-ESM-CHEM	250	1.4 (1.2)	26 % (29 %)	286.7 (+0.3)	NCAR-NCEP
MIROC-ESM	250	1.4 (1.3)	23 % (27 %)	287.2 (-0.1)	ERA-Interim
MIROC5	125	1.5 (1.4)	27 % (30 %)	287.8 (+0.2)	ERA5
MPI-ESM-LR	166	0.8 (0.8)	43 % (44 %)	286.6 (+0.7)	JRA3Q
MPI-ESM-MR	166	0.8 (0.8)	45 % (46 %)	286.9 (+0.6)	JRA3Q
MPI-ESM-P	166	0.8	44 %	286.7	JRA3Q
MRI-CGCM3	100	1.5 (1.4)	30 % (31 %)	286.8 (+0.3)	20CR
NorESM1-ME	192	1.1 (1.1)	32 % (32 %)	285.9 (+0.4)	NCAR-NCEP
NorESM1-M	192	1.0 (1.0)	33 % (36 %)	286.3 (+0.5)	NCAR-NCEP
bcc-csm1-1-m	100	1.5 (1.2)	26 % (29 %)	287.3 (+1.0)	JRA55
bcc-csm1-1	250	1.6 (1.4)	22 % (27 %)	286.8 (+0.9)	JRA55
inmcm4	154	1.8 (1.7)	22 % (22 %)	286.2 (+0.7)	JRA3Q
CMIP6	Δd (km)	RMSD (K)	IF (%)	T (K)	Closest ref.
ACCESS-CM2	136	0.9 (1.0)	36 % (31 %)	287.3 (-0.1)	JRA3Q
ACCESS-ESM1-5	135	0.9 (0.8)	39 % (38 %)	287.6 (+0.3)	ERA40
AWI-CM-1-1-MR	83	0.7 (0.6)	42 % (43 %)	286.9 (+0.6)	ERA40
AWI-ESM-1-1-LR	166	0.9 (0.9)	38 % (37 %)	286.1 (+0.6)	ERA40
BCC-CSM2-MR	100	1.5 (1.3)	26 % (27 %)	287.9 (+0.3)	JRA3Q
BCC-ESM1	250	1.5 (1.7)	28 % (26 %)	287.9 (+0.1)	BerkeleyEarth
CAMS-CSM1-0	100	1.4 (1.3)	29 % (29 %)	287.2 (+0.2)	20CR
CAS-ESM2-0	125	1.1 (1.1)	38 % (36 %)	287.3 (+0.0)	ERA5
CESM2-FV2	192	0.9 (0.8)	39 % (38 %)	287.1 (+0.3)	ERA-Interim
CESM2-WACCM-FV2	192	0.9 (0.8)	39 % (40 %)	287.3 (+0.3)	ERA-Interim
CESM2-WACCM	96	0.7 (0.6)	44 % (45 %)	287.2 (+0.2)	ERA40
CESM2	96	0.7 (0.6)	43 % (45 %)	287.2 (+0.4)	ERA40
CIESM	96	0.7 (0.7)	39 % (40 %)	287.6 (+0.7)	ERA5
CMCC-CM2-SR5	96	0.6 (0.7)	43 % (49 %)	287.2 (+0.7)	MERRA
CMCC-ESM2	96	1.0 (0.7)	32 % (50 %)	286.7 (+1.1)	JRA3Q
CNRM-CM6-1-HR	44	1.0 (1.0)	36 % (35 %)	285.9 (+0.3)	ERA5
CNRM-CM6-1	125	0.9 (0.8)	34 % (33 %)	286.1 (+0.4)	JRA3Q
CNRM-ESM2-1	125	1.0 (0.9)	29 % (29 %)	286.7 (+0.4)	ERA5
CanESM5-1	250	1.2 (1.2)	30 % (31 %)	286.5 (+0.7)	NCAR-NCEP
CanESM5-CanOE	250	1.3 (1.2)	30 % (31 %)	286.4 (+0.8)	NCAR-NCEP
CanESM5	250	1.3 (1.1)	30 % (32 %)	286.5 (+1.0)	NCAR-NCEP
E3SM-1-0	89	0.8 (1.0)	37 % (35 %)	286.9 (+0.2)	ERA5
E3SM-1-1-ECA	89	0.9 (1.2)	40 % (34 %)	286.7 (+0.1)	ERA5
E3SM-1-1	89	0.9 (1.1)	38 % (36 %)	286.9 (+0.0)	ERA5
E3SM-2-0-NARRM	89	1.0 (1.2)	41 % (34 %)	286.6 (-0.1)	ERA5
E3SM-2-0	89	1.0 (1.2)	37 % (36 %)	286.8 (-0.2)	ERA5

EC-Earth3-AerChem	62	1.1 (1.3)	25 % (22 %)	287.3 (+0.0)	ERA5
EC-Earth3-CC	62	1.5 (1.4)	21 % (21 %)	287.5 (+0.6)	ERA5
EC-Earth3-Veg-LR	100	1.9 (1.3)	18 % (21 %)	286.5 (+0.7)	20CR
EC-Earth3-Veg	62	1.5 (1.2)	21 % (23 %)	287.2 (+0.5)	20CR
EC-Earth3	62	1.7 (1.4)	20 % (20 %)	286.8 (+0.5)	20CR
FGOALS-f3-L	99	1.1 (1.1)	37 % (38 %)	286.1 (+0.5)	JRA3Q
FGOALS-g3	188	1.8 (1.7)	24 % (27 %)	285.9 (+0.6)	JRA3Q
FIO-ESM-2-0	96	0.7 (0.6)	45 % (49 %)	286.6 (+0.7)	JRA3Q
GFDL-CM4	99	0.5 (0.5)	49 % (47 %)	285.9 (+0.3)	JRA3Q
GFDL-ESM4	99	0.6 (0.6)	46 % (44 %)	286.5 (+0.3)	JRA3Q
GISS-E2-1-G-CC	198	1.3 (1.4)	33 % (31 %)	287.1 (+0.2)	JRA3Q
GISS-E2-1-G	198	1.4 (1.3)	32 % (32 %)	286.9 (+0.4)	JRA3Q
GISS-E2-1-H	198	1.4 (1.3)	35 % (34 %)	287.3 (+0.6)	ERA5
GISS-E2-2-G	198	2.3 (2.1)	19 % (18 %)	284.8 (+0.6)	JRA3Q
GISS-E2-2-H	198	1.7 (1.6)	30 % (30 %)	285.4 (+0.6)	JRA3Q
HadGEM3-GC31-LL	136	0.7	43 %	287.5	JRA3Q
HadGEM3-GC31-MM	60	1.0	41 %	287.5	ERA5
ICON-ESM-LR	158	1.4 (1.4)	26 % (25 %)	286.9 (+0.8)	MERRA2
IITM-ESM	168	2.2 (2.1)	17 % (19 %)	287.2 (+0.7)	20CR
INM-CM4-8	154	1.5 (1.4)	32 % (33 %)	286.3 (+0.7)	ERA-Interim
INM-CM5-0	154	1.0 (1.0)	34 % (36 %)	286.1 (+0.6)	ERA-Interim
IPSL-CM5A2-INCA	235	1.1 (1.1)	35 % (36 %)	285.9 (+0.8)	ERA40
IPSL-CM6A-LR	157	0.9 (0.9)	38 % (45 %)	285.6 (+1.1)	JRA3Q
KACE-1-0-G	136	1.0 (1.0)	37 % (36 %)	287.8 (-0.3)	ERA5
KIOST-ESM	166	1.7 (1.5)	24 % (26 %)	285.8 (+0.8)	20CR
MCM-UA-1-0	258	1.2 (1.2)	34 % (33 %)	286.7 (+0.5)	NCAR-NCEP
MIROC-ES2L	250	2.1 (2.0)	18 % (19 %)	288.1 (+0.4)	20CR
MIROC6	125	1.7 (1.7)	21 % (21 %)	288.4 (+0.3)	20CR
MPI-ESM-1-2-HAM	166	1.2 (1.2)	33 % (31 %)	286.9 (+0.1)	JRA3Q
MPI-ESM1-2-HR	83	0.6 (0.6)	36 % (37 %)	287.1 (+0.5)	JRA3Q
MPI-ESM1-2-LR	166	0.9 (0.9)	42 % (41 %)	286.7 (+0.6)	MERRA2
MRI-ESM2-0	100	0.7 (0.7)	41 % (38 %)	287.0 (+0.2)	JRA55
NESM3	166	1.0 (1.1)	35 % (32 %)	286.9 (+0.2)	JRA3Q
NorCPM1	192	1.0 (1.4)	32 % (23 %)	286.0 (+0.0)	NCAR-NCEP
NorESM1-F	192	0.8	36 %	287.6	NCAR-NCEP
NorESM2-LM	192	0.9 (0.9)	39 % (41 %)	287.7 (+0.2)	ERA40
NorESM2-MM	96	0.8 (0.8)	46 % (43 %)	287.0 (+0.2)	ERA40
SAM0-UNICON	96	1.1 (1.0)	31 % (35 %)	286.1 (+0.4)	JRA3Q
TaiESM1	96	0.9 (0.9)	41 % (42 %)	287.1 (-0.0)	JRA3Q
UKESM1-0-LL	136	0.7 (0.9)	38 % (31 %)	286.5 (+0.2)	BerkeleyEarth
UKESM1-1-LL	136	0.7 (0.6)	41 % (44 %)	287.0 (+0.4)	BerkeleyEarth

HighResMIP	Δd (km)	RMSD (K)	IF (%)	T (K)	Closest ref.
BCC-CSM2-HR	40	1.1	31 %	287.6	JRA3Q
CMCC-CM2-HR4	96	0.7	45 %	287.7	ERA40
CMCC-CM2-VHR4	24	0.7	35 %	287.2	JRA55
CNRM-CM6-1-HR	44	0.9	37 %	285.9	ERA5
CNRM-CM6-1	125	0.8	38 %	286.1	ERA5
EC-Earth3P-HR	31	0.7	33 %	287.1	ERA5
EC-Earth3P	62	0.7	34 %	287.4	ERA5
ECMWF-IFS-HR	44	1.1	47 %	286.8	ERA5
ECMWF-IFS-LR	88	1.0	34 %	286.0	JRA3Q
ECMWF-IFS-MR	88	1.1	44 %	287.0	ERA5
HadGEM3-GC31-HH	25	0.5	47 %	286.5	ERA40
HadGEM3-GC31-HM	25	0.7	40 %	286.4	JRA3Q
HadGEM3-GC31-LL	136	0.9	34 %	285.8	JRA3Q
HadGEM3-GC31-MM	60	0.7	38 %	286.0	JRA3Q
INM-CM5-H	51	1.1	35 %	287.7	ERA5
MPI-ESM1-2-HR	83	0.8	35 %	287.5	MERRA2
MPI-ESM1-2-XR	42	0.9	33 %	287.1	JRA3Q
km-scale	Δd (km)	RMSD (K)	IF (%)	T (K)	Closest ref.
ICON (ClimateDT)	10	1.6	25 %	287.2	ERA5
ICON (ngc2013)	10	0.7	50 %	288.1	ERA5
ICON (ngc3028)	5	0.9	38 %	286.5	ERA5
ICON (ngc4008)	10	1.7	23 %	287.4	ERA5
IFS-28-F (ngc3)	28	0.7	46 %	287.4	ERA5
IFS-28-N (ngc3)	28	0.6	47 %	287.8	ERA5
IFS-4-F (ngc3)	4	0.6	44 %	288.0	ERA5
IFS-F (ngc4-h)	9	0.5	56 %	287.9	ERA5
IFS-F (ngc4)	9	0.6	43 %	289.1	ERA5
ICON (EERIE-p1)	10	0.9	30 %	285.8	ERA5
IFS-F (EERIE-p1)	9	0.5	60 %	286.1	ERA5
SMILE	Δd (km)	RMSD (K)	IF (%)	T (K)	Closest ref.
r1i1p1f1	166	0.9	41 %	287.3	MERRA2
r2i1p1f1	166	0.9	40 %	287.2	MERRA2
r3i1p1f1	166	0.9	40 %	287.3	MERRA2
r4i1p1f1	166	0.9	40 %	287.3	MERRA2
r5i1p1f1	166	0.9	41 %	287.2	MERRA2
r6i1p1f1	166	0.9	39 %	287.3	MERRA2
r7i1p1f1	166	0.9	41 %	287.2	MERRA2
r8i1p1f1	166	0.9	41 %	287.1	MERRA2
r9i1p1f1	166	0.9	41 %	287.3	MERRA2

r10i1p1f1	166	0.9	41 %	287.3	MERRA2
r11i1p1f1	166	0.9	40 %	287.3	MERRA2
r12i1p1f1	166	0.9	41 %	287.2	MERRA2
r13i1p1f1	166	0.9	40 %	287.2	ERA40
r14i1p1f1	166	0.9	39 %	287.2	MERRA2
r15i1p1f1	166	0.9	40 %	287.2	MERRA2
r16i1p1f1	166	0.8	41 %	287.2	MERRA2
r17i1p1f1	166	0.9	41 %	287.2	MERRA2
r18i1p1f1	166	0.9	41 %	287.2	MERRA2
r19i1p1f1	166	0.9	41 %	287.2	MERRA2
r20i1p1f1	166	0.9	41 %	287.2	MERRA2
r21i1p1f1	166	0.9	41 %	287.3	MERRA2
r22i1p1f1	166	0.9	41 %	287.3	MERRA2
r23i1p1f1	166	0.9	41 %	287.2	MERRA2
r24i1p1f1	166	0.9	39 %	287.3	MERRA2
r25i1p1f1	166	0.9	41 %	287.2	MERRA2
r26i1p1f1	166	0.9	40 %	287.3	MERRA2
r27i1p1f1	166	0.9	40 %	287.2	MERRA2
r28i1p1f1	166	0.9	41 %	287.2	MERRA2
r29i1p1f1	166	0.9	41 %	287.3	MERRA2
r30i1p1f1	166	0.9	40 %	287.2	MERRA2
r31i1p1f1	166	0.9	41 %	287.2	MERRA2
r32i1p1f1	166	0.9	41 %	287.3	MERRA2
r33i1p1f1	166	0.9	40 %	287.2	MERRA2
r34i1p1f1	166	0.9	42 %	287.2	MERRA2
r35i1p1f1	166	0.9	39 %	287.3	MERRA2
r36i1p1f1	166	0.9	40 %	287.3	MERRA2
r37i1p1f1	166	0.9	41 %	287.3	MERRA2
r38i1p1f1	166	0.9	41 %	287.1	ERA40
r39i1p1f1	166	0.9	40 %	287.2	MERRA2
r40i1p1f1	166	0.9	41 %	287.3	MERRA2
r41i1p1f1	166	0.9	40 %	287.2	MERRA2
r42i1p1f1	166	0.9	43 %	287.2	MERRA2
r43i1p1f1	166	0.9	41 %	287.2	MERRA2
r44i1p1f1	166	0.9	40 %	287.2	MERRA2
r45i1p1f1	166	0.8	42 %	287.3	MERRA2
r46i1p1f1	166	0.9	42 %	287.2	MERRA2
r47i1p1f1	166	0.9	40 %	287.2	MERRA2
r48i1p1f1	166	0.9	41 %	287.2	MERRA2
r49i1p1f1	166	0.9	40 %	287.2	MERRA2
r50i1p1f1	166	0.9	40 %	287.2	MERRA2

Reference	Δd (km)	RMSD (K)	IF (%)	T (K)	Closest ref.
ERA5	166	0.1	78 %	287.2	ERA40
ERA40	166	0.1	94 %	287.6	ERA-Interim
MERRA2	166	0.2	85 %	287.4	JRA3Q
MERRA	166	0.3	75 %	287.7	MERRA2
JRA55	166	0.1	91 %	287.6	ERA40
NCAR-NCEP	166	0.3	72 %	287.05	ERA-Interim
ERA-Interim	166	0.1	90 %	287.4	ERA40
20CR	166	0.3	77 %	287.7	JRA55
BerkeleyEarth	166	0.3	63 %	287.55	JRA3Q
JRA3Q	166	0.1	74 %	287.5	ERA5

Table S9: Acknowledgements of model data used in the study. All CMIP data are used from the ETH Zurich archive maintained by the Climate Physics group.

Collection	Acknowledgements
CMIP2	We thank all groups that contributed to CMIP2 for providing their data.
CMIP3	acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy.
CMIP5	We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.
CMIP6	We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. Data downloaded by Urs Beyerle and provided by ETH Zurich CMIP6 next-generation archive [3].
HighResMIP	Same as CMIP6
SMILE	Same as CMIP6; Data provided by the German Climate Computation Center (DKRZ) via Freva.

Table S10: Acknowledgements of observation-based data used in the study.

Dataset	Acknowledgment
20CR (v3)	Provider & Support for the Twentieth Century Reanalysis Project dataset is provided by the U.S. Department of Energy, Office of Science Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and Office of Biological and Environmental Research (BER), by the National Oceanic and Atmospheric Administration Climate Program Office, and by the National Oceanic and Atmospheric Administration Climate Program Office, and by the NOAA Physical Sciences Laboratory
BerkeleyEarth	Data CC BY-NC Berkeley Earth (www.berkeleyearth.org)
ERA40	We thank the data providers for making ERA40 available. ERA40 data were kindly provided by Paula Romanovska from the Potsdam Institute for Climate Impact Research.
ERA-Interim	We thank the data providers for making ERA-Interim available.
ERA5	We thank the Copernicus Climate Change Service information for providing the ERA5 monthly mean data on single levels from 1940 to present (https://doi.org/10.24381/cds.f17050d7)
JRA55	We thank the Japan Meteorological Agency for providing JRA55
JRA3Q	We thank the Japan Meteorological Agency for providing JRA3Q
MERRA	We thank the data providers for making MERRA available. Data provided by the German Climate Computation Center (DKRZ) via Freva.
MERRA2	We thank the Global Modeling and Assimilation Office (GMAO) and Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC) for providing MERRA2.
NCAR-NCEP	We thank the National Centers for Environmental Prediction (NCEP), the National Center for Atmospheric Research (NCAR), and the Department of Energy (DOE) for providing NCAR-NECP.

Supplementary References

[1] Martin B. Andrews et al. "Historical Simulations With HadGEM3-GC3.1 for CMIP6". In: *Journal of Advances in Modeling Earth Systems* 12.6 (2020), e2019MS001995. ISSN: 1942-2466. DOI: 10.1029/2019MS001995.

[2] Lukas Brunner. *supplement for: Three decades of simulating global temperature patterns with coupled global climate models*. 2026. DOI: https://doi.org/10.5281/zenodo.18853482.

[3] Lukas Brunner et al. *The ETH Zurich CMIP6 next Generation Archive : Technical Documentation*. ETH Zurich, Institute for Atmospheric and Climate Science, 2020. DOI: 10.5281/zenodo.3734128.

[4] A. Cherchi et al. "Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model". In: *Journal of Advances in Modeling Earth Systems* 11.1 (2019), pp. 185–209. ISSN: 19422466. DOI: 10.1029/2018MS001369.

[5] G. Danabasoglu et al. "The Community Earth System Model Version 2 (CESM2)". In: *Journal of Advances in Modeling Earth Systems* 12.2 (Feb. 6, 2020), pp. 0–3. ISSN: 1942-2466. DOI: 10.1029/2019MS001916.

[6] Ralf Döscher et al. "The EC-Earth3 Earth System Model for the Coupled Model Intercomparison Project 6". In: *Geoscientific Model Development* 15.7 (Apr. 8, 2022), pp. 2973–3020. ISSN: 1991-9603. DOI: 10.5194/gmd-15-2973-2022.

[7] J.-L. Dufresne et al. "Climate Change Projections Using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5". In: *Climate Dynamics* 40.9–10 (May 24, 2013), pp. 2123–2165. ISSN: 0930-7575. DOI: 10.1007/s00382-012-1636-1.

[8] A. Gettelman et al. "The Whole Atmosphere Community Climate Model Version 6 (WACCM6)". In: *Journal of Geophysical Research: Atmospheres* 124.23 (2019), pp. 12380–12403. ISSN: 21698996. DOI: 10.1029/2019JD030943.

[9] Alexander Sen Gupta et al. "Climate Drift in the CMIP3 Models". In: *Journal of Climate* 25.13 (July 1, 2012), pp. 4621–4640. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/JCLI-D-11-00312.1.

[10] Rein Haarsma et al. "HighResMIP Versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR – Description, Model Computational Performance and Basic Validation". In: *Geoscientific Model Development* 13.8 (Aug. 6, 2020), pp. 3507–3527. ISSN: 1991-959X. DOI: 10.5194/gmd-13-3507-2020.

[11] Hans Hersbach et al. "The ERA5 Global Reanalysis". In: *Quarterly Journal of the Royal Meteorological Society* 146.730 (2020), pp. 1999–2049. ISSN: 1477870X. DOI: 10.1002/qj.3803.

[12] François Massonnet et al. "Using Climate Models to Estimate the Quality of Global Observational Data Sets". In: *Science* 354.6311 (Oct. 28, 2016), pp. 452–455. DOI: 10.1126/science.aaf6369.

[13] Thorsten Mauritsen et al. "Developments in the MPI-M Earth System Model Version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2". In: *Journal of Advances in Modeling Earth Systems* 11.4 (2019), pp. 998–1038. ISSN: 19422466. DOI: 10.1029/2018MS001400.

[14] Gerald A. Meehl et al. "Climate Change Projections in CESM1(CAM5) Compared to CCSM4". In: *Journal of Climate* 26.17 (2013), pp. 6287–6308. ISSN: 08948755. DOI: 10.1175/JCLI-D-12-00572.1.

[15] Gerald A. Meehl et al. "The Coupled Model Intercomparison Project (CMIP)". In: *Bulletin of the American Meteorological Society* 81 (2000), pp. 313–318. ISSN: 00030007. DOI: 10.1175/1520-0477(2001)082<0695:msefiw>2.3.co;2.

[16] A. Molod et al. "Development of the GEOS-5 Atmospheric General Circulation Model: Evolution from MERRA to MERRA2". In: *Geoscientific Model Development* 8.5 (2015), pp. 1339–1356. ISSN: 19919603. DOI: 10.5194/gmd-8-1339-2015.

[17] Kazutoshi Onogi et al. "The JRA-25 Reanalysis". In: *Journal of the Meteorological Society of Japan. Ser. II* 85.3 (2007), pp. 369–432. ISSN: 0026-1165. DOI: 10.2151/jmsj.85.369.

[18] Christopher D. Roberts et al. "Climate Model Configurations of the ECMWF Integrated Forecasting System (ECMWF-IFS Cycle 43r1) for HighResMIP". In: *Geoscientific Model Development* 11.9 (Sept. 11, 2018), pp. 3681–3712. ISSN: 1991-959X. DOI: 10.5194/gmd-11-3681-2018.

[19] Malcolm J. Roberts et al. "Description of the Resolution Hierarchy of the Global Coupled HadGEM3-GC3.1 Model as Used in CMIP6 HighResMIP Experiments". In: *Geoscientific Model Development* 12.12 (Dec. 3, 2019), pp. 4999–5028. ISSN: 1991-959X. DOI: 10.5194/gmd-12-4999-2019.

[20] Benjamin D. Santer et al. "Identification of anthropogenic climate change using a second-generation reanalysis". In: *Journal of Geophysical Research: Atmospheres* 109.D21 (2004). DOI: https://doi.org/10.1029/2004JD005075. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2004JD005075. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JD005075.

[21] Benjamin D. Santer et al. "Using Climate Model Simulations to Constrain Observations". In: *Journal of Climate* 34.15 (Aug. 1, 2021), pp. 6281–6301. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/JCLI-D-20-0768.1.

[22] Keah C. Schuenemann and John J. Cassano. "Changes in Synoptic Weather Patterns and Greenland Precipitation in the 20th and 21st Centuries: 1. Evaluation of Late 20th Century Simulations from IPCC Models". In: *Journal of Geophysical Research: Atmospheres* 114.D20 (2009). ISSN: 2156-2202. DOI: 10.1029/2009JD011705.

[23] Enrico Scoccimarro et al. "Effects of Tropical Cyclones on Ocean Heat Transport in a High-Resolution Coupled General Circulation Model". In: *Journal of Climate* 24.16 (Aug. 15, 2011), pp. 4368–4384. ISSN: 0894-8755, 1520-0442. DOI: 10.1175/2011JCLI4104.1.

[24] Roland Séférian et al. "Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate". In: *Journal of Advances in Modeling Earth Systems* 11.12 (2019), pp. 4182–4227. ISSN: 1942-2466. DOI: 10.1029/2019MS001791.

[25] Øyvind Seland et al. "Overview of the Norwegian Earth System Model (NorESM2) and Key Climate Response of CMIP6 DECK, Historical, and Scenario Simulations". In: *Geoscientific Model Development* 13.12 (Dec. 4, 2020), pp. 6165–6200. ISSN: 1991-959X. DOI: 10.5194/gmd-13-6165-2020.

[26] Laura C. Slivinski et al. "Towards a More Reliable Historical Reanalysis: Improvements for Version 3 of the Twentieth Century Reanalysis System". In: *Quarterly Journal of the Royal Meteorological Society* 145.724 (2019), pp. 2876–2908. ISSN: 1477870X. DOI: 10.1002/qj.3598.

[27] Tongwen Wu et al. "The Beijing Climate Center Climate System Model (BCC-CSM): The Main Progress from CMIP5 to CMIP6". In: *Geoscientific Model Development* 12.4 (Apr. 24, 2019), pp. 1573–1600. ISSN: 1991-959X. DOI: 10.5194/gmd-12-1573-2019.