

Supplementary Information for: GCMagicc v1: a fast generative emulator for multivariate climate-impact ensembles

Nicolai Meinshausen^{1,3}, Malte Meinshausen^{1,2}, Jared Lewis², Zebedee Nicholls^{1,2,4}, Sarah Schöngart^{4,5}, Alistair Self², Xinwei Shen³, Karla Spiller², and Elisabeth Vogel¹

¹ School of Geography, Earth and Atmospheric Sciences, The University of Melbourne, Swanston Street, Parkville, Victoria 3010, Australia

² Climate Resource, Victoria Rd 105, Fitzroy, Victoria 3065, Australia

³ Department of Statistics, University of Washington, Padelford Hall, Seattle, Washington 98195, USA

⁴ Energy, Climate, and Environment Program, International Institute for Applied Systems Analysis, Laxenburg, Austria

⁵ Institute for Atmospheric and Climate Science, ETH Zürich, Zurich, Switzerland

Correspondence: Malte Meinshausen (malte.meinshausen@unimelb.edu.au, malte.meinshausen@climate-resource.com)

Abstract. This Supplementary Information provides the supporting methods, diagnostics, and scenario context for the GCMagicc model presented in the main manuscript. It documents the GCMagicc software release v1.0.1; the shorter identifier v1 is used in the title and in the running text of both documents.

Contents

5	S1 CMIP6 training dataset	4
	S2 Predictors: MAGICC setup and AR6/CMIP7 ScenarioMIP workflows	7
	S2.1 MAGICC version, provenance, and calibration	7
	S2.2 Energy-balance core and predictor time series	8
	S2.3 Predictor variable set exported from MAGICC	8
10	S2.4 (i) Smoothed global predictors.	9
	S2.5 (ii) Forcing components.	9
	S2.6 (iii) Monthly seasonality features.	10
	S2.7 (iv) Ensemble sampling and subsampling (600 → 20/100)	10
	S2.8 (v) ERA5 alignment and replication modes	12
15	S2.9 (vi) Two-pass global-temperature correction	12
	S2.10 Limitations and assumptions	12
	S3 Perfect-model GCMagicc-PM and extrapolation GCMagicc-XS experiments	13
	S3.1 Perfect-model uncertainty experiment (GCMagicc-PM)	13
	S3.2 Held-out high-end scenario extrapolation (GCMagicc-XS)	13

20	S4 Comparison of GCMagicc versus CMIP6 ESMs	24
	S4.1 Fidelity metric diagnostics to compare GCMagicc vs CMIP6	24
	S4.1.1 GlobalTimeseries	24
	S4.1.2 BiasMaps	25
	S4.1.3 SeasonalCycle	25
25	S4.1.4 ZonalMeans	25
	S4.1.5 Variability	25
	S4.1.6 Histograms	25
	S4.1.7 HumidityCoupling	25
	S4.1.8 IndicatorFreq	26
30	S4.1.9 ENSOTelecon	26
	S4.1.10 ENSOCharacteristics	26
	S4.2 Validation: Energy distances and energy distance scores	26
	S4.3 Extended data tables and alternate draws of Figure 3	28
	S5 Emergent constraints from leave-one-family-out cross-validation	44
35	S5.1 Reduced emulator for global-mean warming	44
	S5.2 Leave-one-family-out procedure	44
	S5.3 Warming metric	45
	S5.4 Cross-validation scatter plots	45
	S5.5 Width implied by the GCMagicc vs CMIP6 cross-validation residuals	48
40	S5.6 Summary	49
	S6 Observationally and IPCC AR6 aligned projections	49
	S6.1 Data table for Figure 4	49
	S6.2 New scenarios: Estimating NDC and LT-LEDS future emissions	49
	S6.3 Extended data figure for observational matches of GCMagicc and CMIP6 models	50
45	S7 Suitability of existing climate emulators and ESM workflows for multivariate drought attribution	56
	S7.1 CMIP6 + statistical bias adjustment.	56
	S7.2 MESMER family.	57
	S7.3 MESMER-MAGICC (Beusch et al., 2022).	57
	S7.4 STITCHES (Tebaldi et al., 2022).	57
50	S7.5 ClimaX (Nguyen et al., 2023).	58
	S7.6 Aurora (Bodnar et al., 2025).	58
	S7.7 cBottle (Brenowitz et al., 2025).	58

	S7.8 NeuralGCM (Kochkov et al., 2024).	58
	S7.9 ACE2 (Watt-Meyer et al., 2025).	59
55	S7.10SamudrACE (Duncan et al., 2026).	59
	S7.11LUCIE-3D (Guan et al., 2025b).	59
	S7.12DiffESM (Bassetti et al., 2024).	59
	S7.13Score-based ESM emulator (Bouabid et al., 2026).	60
	S7.14ArchesClimate-SSP (Clyne et al., 2026).	60
60	S7.15GCMagicc (this work).	61
	S8 Drought attribution and drought projections	61
	S8.1 Penman–Monteith SPEI-48	61
	S8.2 Monthly below-threshold probability	62
	S8.3 SMILE inventory and repeated natural-only tails	62
65	S8.4 Predictor sensitivity and SMILE extension	64
	S8.5 Assumptions and limitations	64

S1 CMIP6 training dataset

The model-scenario coverage and ensemble-member counts used for training are listed in Table S1.

Table S1. Number of ensemble runs available in the CMIP6 archive for each model-scenario combination. A dash (–) indicates no runs are available for that combination. Entries in parentheses are available but held out from training, namely all ScenarioMIP runs of UKESM1-0-LL and of the MIROC family, for which only the historical, single-forcing and idealised runs enter the training set. The table also does not list *ssp245*, which is held out from training for every model as the unseen-scenario test case.

	<i>1pctCO2</i>	<i>abrupt-2x</i>	<i>abrupt-4x</i>	<i>hist-GHG</i>	<i>hist-aer</i>	<i>hist-nat</i>	<i>historical</i>	<i>ssp119</i>	<i>ssp126</i>	<i>ssp370</i>	<i>ssp434</i>	<i>ssp460</i>	<i>ssp534-over</i>	<i>ssp585</i>
ACCESS-CM2	1	–	1	3	3	3	10	–	9	5	–	–	1	9
ACCESS-ESM1-5	1	–	2	3	3	3	37	–	30	35	–	–	1	30
CanESM5	6	1	2	50	30	50	50	50	50	50	5	5	5	50
CanESM5-1	2	–	2	–	–	–	50	1	1	1	–	1	–	16
CanESM5-CanOE	1	–	–	–	–	–	3	–	3	3	–	–	–	3
CESM2	–	–	–	3	2	3	1	–	5	5	–	–	–	5
CESM2-WACCM	–	–	–	–	–	–	1	–	1	–	–	–	1	4
EC-Earth3	–	–	2	–	–	–	17	2	9	8	–	–	–	7
EC-Earth3-AerChem	1	–	1	–	–	–	2	–	–	2	–	–	–	–
EC-Earth3-CC	1	–	1	–	–	–	9	–	–	–	–	–	–	1
EC-Earth3-Veg	1	–	1	–	–	–	9	2	6	5	–	–	–	7
EC-Earth3-Veg-LR	1	–	–	–	–	–	1	3	3	3	–	–	–	3
FGOALS-g3	2	–	1	–	–	–	6	1	3	5	1	1	1	4
GFDL-CM4	1	–	1	–	–	2	1	–	–	–	–	–	–	1
GFDL-ESM4	1	–	1	–	–	–	1	1	1	1	–	–	–	1
GISS-E2-1-G	5	4	5	10	10	20	47	6	14	27	7	8	11	15
GISS-E2-1-G-CC	–	–	–	–	–	–	1	–	1	–	–	–	–	–
GISS-E2-1-H	2	3	3	–	–	–	25	2	10	6	2	2	6	10
GISS-E2-2-G	1	2	2	–	–	–	11	–	–	4	–	–	1	5
GISS-E2-2-H	1	1	1	–	–	–	3	–	–	–	–	–	–	–
HadGEM3-GC31-LL	4	1	1	5	5	10	5	–	1	–	–	–	–	4
HadGEM3-GC31-MM	1	–	1	–	–	–	4	–	1	–	–	–	–	4
INM-CM4-8	1	–	1	–	–	–	1	–	1	1	–	–	–	1
INM-CM5-0	1	–	1	–	–	–	10	–	1	5	–	–	–	1
IPSL-CM6A-LR	1	1	11	6	6	5	31	6	5	11	2	7	–	6
IPSL-CM6A-LR-INCA	–	–	1	–	–	–	1	–	–	–	–	–	–	–
IPSL-CM6A-MR1	1	–	1	–	–	–	1	–	–	–	–	–	–	–
MIROC-ES2H	–	–	3	–	–	–	3	–	–	–	–	–	–	–
MIROC-ES2L	1	–	1	–	–	–	31	(10)	(7)	(10)	–	–	–	(9)
MIROC6	1	1	1	10	10	50	50	(50)	(50)	(50)	(1)	(1)	(1)	(50)
SAM0-UNICON	1	–	1	–	–	–	1	–	–	–	–	–	–	–
UKESM1-0-LL	4	–	1	–	–	–	19	(5)	(16)	(16)	(5)	–	(5)	(5)
ERA5	1 run (historical reanalysis)													

Table S2: CMIP6 data-citation references for the model output used in this study, at the model/MIP granularity defined by the CMIP6 Data Citation Guidelines. Each DOI resolves to the WDCC/DKRZ citation landing page for the corresponding CMIP6.*activity_id.institution_id.source_id* collection. Data creators are given in the reference list of this Supplement as the first creator et al.; the complete creator lists are on the DOI landing pages.

Source ID	Activity	DOI
ACCESS-CM2	CMIP	10.22033/ESGF/CMIP6.2281
ACCESS-CM2	DAMIP	10.22033/ESGF/CMIP6.14361
ACCESS-CM2	ScenarioMIP	10.22033/ESGF/CMIP6.2285
ACCESS-ESM1-5	CMIP	10.22033/ESGF/CMIP6.2288
ACCESS-ESM1-5	DAMIP	10.22033/ESGF/CMIP6.14362
ACCESS-ESM1-5	ScenarioMIP	10.22033/ESGF/CMIP6.2291
CanESM5	CFMIP	10.22033/ESGF/CMIP6.1302
CanESM5	CMIP	10.22033/ESGF/CMIP6.1303
CanESM5	DAMIP	10.22033/ESGF/CMIP6.1305
CanESM5	ScenarioMIP	10.22033/ESGF/CMIP6.1317
CanESM5-1	CMIP	10.22033/ESGF/CMIP6.17184
CanESM5-1	ScenarioMIP	10.22033/ESGF/CMIP6.17199
CanESM5-CanOE	CMIP	10.22033/ESGF/CMIP6.10205
CanESM5-CanOE	ScenarioMIP	10.22033/ESGF/CMIP6.10207
CESM2	CMIP	10.22033/ESGF/CMIP6.2185
CESM2	DAMIP	10.22033/ESGF/CMIP6.2187
CESM2	ScenarioMIP	10.22033/ESGF/CMIP6.2201
CESM2-WACCM	CMIP	10.22033/ESGF/CMIP6.10024
CESM2-WACCM	ScenarioMIP	10.22033/ESGF/CMIP6.10026
EC-Earth3	CMIP	10.22033/ESGF/CMIP6.181
EC-Earth3	ScenarioMIP	10.22033/ESGF/CMIP6.251
EC-Earth3-AerChem	CMIP	10.22033/ESGF/CMIP6.639
EC-Earth3-AerChem	ScenarioMIP	10.22033/ESGF/CMIP6.724
EC-Earth3-CC	CMIP	10.22033/ESGF/CMIP6.640
EC-Earth3-CC	ScenarioMIP	10.22033/ESGF/CMIP6.15327
EC-Earth3-Veg	CMIP	10.22033/ESGF/CMIP6.642
EC-Earth3-Veg	ScenarioMIP	10.22033/ESGF/CMIP6.727
EC-Earth3-Veg-LR	CMIP	10.22033/ESGF/CMIP6.643
EC-Earth3-Veg-LR	ScenarioMIP	10.22033/ESGF/CMIP6.728
FGOALS-g3	CMIP	10.22033/ESGF/CMIP6.1783
FGOALS-g3	ScenarioMIP	10.22033/ESGF/CMIP6.2056
GFDL-CM4	CMIP	10.22033/ESGF/CMIP6.1402
GFDL-CM4	DAMIP	10.22033/ESGF/CMIP6.11383
GFDL-CM4	ScenarioMIP	10.22033/ESGF/CMIP6.9242
GFDL-ESM4	CMIP	10.22033/ESGF/CMIP6.1407
GFDL-ESM4	ScenarioMIP	10.22033/ESGF/CMIP6.1414
GISS-E2-1-G	CFMIP	10.22033/ESGF/CMIP6.2061
GISS-E2-1-G	CMIP	10.22033/ESGF/CMIP6.1400
GISS-E2-1-G	DAMIP	10.22033/ESGF/CMIP6.2062
GISS-E2-1-G	ScenarioMIP	10.22033/ESGF/CMIP6.2074
GISS-E2-1-G-CC	CMIP	10.22033/ESGF/CMIP6.11657
GISS-E2-1-G-CC	ScenarioMIP	10.22033/ESGF/CMIP6.17581
GISS-E2-1-H	CFMIP	10.22033/ESGF/CMIP6.13941
GISS-E2-1-H	CMIP	10.22033/ESGF/CMIP6.1421
GISS-E2-1-H	ScenarioMIP	10.22033/ESGF/CMIP6.2080

Source ID	Activity	DOI
GISS-E2-2-G	CFMIP	10.22033/ESGF/CMIP6.11659
GISS-E2-2-G	CMIP	10.22033/ESGF/CMIP6.2081
GISS-E2-2-G	ScenarioMIP	10.22033/ESGF/CMIP6.11671
GISS-E2-2-H	CFMIP	10.22033/ESGF/CMIP6.17582
GISS-E2-2-H	CMIP	10.22033/ESGF/CMIP6.15861
HadGEM3-GC31-LL	CFMIP	10.22033/ESGF/CMIP6.435
HadGEM3-GC31-LL	CMIP	10.22033/ESGF/CMIP6.419
HadGEM3-GC31-LL	DAMIP	10.22033/ESGF/CMIP6.471
HadGEM3-GC31-LL	ScenarioMIP	10.22033/ESGF/CMIP6.10845
HadGEM3-GC31-MM	CMIP	10.22033/ESGF/CMIP6.420
HadGEM3-GC31-MM	ScenarioMIP	10.22033/ESGF/CMIP6.10846
INM-CM4-8	CMIP	10.22033/ESGF/CMIP6.1422
INM-CM4-8	ScenarioMIP	10.22033/ESGF/CMIP6.12321
INM-CM5-0	CMIP	10.22033/ESGF/CMIP6.1423
INM-CM5-0	ScenarioMIP	10.22033/ESGF/CMIP6.12322
IPSL-CM6A-LR	CFMIP	10.22033/ESGF/CMIP6.1522
IPSL-CM6A-LR	CMIP	10.22033/ESGF/CMIP6.1534
IPSL-CM6A-LR	DAMIP	10.22033/ESGF/CMIP6.13801
IPSL-CM6A-LR	ScenarioMIP	10.22033/ESGF/CMIP6.1532
IPSL-CM6A-LR-INCA	CMIP	10.22033/ESGF/CMIP6.13582
IPSL-CM6A-MR1	CMIP	10.22033/ESGF/CMIP6.15922
MIROC-ES2H	CMIP	10.22033/ESGF/CMIP6.901
MIROC-ES2L	CMIP	10.22033/ESGF/CMIP6.902
MIROC-ES2L	ScenarioMIP	10.22033/ESGF/CMIP6.936
MIROC6	CFMIP	10.22033/ESGF/CMIP6.885
MIROC6	CMIP	10.22033/ESGF/CMIP6.881
MIROC6	DAMIP	10.22033/ESGF/CMIP6.894
MIROC6	ScenarioMIP	10.22033/ESGF/CMIP6.898
SAM0-UNICON	CMIP	10.22033/ESGF/CMIP6.1489
UKESM1-0-LL	CMIP	10.22033/ESGF/CMIP6.1569
UKESM1-0-LL	ScenarioMIP	10.22033/ESGF/CMIP6.1567

S2 Predictors: MAGICC setup and AR6/CMIP7 ScenarioMIP workflows

- 70 We use the simple climate model MAGICC to generate probabilistic, scenario-specific global predictors (temperature, energy imbalance/heat uptake, and forcing components) that condition the GCMagicc regional emulator, enabling IPCC-consistent uncertainty propagation at low computational cost (Meinshausen et al., 2011; MAGICC Development Team, 2026; Arias et al., 2021).

S2.1 MAGICC version, provenance, and calibration

- 75 We generate global predictors with MAGICC v7.5.3 driven by the AR6 probabilistic parameter set — the AR6 drawn set `magicc-ar6-0fd0f62-f023edb` — as configured by the AR6 workflow default (Meinshausen et al., 2011). MAGICC is an emulator of coupled climate–carbon-cycle behaviour, with a physically based energy-balance climate module calibrated against more complex models and observations (Meinshausen et al., 2011). We use the AR6-consistent probabilistic parameter

Table S3. Affine normalization coefficients per target variable.

Short name	Long name	a_v	b_v
<i>psl</i>	Sea Level Pressure	10^{-3}	-100
<i>tas</i>	Near-Surface Air Temperature	10^{-1}	-30
<i>pr</i>	Precipitation	$2 \cdot 10^4$	-5
<i>sfcWind</i>	Near-Surface Wind Speed	$1/3$	-2
<i>ts</i>	Surface Temperature	10^{-1}	-30
<i>tasmin</i>	Daily Min. Near-Surface Air Temp.	10^{-1}	-30
<i>tasmax</i>	Daily Max. Near-Surface Air Temp.	10^{-1}	-30
<i>rsds</i>	Surface Downwelling Shortwave Radiation	10^{-1}	-15
<i>hurs</i>	Near-Surface Relative Humidity	10^{-1}	-5
<i>huss</i>	Near-Surface Specific Humidity	10^2	-1

ensemble distributed with MAGICC ($M = 600$ members) to represent uncertainty in global transient response and related
80 parameters, consistent with the assessed warming constraints in IPCC AR6 Working Group I (Arias et al., 2021).

The complete training and inference sequence is shown in Fig. 2 in the main manuscript; the following subsections document each predictor-construction and inference stage.

S2.2 Energy-balance core and predictor time series

MAGICC’s climate module can be summarised as a low-order energy-balance model. MAGICC has a 50-layer upwelling-
85 diffusion-entrainment ocean parameterisation (Meinshausen et al., 2011). In our implementation, MAGICC produces per-scenario, per-ensemble-member global-mean surface air temperature anomaly $T_i(t)$ and a global imbalance or heat-uptake predictor $N_i(t)$, together with ERF components, which are then used as conditioning variables for GCMagicc. Previous studies sometimes also have used emulator chains coupling global simple climate models to spatial emulators, mostly however focused on global-mean temperature only as the coupling variable (Beusch et al., 2022), although an ocean heat uptake variant is
90 explored in their Supplementary.

In GCMagicc and GCMagicc-CE we train the model using smoothed global surface-air-temperature and energy-balance predictors alongside the ERF components. The surface-air-temperature predictor is aligned over the historical period (1940–2024) with ERA5. The 600 MAGICC predictor members are subsampled to either $n = 20$ or $n = 100$ to preserve the central and 5–95% range. A deterministic two-pass correction, described below, aligns the smoothed global GCMagicc *tas* output with
95 its predictor; it is not an iterative convergence loop.

S2.3 Predictor variable set exported from MAGICC

For each scenario s and probabilistic member $i \in \{1, \dots, M\}$, we construct a monthly predictor vector $\mathbf{x}_{s,i}(t, m)$ comprising:

S2.4 (i) Smoothed global predictors.

We export a smoothed global temperature predictor $T_{s,i}^{(W)}(t)$ and a smoothed imbalance or heat-uptake predictor $N_{s,i}^{(W)}(t)$, denoted in code as `tas_smoothed` and `rtmt_smoothed`. For training, the naturally variable CMIP6 monthly global-mean series are filtered with a 21-year-span LOWESS smoother. For inference, the already smooth MAGICC series are filtered with a centred 20-year (240-month) running mean after linear extension by 20 years at each endpoint, followed by trimming to the requested interval. Both operations remove interannual variability so that internal variability is supplied by the trained stochastic component rather than by predictor noise. Conditioning on both $T^{(W)}$ and $N^{(W)}$ provides a compact representation of disequilibrium states, where two members may share the same T but differ in N , which is useful when mapping scenarios with different forcing trajectories, such as overshoot pathways into regional patterns (Beusch et al., 2022; Tebaldi et al., 2025).

S2.5 (ii) Forcing components.

We export ERF components in $W \text{ m}^{-2}$ sufficient to distinguish forcing composition and disequilibrium pathways beyond global-mean temperature alone. The exported set comprises `CO2_ERF`, `GHG_ERF`, `aer_ERF`, `totalO3_ERF`, `stratO3_ERF`, `sol_ERF`, `volc_ERF`, `nat_ERF`, and `other_ERF`, where $F_{\text{nat}}(t) = F_{\text{sol}}(t) + F_{\text{volc}}(t)$. Here `other_ERF` denotes the residual between total ERF and the explicitly tracked components.

For the natural-only SSP2-4.5 experiment used in the drought application, anthropogenic ERF components are zeroed while the archived MAGICC natural-forcing output supplies `sol_ERF` and `volc_ERF`. Neither ERA5 splicing nor the two-pass global-temperature bias correction is applied to the natural-only predictors: MAGICC’s natural-only global `tas` and `rtmt` anomalies are by construction close to zero. For the aerosol-perturbation counterfactual, the aerosol ERF predictor is set to zero while all other forcings are held at their final historical values. As illustration, we provide an aerosol-pattern difference map Δ_{mean} (by default $tas_{\text{max}} - tas_{\text{min}}$). For each of 20 categorical model-index settings ($c = 0, \dots, 19$), the emulator is sampled 100 times under both full-forcing and aerosol-off forcing. The difference between the two 100-realization means is calculated for each setting, and Fig. S1 shows the equally weighted mean of the resulting 20 model-index-specific maps.

The intervention is a *fixed-forcing epoch* contrast rather than a transient run. The predictor matrix is taken from the normalised ERA5 historical cube (not an SSP scenario); all forcing predictors are frozen at their values in the final time step of that record ($x[:, 4:] = x[-1, 4:]$) and the analysed window is the last 120 months, so the seasonal cycle is averaged out while the forcing state is held fixed. The factual and counterfactual runs differ in the single `aer_ERF` column and in nothing else.

Two properties of this design bound how the resulting map should be read. First, `aer_ERF` is a single global forcing time series, not a spatial field, so the intervention perturbs one global number and the spatial structure of the response is entirely the emulator’s learned covariance between global aerosol forcing and regional diurnal temperature range over the training period (Stjern et al., 2020). Aerosol effective radiative forcing is defined over the anthropogenic perturbation relative to pre-industrial conditions, so natural mineral dust and sea salt remain in the background state of both the factual and the counterfactual run and are not removed by zeroing the channel. Any apparent signal over a dust-dominated region is consequently not a dust response

but a region whose diurnal temperature range co-varies with the global anthropogenic aerosol trajectory in the training data; regions with a large climatological diurnal range are the most exposed to this confounding, and the arid subtropical values in Fig. S1 should be read in that light rather than as an attributable local aerosol effect.

The signals over the Indo-Gangetic plain and eastern China are, by contrast, consistent with the observational literature, in which persistent aerosol loading over northern India has been associated with declining surface solar radiation and a falling diurnal temperature range (Mall et al., 2021), and anthropogenic aerosol forcing has been identified as a contributor to the observed diurnal-range decline over eastern China (Liu et al., 2016). That the Indian signal here exceeds the Chinese one is expected for a recent forcing epoch: aerosol optical depth over eastern China has declined markedly since about 2010 under air-quality policy while South Asian loading has continued to rise (Ratnam et al., 2021).

140 **S2.6 (iii) Monthly seasonality features.**

To preserve the seasonality of the regional emulator while using smoothed global predictors, we include month-of-year features: an integer month index $m \in \{1, \dots, 12\}$ and its circular embedding

$$\sin_m = \sin(2\pi m/12), \quad \cos_m = \cos(2\pi m/12). \quad (\text{S1})$$

The full conditioning vector combines long-term forced evolution ($T^{(W)}$, $N^{(W)}$, and ERFs) with a deterministic seasonal phase.

S2.7 (iv) Ensemble sampling and subsampling (600 $\rightarrow$ 20/100)

For each scenario, MAGICC yields $M = 600$ probabilistic predictor realisations $\{\mathbf{x}_{s,i}\}_{i=1}^M$. For the manuscript’s primary ensembles we subsample to $n \in \{20, 100\}$ members for computational tractability while preserving the median end-of-century warming. The selection is median-anchored and stratified. Writing $\Delta T_i = \overline{T}_i(2081\text{--}2100) - \overline{T}_i(1995\text{--}2014)$ for the end-of-century warming of member i , the procedure is:

1. order the M members by ΔT_i ;
2. retain the two central members, whose mean is the full-ensemble median for even M ;
3. draw $(n-2)/2$ further members from below and $(n-2)/2$ from above that central pair, at evenly spaced ranks within each half.

Because M and n are both even, the subsample median equals the full-ensemble median exactly rather than approximately. Even spacing in rank preserves the shape of the distribution as well as its centre. Preserving distributional characteristics rather than only the mean matters because the assessed ranges are themselves distributional statements (Arias et al., 2021), and because distributional fidelity is the relevant criterion in emulator evaluation for risk applications (Watson-Parris et al., 2022). Note though that this procedure only considers distributions in the marginal distribution of surface air temperature and not in the joint distribution of all predictors, so the subsample is not guaranteed to preserve the full multivariate distribution.

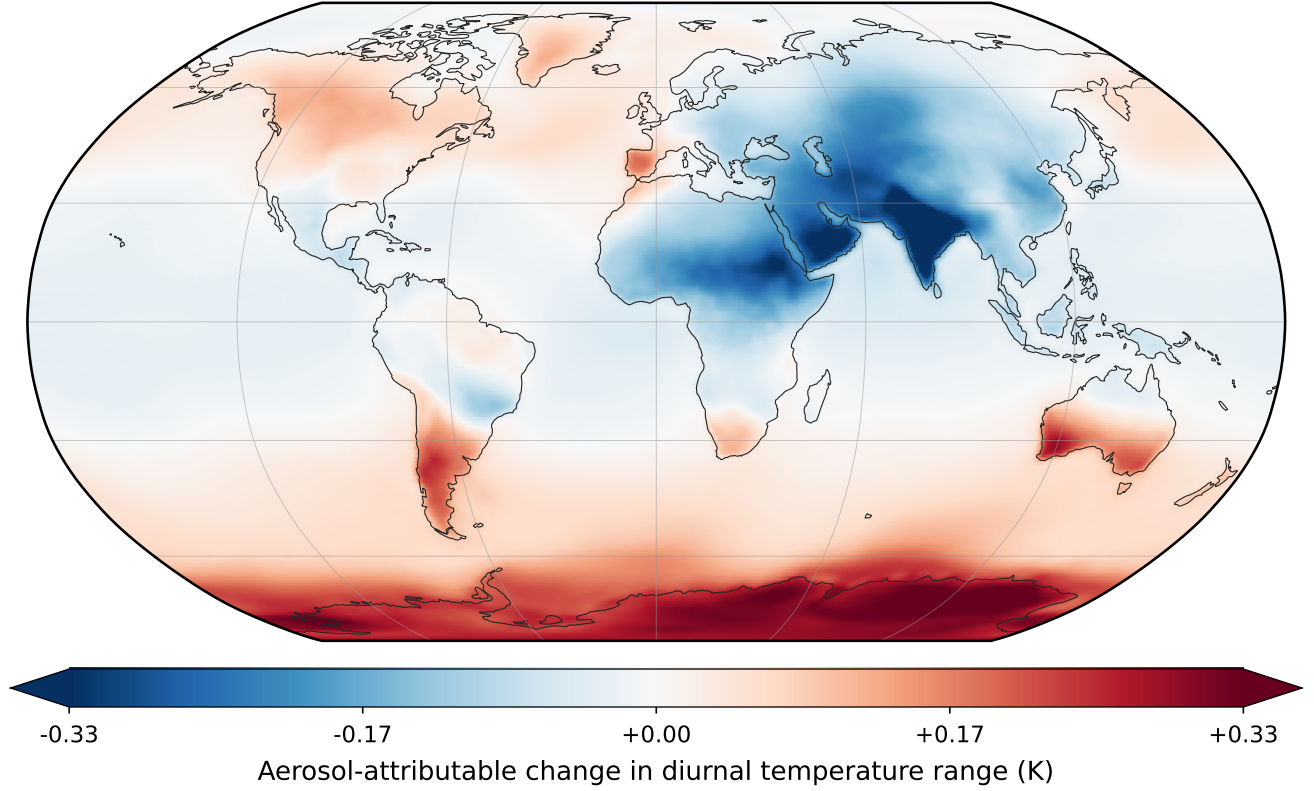


Figure S1. Aerosol-covariate pattern in the diurnal temperature range diagnosed by GCMagicc. Ensemble mean over the 20 Monte Carlo members ($mc = 0-19$, $nsim = 100$ each) of $\Delta_{\text{mean}} = \bar{y}_{\text{factual}} - \bar{y}_{\text{counterfactual}}$ for $tasmax - tasmin$ at $nside = 64$, where the counterfactual sets the aerosol effective-radiative-forcing predictor `aer_ERF` to zero and leaves all other predictors at their factual values. Blue indicates a reduced diurnal temperature range under historical aerosol forcing. The strongest suppression falls over the Indo-Gangetic plain (approximately -0.44 K at 78°E , 25°N), eastern China (-0.18 K) and the Saharan and Arabian dust belt (-0.16 K), consistent with a reduction in surface shortwave radiation damping daytime maxima more than nighttime minima. Area-weighted regional means are -0.110 ± 0.011 K over South and East Asia and -0.129 ± 0.024 K over the Sahara ($\pm$ one standard error across members); the absolute-magnitude response is 3.9 times larger over land than over ocean between 60°S and 70°N . Because `aer_ERF` is a single global forcing time series rather than a spatial field, the map is the emulator’s learned spatial response to a change in global aerosol forcing and need not coincide with the emission distribution. The global mean of the field is 1×10^{-5} K by construction: the diurnal temperature range is a difference of two temperature variables, so a spatially uniform cooling cancels and only the differential daytime–nighttime response remains. The positive localized positive values over the Southern Ocean, parts of the Sahara and Antarctica (including values of $+0.27 \pm 0.03$ K) illustrate the limitations of the approach, because GCMagicc can learn spurious covariate signals associated with the aerosol time series. The figure shows an average across per-member Δ_{mean} maps

S2.8 (v) ERA5 alignment and replication modes

To ensure observational anchoring of historical predictors, we provide an ERA5-aligned mode in which the historical portion of the global predictors is nudged or spliced to match ERA5-derived global-mean temperature anomalies (Copernicus Climate Change Service (C3S), 2019). The splice is a two-sided crossfade: ERA5 is faded in over the five years from 1945 and faded
165 out over the single year ending in 2023, with ERA5 used directly in between and the MAGICC series shifted to the ERA5 mean over the adjoining window outside it. Shift alignment is used for *tas_smoothed*, *rtmt_smoothed*. In the description below, let t_0 denote the last month of the fade-out window, i.e. the last month in which reanalysis information enters the predictors. For each probabilistic member i , we define an anomaly-splicing operator

$$T_i^{\text{ERA5}}(t) = \begin{cases} T_{\text{ERA5}}(t), & t \leq t_0, \\ T_{\text{ERA5}}(t_0) + [T_i(t) - T_i(t_0)], & t > t_0, \end{cases} \quad (\text{S2})$$

170 The same shift splice is applied to *rtmt_smoothed*. Its historical anchor is the global ERA5 net top-of-atmosphere radiation series after the preprocessing stored as *rtmt_smoothed*; the vetted source spans January 1940–December 2025. In the replication setting, the categorical *model_index* is set to the ERA5 category (index 0), allowing an ERA5-flavoured historical climate while retaining the multi-model training distribution for future responses (Copernicus Climate Change Service (C3S), 2019; GCMagicc Development Team, 2026).

175 S2.9 (vi) Two-pass global-temperature correction

Because the regional emulator is trained on gridded fields, its implied global-mean temperature may differ slightly from the conditioning predictor. The released workflow therefore performs exactly two emulator passes. In pass 1, monthly gridded *tas* is reduced to annual cosine-latitude-weighted global means and smoothed with a centred 21-year window. For each year y ,

$$\Delta_y = \overline{T}_{\text{GCMagicc},21}^{(1)}(y) - T_{\text{predictor}}(y). \quad (\text{S3})$$

180 In pass 2, the normalised temperature-predictor column is replaced by

$$x'_{\text{tas_smoothed}}(t) = x_{\text{tas_smoothed}}(t) - \Delta_{y(t)}/10, \quad (\text{S4})$$

and the emulator is run once more with the same ensemble seed. The divisor 10 reverses the predictor normalisation. No other predictor or output field is directly modified.

S2.10 Limitations and assumptions

185 First, the ERF decomposition is only as reliable as the MAGICC configuration and supplied forcing series; scenario-level forcing is a proxy for broad forcing evolution and does not reconstruct each ESM’s diagnosed effective forcing. Second, low-pass filtering deliberately removes predictor-scale variability, leaving internal variability to the learned stochastic component and potentially underrepresenting very-low-frequency modes (Tebaldi et al., 2025; Watson-Parris et al., 2022). Third, ERA5

alignment depends on reanalysis uncertainty and the chosen splice boundary. Finally, the two-pass correction enforces global-
190 mean temperature consistency only; precipitation–energy consistency and other global constraints remain emergent properties.
See also the discussion of limitations in the main manuscript.

S3 Perfect-model GCMagicc-PM and extrapolation GCMagicc-XS experiments

S3.1 Perfect-model uncertainty experiment (GCMagicc-PM)

To quantify the structural uncertainty of GCMagicc-PM (the variant without the MAGICC-derived smoothed global *tas* and
195 *rtmt* indicators), we repeat the GCMagicc-PM training in a perfect-model setup. For each CMIP6 model in turn, a fresh
GCMagicc-PM instance is trained on that model’s historical run alone, treating that CMIP6 model as ground truth. Each
retrained instance is then run forward under future forcing and its end-of-century warming is compared to the corresponding
CMIP6 model’s own end-of-century projection. The spread of these deviations across CMIP6 models defines the perfect-model
uncertainty spread. Combining this spread with the ERA5-informed central estimate from GCMagicc-PM yields ranges that
200 are consistent with, and no narrower than, the IPCC AR6 assessed range for the middle and upper end of the SSP scenarios
(see the small black vertical bars in panel b of Fig. 4).

S3.2 Held-out high-end scenario extrapolation (GCMagicc-XS)

The GCMagicc-XS variant (Methods, Table 2) uses the same reduced predictor set as GCMagicc-PM (no MAGICC-derived
smoothed global indicators) and is trained jointly across the CMIP6 models and ERA5. In the configuration reported here,
205 GCMagicc-XS is trained on *historical*, *hist-GHG*, *hist-aer*, *hist-nat*, *ssp119*, *ssp126*, *ssp370*, *ssp434*, *ssp460* and the idealised
lpctCO2 and *abrupt-2xCO2*; the high-end scenarios *ssp534-over* and *ssp585* and the idealised *abrupt-4xCO2* are held out from
training. Of these, *ssp585* and *abrupt-4xCO2* could be evaluated; *ssp534-over* is provided by too few CMIP6 models to appear
in the normalised evaluation set and is therefore part of the held-out design.

We focus on surface air temperature (Fig. S2) and near-surface specific humidity (Fig. S3) because these are the variables
210 that show the largest relative change under *ssp585* relative to the training-scenario range and therefore provide the most infor-
mative qualitative consistency check under high-end forcing; analogous per-model observed-versus-predicted panels are shown
for precipitation (Fig. S4), daily maximum near-surface air temperature (Fig. S5), daily minimum near-surface air tempera-
ture (Fig. S6), near-surface relative humidity (Fig. S7), sea-level pressure (Fig. S8), surface downwelling shortwave radiation
(Fig. S9), near-surface wind speed (Fig. S10) and surface temperature (Fig. S11). Each per-model panel shows the pooled in-
215 distribution training scenarios (historical and SSP families, in orange) and the held-out *ssp585* (in light blue) on the same pair
of axes. Each plotted point is one monthly global-mean value for one CMIP6 ensemble member (up to 10 per model–scenario)
paired with its GCMagicc-XS ensemble-mean prediction; the visible cloud width along the 1:1 line therefore reflects both the
seasonal cycle and the CMIP6 inter-member spread within that driving model. Skill is near uniform across driving models

with one exception: CESM2-WACCM is flagged as an RMSE outlier on seven of ten variables under held-out *ssp585*, the only
 220 model so flagged more than three times.

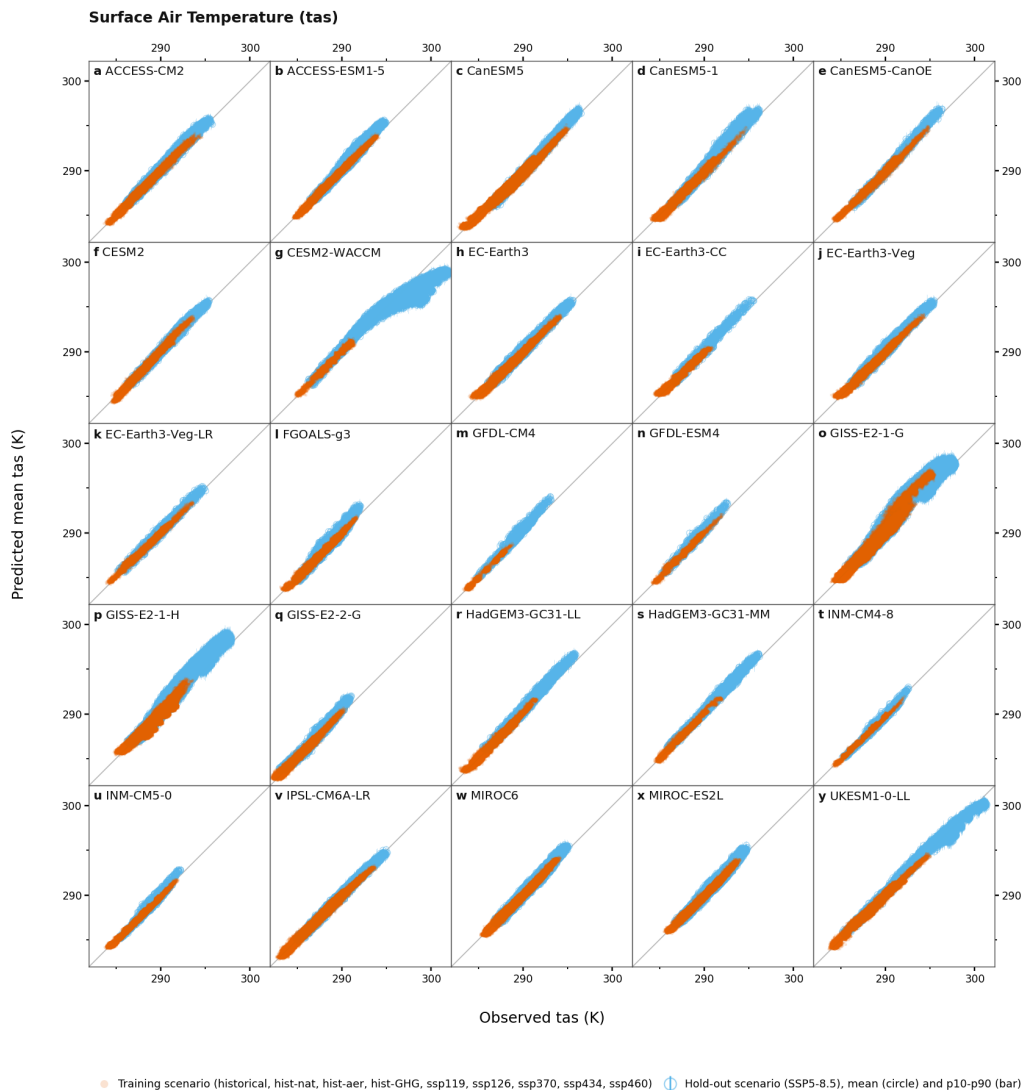


Figure S2. GCMagicc-XS per-model near-surface air temperature (*tas*): observed versus predicted. Twenty-five per-model panels, lettered a–y, one for each CMIP6 driving model for which *ssp585* runs are available (Table S1); the remaining seven models of the training set have no *ssp585* runs and therefore no hold-out case to display. Within each panel, the orange markers show the pooled in-distribution training scenarios (*historical*, *hist-nat*, *hist-aer*, *hist-GHG*, *ssp119*, *ssp126*, *ssp370*, *ssp434*, *ssp460*) and the light blue markers the held-out *SSP5-8.5*, given as the mean (circle) with a p10–p90 bar. The grey diagonal is the 1:1 line, and both axes are monthly global-mean *tas* in K. The panels demonstrate that GCMagicc-XS remains in the correct physical range and reproduces the seasonal cycle under a held-out high-end forcing. Each point pairs a single CMIP6 member with an ensemble-mean prediction, and the spread along the 1:1 line is due to both the seasonal cycle and high-low scenario variations.

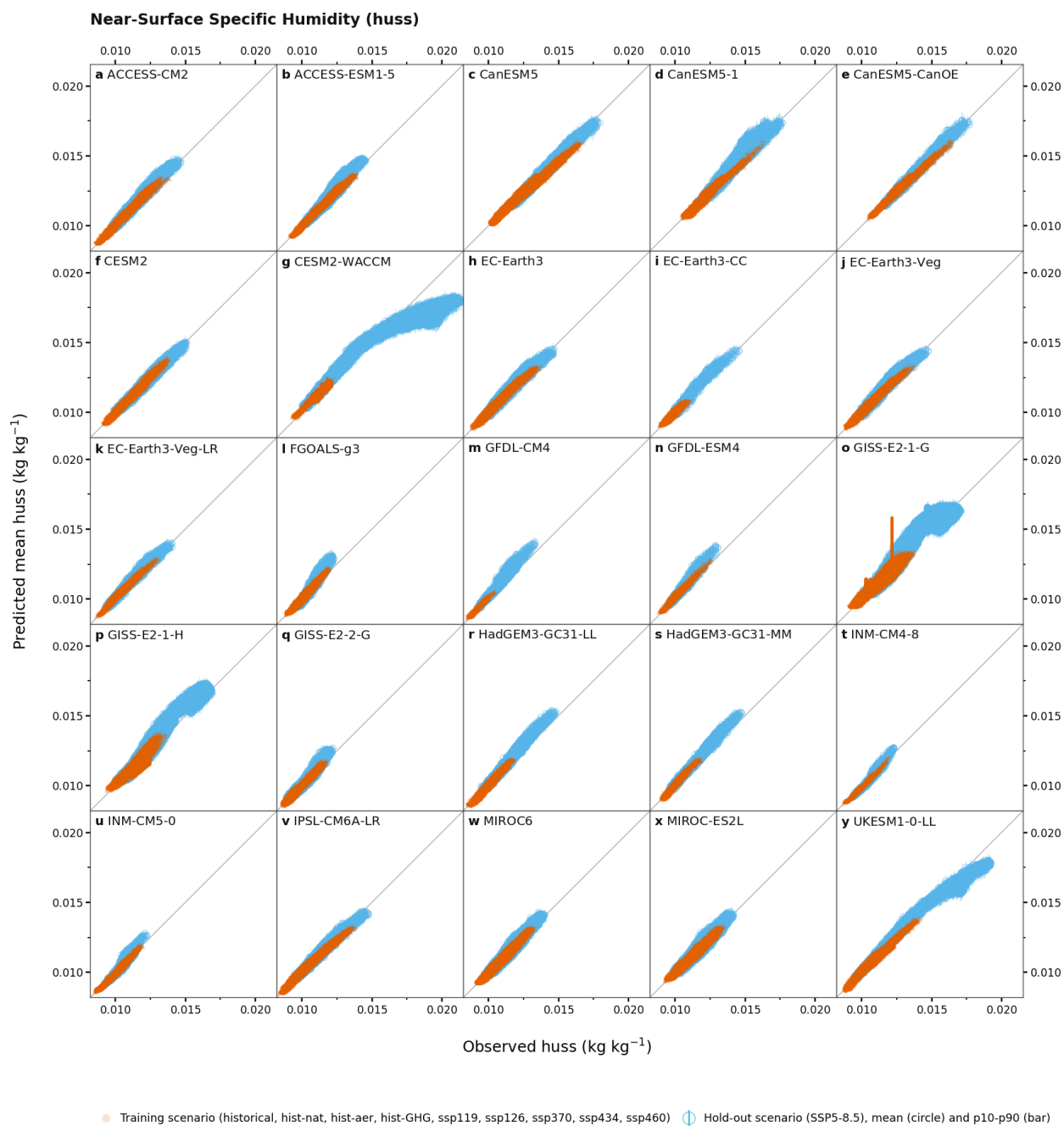


Figure S3. GCMagicc-XS per-model near-surface specific humidity (*huss*): observed versus predicted. As Fig. S2 but for near-surface specific humidity.

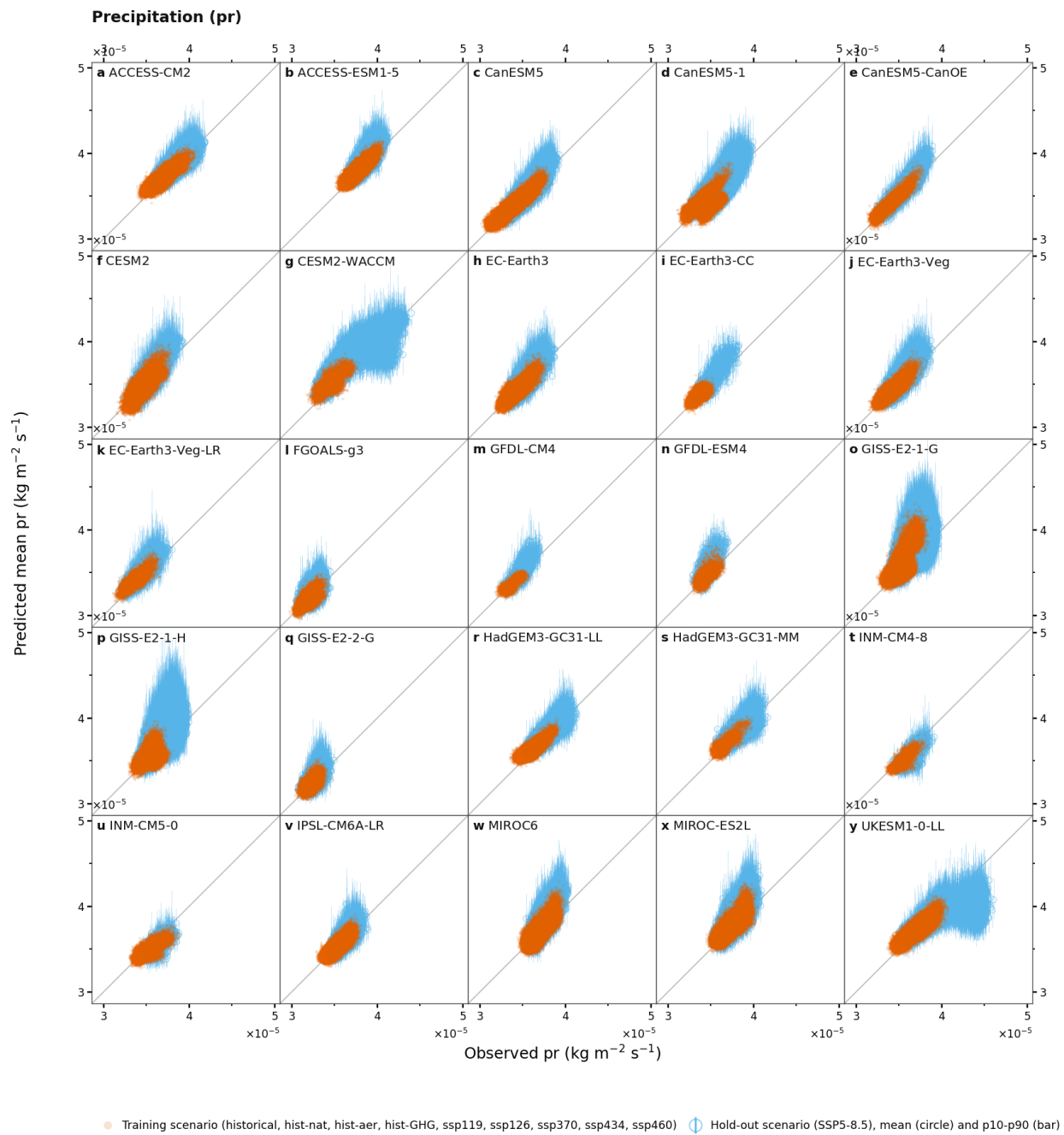


Figure S4. GCMagicc-XS per-model precipitation (*pr*): observed versus predicted. As Fig. S2 but for precipitation.

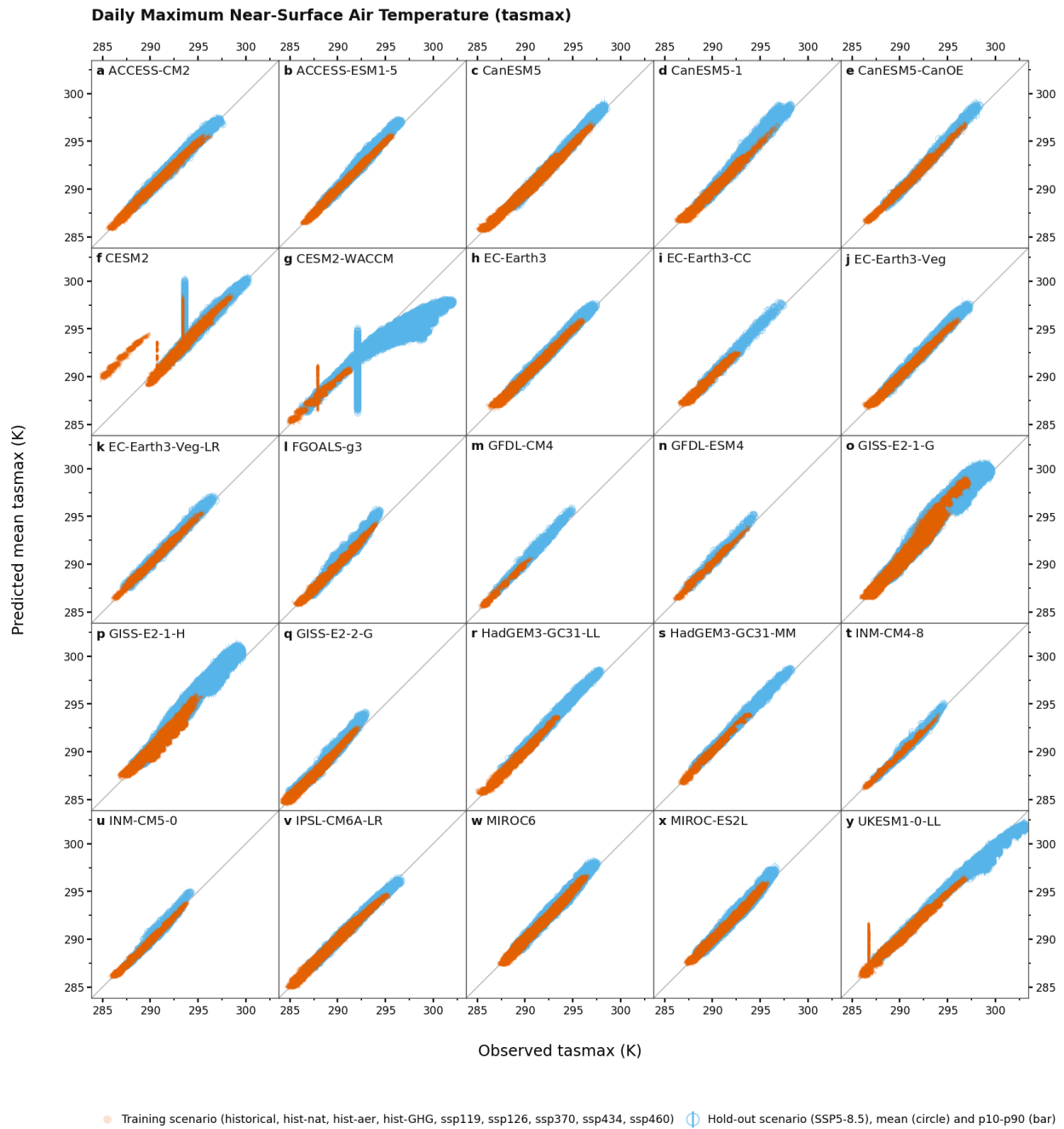


Figure S5. GCMagicc-XS per-model daily maximum near-surface air temperature (*tasmax*): observed versus predicted. As Fig. S2 but for daily maximum near-surface air temperature. Note that some data issues related to the CESM2, CESM2-WACCM and UKESM1-0-LL data and calibrations are visible in the *tasmax* and *tasmin* panels.

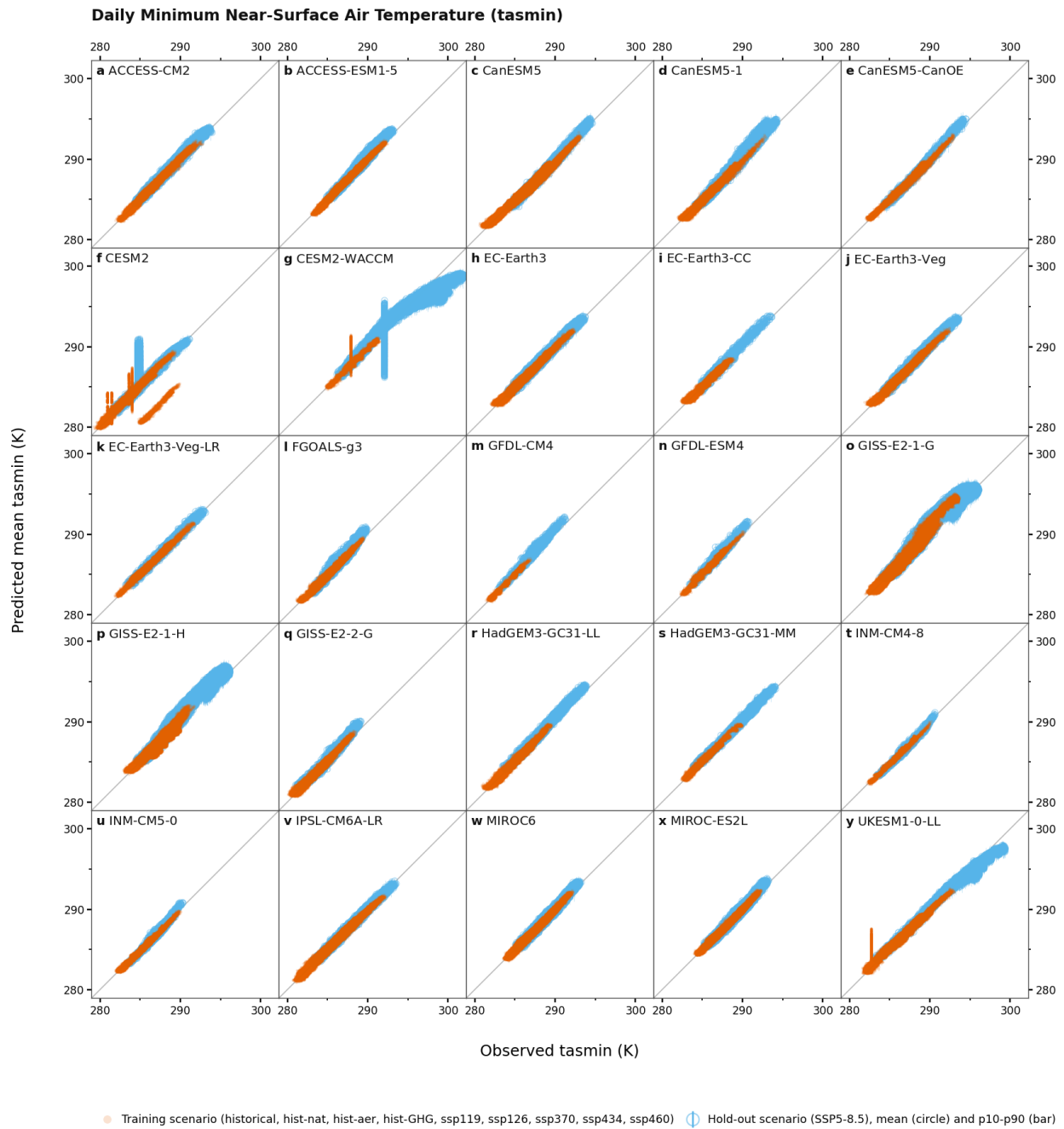


Figure S6. GCMagicc-XS per-model daily minimum near-surface air temperature (*tasmin*): observed versus predicted. As Fig. S2 but for daily minimum near-surface air temperature. Note that some data issues related to the CESM2, CESM2-WACCM and UKESM1-0-LL data and calibrations are visible in the *tasmax* and *tasmin* panels.

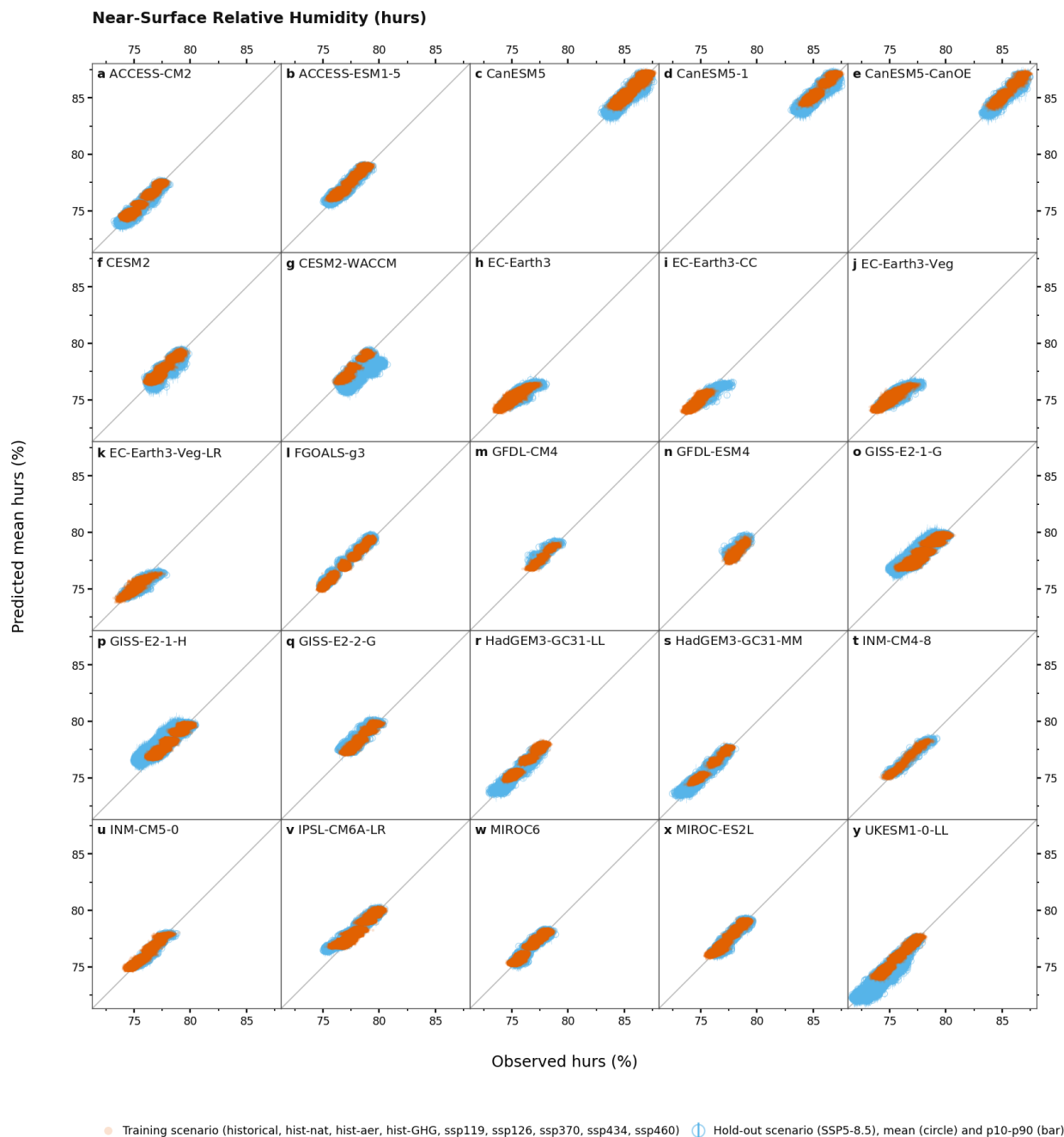


Figure S7. GCMagicc-XS per-model near-surface relative humidity (*hurs*): observed versus predicted. As Fig. S2 but for near-surface relative humidity.

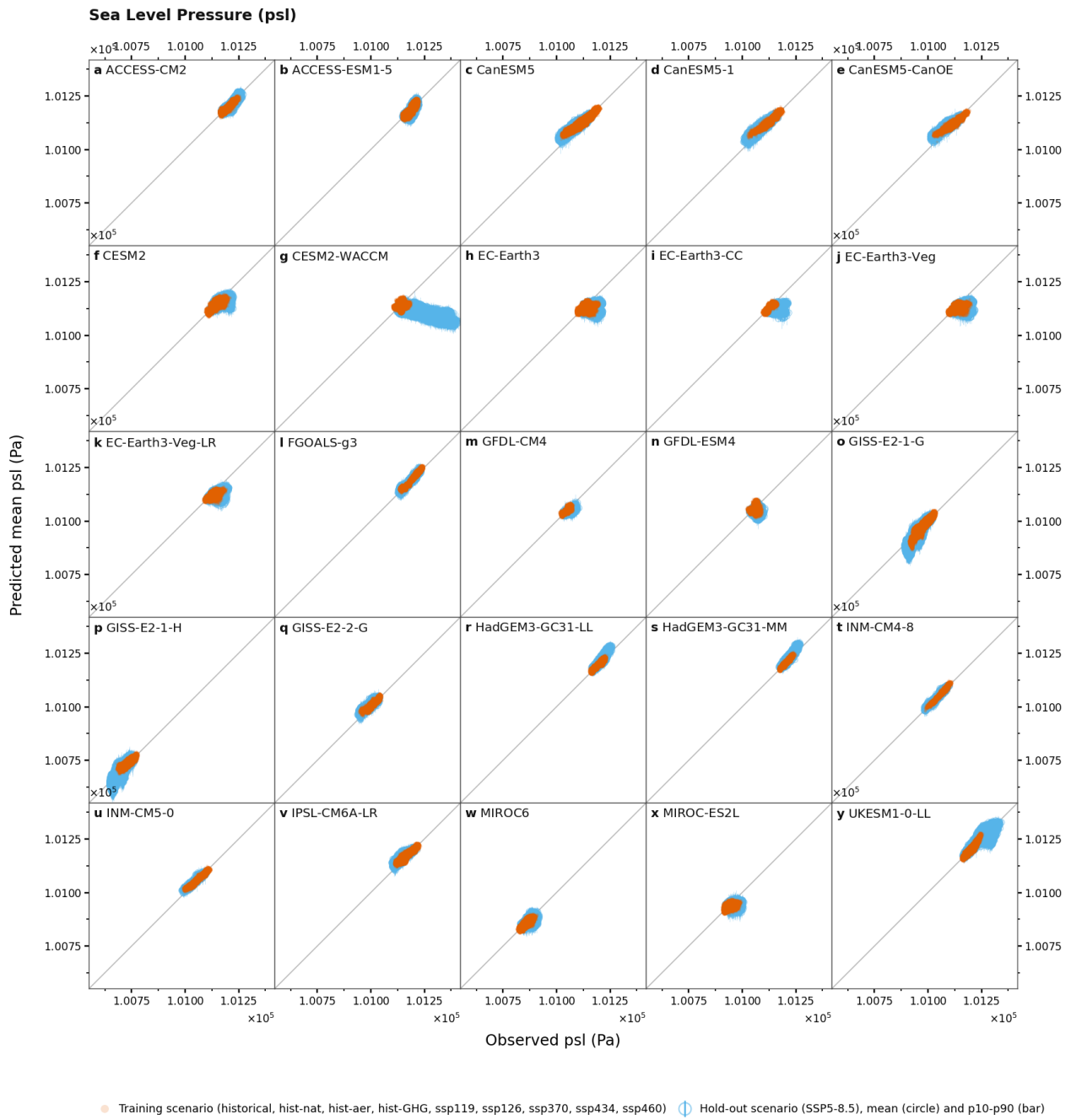


Figure S8. GCMagicc-XS per-model sea-level pressure (*psl*): observed versus predicted. As Fig. S2 but for sea-level pressure.

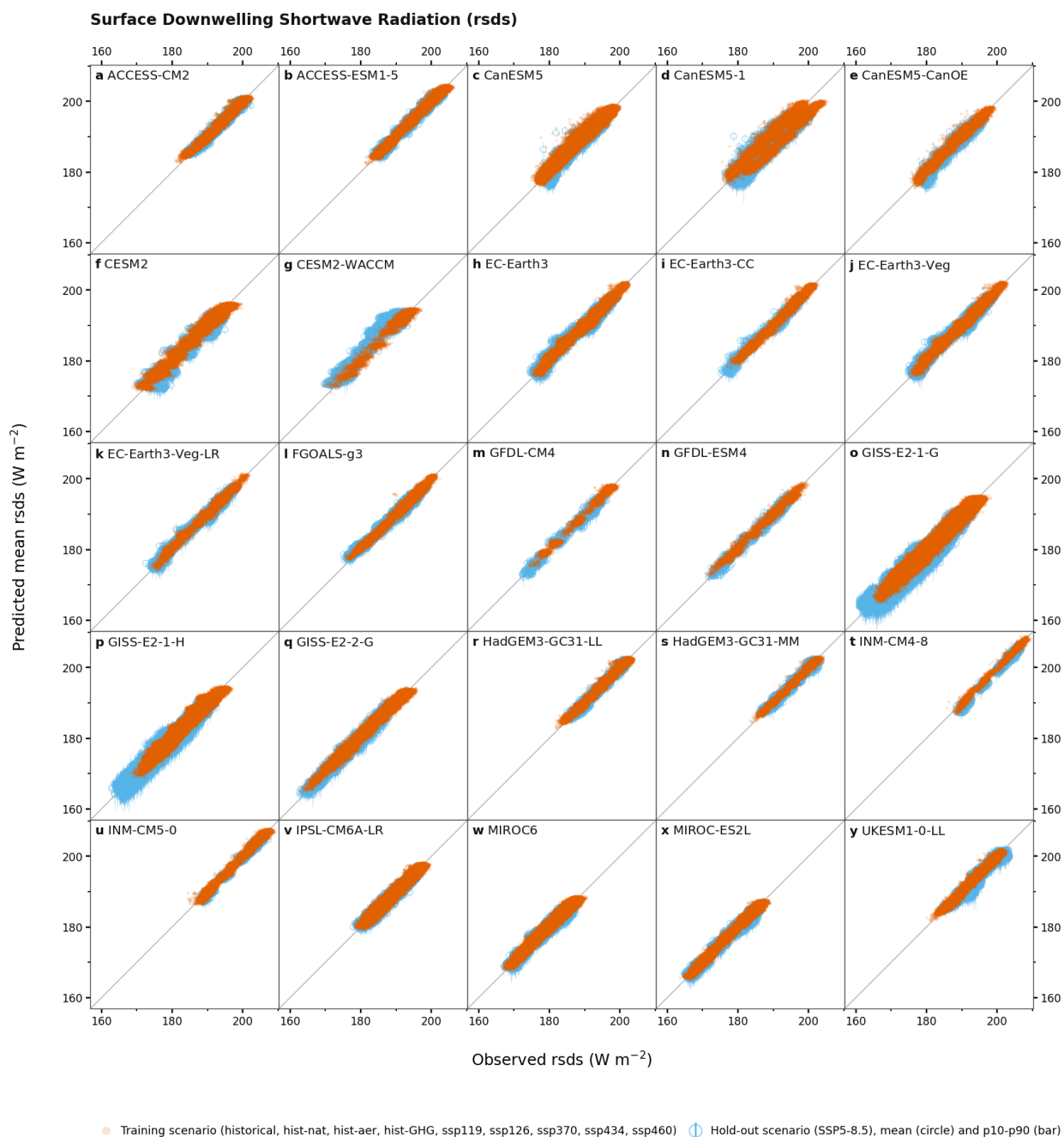


Figure S9. GCMagicc-XS per-model surface downwelling shortwave radiation (*rsds*): observed versus predicted. As Fig. S2 but for surface downwelling shortwave radiation.

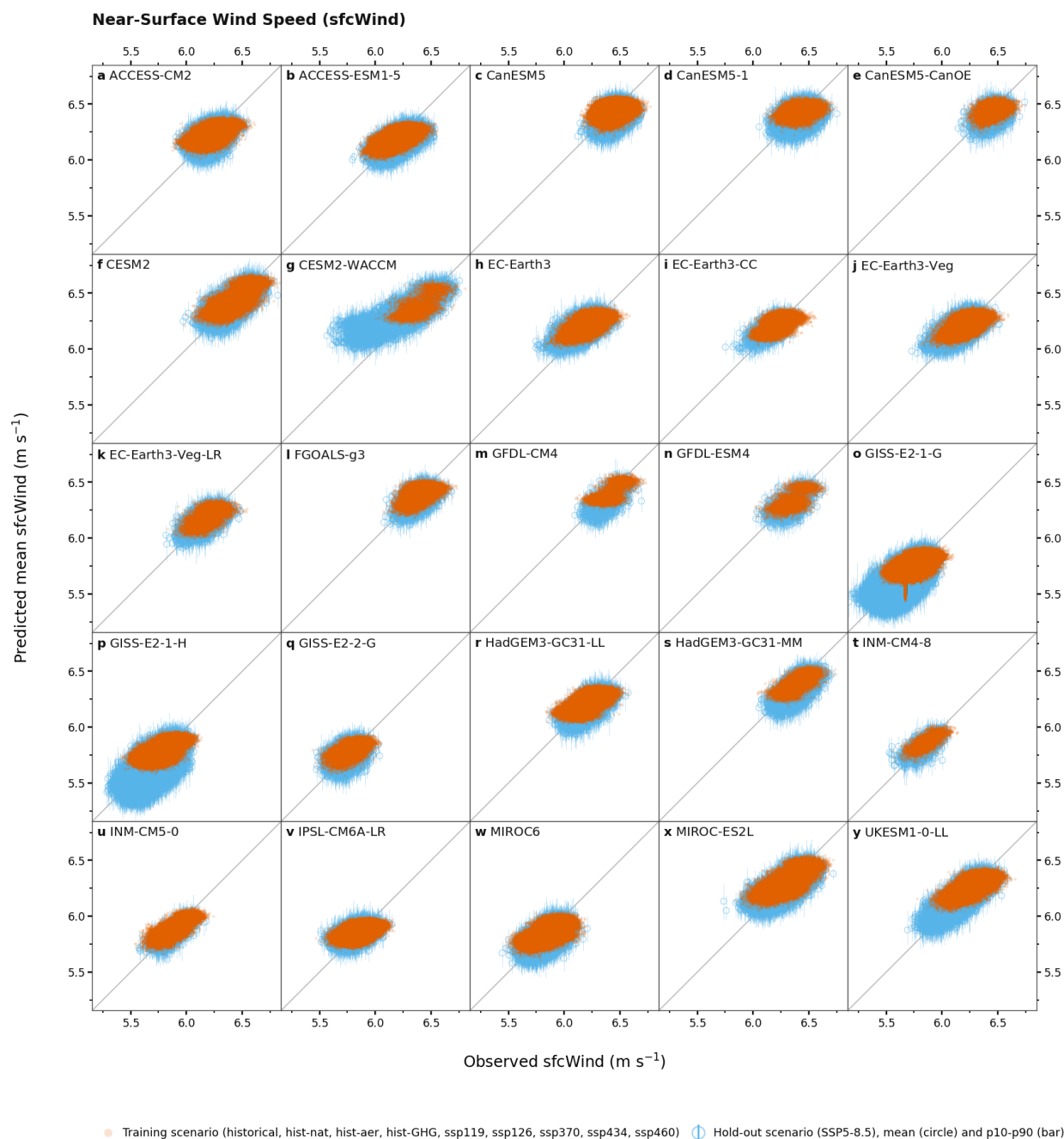


Figure S10. GCMagicc-XS per-model near-surface wind speed (*sfcWind*): observed versus predicted. As Fig. S2 but for near-surface wind speed.

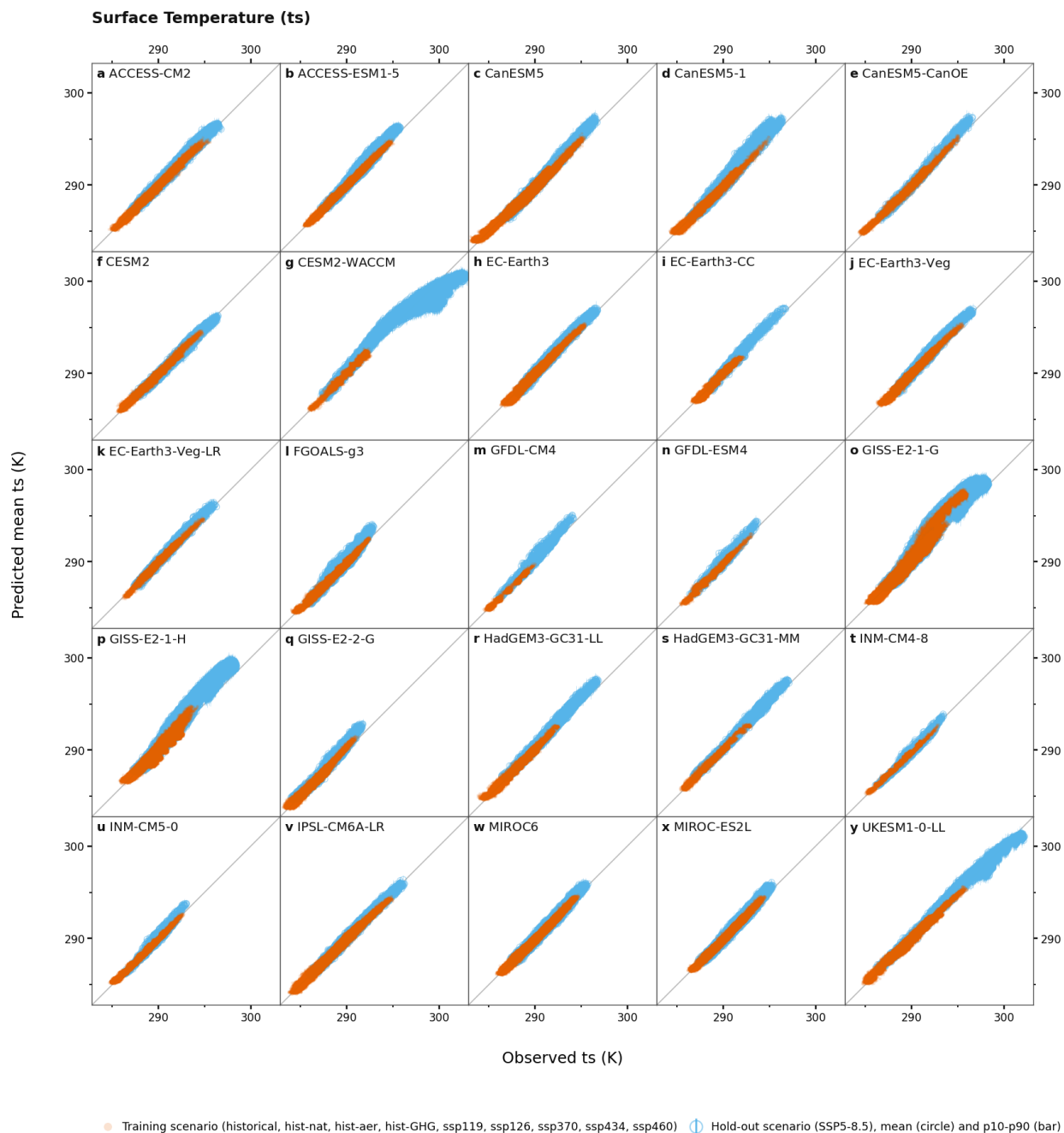


Figure S11. GCMagicc-XS per-model surface temperature (ts): observed versus predicted. As Fig. S2 but for surface temperature.

S4 Comparison of GCMagicc versus CMIP6 ESMs

Systematic evaluation frameworks for climate emulators are emerging. Here we use ten benchmark diagnostics. These are not all independent of the machine-learning training objective: the stochastic generator is trained with an engression objective based on the energy score, and the diagnostics that compare the emulated fields themselves (GlobalTimeseries, BiasMaps, SeasonalCycle, ZonalMeans, Variability and Histograms) evaluate a closely related functional, so a score below one on those diagnostics is a consistency check rather than independent validation. The diagnostics built on derived diagnostics (IndicatorFreq, ENSOTelecon, ENSOCharacteristics and HumidityCoupling) are not targeted by the training objective and carry the independent evidence. The frozen database contains 4,539,079 *gofnc* records across GCMagicc and GCMagicc-CE, of which 3,974,527 belong to the ten diagnostic domains documented here. The database includes 1,265,222 records for the SSP2-4.5 hold-out comparison.

S4.1 Fidelity metric diagnostics to compare GCMagicc vs CMIP6

The validation goodness-of-fit statistics we employ are L_2 distances between a specific CMIP6 run (identified by *source_id* (model), *experiment_id* (scenario), and *member_id* (ensemble member, capped at 4 per scenario)) and a comparator. We use ten diagnostic families across complementary aspects of the climate state. Per-diagnostic record counts in the consolidated database are reported in Table S4 and descriptions are provided below.

All diagnostics first align File A and File B on their common coordinates before computing statistics. Let Ω denote that common support and let w_i denote cosine-latitude area weights where spatial weighting is used, with $w_i = 1$ for unweighted statistics. We write

$$\text{Dev}(A, B) = \frac{\sum_{i \in \Omega} w_i (A_i - B_i)}{\sum_{i \in \Omega} w_i}, \quad \text{RMSE}(A, B) = \left(\frac{\sum_{i \in \Omega} w_i (A_i - B_i)^2}{\sum_{i \in \Omega} w_i} \right)^{1/2}, \quad (\text{S5})$$

and, for histogram-based diagnostics,

$$\|h_A - h_B\|_{L^2} = \left(\sum_b [h_A(z_b) - h_B(z_b)]^2 \Delta z_b \right)^{1/2}. \quad (\text{S6})$$

Temporal means and quantiles are always taken over the diagnostic-specific window encoded in the relevant record. Pearson correlations use the standard covariance normalisation on the aligned monthly series.

S4.1.1 GlobalTimeseries

For each variable the diagnostic forms the global cosine-latitude-weighted mean time series, aligns the two series in time, and computes the root-mean-square difference. Global-annual-mean *tas* is a predictor in the default GCMagicc setup and is therefore not informative for this metric (other than as a consistency check), but the other variables (*hurs*, *huss*, *pr*, *psl*, *rsds*, *sfcWind*) show informative global monthly-resolved evolution over the 21st century.

S4.1.2 BiasMaps

250 BiasMaps investigates regional biases of the gridded $1^\circ \times 1^\circ$ output, for both median patterns and tails (10th and 90th percentiles), with RMSE computed per month or season, variable, 21-year window, scenario and percentile. The diagnostic compares seasonal or annual spatial fields after selecting the diagnostic window and then aggregating over time with the configured mean, median, or percentile. If $F_A^{(s,q)}(x)$ and $F_B^{(s,q)}(x)$ denote the season- s , aggregation- q fields at grid cell x , the stored difference field is $D(x) = F_A^{(s,q)}(x) - F_B^{(s,q)}(x)$. The L^2 norm is the RMSE obtained by applying Eq. S5 to $D(x)$ with cosine-
255 latitude weights. Default windows depend on scenario class (historical, SSP, overlap, or trailing 20-year windows).

S4.1.3 SeasonalCycle

SeasonalCycle computes monthly climatologies for each IPCC AR6 region and compares the resulting 12-month curves, since phase and amplitude of the seasonal cycle matter for agriculture, hydrology and energy demand. The principal statistic is *Seasonal_RMSE*, the RMSE across the 12 monthly climatological values.

260 S4.1.4 ZonalMeans

ZonalMeans forms monthly climatologies, averages them zonally over longitude, and then averages over months to obtain annual zonal profiles for File A and File B. The frozen diagnostic stores *ZonalClim_RMSE* as the unweighted RMSE across the aligned latitude coordinates of those two annual profiles, alongside the difference in cosine-latitude-weighted high-minus-low-latitude gradients.

265 S4.1.5 Variability

Variability is analogous to SeasonalCycle but is applied to monthly standard deviation climatologies rather than monthly means. Before computing standard deviations, each grid point is detrended separately for each calendar month using a 21-year running mean with 10-year trend-preserving edge treatment.

S4.1.6 Histograms

270 Histograms pool all valid values from the chosen time window within each IPCC AR6 region, using the common spatial grid after alignment. The diagnostic first determines a common set of histogram edges from the pooled minima and maxima across both files, then computes regional density histograms h_A and h_B on those common bins. It stores the Kolmogorov-Smirnov distance $\sup_z |F_A(z) - F_B(z)|$ and the histogram L^2 distance given by Eq. S6, the latter being used for the energy distances.

S4.1.7 HumidityCoupling

275 Humidity coupling evaluates the co-variability between global-mean temperature and either relative humidity (*hurs*) or specific humidity (*huss*). The diagnostics computes grid-cell correlations between monthly *tas* anomalies and humidity anomalies,

forms the correlation-difference map, and provides the weighted mean and weighted RMSE of that map, the latter again being used for the energy distances.

S4.1.8 IndicatorFreq

280 IndicatorFreq analyses six monthly climate indices: Niño-3.4, Niño-3, Niño-4, SOI, PDO, and NAO. Niño indices are built from area-weighted tropical Pacific SST or near-surface temperature anomalies; SOI is the Darwin-minus-Tahiti sea-level-pressure anomaly difference; PDO is the first principal component from an SVD of North Pacific SST anomalies weighted by $\sqrt{\cos\phi}$; and NAO is the Azores-minus-Iceland sea-level-pressure anomaly difference. All six indices use the same post-processing: a centred 3-month running mean followed by removal of low-frequency variability with a fourth-order Butterworth
285 low-pass filter with a 30-year cutoff, subtracted from the smoothed series. The diagnostic estimates power spectral density with Welch’s method, a Hann window, and $n_{\text{perseg}} = \min(256, \lfloor N/2 \rfloor)$ for series length $N \geq 32$ (otherwise the full series length), interpolates $\log_{10}$ PSD to 150 common log-spaced frequencies, and we use the resulting RMSE between CMIP6 and GCMagicc for the energy distances.

S4.1.9 ENSOTelecon

290 ENSO teleconnections are diagnosed by regressing DJF anomalies of each target variable (tas, pr) on the 1-year lagged OND Niño-3.4 index, producing a slope map for File A and for File B. The anomaly preprocessing uses smoothing and detrending prior to the seasonal regression. The RMSE is computed from the map difference after restricting latitude to $[-60^\circ, 60^\circ]$ and applying cosine-latitude weighting.

S4.1.10 ENSOCharacteristics

295 ENSO characteristics are derived from a Niño-3.4 index built from monthly anomalies over 5°S – 5°N and 190°E – 240°E , followed by linear detrending and triangular 5-month smoothing. The lifecycle diagnostic regresses each lagged monthly anomaly on December Niño-3.4 anomalies over lags $-36, \dots, +36$ months; *ENSO_lifecycle_RMSE* is the RMSE between the two resulting slope vectors over common lags. Event duration is the number of contiguous lags around zero for which the absolute lifecycle slope exceeds 0.25, and the diagnostic stores both the signed duration difference and its absolute value as
300 *ENSO_duration_RMSE*. ENSO amplitude is the sample standard deviation of DJF means of the smoothed Niño-3.4 series, again stored both as a signed difference and as *ENSO_amplitude_RMSE*.

S4.2 Validation: Energy distances and energy distance scores

The database comparison keys used here are ‘goodness-of-fit’ GOF statistics with key letters, as explained in the following table.

Table S4. Number of CMIP6 ESM and GCMagicc comparisons we performed. The database holds 4,539,079 *gofnc* records in total, of which 3,974,527 fall in these ten diagnostic domains, 1,265,222 belong to the SSP2-4.5 hold-out comparison, and 649,058 enter the observation-referenced displays of Figs. S26–S28. Note that the below counts comprise RMSE, but also amplitude differences, phase-shifts, and absolute differences.

Diagnostic	GCMagicc	GCMagicc-CE
BiasMaps	156,270	91,770
ENSOCharacteristics	2,950	2,045
ENSOTelecon	2,092	1,316
GlobalTimeseries	10,298	6,618
Histograms	597,284	343,244
HumidityCoupling	2,368	1,636
IndicatorFreq	10,638	5,004
SeasonalCycle	892,446	507,906
Variability	890,706	431,520
ZonalMeans	10,238	8,178
Total	2,575,290	1,399,237

Table S5. Database comparison keys and corresponding data-source comparisons. Here, N denotes a GCMagicc member or configuration in the legacy database code, C the paired CMIP6 model/member, Z a different randomly selected CMIP6 model, and O an observational or reanalysis product.

Key	Comparison
GOFNC	Paired GCMagicc and CMIP6 outputs
GOFNN	GCMagicc members or configurations
GOFCC	Initial-condition members of the same CMIP6 model
GOF CZ	Target CMIP6 model and another CMIP6 model
GOF ZZ	Members within the other CMIP6 models
GOF OC	Observations/reanalysis and CMIP6
GOF ON	Observations/reanalysis and GCMagicc

305 For random vectors X and Y with independent copies X' and Y' , the energy distance between the distribution P_X and P_Y of X and Y respectively is

$$\text{ED}(P_X, P_Y) = 2\mathbb{E}\|X - Y\|_2 - \mathbb{E}\|X - X'\|_2 - \mathbb{E}\|Y - Y'\|_2.$$

By construction, ED reflects distributional agreement rather than matching a single moment.

While these energy distances all approach zero when the emulator matches the original ESM distribution, their magnitudes
 310 are not directly comparable across diagnostics or variables. We therefore define the energy-distance score as

$$S(M, C; Z) = \frac{\text{ED}(M, C)}{\text{ED}(C, Z)},$$

where $\text{ED}(M, C)$ is the energy distance between the calibrated GCMagicc M and a CMIP6 model C , and $\text{ED}(C, Z)$ is the
 energy distance between that CMIP6 model C and a random selection of other CMIP6 models Z for the same experiment.
 $S < 1$ indicates that the emulator is closer to a given ESM than the ESMs are to each other for that metric and experiment
 315 (green area in the last column in Figure 3).

For observational data $O_i, i = 1, \dots, n$ with empirical distribution $P_{n,O}$ and model data $M_i, i = 1, \dots, n$ with empirical
 distribution $P_{n,M}$ and an independent set of draws $M'_i, i = 1, \dots, n$ generated from the same model, we get with the definition
 above

$$\text{ED}(P_{n,O}, P_{n,M}) = n^{-1} \sum_{i=1}^n (2\|O_i - M_i\|_2 - \|M_i - M'_i\|_2 - \|O_i - O'_i\|_2),$$

320 Since the real world provides only a single realisation but the term $\|O_i - O'_i\|_2$ is independent of the model M , it can safely be
 dropped to get

$$\text{ED}_{\text{obs}}(P_{n,O}, P_{n,M}) = n^{-1} \sum_{i=1}^n (2\|O_i - M_i\|_2 - \|M_i - M'_i\|_2),$$

Because there is no second observational ensemble member, this quantity does not approach zero under a perfect distributional
 match; instead it asymptotically approaches $\mathbb{E}\|O - O'\|_2$, where O and O' denote two hypothetical ensemble members of the
 325 real world (Figure 4).

S4.3 Extended data tables and alternate draws of Figure 3

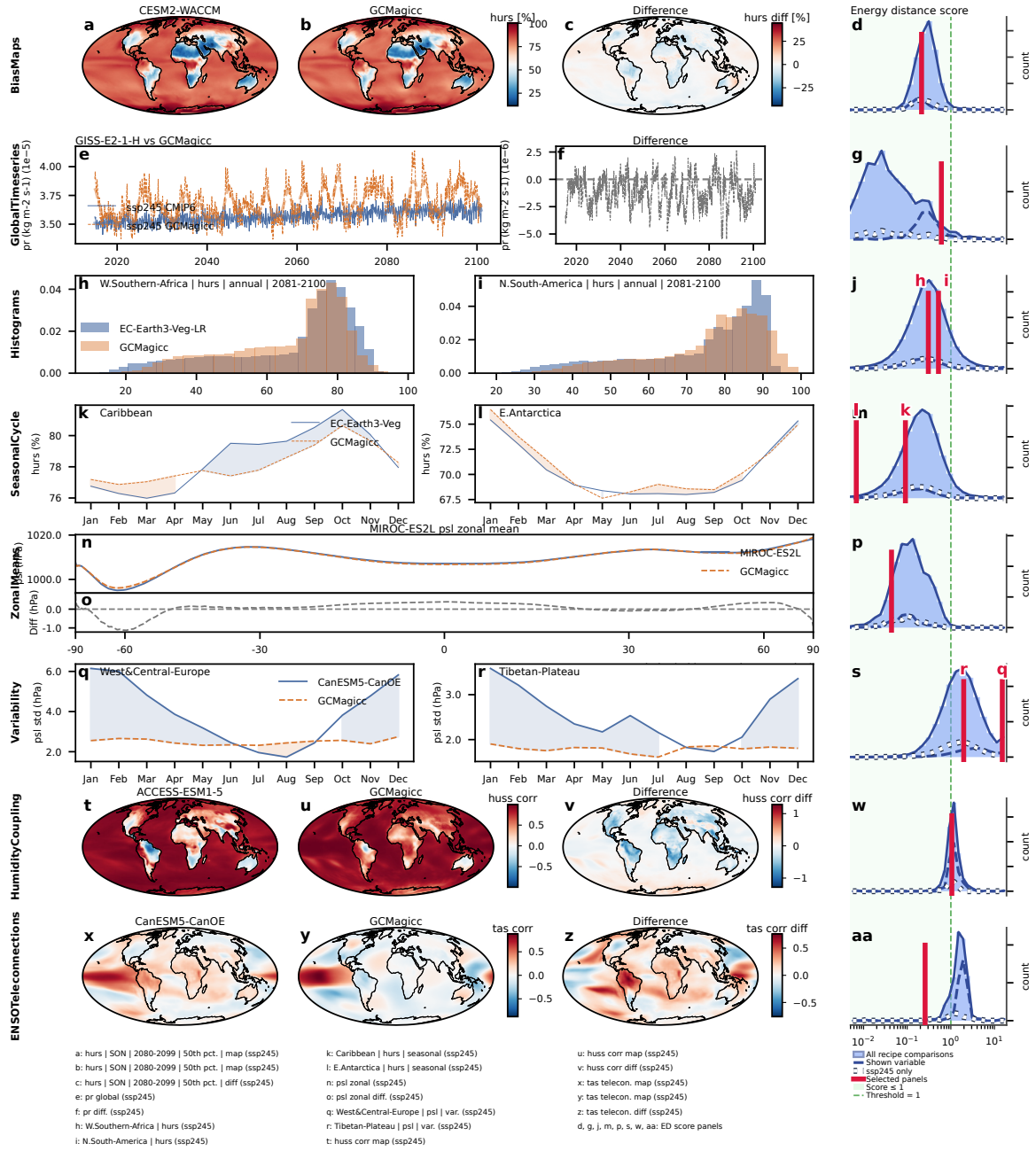
The following data tables provide the row-specific selections, energy-distance-score summaries, and 10 alternate random draws
 of Figure 3 in the main manuscript. Each draw is an independent random selection of metric diagnostics from the same pool,
 hence directly comparable to the main text Figure 3. The alternate draws are numbered 001–010.

Table S6. Row-specific selections and validation metrics corresponding to the retained rows in Figure 3. Energy-distance scores computed from the smallest samples ($n = 1$ to $n = 4$) can take negative values under the estimator used here; these are retained rather than truncated. Given the close agreement between ensemble members at these sample sizes this does not materially affect the relative comparisons reported.

Panel	diagnostic	Selected case	Metrics summary
a-d	BiasMaps	source_id: ACCESS-CM2 member_id: r1i1p1f1 experiment: ssp245 variable: tas metrictype: Map_RMSE_2080to2099_annual_perc90	SCOREEDISTC: 0.339; GOFNC: 0.973254 (n=1); GOFNN: 0.578135 (n=2); GOFCC: 0.309362 (n=3); GOF CZ: 1.88792 (Z n=4); GOFZZ: 0.340901 (Z n=4); Z coverage: 4/5 (partial available case); EDISTNC: 1.059; EDISTCZ: 3.126 (n=4)
e-g	GlobalTimeseries	source_id: IPSL-CM6A-LR member_id: r1i1p1f1 variable: psl scenario: ssp245	SCOREEDISTC: mean=0.053 (n=1); GOFNC: 11.8997 (n=1); GOFNN: 12.8161 (n=1); GOFCC: 4.99083 (n=3); GOF CZ: 62.3034 (Z n=5); GOFZZ: 5.86809 (Z n=5); Z coverage: 5/5 (complete); EDISTNC: 5.993; EDISTCZ: 113.748 (n=5)
h-j	Histograms	metrictype: RMSE_global source_id: ACCESS-CM2 member_id: r1i1p1f1 variable: tasmax scenario: ssp245	ARABIAN-SEA: SCOREEDISTC: 1.434; GOFNC: 0.106665 (n=1); GOFNN: 0.0641687 (n=2); GOFCC: 0.0146257 (n=3); GOF CZ: 0.0597889 (Z n=3); ARABIAN-SEA: GOFZZ: 0.0111461 (Z n=3); Z coverage: 3/4 (partial available case); EDISTNC: 0.135; EDISTCZ: 0.094 (n=3); RUSSIAN-ARCTIC: SCOREEDISTC: 1.563; GOFNC: 0.0518836 (n=1); GOFNN: 0.0154951 (n=2); GOFCC: 0.00777645 (n=3); GOF CZ: 0.034306 (Z n=3); RUSSIAN-ARCTIC: GOFZZ: 0.00932748 (Z n=3); Z coverage: 3/4 (partial available case); EDISTNC: 0.080; EDISTCZ: 0.052 (n=3)
k-m	SeasonalCycle	source_id: HadGEM3-GC31-LL member_id: r1i1p1f3 variable: ts scenario: ssp245	BAY-OF-BENGAL: SCOREEDISTC: 0.329; GOFNC: 0.318403 (n=1); GOFNN: 0.229299 (n=1); GOFCC: 0.0937754 (n=3); GOF CZ: 0.556747 (Z n=3); BAY-OF-BENGAL: GOFZZ: 0.0648674 (Z n=3); Z coverage: 3/4 (partial available case); EDISTNC: 0.314; EDISTCZ: 0.955 (n=3); NEW-ZEALAND: SCOREEDISTC: 0.702; GOFNC: 0.488596 (n=1); GOFNN: 0.280022 (n=1); GOFCC: 0.130764 (n=3); GOF CZ: 0.521203 (Z n=3); NEW-ZEALAND: GOFZZ: 0.104618 (Z n=3); Z coverage: 3/4 (partial available case); EDISTNC: 0.566; EDISTCZ: 0.807 (n=3)
n-p	ZonalMeans	shown panels: k = Bay-of-Bengal; l = New-Zealand source_id: IPSL-CM6A-LR member_id: r1i1p1f1 variable: tas scenario: ssp245	SCOREEDISTC: 0.138; GOFNC: 0.374453 (n=1); GOFNN: 0.054541 (n=1); GOFCC: 0.150019 (n=3); GOF CZ: 2.31859 (Z n=5); GOFZZ: 0.54091 (Z n=5); Z coverage: 5/5 (complete); EDISTNC: 0.544; EDISTCZ: 3.946 (n=5)
q-s	Variability	metrictype: ZonalClim_RMSE_tas_full source_id: IPSL-CM6A-LR member_id: r1i1p1f1 variable: tasmax scenario: ssp245	EQUATORIAL-ATLANTIC-OCEAN: SCOREEDISTC: 12.874; GOFNC: 0.418542 (n=1); GOFNN: 0.0530612 (n=1); GOFCC: 0.026107 (n=1); GOF CZ: 0.0632225 (Z n=5); EQUATORIAL-ATLANTIC-OCEAN: GOFZZ: 0.0414671 (Z n=5); Z coverage: 5/5 (complete); EDISTNC: 0.758; EDISTCZ: 0.059 (n=5); N-PACIFIC-OCEAN: SCOREEDISTC: 5.301; GOFNC: 0.281592 (n=1); GOFNN: 0.0227166 (n=1); GOFCC: 0.0196799 (n=1); GOF CZ: 0.0750647 (Z n=5); N-PACIFIC-OCEAN: GOFZZ: 0.0322043 (Z n=5); Z coverage: 5/5 (complete); EDISTNC: 0.521; EDISTCZ: 0.098 (n=5)
t-w	HumidityCoupling	region 1: EQUATORIAL-ATLANTIC-OCEAN; metrictype: Variability_RMSE_EQUATORIAL-ATLANTIC-OCEAN_full region 2: N-PACIFIC-OCEAN; metrictype: Variability_RMSE_N-PACIFIC-OCEAN_full shown panels: q = Equatorial.Atlantic-Ocean; r = N.Pacific-Ocean source_id: CESM2-WACCM member_id: r2i1p1f1 experiment: ssp245 variable: hurs	SCOREEDISTC: 0.888; GOFNC: 0.372297 (n=1); GOFNN: 0.297419 (n=1); GOFCC: 0.04765 (n=3); GOF CZ: 0.28067 (Z n=5); GOFZZ: 0.0637613 (Z n=5); Z coverage: 5/5 (complete); EDISTNC: 0.400; EDISTCZ: 0.450 (n=5)
x-aa	ENSOTeleconnections	metrictype: TempHum_RMSE_Diff_hurs_full source_id: EC-Earth3-Veg member_id: r12i1p1f1 experiment: ssp245 variable: pr metrictype: Map_RMSE_2015to2100_annual	SCOREEDISTC: 1.357; GOFNC: 0.356968 (n=1); GOFNN: 0.311549 (n=1); GOFCC: 0.139724 (n=3); GOF CZ: 0.235148 (Z n=4); GOFZZ: 0.13702 (Z n=4); Z coverage: 4/5 (partial available case); EDISTNC: 0.263; EDISTCZ: 0.194 (n=4)

Table S7. Energy-distance score summaries corresponding to the histogram panels in Figure 3. Entries give the median (10th–90th percentile range) and sample size. Three diagnostics have median scores above one over all records: Variability (1.6), HumidityCoupling (1.1) and ENSOTelecon (1.6). For these diagnostics GCMagicc is, at the median, further from the target CMIP6 model than a randomly chosen other CMIP6 model is, so the statement that GCMagicc matches every considered CMIP6 model better than an average other CMIP6 model holds for the mean percentile rank across the displayed diagnostic set and not for the individual diagnostics.

Panel	diagnostic	Variable subset	Energy-distance score			<i>n</i>			Selected example
			all	var-subset	ssp245	all	var-subset	ssp245	
d	BiasMaps	tas	0.27 (0.13–0.53)	0.26 (0.13–0.45)	0.22 (0.12–0.44)	33750	3375	4050	0.34
g	GlobalTimeseries	psl	0.013 (-0.078–0.21)	0.022 (0.0078–0.068)	0.0047 (-0.12–0.15)	2480	248	270	0.053
j	Histograms	tasmax	0.26 (0.025–0.97)	0.24 (-0.013–0.85)	0.21 (-0.0079–0.82)	124120	12412	15080	h: 1.4; i: 1.6
m	SeasonalCycle	ts	0.18 (0.018–0.66)	0.18 (0.011–0.62)	0.14 (-0.0067–0.51)	107880	10788	15080	k: 0.33; l: 0.70
p	ZonalMeans	tas	0.11 (0.021–0.37)	0.11 (-0.0063–0.31)	0.084 (0.0076–0.28)	2440	244	270	0.14
s	Variability	tasmax	1.6 (0.37–6.1)	1.8 (0.25–6.9)	1.4 (0.13–5)	89262	8932	15602	q: 13; r: 5.3
w	HumidityCoupling	hurs	1.1 (0.73–1.6)	1.1 (0.69–1.6)	1 (0.63–1.6)	496	248	54	0.89
aa	ENSOTelecon	pr	1.6 (0.85–2.3)	1.4 (0.78–2.1)	1.2 (0.34–2.7)	424	212	52	1.4



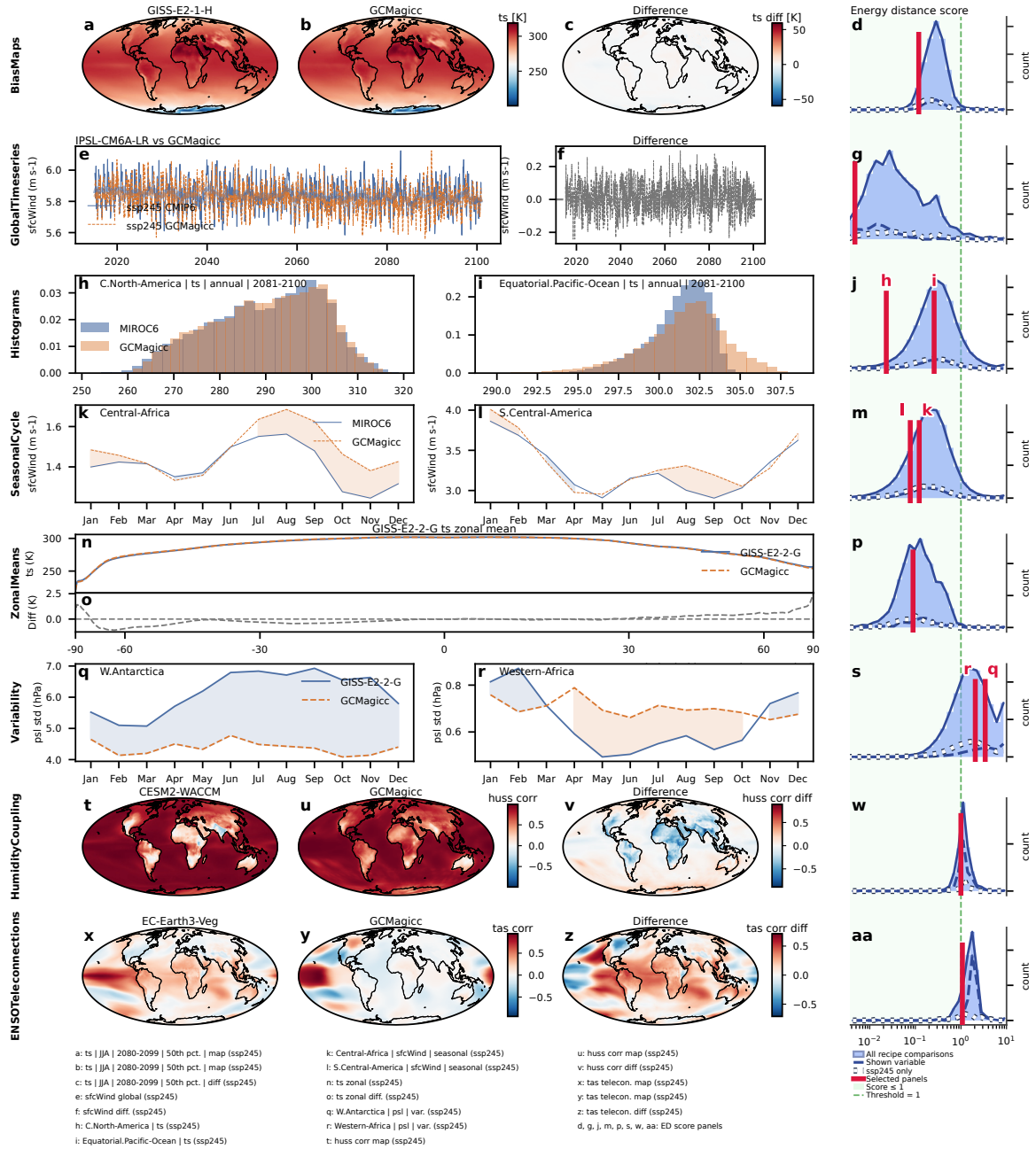


Figure S14. Fidelity evaluation of GCMagicc vs CMIP6. As Figure 3 in the manuscript, but random alternate draws of metric diagnostics

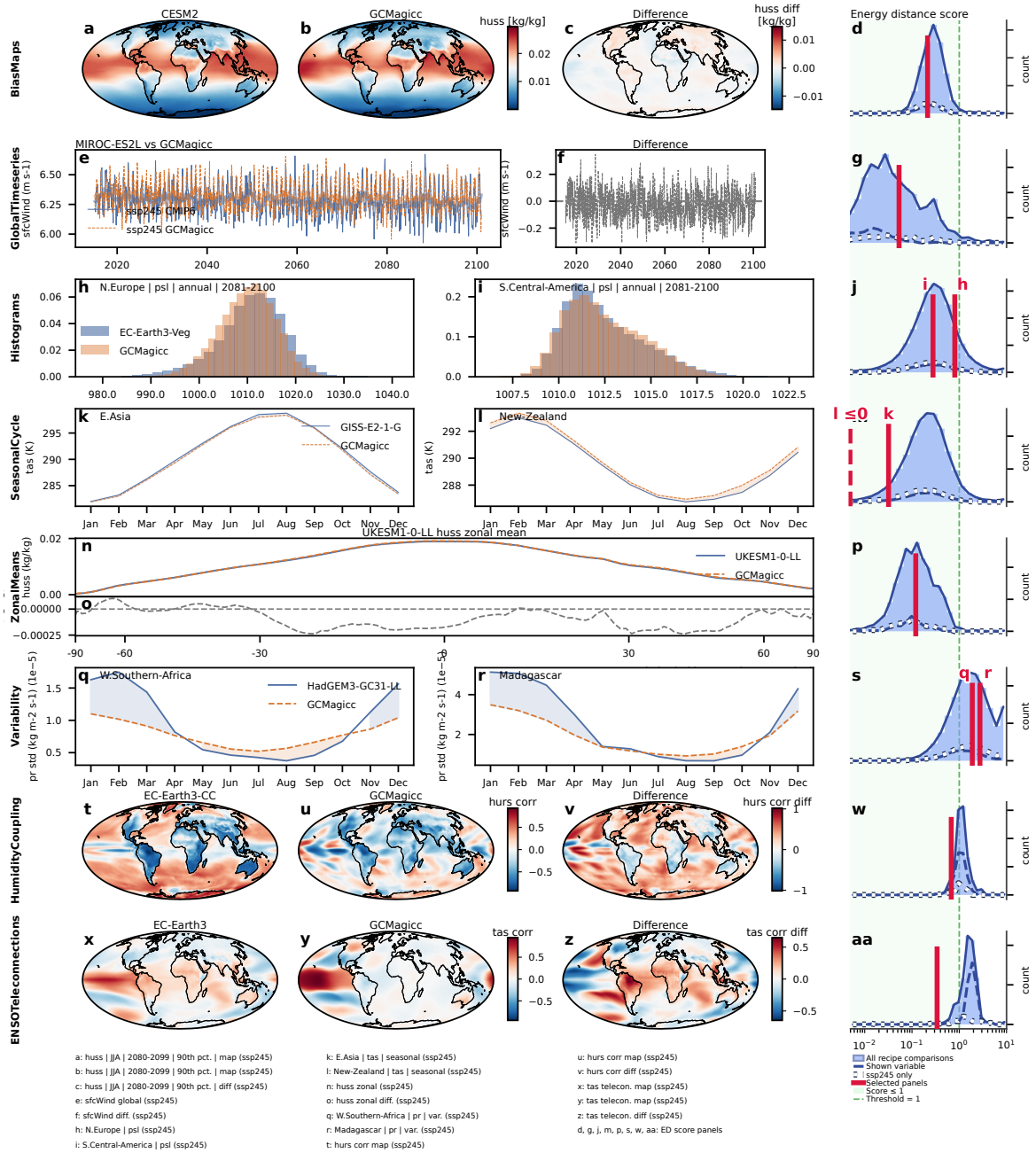


Figure S15. Fidelity evaluation of GCMagicc vs CMIP6. As Figure 3 in the manuscript, but random alternate draws of metric diagnostics

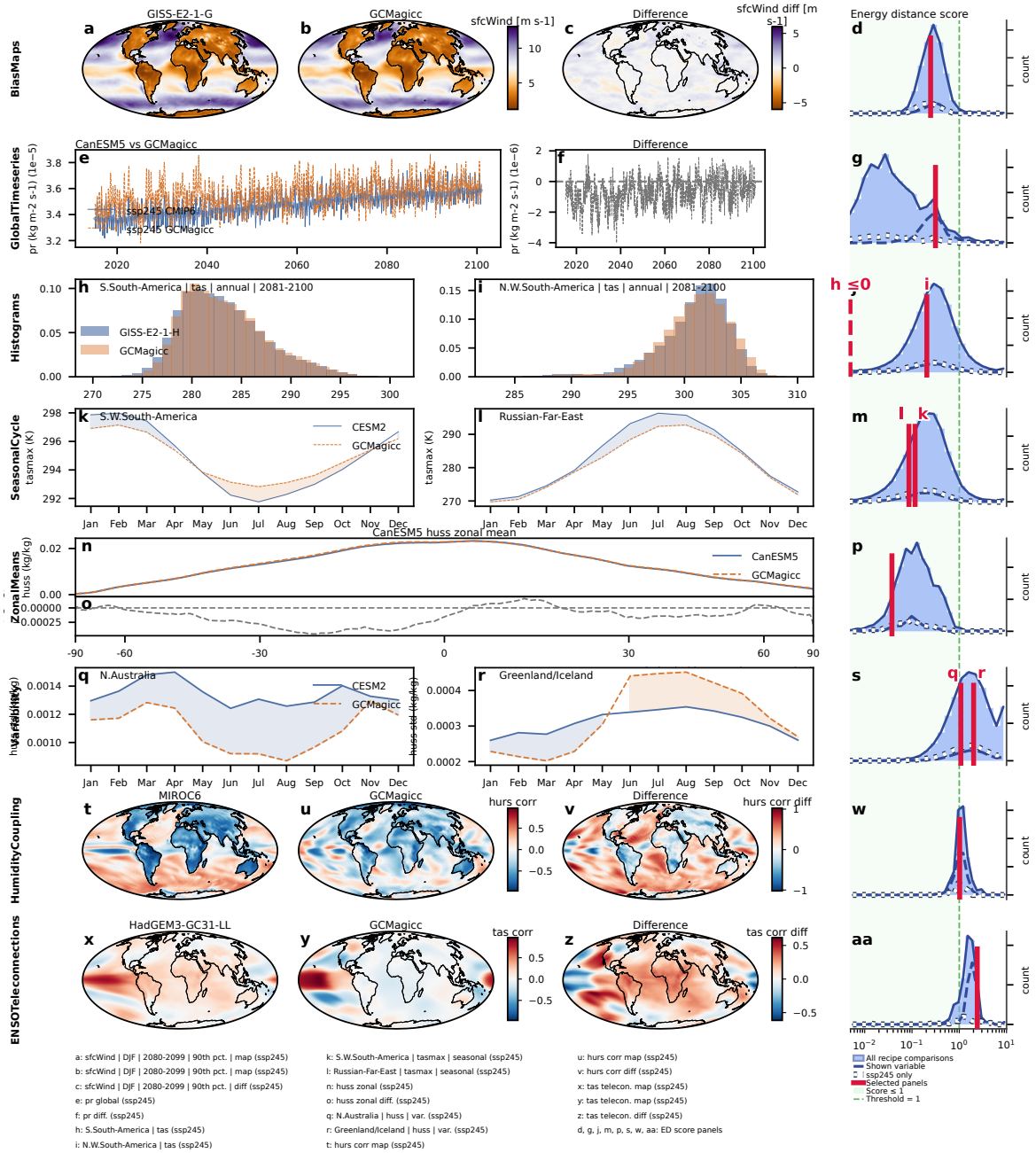


Figure S16. Fidelity evaluation of GCMagicc vs CMIP6. As Figure 3 in the manuscript, but random alternate draws of metric diagnostics

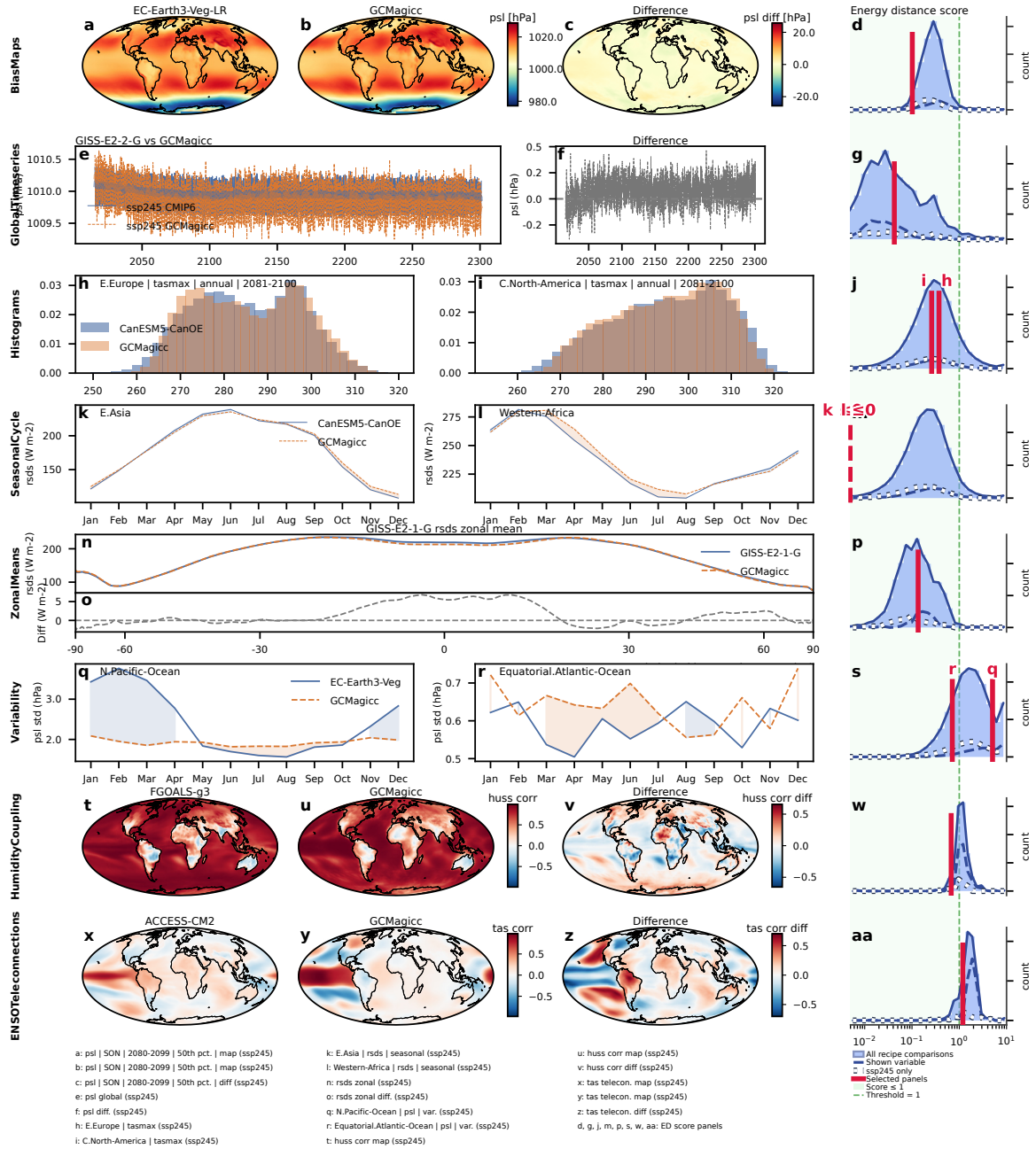


Figure S17. Fidelity evaluation of GCMagicc vs CMIP6. As Figure 3 in the manuscript, but random alternate draws of metric diagnostics

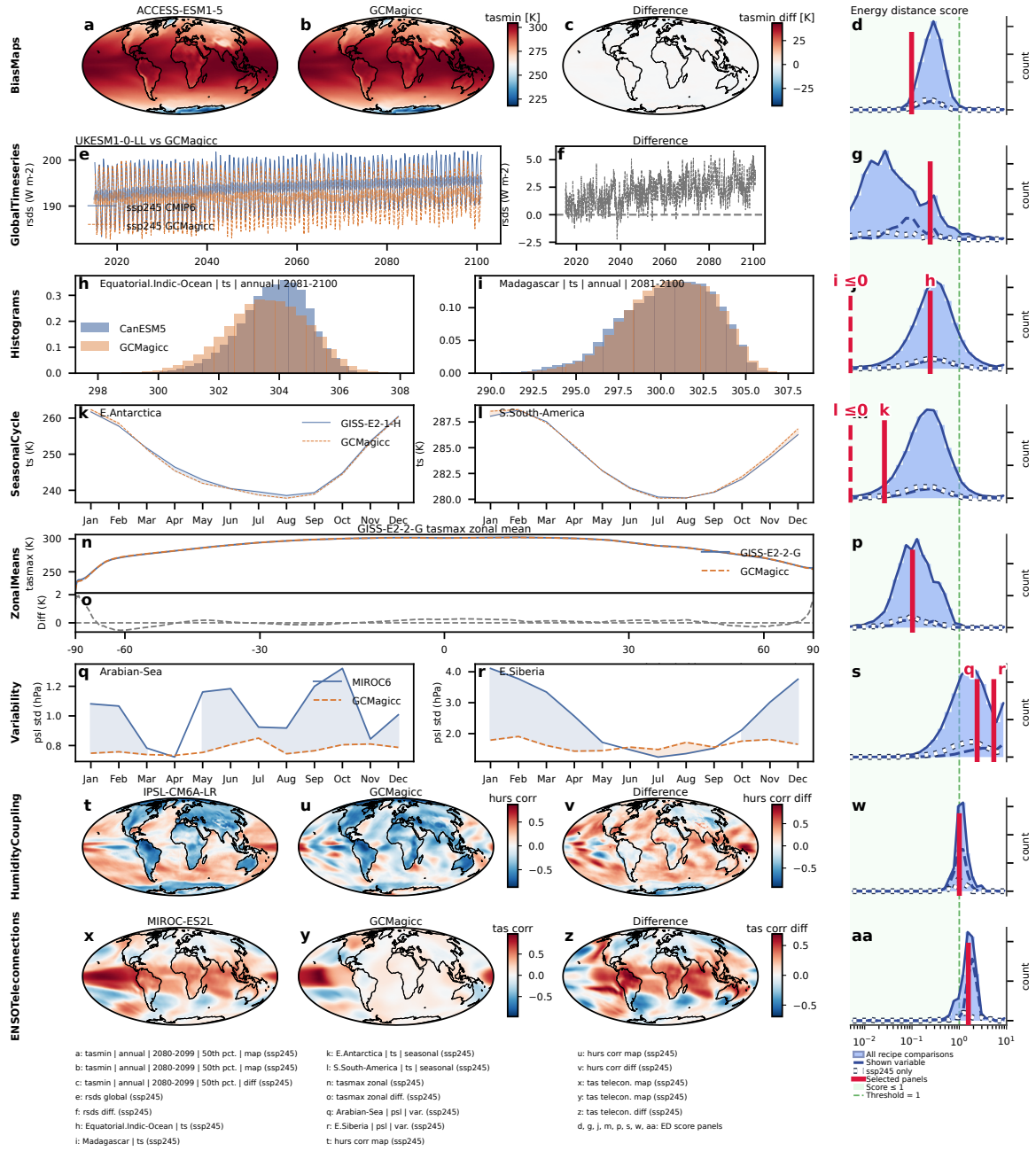


Figure S18. Fidelity evaluation of GCMagicc vs CMIP6. As Figure 3 in the manuscript, but a random alternate draw of metric diagnostics.

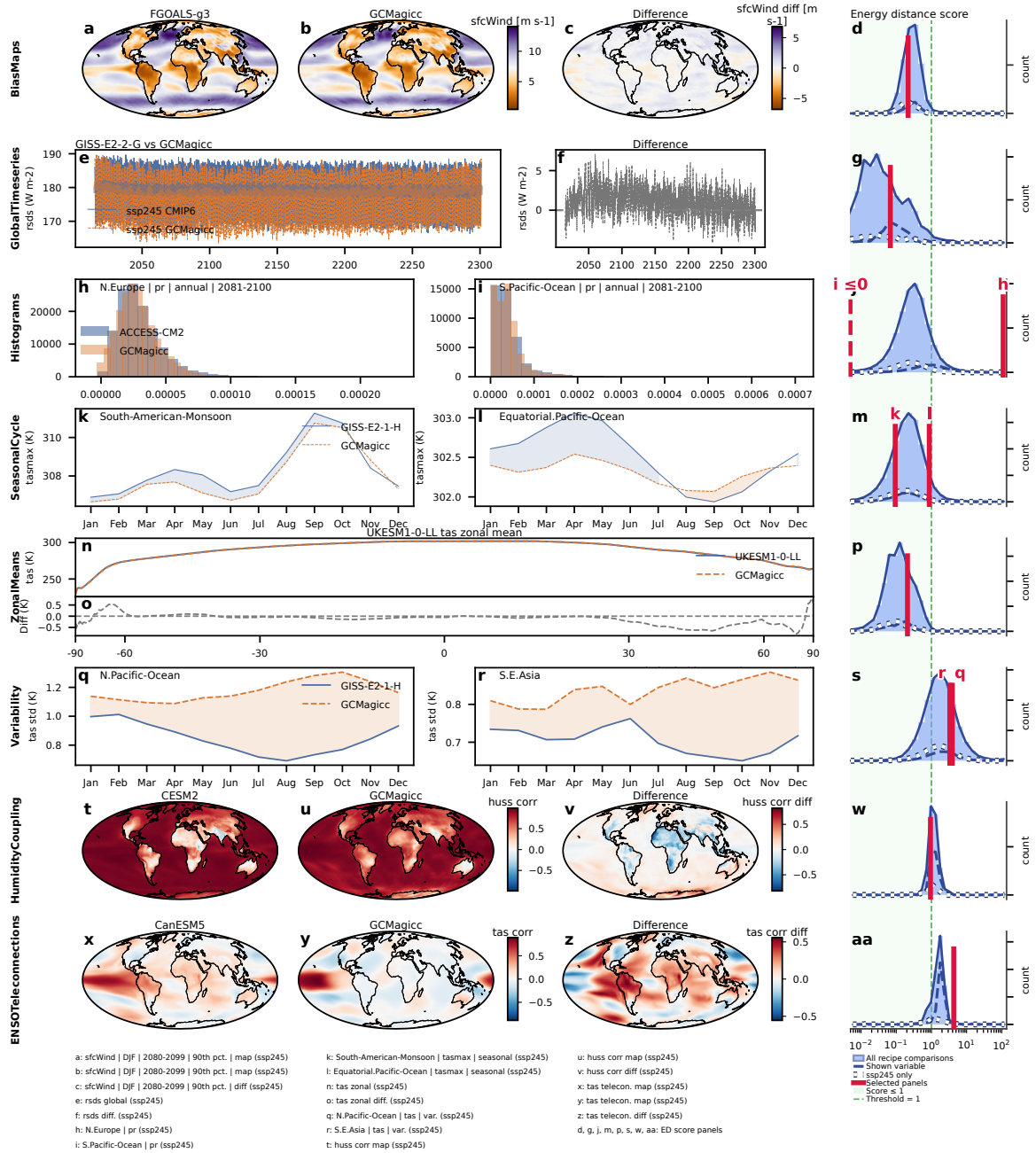


Figure S19. Fidelity evaluation of GCMagicc vs CMIP6. As Figure 3 in the manuscript, but random alternate draws of metric diagnostics

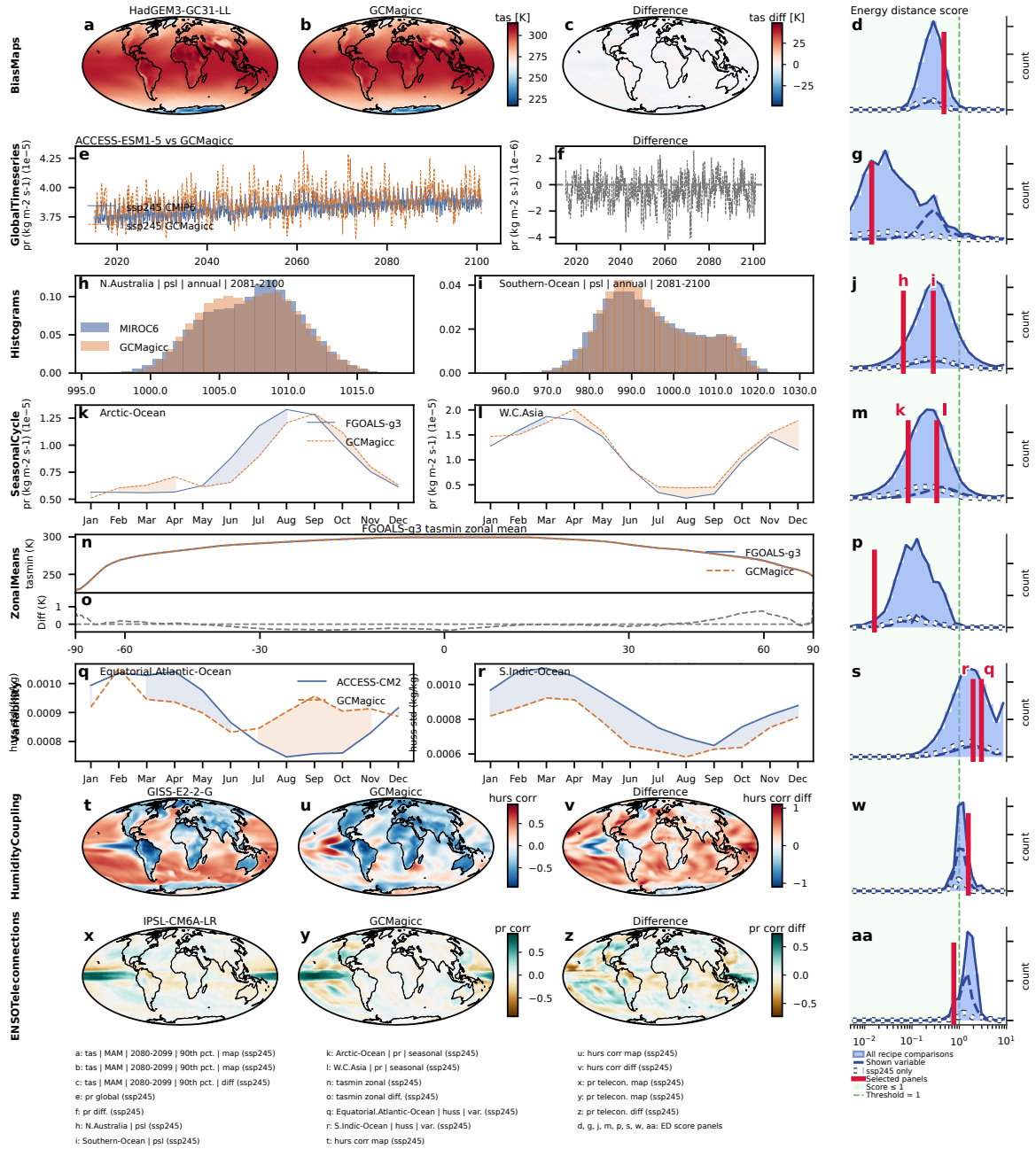


Figure S20. Fidelity evaluation of GCMagicc vs CMIP6. As Figure 3 in the manuscript, but random alternate draws of metric diagnostics

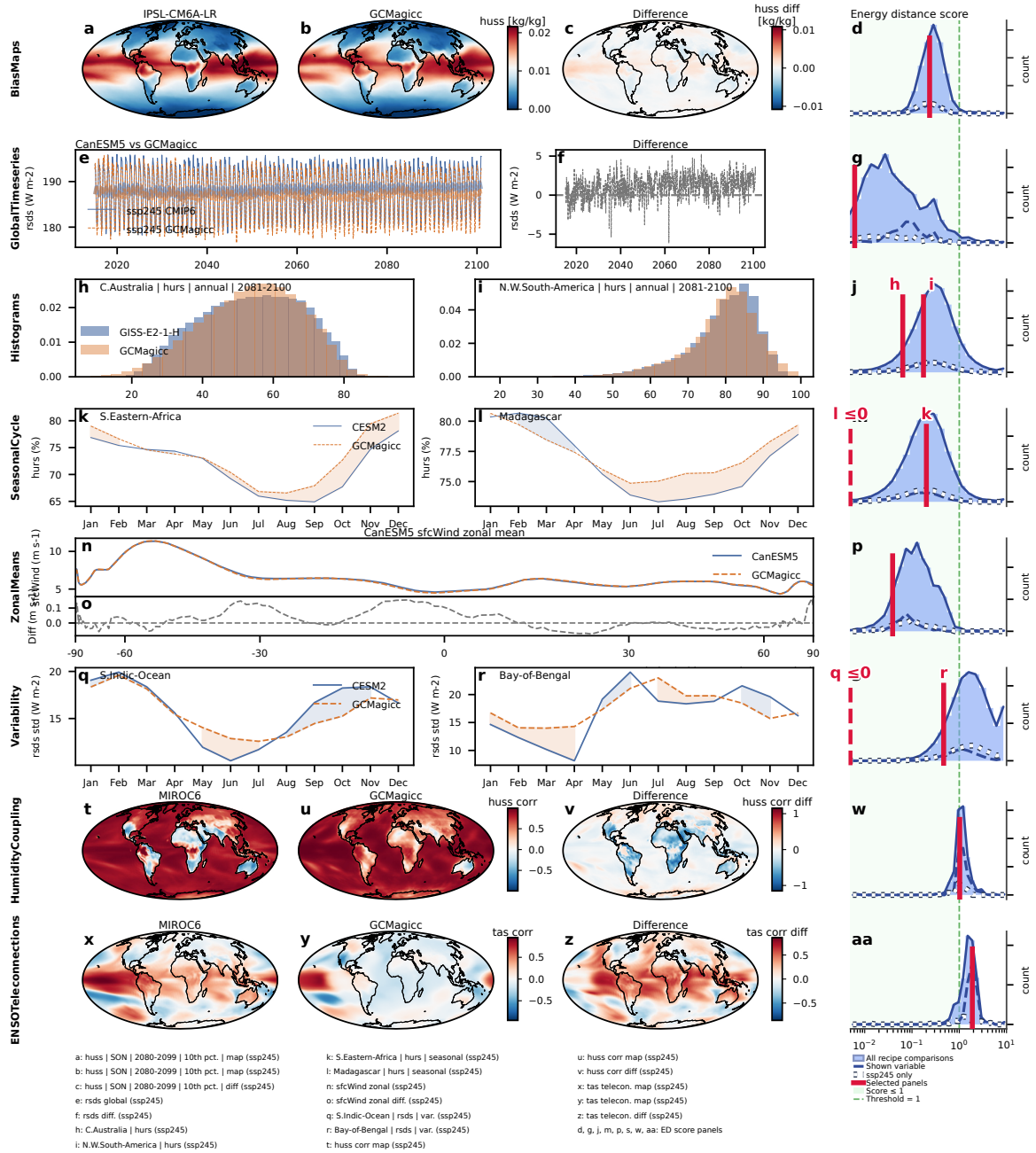
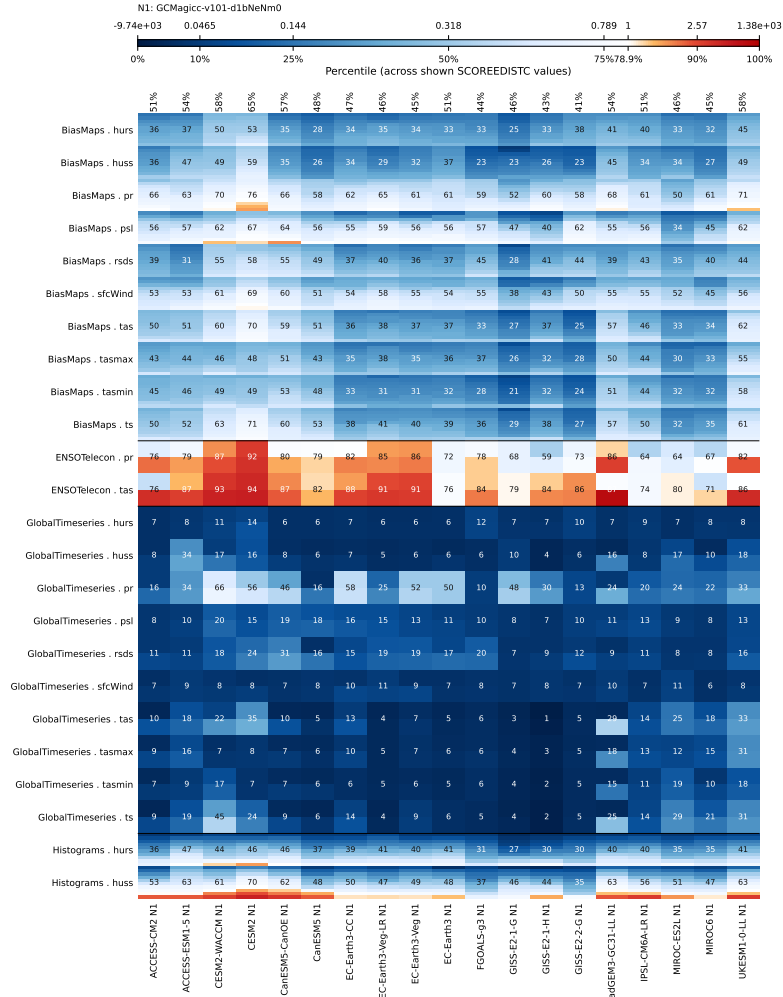


Figure S21. Fidelity evaluation of GCMagicc vs CMIP6. As Figure 3 in the manuscript, but random alternate draws of metric diagnostics

SCOREEDISTC ssp245 only (part 1 of 3)



SCOREDISTC ssp245 only (part 2 of 3)

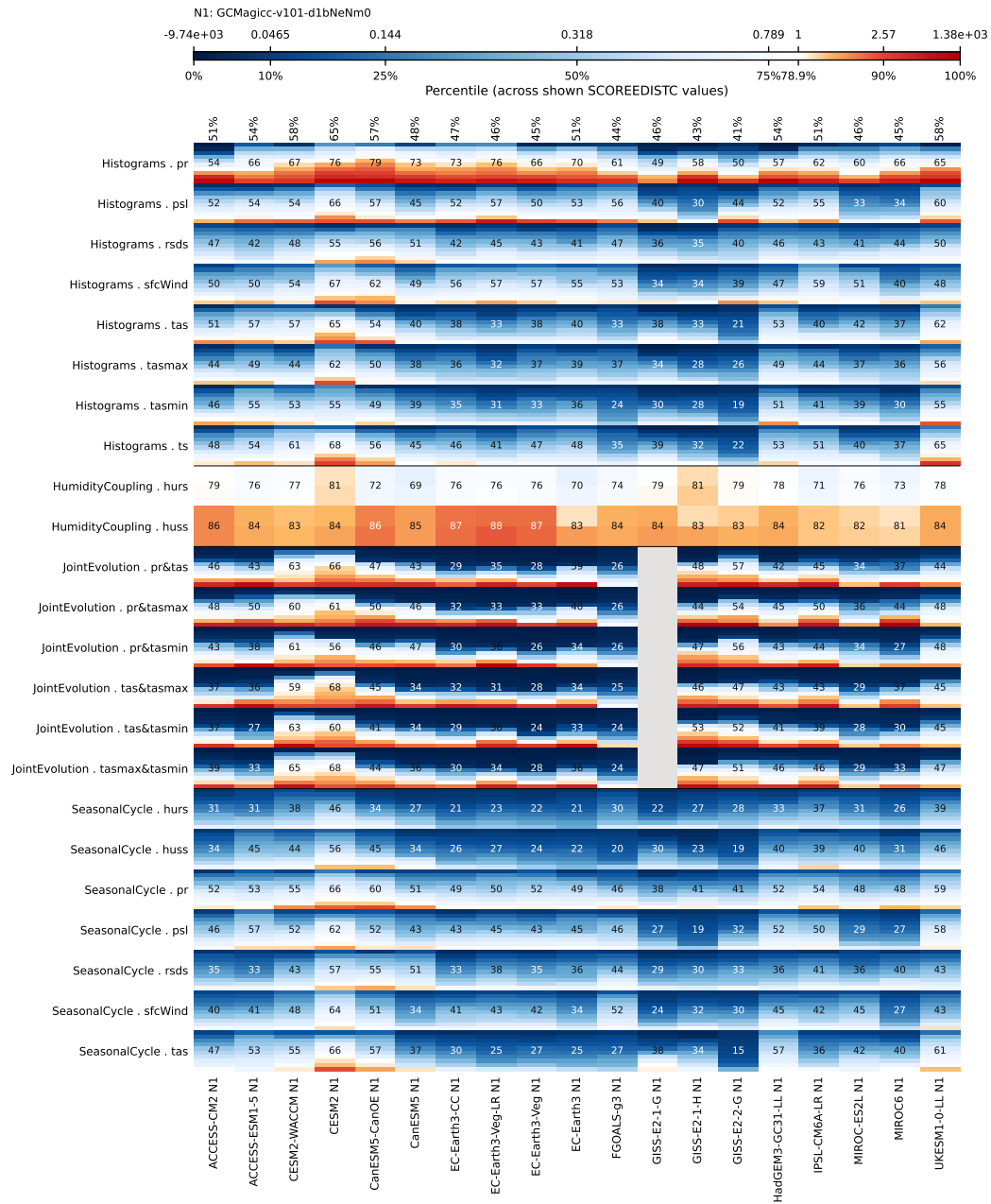


Figure S23. Part 2 Energy distance scores for the SSP2-4.5 hold-out scenario. As Figure S22, continued.

SCOREEDISTC ssp245 only (part 3 of 3)

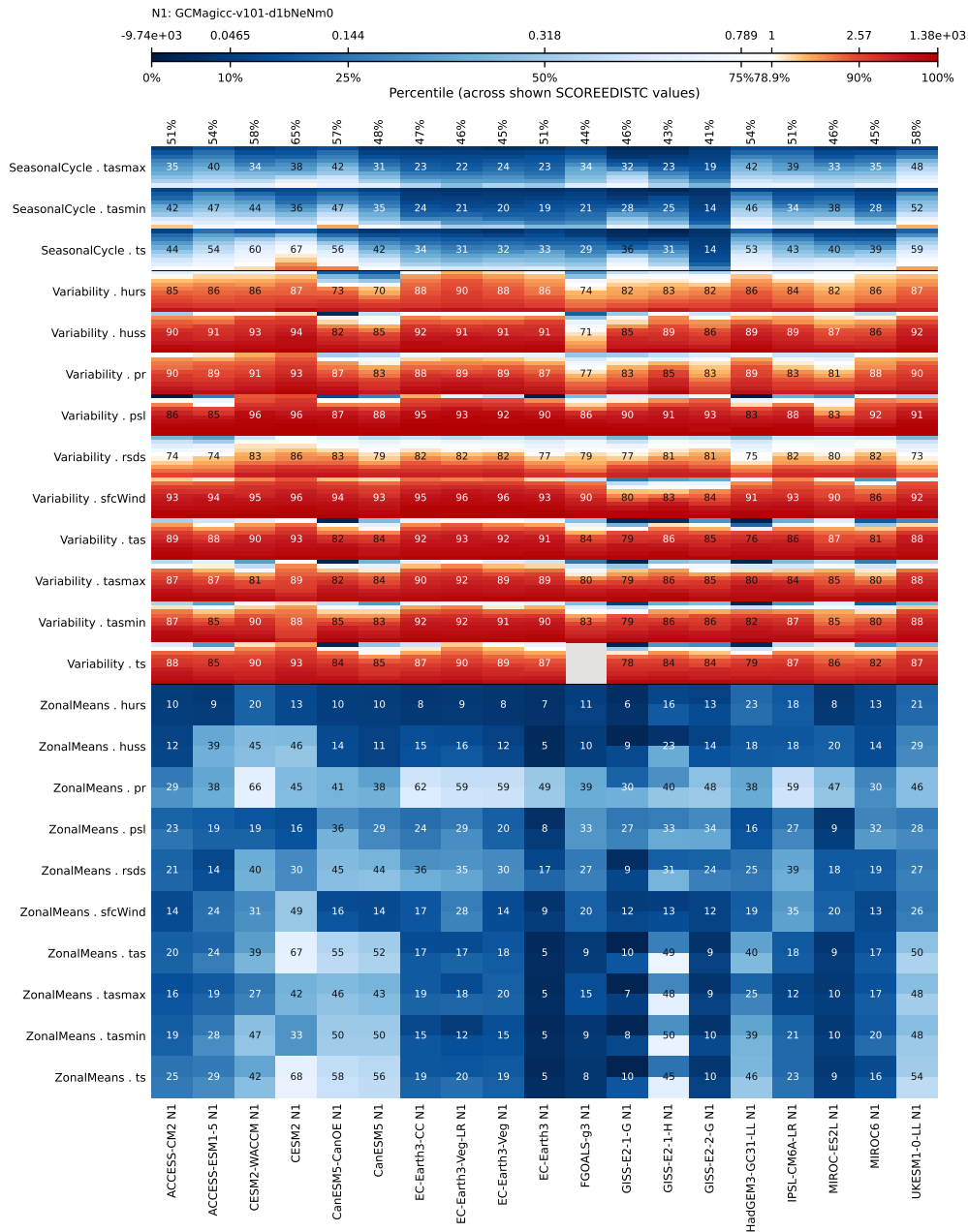


Figure S24. Part 3 Energy distance scores for the SSP2-4.5 hold-out scenario. As Figure S22, continued.

S5.1 Reduced emulator for global-mean warming

This section uses the GCMagicc-PM architecture (Methods, Table 2) operated at HEALPix resolution $n_{side} = 1$ only (12 equal-area pixels), giving effectively a global-mean representation. It retains the bias model b_1 and the stochastic generator g_1 described in the main-text Methods, but omits the common-effect model e_1 . That is, the prediction at $n_{side} = 1$ reduces to

$$335 \quad \hat{y}_t^{(1)} = b_1(x_t^{model}, x_t^{season}) + g_1(x_t, z_t), \quad (S7)$$

where z_t includes autoregressive lags with $\mathcal{L}_1 = \{1, \dots, 6\}$.

The forcing vector x_t retains all ERF components and the model index, month, and time covariates, but excludes the smoothed global-mean indicators *tas_smoothed* and *rtmt_smoothed*. These two covariates encode externally prescribed global temperature and radiative-imbalance trajectories from MAGICC; removing them forces the emulator to predict global-mean
340 warming from the raw ERF forcings alone, without anchoring to a pre-specified warming pathway. This makes the emulator’s warming prediction genuinely out-of-sample with respect to the global temperature trajectory.

S5.2 Leave-one-family-out procedure

The 32 CMIP6 source identifiers are partitioned into 13 model families defined by the first four characters of each model name (Table S8). For each family $\mathcal{F}_k$, the reduced emulator is retrained on data from all families $\mathcal{F}_{j \neq k}$ plus ERA5, together with the
345 historical runs of family $\mathcal{F}_k$ itself, from which its model-index embedding is fitted; no scenario run of $\mathcal{F}_k$ enters the training set. A separate emulator is also trained on ERA5 data alone (the ERA5 model).

We note explicitly that ERA5 is retained in the training data of every fold, and that the constrained predictions are produced by an emulator fitted to that same observational record. The cross-validation therefore measures how well the emulator generalises to a withheld *model* family, not how well it would predict a withheld real world: the residuals r_i quantify within-sample
350 spread with respect to the observational record rather than out-of-sample observational error. This section is accordingly presented as a limitation rather than as a constrained projection. Its purpose is to test whether the reduced emulator, once its own cross-validated uncertainty is accounted for, could deliver end-of-century ranges narrower than those assessed by IPCC AR6. It cannot.

To produce ERA5 warming estimates for each scenario, the ERA5-trained emulator is run on the same concatenated
355 historical-scenario forcing sequences used above, but with the model index set to $c_t = 0$ (ERA5). This gives the emulator’s best guess for how the real climate system would warm under each scenario’s forcing pathway.

Table S8. Model families used in leave-one-out cross-validation, defined by the first four characters of the CMIP6 source identifier.

Family prefix	Members
ACCE	ACCESS-CM2, ACCESS-ESM1-5
CanE	CanESM5, CanESM5-1, CanESM5-CanOE
CESM	CESM2, CESM2-WACCM
EC-E	EC-Earth3, EC-Earth3-AerChem, EC-Earth3-CC, EC-Earth3-Veg, EC-Earth3-Veg-LR
FGOA	FGOALS-g3
GFDL	GFDL-CM4, GFDL-ESM4
GISS	GISS-E2-1-G, GISS-E2-1-G-CC, GISS-E2-1-H, GISS-E2-2-G, GISS-E2-2-H
HadG	HadGEM3-GC31-LL, HadGEM3-GC31-MM
INM-	INM-CM4-8, INM-CM5-0
IPSL	IPSL-CM6A-LR, IPSL-CM6A-LR-INCA, IPSL-CM6A-MR1
MIRO	MIROC-ES2H, MIROC-ES2L, MIROC6
SAM0	SAM0-UNICON
UKES	UKESM1-0-LL

S5.3 Warming metric

Global-mean warming ΔT is defined following the IPCC AR6 convention in Table SPM.1 of IPCC AR6 Summary for Policy Makers (IPCC, 2021). Let $\bar{T}(a, b)$ denote the temporal mean of global-mean *tas* over the period $[a, b]$. Then

360
$$\Delta T = \bar{T}(2081, 2100) - \bar{T}(1995, 2014) + 0.85^\circ\text{C}, \tag{S8}$$

where 0.85°C is the IPCC AR6 central estimate of observed warming from the 1850–1900 pre-industrial baseline to the 1995–2014 reference period (IPCC, 2021). This shifts the anomaly so that ΔT is measured relative to 1850–1900. The global mean at *nside* = 1 is computed as the unweighted average over the 12 equal-area HEALPix pixels.

Four quantities are recorded for each model *c*, scenario, and stochastic replicate:

- 365
- ΔT_c^{true} : warming from the CMIP6 ground-truth *tas* of run 1,
 - $\Delta T_c^{\text{true}, 2\text{nd}}$: warming from a second independent run of the same model and scenario (if available; NaN otherwise),
 - ΔT_c^{hat} : warming from the leave-one-out emulator using the model’s forcing data and its own model index, whose em-
bedding was fitted on that model’s historical run only,
 - ΔT_c^{ERA} : warming from the ERA5-trained emulator using the model’s forcing data and ERA5 model index.

370 Each model-scenario pair is sampled with multiple stochastic replicates (3 per combination) to capture generator variability.

S5.4 Cross-validation scatter plots

Four panels present the cross-validation and its illustrative uncertainty consequence (Fig. S25).

Panel (a): observed versus observed. Each point plots $\Delta T_c^{\text{true},2\text{nd}}$ (x-axis) against ΔT_c^{true} (y-axis). Both axes show actual CMIP6 warming from two independent runs of the same model for the end of the 21st century (2081-2100). Points that lie close to the identity line indicate that the two runs agree well; scatter away from the diagonal reflects internal variability and ensemble spread. This panel serves as a baseline: it shows the irreducible inter-run variability that no emulator can improve upon.

Panel (c): leave-one-family-out estimate versus observed. Each point plots ΔT_c^{hat} (x-axis) against ΔT_c^{true} (y-axis). The x-axis now shows the emulator's leave-one-out prediction rather than a second CMIP6 run. Points near the identity line indicate that the emulator, despite never having seen any member of the held-out family, recovers that family's warming well. Short vertical rug ticks along the x-axis of panel c and in panel d mark the ERA5-trained emulator's warming predictions ΔT_c^{ERA} for the same scenario, summarising the range of emulated ERA5 warming across the available forcing sequences.

All 799 paired cases are shown in the scatter panels with low point opacity so that overlapping cases appear as density while individual model colours and markers remain visible. In panel (a), the dotted lines show one sample standard deviation of the paired difference $\Delta T^{\text{true}} - \Delta T^{\text{true},2\text{nd}}$ on either side of the identity line (0.24°C ; $n = 799$). The model legend uses higher-opacity marker examples for legibility.

Panel (b): scenario-specific illustrative distributions. Panel (b) shows illustrative warming distributions for SSP1-1.9, SSP1-2.6, SSP3-7.0, SSP4-3.4, SSP4-6.0 and SSP5-8.5, obtained by adding resampled leave-one-family-out residuals to the ERA5-conditioned GCMagicc-PM estimates by the procedure of Section S5.5. Nested intervals show the 2.5–97.5 %, 5–95 % and 10–90 % ranges, and the annotations give the 10th, 50th and 90th percentiles. These distributions carry the section's central negative result: their width is approximately 2.3°C and exceeds the corresponding IPCC WG1 AR6 assessed range for every scenario for which one is assessed except SSP5-8.5, where it is marginally narrower (2.32 against 2.40°C); SSP4-3.4 and SSP4-6.0 have no assessed range. The constraint therefore delivers no reduction in uncertainty except at the top of the scenario range, where the near-invariant width happens to fall just inside the widest assessed range.

Panel (d): individual ERA5-conditioned estimates. Panel (d) shows the individual ERA5-conditioned GCMagicc-PM warming estimates for each scenario, that is the set $\{e_j\}$ that enters the sampling in Eq. S9 before the residual pool is added.

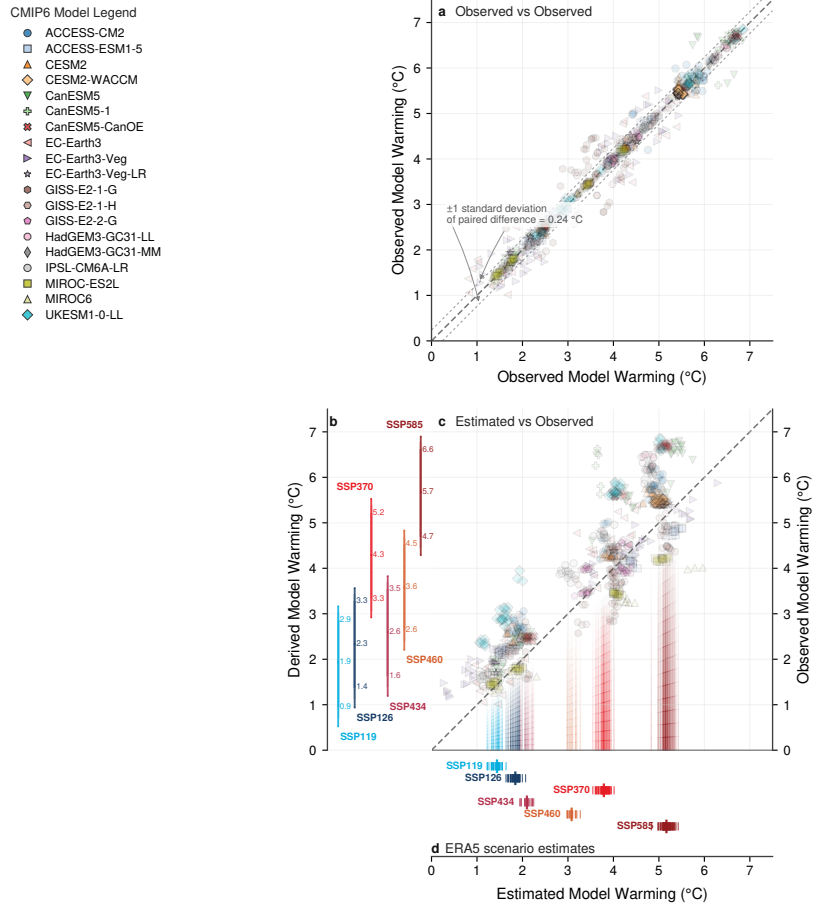


Figure S25. Leave-one-family-out evaluation and illustrative estimates of end-of-century global warming using GCMagicc-PM. (a) Agreement between warming simulated by two independent CMIP6 realisations for 799 paired cases from 19 models; colours and symbols identify CMIP6 models, the dashed line denotes the one-to-one relationship, and the dotted lines show ± 1 sample standard deviation of the paired difference (0.24 °C). (b) Scenario-specific illustrative warming distributions for SSP1-1.9, SSP1-2.6, SSP3-7.0, SSP4-3.4, SSP4-6.0, and SSP5-8.5, obtained by adding resampled leave-one-family-out residuals to ERA5-conditioned GCMagicc-PM estimates. Nested intervals show the 2.5–97.5 %, 5–95 %, and 10–90 % ranges; annotations give the 10th, 50th, and 90th percentiles. (c) Leave-one-family-out GCMagicc-PM estimates versus the corresponding CMIP6-simulated warming; the dashed line denotes the one-to-one relationship and the scenario-coloured vertical plumes show ERA5-conditioned estimates. (d) Individual ERA5-conditioned GCMagicc-PM warming estimates for each scenario. Warming is expressed relative to 1850–1900 and constructed as the 2081–2100 mean minus the 1995–2014 mean plus 0.85 °C, the IPCC AR6 central estimate of warming in 1995–2014 relative to 1850–1900 (IPCC, 2021). The illustrative distributions use 2,000 independent draws with replacement of an ERA5-conditioned estimate and a pooled residual.

S5.5 Width implied by the GCMagicc vs CMIP6 cross-validation residuals

The scatter plots establish that the leave-one-out emulator captures inter-model warming differences. To express this as the width the emulator implies for real-world warming (here proxied by ERA5), we adopt the emergent-constraint sampling procedure described below.

Let $r_i = \Delta T_i^{\text{true}} - \Delta T_i^{\text{hat}}$ denote the cross-validated residual for observation i , pooled across all models, scenarios, and replicates in the SSP trends data. Let $\{e_j\}$ denote the set of ERA5-trained emulator warming predictions ΔT_j^{ERA} for a given target scenario. The residuals capture the typical discrepancy between the emulator’s warming prediction and the true GCM warming, while the ERA5 predictions give the emulator’s best guess for observed warming.

To construct an illustrative resampling distribution for warming under the target scenario, we draw $N_{\text{sim}} = 2,000$ samples

$$\widetilde{\Delta T}_n = e_{j_n} + r_{i_n}, \quad j_n \sim \text{Uniform}\{e_j\}, \quad i_n \sim \text{Uniform}\{r_i\}, \tag{S9}$$

where j_n and i_n are drawn independently with replacement. Each sample $\widetilde{\Delta T}_n$ combines a random ERA5 prediction with a random residual, so that the distribution of $\widetilde{\Delta T}$ reflects both the emulator’s uncertainty about ERA5 warming and the systematic prediction error observed across the model ensemble. Because r_i is computed from single generator draws and e_j is itself a single draw, the generator’s sampling variance enters the width twice; the generator’s sampling variance therefore inflates the width relative to the emulator-implied uncertainty. A second limitation is that the procedure does not capture the warm bias of residuals for low scenarios and resulting illustrative uncertainty ranges for the low scenarios are spuriously low.

From the N_{sim} samples, we extract the quantiles listed in Table S9. These define symmetric resampling ranges: the 5%–95% range spans 90% of the samples, the 2.5%–97.5% range 95%, and so on.

Table S9. Emergent-constraint prediction quantiles for global-mean warming ΔT (°C above 1850–1900) under six SSP scenarios, period 2081–2100. Obtained from $N_{\text{sim}} = 2,000$ Monte Carlo samples combining ERA5 emulator predictions with cross-validated residuals. These quantiles demonstrate that the emulator’s own cross-validated uncertainty is not smaller than the AR6 assessed ranges; they are not offered as calibrated prediction intervals for real-world warming. The residual pool is not stratified by scenario, so the 5–95 % width is near-invariant across scenarios (2.30–2.32 °C here) by construction, whereas the assessed AR6 widths widen with forcing. Read the table as an illustrative bound on achievable uncertainty rather than as a scenario-resolved projection, noting that the double-counted generator variance biases it up. Two consequences follow: the intervals are too wide for the low scenarios — the SSP1-1.9 5 % bound of 0.71 °C lies below already-observed warming — and the outermost 2.5 % and 97.5 % quantiles are reported at a precision finer than $N_{\text{sim}} = 2,000$ samples resolve.

Quantile	SSP1-1.9	SSP1-2.6	SSP3-7.0	SSP4-3.4	SSP4-6.0	SSP5-8.5
2.5%	0.52	0.94	2.92	1.19	2.21	4.29
5%	0.71	1.14	3.07	1.39	2.38	4.46
10%	0.95	1.39	3.32	1.63	2.65	4.69
90%	2.86	3.28	5.21	3.54	4.54	6.61
95%	3.03	3.44	5.39	3.70	4.69	6.78
97.5%	3.16	3.57	5.53	3.83	4.83	6.90

415 For instance, under SSP5-8.5 the 5–95% range of the resampled distribution spans $[4.46, 6.78]^{\circ}\text{C}$ and under SSP1-2.6 $[1.14, 3.44]^{\circ}\text{C}$; these are emulator-implied widths, not calibrated prediction intervals for real-world warming, and the SSP5-8.5 lower bound lies above both the AR6 assessed median and this study’s own SSP5-8.5 ensemble median, which shows how coarsely the unstratified residual pool locates the distribution.

S5.6 Summary

420 The leave-one-family-out cross-validation demonstrates that the reduced $n_{\text{side}} = 1$ emulator, operating without the common-effect model and without externally prescribed global temperature trajectories, captures the spread in end-of-century warming across CMIP6 model families from ERF forcings alone. Combining the same emulator’s ERA5-forced estimates with the cross-validated residual distribution yields illustrative intervals that are as wide as, or wider than, the corresponding IPCC AR6 assessed ranges for every scenario with an assessed range except SSP5-8.5, where the near-invariant width is marginally
425 narrower. The purpose of this exercise is precisely that negative result: GCMagicc-PM does not deliver a reduction in end-of-century warming uncertainty relative to the assessed ranges, and the projections in this paper should not be read as offering one. Note that our main application GCMagicc workflow has a different setup. It uses the MAGICC simple climate model to impose IPCC AR6 calibrated ensembles for GCMagicc, hence - by construction - the IPCC AR6 assessed uncertainty distributions for *tas* are honoured.

430 S6 Observationally and IPCC AR6 aligned projections

S6.1 Data table for Figure 4

Table S10 tabulates the values plotted in panel b, and Table S11 decodes the CMIP6 line numbers of panels c1 and c2 (24 of the 32 CMIP6 models; the eight not shown are those without the ensemble coverage the panel requires).

S6.2 New scenarios: Estimating NDC and LT-LEDS future emissions

435 **Country-level NDC and LT-LEDS emission quantification.** Country-level greenhouse gas (GHG) emission trajectories were constructed in a bottom-up fashion for all 196 countries using officially submitted inventory data, NDCs, and LT-LEDS. Historical emission data were drawn from the PRIMAP-hist dataset (Gütschow et al., 2025, 2016) to anchor national time series, while NDC targets, expressed as absolute emissions levels, reductions relative to historical or BAU levels, emissions-intensity targets, per-capita targets, or multi-year budgets, were implemented while taking into account the gas-by-gas and sectoral
440 coverage of those targets. Emissions from sectors and gases not covered by a country’s NDC were assumed to follow NGFS GCAM Current Policies growth rates (Richters et al., 2024), with LULUCF projections treated separately where reported data were available. For countries with long-term targets, including net-zero CO₂ or GHG targets, emissions were linearly interpolated between the 2030 or 2035 NDC level (whichever the latest NDC target year by a country) and the stated long-term target year.

445 **Infilling, harmonisation, and gas-by-gas extension to 2100.** To construct complete, gas-by-gas global emission pathways extending to 2100, the country-aggregated GHG totals were complemented, harmonised and infilled. International aviation and shipping emissions were added to aggregated country emissions. LULUCF emissions were adjusted to approximate directly induced anthropogenic sinks (as countries report a mix of directly anthropogenic and non-anthropogenic LULUCF sinks), in line with IPCC AR6 assessed GHG emission scenarios, by subtracting indirectly induced LULUCF CO₂ emissions derived
450 from the NGFS downscaled GCAM pathways(Richters et al., 2024). The resulting global GHG emissions were harmonised via the Aneris python package (Gidden et al., 2018) and infilling/extension algorithms (via Silicone) (Lamboll et al., 2020). Beyond 2060, an adapted equal-quantile-walk approach was applied using the IPCC AR6 scenario database. Also, in order to split up the GHG emissions into separate individual gases (CO₂, CH₄, N₂O, and others) in a manner consistent with the characteristics of that IPCC AR6 scenario ensemble, we used an adapted equal quantile walk approach. This yielded complete,
455 multi-gas emission pathways from the present day to 2100 for each scenario variant. For more details on the NDC and LT-LEDs quantification methods, see Meinshausen et al. (2022).

S6.3 Extended data figure for observational matches of GCMagicc and CMIP6 models

The CMIP6 subset shown differs between evaluation displays (Fig. 4d and Figs. S26, S27 and S28) because a model is included in a display only where the frozen evaluation database provides every diagnostic–variable row that the display shows; the
460 remaining training models lack at least one row, sometimes due to lack of raw data for the investigated scenarios. The specific models entering each display can be read from that display’s own axis labels and legend.

Scenario	IPCC AR6			CMIP6			GCMagicc n20			n100 AR6			n100 CMIP7-ScenarioMIP		
	p5	med	p95	p5	med	p95	p5	med	p95	p5	med	p95	p5	med	p95
ssp119	1.00	1.40	1.80	0.99	1.63	2.35	1.21	1.44	1.90	1.21	1.44	1.85	1.22	1.45	1.94
ssp126	1.30	1.80	2.40	1.46	2.29	2.96	1.45	1.82	2.42	1.48	1.81	2.40	1.50	1.88	2.52
ssp245	2.10	2.70	3.50	2.25	3.09	4.07	2.25	2.71	3.66	2.21	2.71	3.60	2.23	2.79	3.80
ssp370	2.80	3.60	4.60	3.21	4.30	5.66	3.17	3.78	4.75	3.07	3.71	4.77	3.11	3.76	4.95
ssp434	–	–	–	1.63	2.53	3.77	1.89	2.19	2.71	1.88	2.19	2.84	1.89	2.24	2.88
ssp460	–	–	–	2.17	3.25	4.68	2.58	3.07	4.16	2.57	3.09	4.10	2.57	3.18	4.12
ssp534-over	–	–	–	1.38	2.45	3.67	1.81	2.16	2.95	1.78	2.19	2.99	1.75	2.26	3.12
ssp585	3.30	4.40	5.70	3.67	5.13	6.68	3.76	4.51	6.08	3.56	4.44	5.94	3.65	4.55	6.17
SSP scenarios	1.00	2.78	5.70	0.99	3.09	6.68	1.21	2.71	6.08	1.21	2.70	5.94	1.22	2.76	6.17
PA-all-l	–	–	–	–	–	–	1.45	1.66	2.19	1.44	1.66	2.08	1.44	1.71	2.21
PA-all-u	–	–	–	–	–	–	1.94	2.32	3.06	1.92	2.29	3.01	1.94	2.35	3.19
PA-woUS-l	–	–	–	–	–	–	1.55	1.77	2.13	1.49	1.78	2.35	1.51	1.81	2.52
PA-woUS-u	–	–	–	–	–	–	2.05	2.42	3.28	2.01	2.39	3.19	2.05	2.50	3.25
SSP2-com	–	–	–	–	–	–	1.67	1.99	2.82	1.65	2.04	2.78	1.66	2.07	2.80
CP_GCAM	–	–	–	–	–	–	2.42	2.87	3.76	2.40	2.83	3.67	2.38	2.92	3.79
CP_MSG	–	–	–	–	–	–	2.25	2.76	3.53	2.22	2.73	3.53	2.25	2.80	3.67
CP_RMD	–	–	–	–	–	–	2.40	2.94	3.98	2.36	2.91	3.90	2.38	2.99	4.06
NDCs + others	–	–	–	–	–	–	1.45	2.34	3.98	1.44	2.33	3.90	1.44	2.39	4.06
VL	–	–	–	–	–	–	1.17	1.56	2.07	1.22	1.54	2.07	1.21	1.59	2.21
LN	–	–	–	–	–	–	1.38	1.64	2.20	1.41	1.66	2.17	1.44	1.72	2.23
L	–	–	–	–	–	–	1.40	1.75	2.17	1.47	1.76	2.33	1.50	1.80	2.45
M	–	–	–	–	–	–	2.38	2.80	3.63	2.35	2.76	3.56	2.36	2.82	3.65
ML	–	–	–	–	–	–	2.00	2.34	2.96	1.97	2.33	3.05	1.97	2.40	3.17
HL	–	–	–	–	–	–	2.34	2.89	3.77	2.29	2.87	3.90	2.39	2.90	4.04
H	–	–	–	–	–	–	2.77	3.30	4.22	2.71	3.24	4.23	2.77	3.31	4.37
CMIP7 scenarios	–	–	–	–	–	–	1.17	2.33	4.22	1.22	2.31	4.23	1.21	2.36	4.37

Table S10. Values from Figure 4, panel b, relative to 1850–1900 in °C. Each bold category row reports the minimum p5, mean median, and maximum p95 over available scenarios; dataset coverage may differ. Values are changes from 1995–2014 to 2081–2100 plus the IPCC AR6 estimate of 0.85 °C warming from 1850–1900 to 1995–2014 (IPCC, 2021).

No.	CMIP6 model
1	ACCESS-CM2
2	ACCESS-ESM1-5
3	CESM2
4	CESM2-WACCM
5	CanESM5
6	CanESM5-1
7	CanESM5-CanOE
8	EC-Earth3
9	EC-Earth3-CC
10	EC-Earth3-Veg
11	EC-Earth3-Veg-LR
12	FGOALS-g3
13	GFDL-CM4
14	GFDL-ESM4
15	GISS-E2-1-G
16	GISS-E2-1-H
17	GISS-E2-2-G
18	HadGEM3-GC31-LL
19	INM-CM4-8
20	INM-CM5-0
21	IPSL-CM6A-LR
22	MIROC-ES2L
23	MIROC6
24	UKESM1-0-LL

Table S11. Line-number mapping for panels c1 and c2 of Fig. 4: the 24 CMIP6 models plotted there, numbered as in the panel. The eight remaining models of the 32 used elsewhere in this study are not plotted in these panels.

EDISTO all (part 1 of 3)

metrics.sqlite rows analysed: 649,058 (BiasMaps: 71,550; ENSOTelecon: 956; GlobalTimeseries: 4,760; Histograms: 276,080; HumidityCoupling: 952; SeasonalCycle: 276,080; Variability: 13,920; ZonalMeans: 4,760)

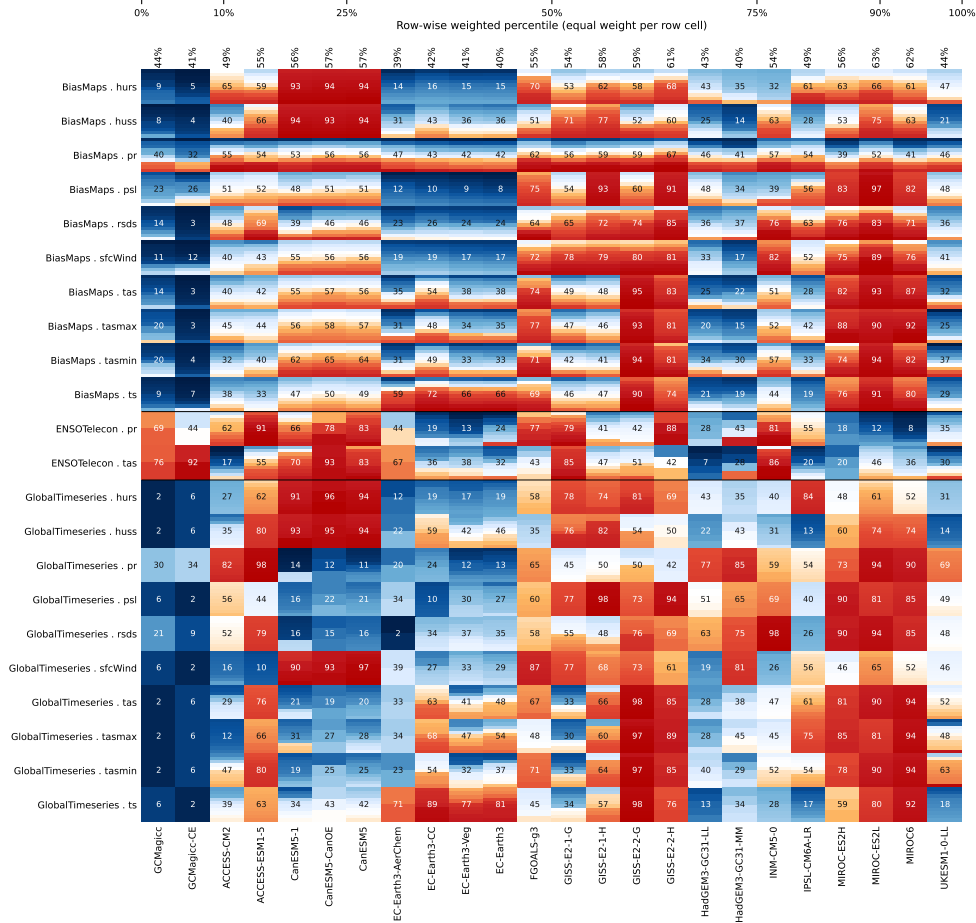


Figure S26. Observation-referenced energy-distance performance across diagnostics and variables (Part 1 of 3). Columns show two GCMagicc configurations - GCMagicc and GCMagicc-CE - followed by 22 CMIP6 models (the models with complete diagnostic coverage in the frozen evaluation database for all displayed diagnostic-variable rows). Rows represent combinations of diagnostic and climate variable. Part 1 contains BiasMaps, ENSOTelecon and GlobalTimeseries; Part 2 contains Histograms, HumidityCoupling and SeasonalCycle except surface temperature (*ts*); Part 3 contains SeasonalCycle-*ts*, Variability and ZonalMeans. Together, the panels summarise 649,058 evaluations in 64 diagnostic-variable rows and 24 columns in total (two GCMagicc configurations plus the 22 CMIP6 models). For the CMIP6 columns, the observation-referenced energy-distance measure is $EDISTO_C = 2GOF_{OC} - \overline{GOF_{CC}}$; for GCMagicc it is $EDISTO_N = 2GOF_{ON} - \overline{GOF_{NN}}$. Here, *O*, *C* and *N* denote the observational reference (ERA5), CMIP6 and GCMagicc, respectively, and the within-model terms correct for internal ensemble spread. Lower energy-distance values indicate a smaller discrepancy to observations.

EDISTO all (part 2 of 3)

metrics.sqlite rows analysed: 649,058 (BiasMaps: 71,550; ENSOTelecon: 956; GlobalTimeSeries: 4,760; Histograms: 276,080; HumidityCoupling: 952; SeasonalCycle: 276,080; Variability: 13,920; ZonalMeans: 4,760)

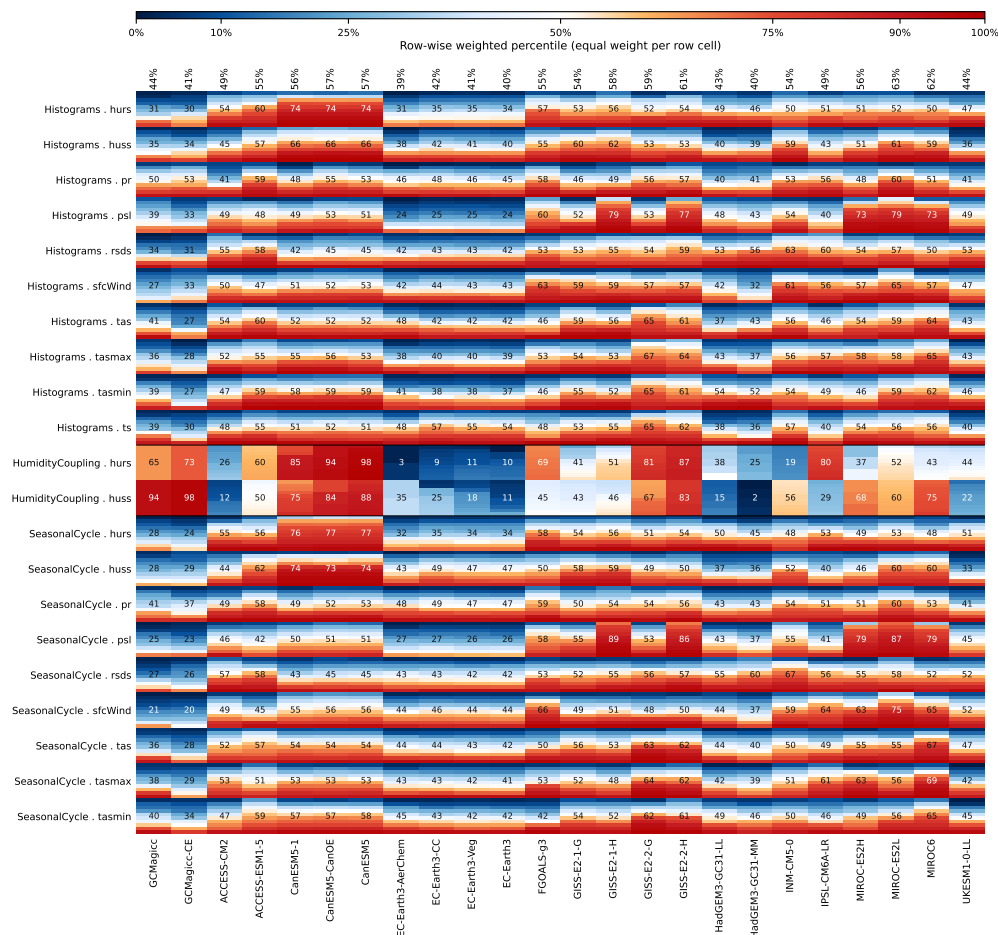


Figure S27. Observation-referenced energy-distance performance (continued; Part 2 of 3). Because the underlying diagnostics have different numerical scales, energy-distance values are converted separately within every diagnostic–variable row to weighted percentiles. Each model cell is assigned equal total weight, irrespective of how many underlying records it contains. Blue therefore denotes a relatively low energy-distance percentile and closer agreement with ERA5, white denotes approximately the row median, and red denotes a relatively high energy-distance percentile and larger discrepancy. Colours should consequently be compared between models within a row rather than interpreted as absolute differences between different diagnostics. Each cell is subdivided vertically into multiple horizontal colour bands. These bands represent the distribution of individual evaluations contributing to that diagnostic–variable–model combination, which can differ by diagnostic subtype, region, season or time window, source member and comparison member. Records are ordered from lower percentiles at the top to higher percentiles at the bottom in each cell. When ten or fewer records contribute, each record is displayed; for larger samples, ten representative mid-bin order statistics are shown.

EDISTO all (part 3 of 3)

metrics.sqlite rows analysed: 649,058 (BiasMaps: 71,550; ENSOTelecon: 956; GlobalTimeseries: 4,760; Histograms: 276,080; HumidityCoupling: 952; SeasonalCycle: 276,080; Variability: 13,920; ZonalMeans: 4,760)

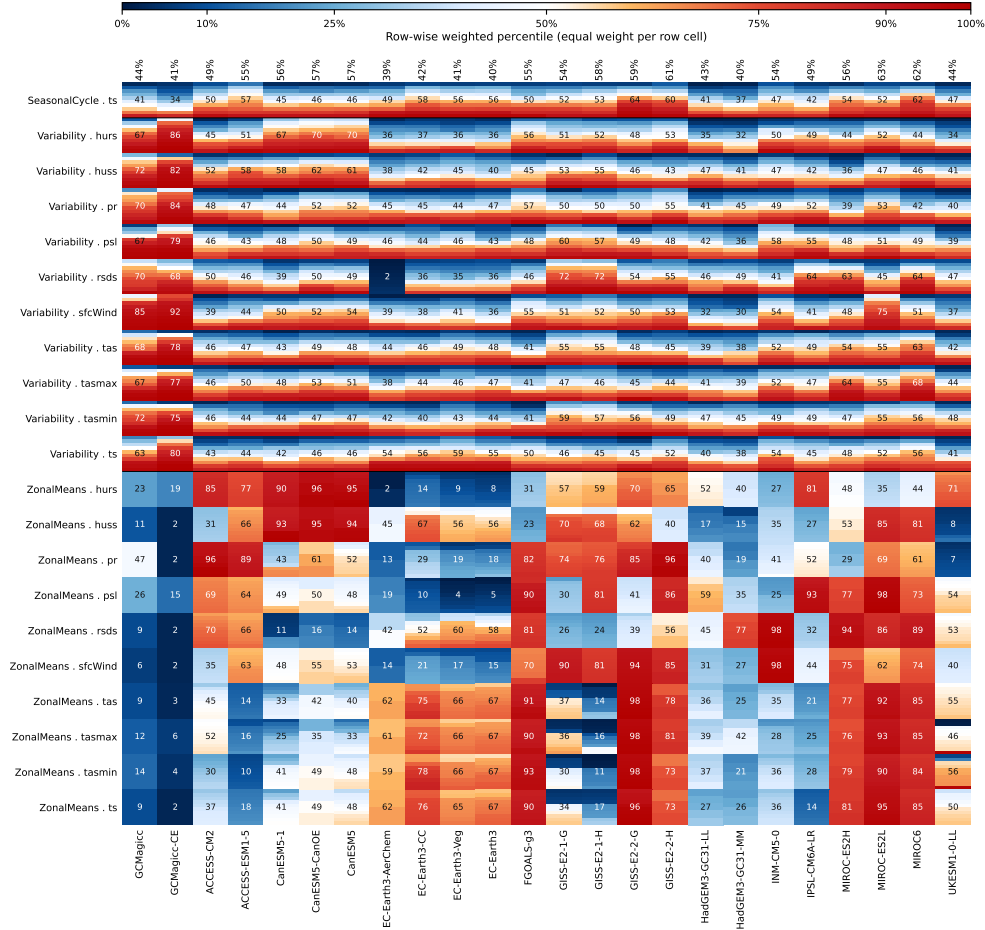


Figure S28. Observation-referenced energy-distance performance (continued; Part 3 of 3). The bands therefore expose within-cell heterogeneity that would be hidden by a single mean colour. The integer printed in each cell is the rounded mean percentile calculated from all contributing records, not the number of records. Percentages above the columns give the corresponding mean percentile across all displayed records for that model; lower percentages indicate better overall relative performance across the included rows. Thin horizontal separators delineate diagnostic groups. These files exclude the explicit oscillation-index variables NAO, SOI, Niño-3, Niño-3.4, Niño-4 and PDO, and therefore omit the ENSOCharacteristics and IndicatorFreq row groups; the precipitation- and temperature-based ENSOTelecon diagnostics remain included. The three parts differ only in which rows are displayed and share the same normalisation and colour scale. Because the visualisation is percentile-based, it describes relative model ordering and within-cell spread, not the absolute magnitude or statistical significance of energy distances or their differences.

S7 Suitability of existing climate emulators and ESM workflows for multivariate drought attribution

The drought-attribution and projection analyses presented in the main manuscript impose seven concurrent requirements on any candidate emulator or ESM-based workflow:

- 465 (i) joint emulation of at least the seven variables required to evaluate Penman–Monteith potential evapotranspiration (*tas*, *tasmin*, *tasmax*, *hurs* or *huss*, *sfcWind*, *rsds*, *pr*);
- (ii) free-running operation without prescribed sea-surface temperatures or ocean state, so that all-forcing and counterfactual natural-only forcing climates remain distinguishable;
- (iii) decomposed effective radiative forcing inputs (GHG, aerosol, solar, volcanic, ozone), so that natural-only (and potentially
470 aerosol-perturbed) counterfactuals can be constructed rather than only a CO₂ or total forcing implementation;
- (iv) a stochastic, generative formulation supporting $\mathcal{O}(100)$ independent realisations per scenario at low marginal cost, as needed for tail-event probability estimation;
- (v) temporally coherent multi-decadal output with stable low-frequency variability sufficient for SPEI-48, which accumulates 48 months of climatic water balance;
- 475 (vi) present-day observational alignment so that inferred event probabilities are anchored to the realised climate;
- (vii) applicability to new scenarios without re-training, since the most decision-relevant scenario space (CMIP7, NDC-aligned and current-policy pathways) lags both ESM simulation and any per-scenario ML retraining cycle.

To our knowledge, no published emulator or off-the-shelf workflow satisfies all seven requirements simultaneously (Table 1); this includes multi-resolution compound-hazard emulators such as MERCURY (Nath et al., 2025), which is not tabulated here.

- 480 To our knowledge, the best toolset at present is the conventional ESM-driven pipeline using CanESM5, MIROC6, and GISS-E2-1-G historical, natural-only, and SSP2-4.5 ensembles. Our inventory contains 65 / 50 / 50 historical / SSP2-4.5 / historical-natural CanESM5 members, 50 / 50 / 50 MIROC6 members, and 46 / 22 / 20 GISS-E2-1-G members after exclusion of one incomplete historical file. The quantitative comparison uses the monthly Penman–Monteith protocol in Section S8.2; a single-model large ensemble samples internal variability conditional on that model but does not estimate model-structural
485 uncertainty (see Section 5.5).

In the following, we briefly discuss the illustrative entries in our comparison table of different approaches, as shown in the main manuscript Table 1. The table is not exhaustive, but it illustrates the diversity of approaches and their relative strengths and weaknesses with respect to the seven requirements. It is far from a comprehensive review of these approaches, and we do not attempt to provide a complete list of all existing climate emulators.

490 S7.1 CMIP6 + statistical bias adjustment.

The conventional ESM-driven attribution workflow combines CMIP6 ensembles with statistical bias adjustment. ESMs provide the Penman–Monteith variables and multi-decadal simulations, but complete single-forcing archives, natural-only extensions,

and large ensembles are available for relatively few models. Bias adjustment can also modify multivariate dependence among PM drivers. Applicability to a new scenario remains conditional on an upstream ESM simulation, motivating the partial column 'New scen.' rating in Table 1.

S7.2 MESMER family.

MESMER (Beusch et al., 2020), MESMER-M (Nath et al., 2022), and MESMER-X (Quilcaille et al., 2022) are stochastic pattern-scaling (and, for MESMER-X, distribution-scaling) emulators trained per-ESM. MESMER and MESMER-X are conditioned on a global *tas* trajectory; MESMER-M is conditioned instead on local yearly temperatures. They produce stochastic ensembles via AR(1) processes with spatially correlated innovations (MESMER, MESMER-M) or draws from conditional generalized extreme value (GEV) distributions (MESMER-X), and are stable on multi-decadal timescales. Recent extensions add precipitation (MESMER-M-TP), fire-weather indices and soil moisture (extended MESMER-X (Quilcaille et al., 2023)), but the family does not produce the full Penman–Monteith driver set (*tasmin/tasmax*, *hurs/huss*, *sfcWind*, *rsds* are absent). Volcanic forcing enters via a stratospheric-AOD response correction (Beusch et al., 2020), which is a single-variable adjustment; the partial (◦) column 'Dec.ERF' rating reflects this limited form of forcing handling. Scenario transferability without retraining is most clearly established for MESMER-X (GMT-conditioned scaling); MESMER and MESMER-M flag full inter-scenario transfer as work-in-progress, but no retraining step is technically required to apply trained parameters to a new GMT trajectory.

S7.3 MESMER-MAGICC (Beusch et al., 2022).

Couples MAGICC to MESMER, so the global response is by construction tied to the IPCC AR6 calibration of MAGICC. The published chain is emission-driven and passes only the forced global-mean temperature change to MESMER - with explorations of adding ocean heat uptake as predictor (Beusch et al., 2022); the decomposed effective radiative forcings handled internally by MAGICC do not reach the regional emulator, hence the partial (◦) rating on column 'Dec.ERF'. Output remains restricted to MESMER's variable set (annual-mean *tas* in the published v0.8.3), and the spatial patterns inherit from the CMIP6 ESMs that MESMER was calibrated against. Recent additions include a combination of MAGICC and MESMER-M-TP to monthly heat and drought indicators, the latter being based on the Thornthwaite PET formulation (Thornthwaite, 1948) that depends on temperature and precipitation (Schöngart et al., 2025).

S7.4 STITCHES (Tebaldi et al., 2022).

The STITCHES approach 're-stitches' existing CMIP6 simulation segments (default nine-year windows) under user-specified GSAT trajectories. It inherits all archived variables and source-ESM bias. The nine-year window disrupts low-frequency variability modes such as NAO, PDO, and ENSO, and the method cannot extend to GSAT trajectories outside the envelope available in the archive. No retraining is required to construct new scenarios within that envelope, although scenarios with different

aerosol loadings would be considered the same as long as GSAT is identical, hence the partial (○) column 'New scen.' rating; the limitation is reach and forcing specific responses, not retraining overhead, given that there is no training step per se.

525 **S7.5 ClimaX (Nguyen et al., 2023).**

A foundation model with weather and climate configurations, pretrained on five CMIP6 sources and fine-tuned on ERA5 for weather tasks. In its documented ClimateBench configuration, ClimaX uses the preceding ten years of CO₂, SO₂, BC, and CH₄ forcing fields and predicts annual-mean surface temperature, diurnal temperature range, precipitation, and 90th-percentile precipitation (Watson-Parris et al., 2022; Nguyen et al., 2023). Surface shortwave radiation, surface wind, and humidity are
530 not outputs of this configuration. The published ClimateBench implementation is deterministic, rather than emphasizing the distributional match.

S7.6 Aurora (Bodnar et al., 2025).

A foundation model pre-trained on a six-dataset mixture (ERA5, CMCC, IFS-HR, HRES Forecasts, GFS Analysis, GFS Forecasts) and demonstrated for medium-range weather forecasts up to roughly ten days, plus fine-tuned variants for air quality,
535 ocean waves and tropical-cyclone tracks. Aurora's standard surface output set comprises 2 m temperature, 10 m wind components and mean sea-level pressure, omitting humidity, surface shortwave radiation, precipitation, and daily temperature extremes, which is why we assign the partial (○) rating in the column 'Penman-M vars'. The published paper does not document climate-scale free-running stability or a mechanism for accepting time-varying greenhouse-gas, aerosol, or solar forcings. The model is deterministic in the strong sense – one forward pass yields a single forecast with no stochastic generator – and the
540 authors flag probabilistic ensembles as an open problem to be addressed either by a probabilistic fine-tuning or by ensembles of separately trained deterministic models; it therefore receives × on column 'Stoch'.

S7.7 cBottle (Brenowitz et al., 2025).

A generative diffusion-based foundation model trained on hourly ERA5 and ICON kilometre-scale simulations, conditioned on prescribed sea-surface temperature, solar position, and a dataset label. It samples atmospheric states and outputs temperature,
545 winds, pressure, precipitation, radiative/cloud variables, and humidity on atmospheric levels, but the primary release does not document a dedicated 2-m relative- or specific-humidity output. It does not accept anthropogenic forcing inputs or run as a continuous time-stepping simulator; Table 1 therefore treats its Penman–Monteith variable coverage as partial.

S7.8 NeuralGCM (Kochkov et al., 2024).

A hybrid neural–physical atmospheric model trained on 40 years of ERA5. In climate mode, it is run AMIP-style with prescribed sea-surface temperatures; multi-decadal stability is partially demonstrated, with 22 of 37 initial 40-year runs completing
550 stably. It supports stochastic ensembles via learned space-time-correlated Gaussian random fields. Prognostic variables are 3D

temperature, humidity components, winds, and surface pressure; the published standard output set does not include surface downwelling shortwave radiation.

S7.9 ACE2 (Watt-Meyer et al., 2025).

555 An atmosphere-only emulator trained on ERA5, with a 1000-year stable AMIP-style rollout demonstrated. ACE2 requires a prescribed sea-surface-temperature field; the only free-running configuration shown is a short-horizon (~ 10 -day) persisted-SST forecast mode, motivating the partial ($\circ$) column 'Free-run' rating in this isolated form rather than a flat $\times$. ACE2 accepts time-varying SST and a global-mean CO_2 scalar but no other separable anthropogenic forcing components. It is deterministic and outputs (at 6-hourly resolution) 2-m temperature, specific humidity, 10-m winds, surface pressure, surface temperature, 560 surface radiative fluxes, and precipitation.

S7.10 SamudrACE (Duncan et al., 2026).

Couples the ACE2 atmosphere–land emulator to the Samudra ocean emulator, providing a free-running coupled ML climate model. It was trained on a GFDL CM4 pre-industrial control simulation with constant forcing. Atmospheric output inherits the ACE2 variable set. Related coupled designs include ACE2-SOM, which couples ACE2 to a slab ocean and learns equilibrium 565 responses across prescribed CO_2 levels (Clark et al., 2025), and DLESyM, which couples learned atmosphere and ocean components for stable simulations of the observed climate (Cresswell-Clay et al., 2025); neither is scored as a separate row in Table 1.

S7.11 LUCIE-3D (Guan et al., 2025b).

A free-running three-dimensional atmospheric emulator trained on 30 years of ERA5, with optional prescribed-SST AMIP 570 mode. It accepts CO_2 as a forcing variable and demonstrates 40-year free-running rollouts. The LUCIE-3D forward model is deterministic – no stochastic perturbation is injected during the autoregressive rollout – but the original LUCIE (Guan et al., 2025a) is explicitly designed for $O(1000)$ -member ensembles, generated by stochastic weight averaging (an ensemble of parameter snapshots saved during extended stochastic-gradient-descent training) rather than by a stochastic internal-variability generator; in a chaotic rollout these parameter-space differences diverge into ensemble spread. Because this delivers large 575 ensembles suitable for tail-risk assessment, albeit through parameter-space spread and not demonstrated in the cited three-dimensional forced-response configuration, we assign the partial ($\circ$) column 'Stoch' rating rather than $\times$. Its output set comprises temperature, specific humidity, and horizontal wind components on log surface pressure levels, and total precipitation.

S7.12 DiffESM (Bassetti et al., 2024).

A 3D diffusion model that downscales monthly mean ESM output for temperature and precipitation only to daily values. 580 DiffESM is therefore not a self-contained climate emulator: it requires an existing ESM simulation as input, and inherits its forcing dependence, stability, and scenario coverage from the ESM output it is fed. Observational alignment is not directly

evaluated in the DiffESM paper (which compares only to held-out ESM members), so any present-day alignment is inherited rather than validated; the partial (○) column 'Obs' rating reflects this inherited-but-unvalidated status.

S7.13 Score-based ESM emulator (Bouabid et al., 2026).

585 This score-based diffusion emulator operates on an equal-area HEALPix mesh as in GCMagicc and learns the joint monthly distribution of near-surface air temperature, precipitation, relative humidity, and near-surface wind speed. It is trained separately for each of MPI-ESM1-2-LR, MIROC6, and ACCESS-ESM1-5 using large initial-condition ensembles, rather than being trained jointly across ESMs. Its four demonstrated outputs cover part of the Penman–Monteith driver set, but omit daily minimum and maximum temperature and surface downwelling shortwave radiation, motivating the partial (○) column 'Penman-M
590 vars' rating. The diffusion formulation draws independent samples conditioned on global-mean surface-temperature anomaly and calendar month, efficiently supporting large stochastic ensembles and reproducing cross-variable statistics and distribution tails; this supports ✓ on column 'Stoch'. However, it neither dynamically advances a climate state nor uses a prescribed or prognostic ocean state, and the authors explicitly note that its instantaneous formulation cannot generate temporally consistent samples. It therefore receives × for both 'Free-run' and 'Multi-dec.' and cannot directly supply temporally coherent 48-month
595 sequences for SPEI-48. The conditioning does not separate GHG, aerosol, solar, volcanic, or ozone forcing contributions, giving × for 'Dec.ERF'. Training and validation are against each parent CMIP6 ESM rather than observations, while bias-aware training or transfer learning is identified as future work, giving × for 'Obs'. The model successfully generates the held-out SSP2-4.5 and SSP3-7.0 pathways without retraining, supporting ✓ for 'New scen.'; this demonstrated transfer is interpolation within the SSP1-2.6 to SSP5-8.5 training range, requires an externally supplied global-temperature trajectory, and is not
600 claimed to extrapolate beyond that range. These task-specific ratings do not diminish the method's demonstrated strengths in multivariate distributional and tail fidelity, and they do not imply that GCMagicc has perfect temporal fidelity: GCMagicc's remaining low-frequency and spectral limitations are discussed in the main manuscript Limitations section.

S7.14 ArchesClimate-SSP (Clyne et al., 2026).

A flow-matching generative emulator of the IPSL-CM6A-LR model: a deterministic mean-state prediction is followed by a
605 stochastic residual, and the climate state is advanced autoregressively as a coupled atmosphere–ocean system (ocean temperature *thetao* is predicted rather than prescribed, motivating the free-running ✓ rating). Conditioning forcings comprise CO₂, CH₄ and N₂O, multi-species sulphate, nitrate and black-carbon aerosols, solar irradiance in six wavelength bands and a ten-level ozone profile, but no separable volcanic channel (assumed implicit in the aerosol fields); this is the most complete forcing decomposition among the machine-learning emulators considered here, yet the missing volcanic component motivates
610 the partial (○) column 'Dec.ERF' rating rather than a full ✓. Surface output includes *tas*, *pr*, *evspsbl* and *uas/vas* (so *sfcWind* is derivable), with near-surface humidity inferable from the lowest level of the three-dimensional *hus* field; *tasmin/tasmax* and surface shortwave radiation (*rsds*) are absent, giving the partial (○) column 'Penman-M vars' rating on par with LUCIE-3D. The model is generative, but only five- to ten-member ensembles are demonstrated and the stochastic component is a residual around a deterministic mean, so the partial (○) column 'Stoch' rating is assigned rather than the large-ensemble ✓.

615 Free-running rollouts of 85 years (2015–2100) and a 100-year abrupt-4×CO₂ control are numerically stable, supporting the multi-decadal (✓) column ‘Multi-dec.’ rating; we note, however, that the out-of-distribution SSP5-8.5 scenario exhibits a systematic warm drift of order 2°C by 2100 and that an ozone-ablated variant destabilises, so this stability is best established within, rather than beyond, the training forcing envelope. Training and validation use the single ESM IPSL-CM6A-LR and the statistical emulator MESMER-M rather than observations or reanalysis, hence the × on column ‘Obs’. Scenario transfer is the
 620 paper’s headline result: trained on *historical*, *piControl*, SSP1-1.9, SSP1-2.6, SSP2-4.5, SSP3-7.0 and SSP4-6.0, it generates the held-out SSP4-3.4, SSP5-3.4 and SSP5-8.5 scenarios without retraining, supporting the ✓ on column ‘New scen.’.

S7.15 GCMagicc (this work).

The partial (○) column ‘Free-run’ rating reflects that the free-running claim holds for the coupled MAGICC+GCMagicc package: GCMagicc and GCMagicc-CE are conditioned on MAGICC global temperature and heat uptake, making them less free-
 625 running than GCMagicc-PM and GCMagicc-XS. The partial (○) column ‘Multi-dec.’ rating reflects that the amplitude and frequency spectra of the low-frequency modes are not yet at parity with the range across CMIP6 ESMs (main manuscript Limitations section). The ✓[†] on column ‘Obs.’ records that alignment is imposed for global-mean *tas* by construction, while the regional and multivariate alignment is learned and evaluated separately (main manuscript Section 5.1).

S8 Drought attribution and drought projections

630 S8.1 Penman–Monteith SPEI-48

We quantify multi-year meteorological drought with the 48-month Standardised Precipitation Evapotranspiration Index (SPEI-48) (Vicente-Serrano et al., 2010; Beguería et al., 2014). Monthly climatic water balance is $WB_m = P_m - PET_m$ and its trailing accumulation is

$$WB_m^{(48)} = \sum_{j=0}^{47} WB_{m-j}. \quad (\text{S10})$$

635 Potential evapotranspiration is calculated only with the FAO-56 Penman–Monteith reference-evapotranspiration formulation (Allen et al., 1998; Penman, 1948; Monteith, 1965):

$$ET_0 = \frac{0.408 \Delta (R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34 u_2)}. \quad (\text{S11})$$

Here Δ is the slope of the saturation-vapour-pressure curve, R_n is net radiation, G is soil heat flux, γ is the psychrometric constant, T is air temperature, u_2 is wind speed, and $e_s - e_a$ is vapour-pressure deficit. Daily values are converted to monthly
 640 totals. The required monthly drivers are precipitation, mean/minimum/maximum air temperature, relative humidity, surface downwelling shortwave radiation, surface wind, and pressure.

For every product, SPEI is standardised separately for each calendar month and grid cell with a 1991–2010 baseline and a generalised logistic (GLO) distribution fitted by unbiased probability-weighted moments (Beguería et al., 2014; Hosking,

1990). The GCMagicc family fit is pooled across all ensemble members and is shared by its all-forcing and natural-only ensembles. Each SMILE uses a fit pooled across all eligible historical members of that model and applies the same fitted parameters to every all-forcing and natural-only member. Surface downwelling shortwave radiation is used without an ERA5 offset. Grid-cell SPEI is aggregated over Iran with cosine-latitude weights and finite-point renormalisation:

$$\overline{SPEI}_R(t) = \frac{\sum_{i \in R} SPEI_i(t) \cos(\varphi_i)}{\sum_{i \in R} \cos(\varphi_i)}. \quad (\text{S12})$$

S8.2 Monthly below-threshold probability

For each ERA5 year y and calendar month m in 2021–2025, probability is estimated from the ensemble-member cross-section:

$$\hat{p}_{y,m} = \frac{1}{M} \sum_{i=1}^M \mathbb{I} \left[\overline{SPEI}_{y,m}^{(i)} \leq \overline{SPEI}_{y,m}^{ERA5} \right]. \quad (\text{S13})$$

Conditional on the model and forcing state, the M members are the primary independent sampling units. The analysis therefore yields 60 same-year/month probabilities for 2021–2025. For the 2041–2060 diagnostic, each of the same 60 ERA5 thresholds is compared with the distribution for that calendar month pooled across the 20 future years and all members. In addition, the pooled comparison estimates the fraction of all modeled values in each window at or below the latest ERA5 threshold from December 2025. A zero count is reported as an estimated probability of 0%; the empirical resolution of an estimate is $1/n$, and zero is not evidence that the event is physically impossible.

Table S12 reports the pooled latest-threshold probabilities and the medians and descriptive 5–95% ranges over the 60 month-matched probabilities. These ranges represent seasonal and interannual variation and are not confidence intervals. SPEI-48 autocorrelation does not invalidate a probability calculated across independent members at a fixed year and month. It does mean, however, that neither the 60 diagnostics nor the nominal $60M$ recent values (or $20M$ values per future calendar month) can be treated as independent observations. Any inferential resampling must therefore resample members as complete trajectories, preserving all within-member autocorrelation.

The recent comparison is structurally model-dependent. GCMagicc and CanESM5 show higher all-forcing than natural-only probabilities, whereas MIROC6 and GISS-E2-1-G show little or no all-forcing–natural-only separation for current times, while presenting that separation for the future. All three SMILEs show higher all-forcing drought probabilities in 2041–2060 than in 2021–2025, although the magnitude differs substantially among models.

S8.3 SMILE inventory and repeated natural-only tails

The live inventory contains 65 / 50 / 50 historical / SSP2-4.5 / historical-natural CanESM5 files, 50 / 50 / 50 MIROC6 files, and 46 / 22 / 20 GISS-E2-1-G files. All-forcing diagnostics use the 50, 50, and 22 SSP2-4.5 members, respectively. Natural-only CanESM5 and MIROC6 output ends in 2020 and GISS-E2-1-G output ends in 2014. To display and diagnose later periods, each member’s final five complete years is repeated member-wise. Natural-only comparisons after the observed endpoints are

Table S12. Estimated probabilities of modeled SPEI48 being at or below ERA5 thresholds for Iran, Islamic Republic of. The pooled comparison uses the latest ERA5 value (2025-12; SPEI48=-1.871). Each estimate is the fraction of modeled values at or below the corresponding ERA5 threshold. The month-matched comparison reports the median [5th–95th percentile] across the 60 ERA5 months in 2021–2025.

Panel (window)	Product	Forcing state	Members	Probability at/below latest ERA5, % (count/ <i>N</i>)	Month-matched probability, % median [<i>P</i> ₅ – <i>P</i> ₉₅] (<i>n</i>)
d (2021-2025)	GCMagicc	Factual	100	11.25% (675/6,000)	29.00% [13.00–57.05] (n=100)
d (2021-2025)	GCMagicc	Natural-only	100	0.35% (21/6,000)	0.00% [0.00–7.05] (n=100)
d (2021-2025)	CanESM5	Factual	50	28.10% (843/3,000)	44.00% [26.00–76.10] (n=50)
d (2021-2025)	CanESM5	Natural-only	50	0.00% (0/3,000)	0.00% [0.00–0.00] (n=50)
d (2021-2025)	MIROC6	Factual	50	0.37% (11/3,000)	2.00% [0.00–18.00] (n=50)
d (2021-2025)	MIROC6	Natural-only	50	0.63% (19/3,000)	2.00% [0.00–30.00] (n=50)
d (2021-2025)	GISS-E2-1-G	Factual	22	6.21% (82/1,320)	13.64% [8.86–45.45] (n=22)
d (2021-2025)	GISS-E2-1-G	Natural-only	20	7.08% (85/1,200)	15.00% [5.00–55.25] (n=20)
e (2041-2060)	GCMagicc	Factual	100	58.11% (13,947/24,000)	75.10% [58.17–91.41] (n=2,000)
e (2041-2060)	GCMagicc	Natural-only	100	0.30% (73/24,000)	1.40% [0.25–6.39] (n=2,000)
e (2041-2060)	CanESM5	Factual	50	82.49% (9,899/12,000)	91.20% [82.75–96.91] (n=1,000)
e (2041-2060)	CanESM5	Natural-only	50	0.00% (0/12,000)	0.00% [0.00–0.00] (n=1,000)
e (2041-2060)	MIROC6	Factual	50	9.69% (1,163/12,000)	24.55% [10.32–50.83] (n=1,000)
e (2041-2060)	MIROC6	Natural-only	50	0.63% (76/12,000)	2.40% [0.80–18.88] (n=1,000)
e (2041-2060)	GISS-E2-1-G	Factual	22	31.44% (1,660/5,280)	49.55% [31.07–76.16] (n=440)
e (2041-2060)	GISS-E2-1-G	Natural-only	20	7.08% (340/4,800)	16.00% [7.00–49.20] (n=400)

Notes. Probabilities are empirical fractions of modeled SPEI values less than or equal to the corresponding ERA5 threshold. For panel d, each monthly probability compares ERA5 with the ensemble-member cross-section at the same year and calendar month; for panel e, it compares ERA5 with all member-years in 2041–2060 for the same calendar month. A zero count is reported as an estimated probability of 0%; the empirical resolution of an estimate is 1/*n*. The 5th–95th percentile range describes month-to-month variation, not a confidence interval. Factual denotes SSP2-4.5 for the GCMagicc product and historical plus SSP2-4.5 for the SMILEs; natural-only denotes SSP2-4.5-nat for the GCMagicc product and hist-nat for the SMILEs. For the SMILE natural-only rows, the final five complete hist-nat years are repeated member-wise through 2100; all panel d/e values are therefore synthetic repetitions and add no temporal degrees of freedom. Nominal member-month counts are not independent sample sizes because values share ensemble trajectories and SPEI48 is autocorrelated.

675

explicitly flagged as synthetic tails and are interpreted as stationary diagnostic extensions, not future natural-forcing simulations.

S8.4 Predictor sensitivity and SMILE extension

The predictor-sensitivity experiment is shown at the bottom of Fig. S29. For each member, one case leaves all drivers evolving, one clamps all drivers after 2001, six clamp one driver group, and six allow only one group to evolve. For each single-member CMIP6 ESM, we compute a member-specific climatology and SPEI baseline over 1975–2000. The three added SMILE
680 columns instead apply each model’s shared pooled 1991–2010 historical GLO fit, raw *rsds*, and cosine-latitude regional weighting to every member and sensitivity case.

S8.5 Assumptions and limitations

SPEI-48 diagnoses multi-year atmospheric supply–demand anomalies rather than plant-available soil water. Results remain conditional on the 1991–2010 GLO fit, Penman–Monteith assumptions, reanalysis uncertainty, cosine-latitude aggregation,
685 forcing construction, and each model’s representation of internal variability and long-term change. The monthly cross-member estimator substantially limits the direct effect of strong SPEI-48 autocorrelation on each probability estimate, because members rather than adjacent months are the independent units.

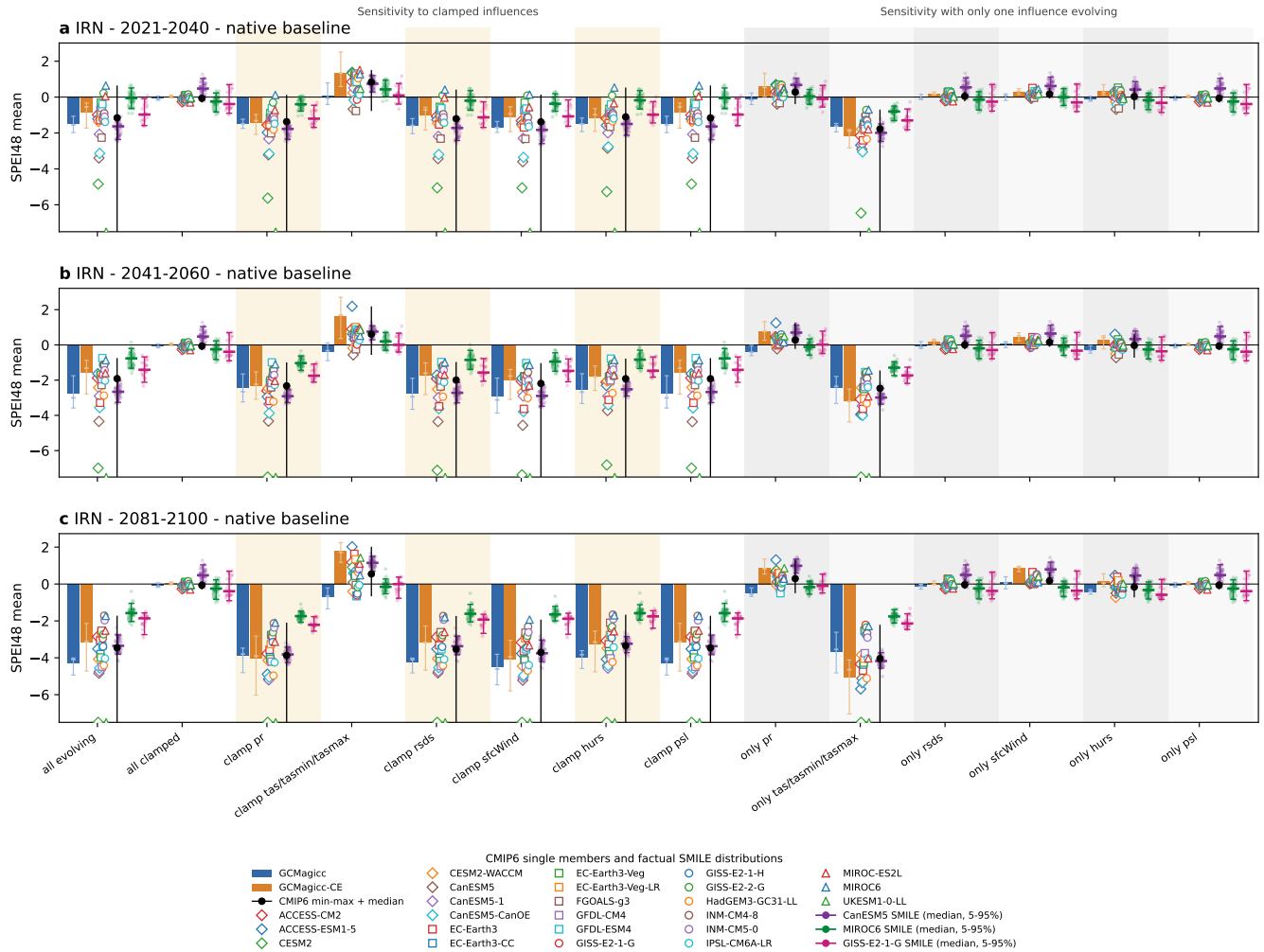


Figure S29. Penman–Monteith SPEI-48 sensitivity to evolving meteorological drivers. Iranian projections are summarised for 2021–2040, 2041–2060, and 2081–2100. Cases on the left clamp one driver group after 2001; cases on the right retain evolution in only one group. Existing GCMagicc summaries and single-member CMIP6 dots retain their native member-specific baseline treatment. Values beyond the plotted range are clipped to the axis floor at -7.5 . Clipping affects two models: UKESM1-0-LL in 21 of its 42 period–case cells, spanning outliers of -23.5 to -28.8 SPEI-48, and CESM2 in 9 cells, spanning -7.5 to -9.0 , possibly because of variability or limited robustness in the single-member baseline estimate. Each CMIP6 model is identified by a unique combination of outline colour and marker shape. CanESM5, MIROC6, and GISS-E2-1-G factual SMILE columns show faint member points, the ensemble median, and 5–95% whiskers, using each model’s shared pooled 1991–2010 historical GLO fit.

References

- Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration — Guidelines for computing crop water requirements, FAO
690 Irrigation and Drainage Paper 56, Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, ISBN 92-5-104219-5,
<https://www.fao.org/4/x0490e/x0490e00.htm>, FAO-56 (original 1998 release)., 1998.
- Arias, P. A., Bellouin, N., Coppola, E., Jones, R. G., Krinner, G., Marotzke, J., Naik, V., Palmer, M. D., Plattner, G.-K., Rogelj, J., Rojas,
M., Sillmann, J., Storelmo, T., Thorne, P. W., Trewin, B., Achuta Rao, K., Adhikary, B., Allan, R. P., Armour, K., Bala, G., Barimalala,
R., Berger, S., Canadell, J. G., Cassou, C., Cherchi, A., Collins, W., Collins, W. D., Connors, S. L., Corti, S., Cruz, F., Dentener, F. J.,
695 Dereczynski, C., Di Luca, A., Diongue Niang, A., Doblas-Reyes, F. J., Dosio, A., Douville, H., Engelbrecht, F., Eyring, V., Fischer, E.,
Forster, P., Fox-Kemper, B., Fuglestedt, J. S., Fyfe, J. C., Gillett, N. P., Goldfarb, L., Gorodetskaya, I., Gutierrez, J. M., Hamdi, R.,
Hawkins, E., Hewitt, H. T., Hope, P., Islam, A. S., Jones, C., Kaufman, D. S., Kopp, R. E., Kosaka, Y., Kossin, J., Krakovska, S., Lee,
J.-Y., Li, J., Mauritsen, T., Maycock, T. K., Meinshausen, M., Min, S.-K., Monteiro, P. M. S., Ngo-Duc, T., Otto, F., Pinto, I., Pirani,
A., Raghavan, K., Ranasinghe, R., Ruane, A. C., Ruiz, L., Sallée, J.-B., Samset, B. H., Sathyendranath, S., Seneviratne, S. I., Sörensson,
700 A. A., Szopa, S., Takayabu, I., Tréguier, A.-M., van den Hurk, B., Vautard, R., von Schuckmann, K., Zaehle, S., Zhang, X., and Zickfeld,
K.: Technical Summary, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment
Report of the Intergovernmental Panel on Climate Change*, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan,
C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K.,
Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., pp. 35–144, Cambridge University Press, Cambridge, UK and New York, NY, USA,
705 <https://doi.org/10.1017/9781009157896.002>, 2021.
- Bassetti, S., Hutchinson, B., Tebaldi, C., and Kravitz, B.: DiffESM: Conditional Emulation of Temperature and Precipitation in
Earth System Models With 3D Diffusion Models, *Journal of Advances in Modeling Earth Systems*, 16, e2023MS004194,
<https://doi.org/10.1029/2023MS004194>, 2024.
- Beguiría, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B.: Standardized precipitation evapotranspiration index (SPEI) revisited: pa-
710 rameter fitting, evapotranspiration models, tools, datasets and drought monitoring, *International Journal of Climatology*, 34, 3001–3023,
<https://doi.org/10.1002/joc.3887>, 2014.
- Beusch, L., Gudmundsson, L., and Seneviratne, S. I.: Emulating Earth system model temperatures with MESMER: from global mean
temperature trajectories to grid-point-level realizations on land, *Earth System Dynamics*, 11, 139–159, <https://doi.org/10.5194/esd-11-139-2020>, 2020.
- 715 Beusch, L., Nicholls, Z., Gudmundsson, L., Hauser, M., Meinshausen, M., and Seneviratne, S. I.: From emission scenarios to spatially
resolved projections with a chain of computationally efficient emulators: coupling of MAGICC (v7.5.1) and MESMER (v0.8.3), *Geosci-
entific Model Development*, 15, 2085–2103, <https://doi.org/10.5194/gmd-15-2085-2022>, 2022.
- Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Allen, A., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J. A., Dong, H., Gupta, J. K.,
Thambiratnam, K., Archibald, A. T., Wu, C.-C., Heider, E., Welling, M., Turner, R. E., and Perdikaris, P.: A Foundation Model for the
720 Earth System, *Nature*, 641, 1180–1187, <https://doi.org/10.1038/s41586-025-09005-y>, 2025.
- Bouabid, S., Souza, A. N., and Ferrari, R.: Score-Based Generative Emulation of Impact-Relevant Earth System Model Outputs, *Journal of
Advances in Modeling Earth Systems*, 18, e2025MS005558, <https://doi.org/10.1029/2025MS005558>, 2026.
- Boucher, O. et al.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 CFMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.1522>, downloaded 1 July 2025, 2018a.

- 725 Boucher, O. et al.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1534>, downloaded 1 July 2025, 2018b.
- Boucher, O. et al.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 DAMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.13801>, downloaded 1 July 2025, 2018c.
- Boucher, O. et al.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1532>, downloaded 1 July 2025, 2019.
- 730 Boucher, O. et al.: IPSL IPSL-CM6A-LR-INCA model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.13582>, downloaded 1 July 2025, 2021.
- Boucher, O. et al.: IPSL IPSL-CM6A-MR1 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.15922>, downloaded 1 July 2025, 2023.
- 735 Brenowitz, N. D., Ge, T., Subramaniam, A., Manshausen, P., Gupta, A., Hall, D. M., Mardani, M., Vahdat, A., Kashinath, K., and Pritchard, M. S.: Climate in a Bottle: Towards a Generative Foundation Model for the Kilometer-Scale Global Atmosphere, arXiv preprint arXiv:2505.06474, <https://doi.org/10.48550/arXiv.2505.06474>, preprint, 2025.
- Clark, S. K., Watt-Meyer, O., Kwa, A., McGibbon, J., Henn, B., Perkins, W. A., Wu, E., Harris, L. M., and Bretherton, C. S.: ACE2-SOM: Coupling an ML Atmospheric Emulator to a Slab Ocean and Learning the Sensitivity of Climate to Changed CO₂, Journal of Geophysical Research: Machine Learning and Computation, 2, e2024JH000 575, <https://doi.org/10.1029/2024JH000575>, 2025.
- 740 Clyne, G., Kaltenborn, J., Nowack, P., Monteleoni, C., and Charantonis, A.: Emulating the Forced Response of Climate Models with Flow Matching, arXiv preprint arXiv:2605.16929, <https://doi.org/10.48550/arXiv.2605.16929>, preprint, 2026.
- Cole, J. N. et al.: CCCma CanESM5 model output prepared for CMIP6 CFMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1302>, downloaded 1 July 2025, 2019.
- 745 Copernicus Climate Change Service (C3S): ERA5 monthly averaged data on single levels from 1940 to present, Copernicus Climate Change Service (C3S) Climate Data Store, <https://cds.climate.copernicus.eu/>, <https://doi.org/10.24381/cds.f17050d7>, last access: 7 April 2026, 2019.
- Cresswell-Clay, N., Liu, B., Durran, D., Liu, Z., Espinosa, Z. I., Moreno, R., and Karlbauer, M.: A Deep Learning Earth System Model for Efficient Simulation of the Observed Climate, AGU Advances, 6, e2025AV001 706, <https://doi.org/10.1029/2025AV001706>, 2025.
- 750 Danabasoglu, G.: NCAR CESM2 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2185>, downloaded 1 July 2025, 2019a.
- Danabasoglu, G.: NCAR CESM2 model output prepared for CMIP6 DAMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2187>, downloaded 1 July 2025, 2019b.
- Danabasoglu, G.: NCAR CESM2 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2201>, downloaded 1 July 2025, 2019c.
- 755 Danabasoglu, G.: NCAR CESM2-WACCM model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.10024>, downloaded 1 July 2025, 2019d.
- Danabasoglu, G.: NCAR CESM2-WACCM model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.10026>, downloaded 1 July 2025, 2019e.
- 760 Dix, M. et al.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2281>, downloaded 1 July 2025, 2019a.

- Dix, M. et al.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2285>, downloaded 1 July 2025, 2019b.
- Dix, M. et al.: CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 DAMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.14361>, downloaded 1 July 2025, 2020.
- 765 Duncan, J. P. C., Wu, E., Dheeshjith, S., Subel, A., Arcomano, T., Clark, S. K., Henn, B., Kwa, A., McGibbon, J., Perkins, W. A., Gregory, W., Fernandez-Granda, C., Busecke, J., Watt-Meyer, O., Hurlin, W. J., Adcroft, A., Zanna, L., and Bretherton, C.: SamudrACE: Fast and Accurate Coupled Climate Modeling with 3D Ocean and Atmosphere Emulators, *Geophysical Research Letters*, 53, e2025GL119340, <https://doi.org/10.1029/2025GL119340>, 2026.
- 770 (EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth3 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.181>, downloaded 1 July 2025, 2019a.
- (EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth3 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.251>, downloaded 1 July 2025, 2019b.
- (EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth3-Veg model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.642>, downloaded 1 July 2025, 2019c.
- 775 (EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth3-Veg model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.727>, downloaded 1 July 2025, 2019d.
- (EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth3-AerChem model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.639>, downloaded 1 July 2025, 2020a.
- 780 (EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth3-AerChem model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.724>, downloaded 1 July 2025, 2020b.
- (EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth-3-CC model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.640>, downloaded 1 July 2025, 2020c.
- (EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth3-Veg-LR model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.643>, downloaded 1 July 2025, 2020d.
- 785 (EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth3-Veg-LR model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.728>, downloaded 1 July 2025, 2020e.
- (EC-Earth), E.-E. C.: EC-Earth-Consortium EC-Earth3-CC model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.15327>, downloaded 1 July 2025, 2021.
- 790 for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2.1G model output prepared for CMIP6 CFMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2061>, downloaded 1 July 2025, 2018a.
- for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2.1G model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1400>, downloaded 1 July 2025, 2018b.
- for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2.1G model output prepared for CMIP6 DAMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2062>, downloaded 1 July 2025, 2018c.
- 795 for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2.1H model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1421>, downloaded 1 July 2025, 2018d.
- for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2-1-G-CC model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.11657>, downloaded 1 July 2025, 2019a.

for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2.1H model output prepared for CMIP6 CFMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.13941>, downloaded 1 July 2025, 2019b.

for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2-2-G model output prepared for CMIP6 CFMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.11659>, downloaded 1 July 2025, 2019c.

for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2-2-G model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2081>, downloaded 1 July 2025, 2019d.

for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2.2H model output prepared for CMIP6 CFMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.17582>, downloaded 1 July 2025, 2019e.

for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2.2H model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.15861>, downloaded 1 July 2025, 2019f.

for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2.1G model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2074>, downloaded 1 July 2025, 2020a.

for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2.1H model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2080>, downloaded 1 July 2025, 2020b.

for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2-2-G model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.11671>, downloaded 1 July 2025, 2021.

for Space Studies (NASA/GISS), N. G. I.: NASA-GISS GISS-E2-1-G-CC model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.17581>, downloaded 1 July 2025, 2022.

GCMagicc Development Team: GCMagicc software and documentation, <https://gcmagicc.org>, last access: 7 April 2026, 2026.

Gidden, M. J., Fujimori, S., van den Berg, M., Klein, D., Smith, S. J., van Vuuren, D. P., and Riahi, K.: A methodology and implementation of automated emissions harmonization for use in Integrated Assessment Models, *Environmental Modelling & Software*, 105, 187–200, <https://doi.org/10.1016/j.envsoft.2018.04.002>, 2018.

Good, P.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.10845>, downloaded 1 July 2025, 2019.

Good, P. et al.: MOHC UKESM1.0-LL model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1567>, downloaded 1 July 2025, 2019.

Guan, H., Arcomano, T., Chattopadhyay, A., and Maulik, R.: LUCIE: A Lightweight Uncoupled Climate Emulator With Long-Term Stability and Physical Consistency, *Journal of Advances in Modeling Earth Systems*, 17, e2025MS005152, <https://doi.org/10.1029/2025MS005152>, 2025a.

Guan, H., Arcomano, T., Chattopadhyay, A., and Maulik, R.: LUCIE-3D: A Three-Dimensional Climate Emulator for Forced Responses, *EGUsphere*, <https://doi.org/10.5194/egusphere-2025-4305>, preprint, in discussion for Geoscientific Model Development, 2025b.

Guo, H. et al.: NOAA-GFDL GFDL-CM4 model output, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1402>, downloaded 1 July 2025, 2018a.

Guo, H. et al.: NOAA-GFDL GFDL-CM4 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.9242>, downloaded 1 July 2025, 2018b.

Gütschow, J., Jeffery, L., Gieseke, R., Gebel, R., Stevens, D., Krapp, M., and Rocha, M.: The PRIMAP-hist national historical emissions time series, *Earth System Science Data*, 8, 571–603, <https://doi.org/10.5194/essd-8-571-2016>, 2016.

- Gütschow, J., Busch, D., and Pflüger, M.: The PRIMAP-hist national historical emissions time series (1750–2024) v2.7, Zenodo [data set], <https://doi.org/10.5281/zenodo.17090760>, 2025.
- Hajima, T. et al.: MIROC MIROC-ES2L model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set],
840 <https://doi.org/10.22033/ESGF/CMIP6.902>, downloaded 1 July 2025, 2019.
- Hosking, J. R. M.: L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics, *Journal of the Royal Statistical Society: Series B (Methodological)*, 52, 105–124, <https://doi.org/10.1111/j.2517-6161.1990.tb01775.x>, 1990.
- IPCC: Summary for Policymakers, in: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E.,
845 Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., pp. 3–32, Cambridge University Press, <https://doi.org/10.1017/9781009157896.001>, assessment published in 2021; version of record published by Cambridge University Press in 2023, 2021.
- Jackson, L.: MOHC HadGEM3-GC31-MM model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set],
850 <https://doi.org/10.22033/ESGF/CMIP6.10846>, downloaded 1 July 2025, 2020.
- John, J. G. et al.: NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1414>, downloaded 1 July 2025, 2018.
- Jones, G.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 DAMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.471>, downloaded 1 July 2025, 2019.
- 855 Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., Hatfield, S., Battaglia, P., Sanchez-Gonzalez, A., Willson, M., Brenner, M. P., and Hoyer, S.: Neural general circulation models for weather and climate, *Nature*, 632, 1060–1066, <https://doi.org/10.1038/s41586-024-07744-y>, 2024.
- Krasting, J. P. et al.: NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1407>, downloaded 1 July 2025, 2018.
- 860 Lamboll, R. D., Nicholls, Z. R. J., Kikstra, J. S., Meinshausen, M., and Rogelj, J.: Silicone v1.0.0: an open-source Python package for inferring missing emissions data for climate change research, *Geoscientific Model Development*, 13, 5259–5275, <https://doi.org/10.5194/gmd-13-5259-2020>, 2020.
- Li, L.: CAS FGOALS-g3 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1783>, downloaded 1 July 2025, 2019a.
- 865 Li, L.: CAS FGOALS-g3 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2056>, downloaded 1 July 2025, 2019b.
- Liu, L., Li, Z., Yang, X., Gong, H., Li, C., and Xiong, A.: The long-term trend in the diurnal temperature range over Asia and its natural and anthropogenic causes, *Journal of Geophysical Research: Atmospheres*, 121, 3519–3533, <https://doi.org/10.1002/2015JD024549>, 2016.
- MAGICC Development Team: MAGICC: The Model for the Assessment of Greenhouse-gas Induced Climate Change (downloads, documentation, and probabilistic distributions), <https://magicc.org>, last access: 7 April 2026, 2026.
- 870 Mall, R. K., Chaturvedi, M., Singh, N., Bhatla, R., Singh, R. S., Gupta, A., and Niyogi, D.: Evidence of asymmetric change in diurnal temperature range in recent decades over different agro-climatic zones of India, *International Journal of Climatology*, 41, 2597–2610, <https://doi.org/10.1002/joc.6978>, 2021.

- Meinshausen, M., Raper, S. C. B., and Wigley, T. M. L.: Emulating Coupled Atmosphere-Ocean and Carbon Cycle Models with
875 a Simpler Model, *MAGICC6 – Part 1: Model Description and Calibration, Atmospheric Chemistry and Physics*, 11, 1417–1456,
<https://doi.org/10.5194/acp-11-1417-2011>, 2011.
- Meinshausen, M., Lewis, J., McGlade, C., Gütschow, J., Nicholls, Z., Burdon, R., Cozzi, L., and Hackmann, B.: Realization of Paris Agree-
ment pledges may limit warming just below 2 C, *Nature*, 604, 304–309, <https://doi.org/10.1038/s41586-022-04553-z>, 2022.
- Monteith, J. L.: Evaporation and environment, in: *The State and Movement of Water in Living Organisms*, edited by Fogg, G. E., vol. 19
880 of *Symposia of the Society for Experimental Biology*, pp. 205–234, Cambridge University Press, Cambridge, UK, <https://repository.rothamsted.ac.uk/item/8v5v7/evaporation-and-environment>, 1965.
- Nath, S., Lejeune, Q., Beusch, L., Seneviratne, S. I., and Schleussner, C.-F.: MESMER-M: an Earth system model emulator for spatially
resolved monthly temperature, *Earth System Dynamics*, 13, 851–877, <https://doi.org/10.5194/esd-13-851-2022>, 2022.
- Nath, S., Carreau, J., Kornhuber, K., Pflleiderer, P., Schleussner, C.-F., and Naveau, P.: MERCURY: A Fast and Versatile Multi-Resolution
885 Based Global Emulator of Compound Climate Hazards, *Journal of Advances in Modeling Earth Systems*, 17, e2024MS004905,
<https://doi.org/10.1029/2024MS004905>, 2025.
- Nguyen, T., Brandstetter, J., Kapoor, A., Gupta, J. K., and Grover, A.: ClimaX: A foundation model for weather and climate, in: *Proceedings
of the 40th International Conference on Machine Learning*, edited by Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and
Scarlett, J., vol. 202 of *Proceedings of Machine Learning Research*, pp. 25 904–25 938, PMLR, [https://proceedings.mlr.press/v202/](https://proceedings.mlr.press/v202/nguyen23a.html)
890 [nguyen23a.html](https://proceedings.mlr.press/v202/nguyen23a.html), 2023.
- Ogura, T. et al.: MIROC MIROC6 model output prepared for CMIP6 CFMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.885>, downloaded 1 July 2025, 2019.
- Park, S. et al.: SNU SAM0-UNICON model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.1489>, downloaded 1 July 2025, 2019.
- 895 Penman, H. L.: Natural evaporation from open water, bare soil and grass, *Proceedings of the Royal Society of London. Series A, Mathematical
and Physical Sciences*, 193, 120–145, <https://doi.org/10.1098/rspa.1948.0037>, 1948.
- Ploshay, J. et al.: NOAA-GFDL GFDL-CM4 model output prepared for CMIP6 DAMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.11383>, downloaded 1 July 2025, 2018.
- Quilcaille, Y., Gudmundsson, L., Beusch, L., Hauser, M., and Seneviratne, S. I.: Showcasing MESMER-X: spatially resolved
900 emulation of annual maximum temperatures of Earth system models, *Geophysical Research Letters*, 49, e2022GL099012,
<https://doi.org/10.1029/2022GL099012>, 2022.
- Quilcaille, Y., Gudmundsson, L., and Seneviratne, S. I.: Extending MESMER-X: A Spatially Resolved Earth System Model Emulator for
Fire Weather and Soil Moisture, *Earth System Dynamics*, 14, 1333–1362, <https://doi.org/10.5194/esd-14-1333-2023>, 2023.
- Ratnam, M. V., Prasad, P., Raj, S. T. A., Raman, M. R., and Basha, G.: Changing patterns in aerosol vertical distribution over South and East
905 Asia, *Scientific Reports*, 11, 308, <https://doi.org/10.1038/s41598-020-79361-4>, 2021.
- Richters, O., Kriegl, E., Bertram, C., Cui, R., Edmonds, J., Fawcett, A., Fuhrman, J., George, M., Hackstock, P., Hurst, I., Ju, Y., Kotz, M.,
Liadze, I., Min, J., Piontek, F., Sanchez Juanino, P., Sferri, F., Stevanovic, M., van Ruijven, B., Weigmann, P., Wenz, L., Westphal, M. I.,
Zwerling, M., Abrahão, G., Baumstark, L., Bresch, D. N., Chen, D. M.-C., Dietrich, J. P., Durga, S., Fricko, O., Hasse, R., Hoppe, J.,
Humpenöder, F., Iyer, G., Javaid, A., Joshi, S., Kikstra, J., Kishimoto, P., Klein, D., Koch, J., Krey, V., Kropf, C. M., Lewis, J., Lochner,
910 E., Luderer, G., Maczek, F., Mandaroux, R., Mastrucci, A., Meinshausen, M., Meng, M., Merfort, A., Nicholls, Z., Patel, P., Pehl, M.,
Pelz, S., Popp, A., Rüter, T., Sauer, I., Schreyer, F., Ünlü, G., von Jeetze, P., Zhao, A., and Zhao, X.: NGFS Phase 5 Scenario Explorer,

- <https://data.ene.iiasa.ac.at/ngfs/>, published 5 November 2024; updated 20 March 2026 (v5.1). Produced in partnership with PIK, IIASA, UMD, Climate Analytics, and NIESR. Data protected by EU Sui generis database rights., 2024.
- Ridley, J. et al.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set],
 915 <https://doi.org/10.22033/ESGF/CMIP6.419>, downloaded 1 July 2025, 2018.
- Ridley, J. et al.: MOHC HadGEM3-GC31-MM model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.420>, downloaded 1 July 2025, 2019.
- Schöngart, S., Nicholls, Z., Hoffmann, R., Pelz, S., and Schleussner, C.-F.: High-income groups disproportionately contribute to climate extremes worldwide, *Nature Climate Change*, 15, 627–633, <https://doi.org/10.1038/s41558-025-02325-x>, 2025.
- 920 Shiogama, H.: MIROC MIROC6 model output prepared for CMIP6 DAMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.894>, downloaded 1 July 2025, 2019.
- Shiogama, H. et al.: MIROC MIROC6 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.898>, downloaded 1 July 2025, 2019.
- Stjern, C. W., Samset, B. H., Boucher, O., Iversen, T., Lamarque, J.-F., Myhre, G., Shindell, D., and Takemura, T.: How aerosols and green-
 925 house gases influence the diurnal temperature range, *Atmospheric Chemistry and Physics*, 20, 13 467–13 480, <https://doi.org/10.5194/acp-20-13467-2020>, 2020.
- Swart, N. C. et al.: CCCma CanESM5.1 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.17184>, downloaded 1 July 2025, 2019a.
- Swart, N. C. et al.: CCCma CanESM5.1 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set],
 930 <https://doi.org/10.22033/ESGF/CMIP6.17199>, downloaded 1 July 2025, 2019b.
- Swart, N. C. et al.: CCCma CanESM5-CanOE model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.10205>, downloaded 1 July 2025, 2019c.
- Swart, N. C. et al.: CCCma CanESM5-CanOE model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.10207>, downloaded 1 July 2025, 2019d.
- 935 Swart, N. C. et al.: CCCma CanESM5 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.1303>, downloaded 1 July 2025, 2019e.
- Swart, N. C. et al.: CCCma CanESM5 model output prepared for CMIP6 DAMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.1305>, downloaded 1 July 2025, 2019f.
- Swart, N. C. et al.: CCCma CanESM5 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set],
 940 <https://doi.org/10.22033/ESGF/CMIP6.1317>, downloaded 1 July 2025, 2019g.
- Tachiiri, K. et al.: MIROC MIROC-ES2L model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.936>, downloaded 1 July 2025, 2019.
- Tang, Y. et al.: MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.1569>, downloaded 1 July 2025, 2019.
- 945 Tatebe, H. et al.: MIROC MIROC6 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set],
<https://doi.org/10.22033/ESGF/CMIP6.881>, downloaded 1 July 2025, 2018.
- Tebaldi, C., Snyder, A., and Dorheim, K.: STITCHES: creating new scenarios of climate model output by stitching together pieces of existing simulations, *Earth System Dynamics*, 13, 1557–1609, <https://doi.org/10.5194/esd-13-1557-2022>, 2022.

- Tebaldi, C., Selin, N. E., Ferrari, R., and Flierl, G.: Emulators of Climate Model Output, *Annual Review of Environment and Resources*, 50, 709–737, <https://doi.org/10.1146/annurev-environ-012125-085838>, 2025.
- 950 Thornthwaite, C. W.: An Approach toward a Rational Classification of Climate, *Geographical Review*, 38, 55–94, <https://doi.org/10.2307/210739>, 1948.
- Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index, *Journal of Climate*, 23, 1696–1718, <https://doi.org/10.1175/2009JCLI2909.1>, 2010.
- 955 Volodin, E. et al.: INM INM-CM4-8 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1422>, downloaded 1 July 2025, 2019a.
- Volodin, E. et al.: INM INM-CM4-8 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.12321>, downloaded 1 July 2025, 2019b.
- Volodin, E. et al.: INM INM-CM5-0 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.1423>, downloaded 1 July 2025, 2019c.
- 960 Volodin, E. et al.: INM INM-CM5-0 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.12322>, downloaded 1 July 2025, 2019d.
- Watanabe, S. et al.: MIROC MIROC-ES2H model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.901>, downloaded 1 July 2025, 2021.
- 965 Watson-Parris, D., Rao, Y., Olivié, D., Seland, Ø., Nowack, P., Camps-Valls, G., Stier, P., Bouabid, S., Dewey, M., Fons, E., Gonzalez, J., Harder, P., Jeggle, K., Lenhardt, J., Manshausen, P., Novitasari, M., Ricard, L., and Roesch, C.: ClimateBench v1.0: A benchmark for data-driven climate projections, *Journal of Advances in Modeling Earth Systems*, 14, e2021MS002954, <https://doi.org/10.1029/2021MS002954>, 2022.
- Watt-Meyer, O., Henn, B., McGibbon, J., Clark, S. K., Kwa, A., Perkins, W. A., Wu, E., Harris, L., and Bretherton, C. S.: ACE2: Accu-
 970 rately Learning Subseasonal to Decadal Atmospheric Variability and Forced Responses, *npj Climate and Atmospheric Science*, 8, 205, <https://doi.org/10.1038/s41612-025-01090-0>, 2025.
- Webb, M.: MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CFMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.435>, downloaded 1 July 2025, 2019.
- Ziehn, T. et al.: CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2288>, downloaded 1 July 2025, 2019a.
- 975 Ziehn, T. et al.: CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 ScenarioMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.2291>, downloaded 1 July 2025, 2019b.
- Ziehn, T. et al.: CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 DAMIP, Earth System Grid Federation [data set], <https://doi.org/10.22033/ESGF/CMIP6.14362>, downloaded 1 July 2025, 2020.