Supporting Information for

Selecting CMIP6 GCMs for CORDEX Dynamical Downscaling over Southeast Asia Using a Standardised Benchmarking Framework

Phuong Loan Nguyen[1,2*], Lisa V. Alexander[1,2], Marcus J. Thatcher[3], Son C. H. Truong[3], Rachael N. Isphording[1,2], John L. McGregor[3]

[1]ARC Centre of Excellence of Climate Extremes and Climate Change Research Centre, UNSW Sydney, Sydney, New South Wales, Australia

[2]ARC Centre of Excellence for the Weather of the 21st Century

[3]CSIRO Environment, Commonwealth Scientific and Industrial Research Organisation, Aspendale, Victoria, Australia

[*]*Corresponding author*: Phuong Loan Nguyen (phuongloan.nguyen@unsw.edu.au)

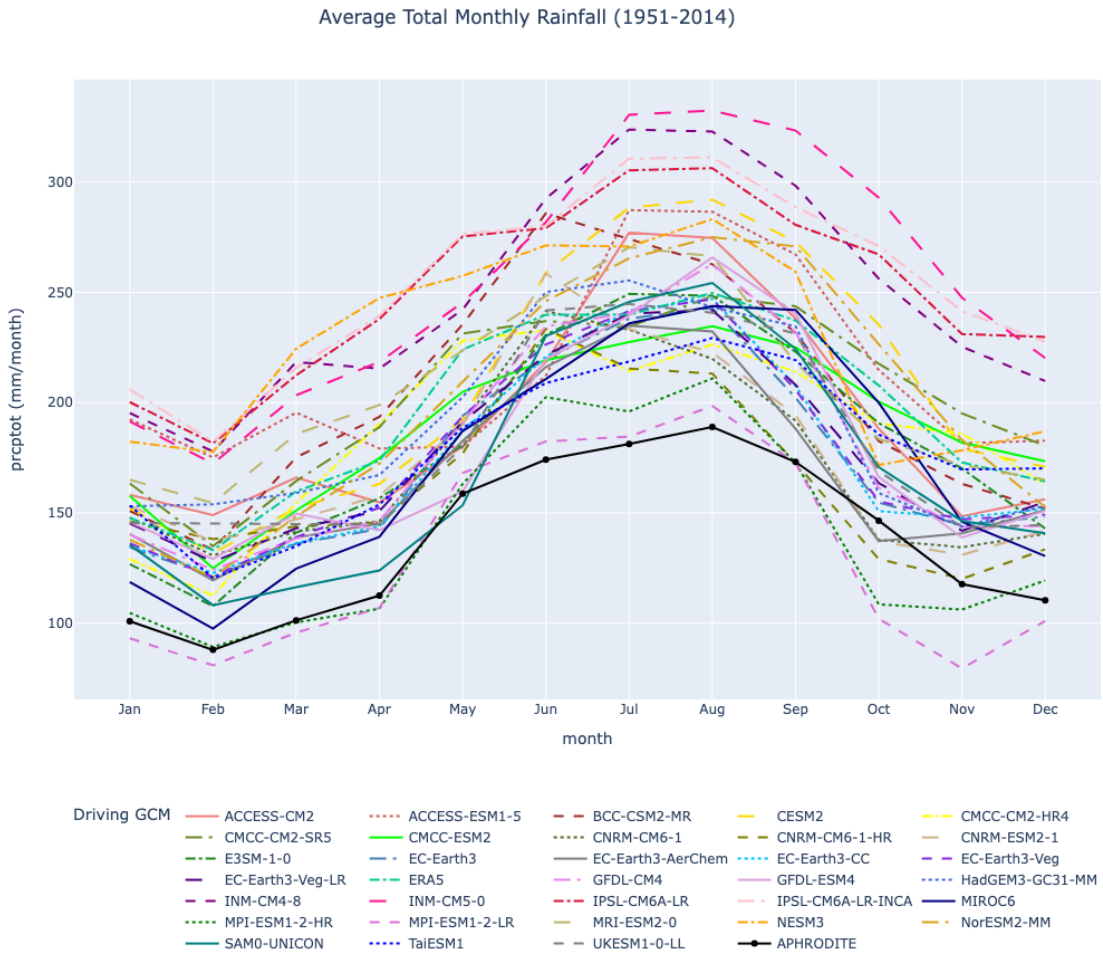

Figure s1. The climatological (1951-2014) regionally averaged total monthly rainfall (prcptot) over Southeast Asia. The APHRODITE data used as the benchmark is shown in black.

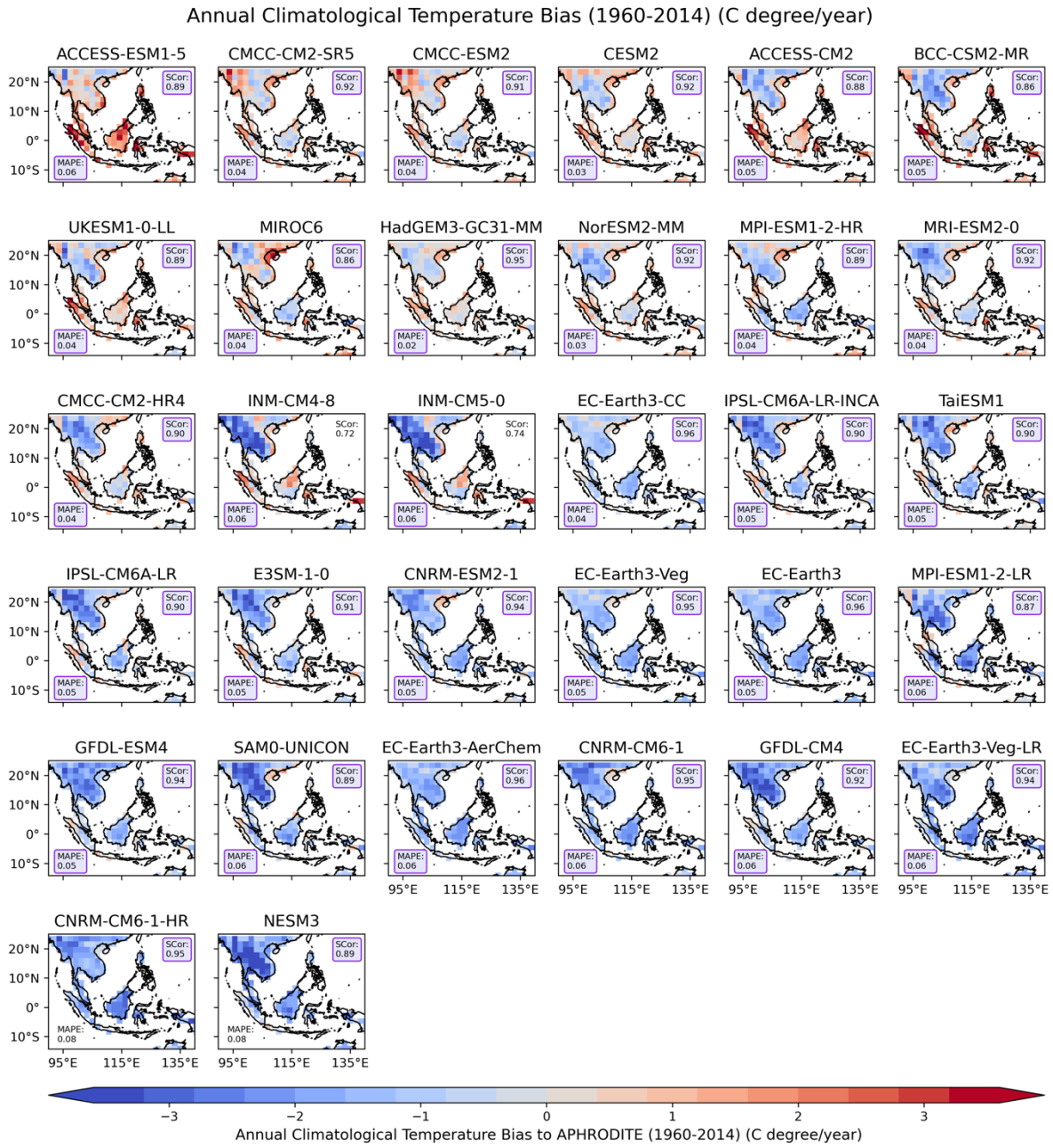

Figure s2. The annual climatological (1960-2014) bias (in C degree/year) for each model against the APHRODITE observational product, ranked hottest to coldest based on regionally-average of the bias. The mean absolute percentage error (MAPE) and spatial correlation (Scor) calculated against APHRODITE are plotted in the bottom left and upper right corners respectively. Values highlighted in purple-coloured boxes indicate values that meet our defined benchmarking thresholds (e.g., Scor >=0.85 and MAPE <=0.06) for this informed hypothetical. All analyses are considered at the coarsest CMIP6 GCM (i.e., NESM3, ~ 216km)

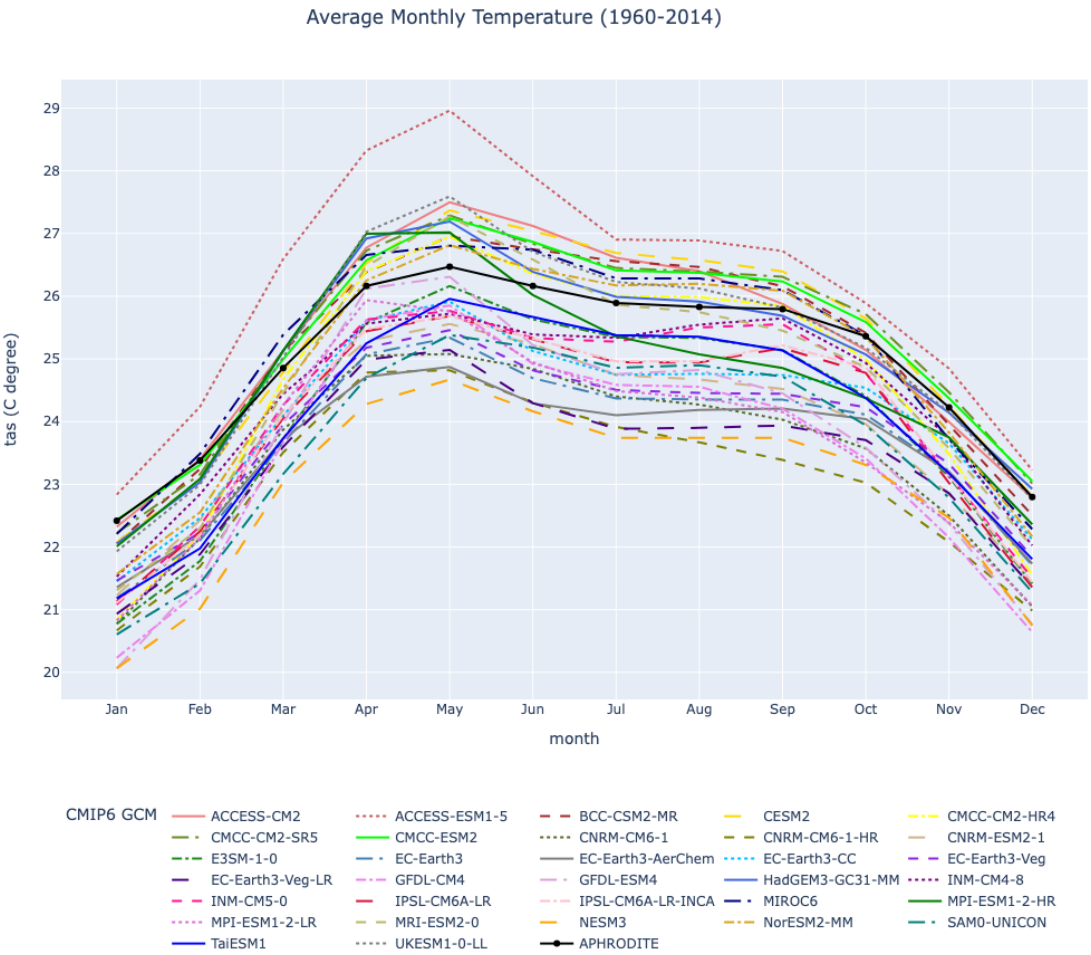

Figure s3. The climatological (1951-2014) regionally averaged monthly temperature (tas) over the mainland of Southeast Asia. The APHRODITE data used as the benchmark is shown in black.

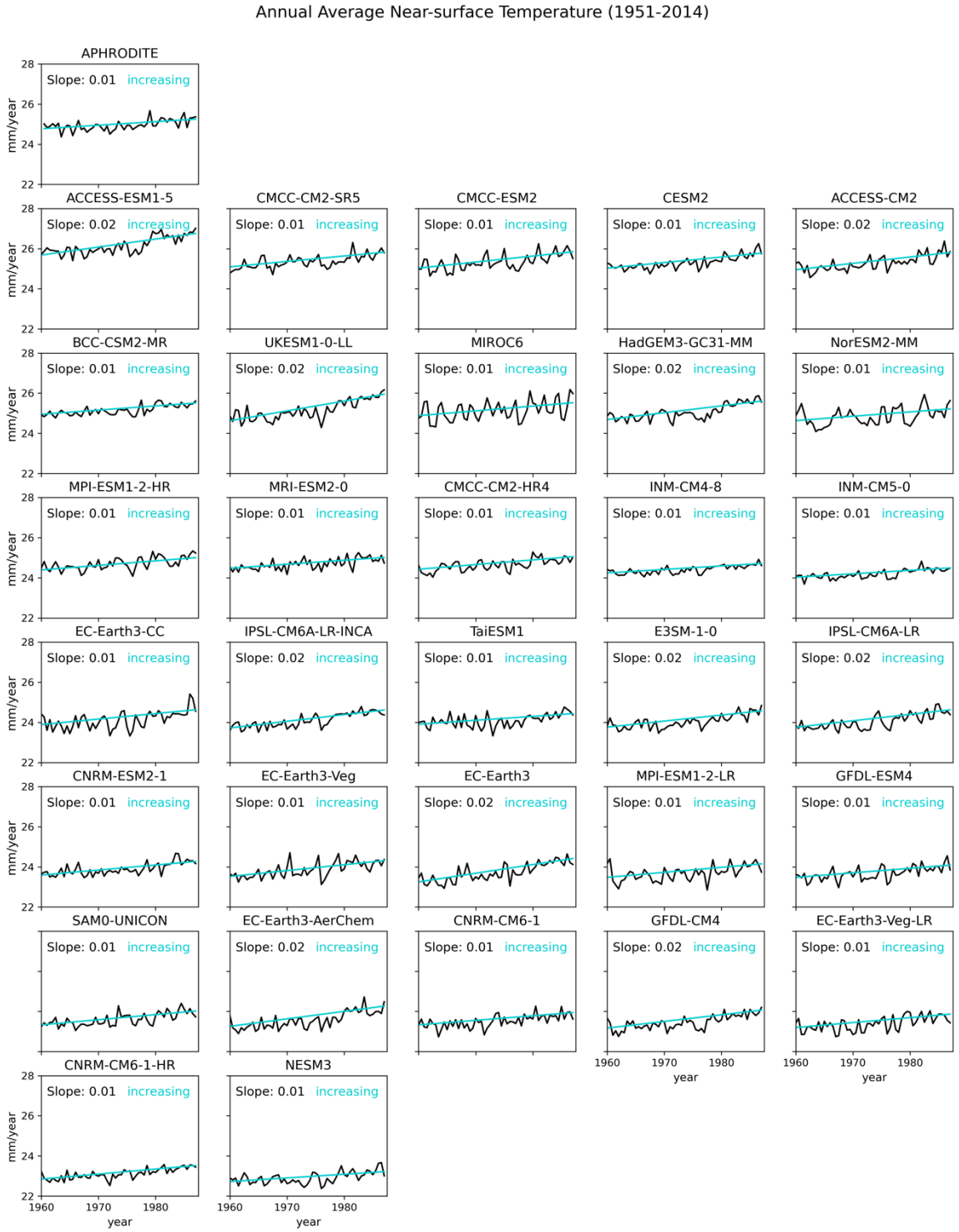

Figure s4. The observed (top row) and modelled annual average temperature across the land of Southeast Asia for the period 1960-2014. The direction of the observed Theil-Sen trend is the benchmark (top row). The Theil-Sen trend line for each of the simulations is plotted in grey if the models fail the benchmark and in purple if they pass. The magnitude of the trend is noted in the bottom left corner and the results of the Mann-Kendall significance test (Hussain et al., 2019) is noted in the bottom right corner. Models are sorted based on the magnitude of the patial average to match the order of Figure s2. All models pass the benchmark.

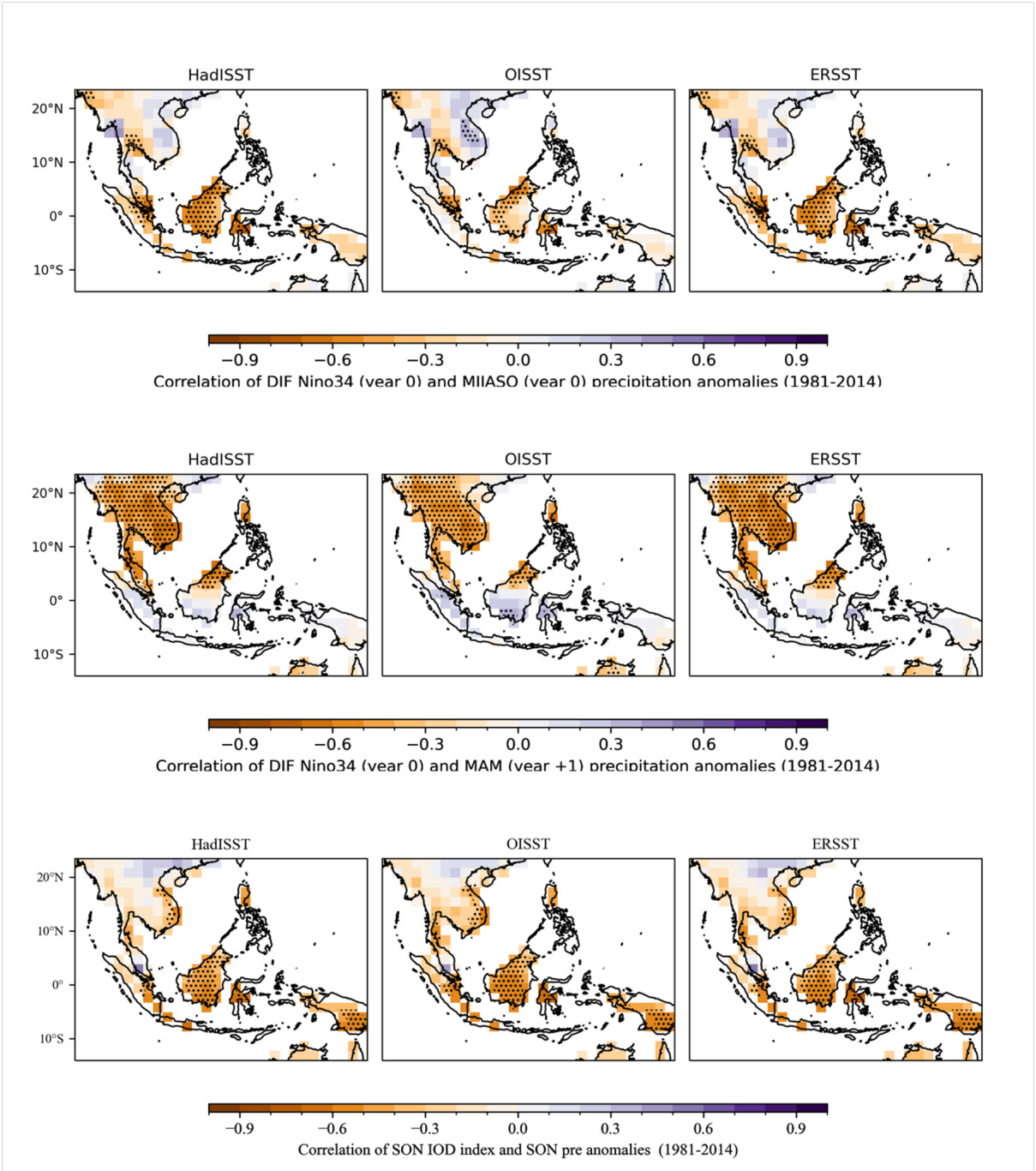

Figure s5. Teleconnection between observed precipitation and modes of variability with different SST datasets: HadISST, OISST and ERSST. The stippling indicates the grid points where the correlation coefficient is statistically significant at 95% confidence level according to the Student t-test.

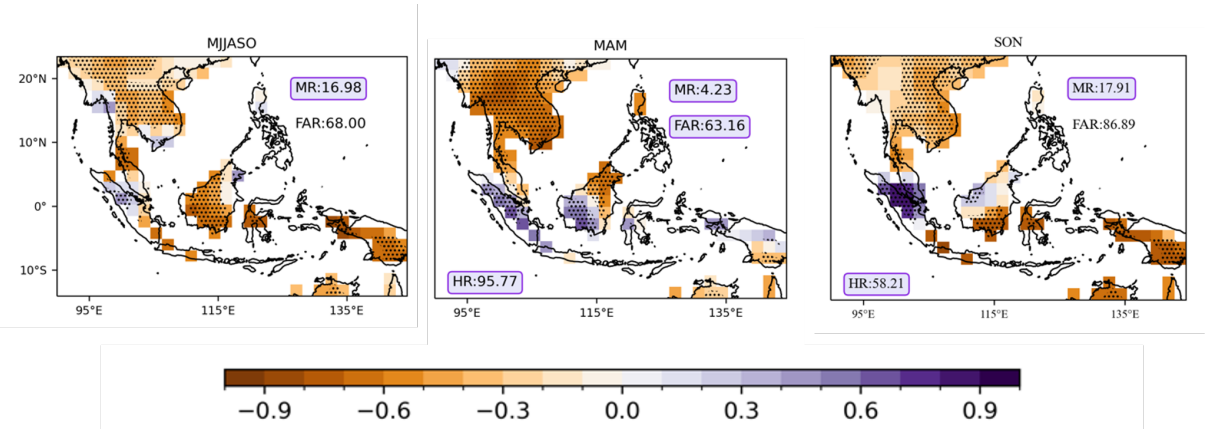

Figure s6. The teleconnection of precipitation and modes of variability in NorESM1-MM. The stippling indicates the grid points where the correlation coefficient is statistically significant at 95% confidence level according to the Student t-test. The Hite Rate (HR), Miss Rate (MR) and False Alarm Rate (FAR) calculated against ERA5 are plotted in the bottom left and upper right corners respectively. Values highlighted in purple-coloured boxes indicate values that meet our defined benchmarking thresholds for this informed hypothetical. All analyses are considered at the coarsest CMIP6 GCM (i.e., NESM3, ~ 216km).

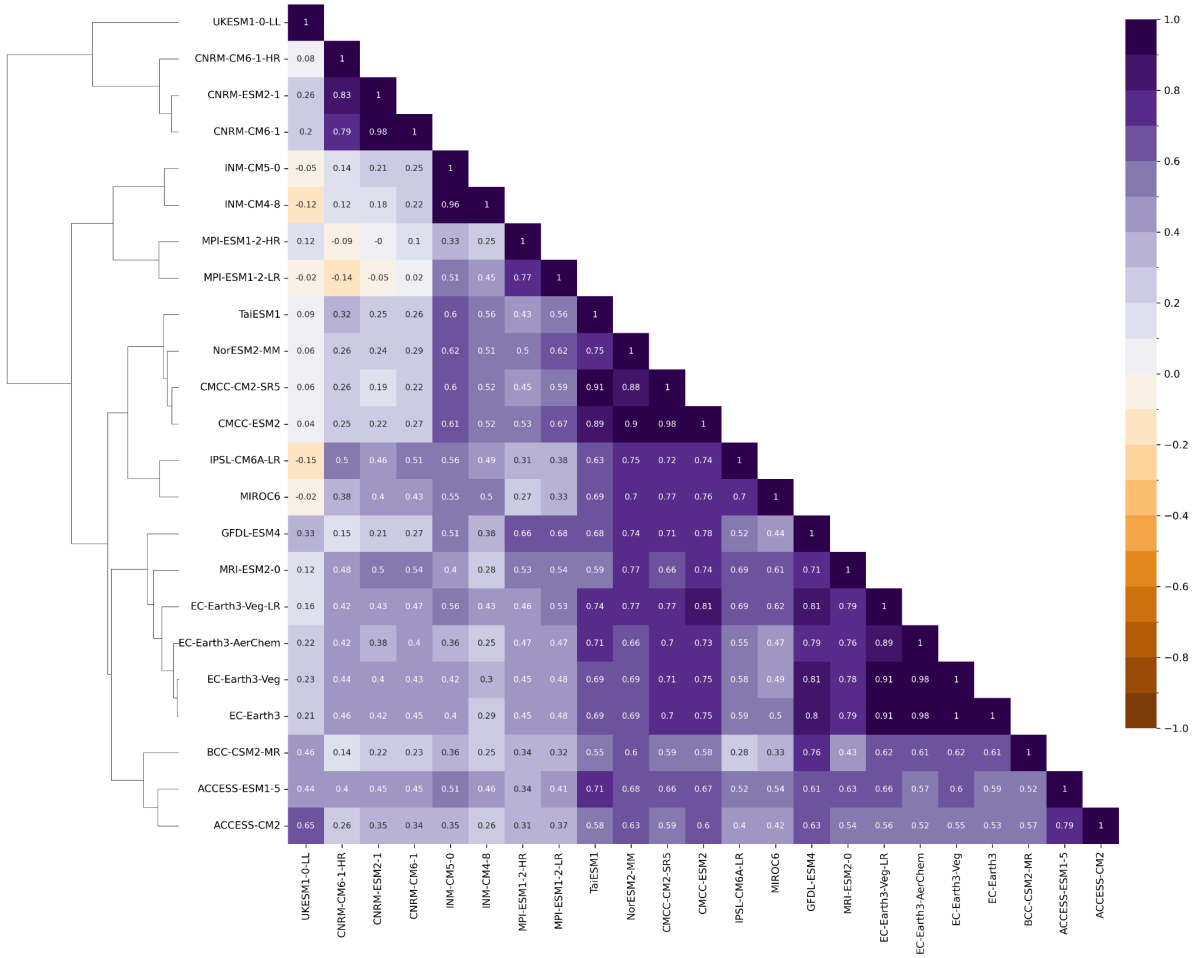

Figure s7: Dendrogram with hierarchical clustering applied for matrix of spatial correlation coefficient between CMIP6 climate models for the climatological (1951-2014) biases (relative to APHRODITE, Figure 2) in total precipitation during the wet season (MJJASO). The matrix is plotted for GCMs that simulated at least monthly tas (near-surface air temperature) and pr (precipitation) for the SSP-3.70 scenario only. Models are clustered with the Ward's linkage criterion.

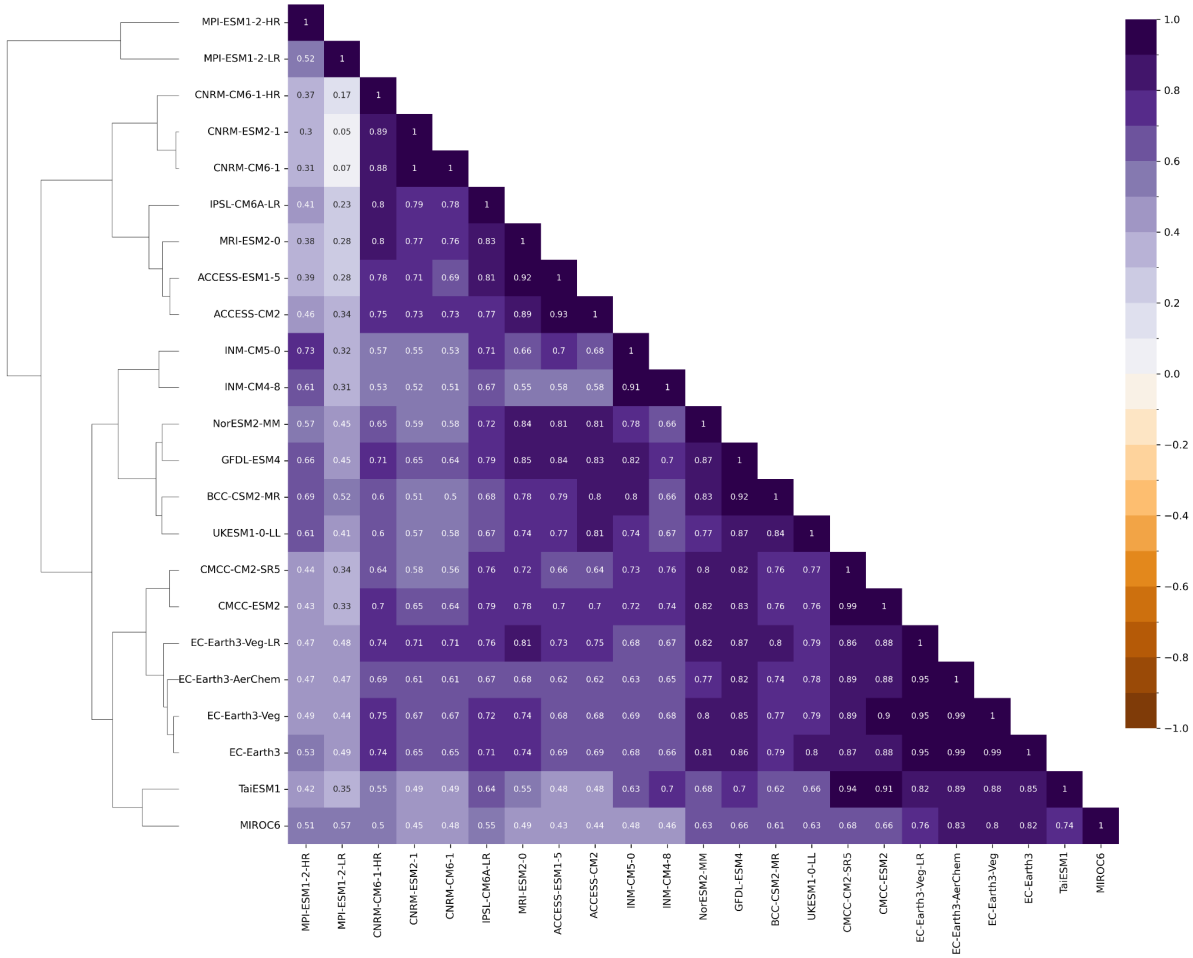

Figure s8: Similar to Figure s7 but for the dry season (November – April, NDJFMA).

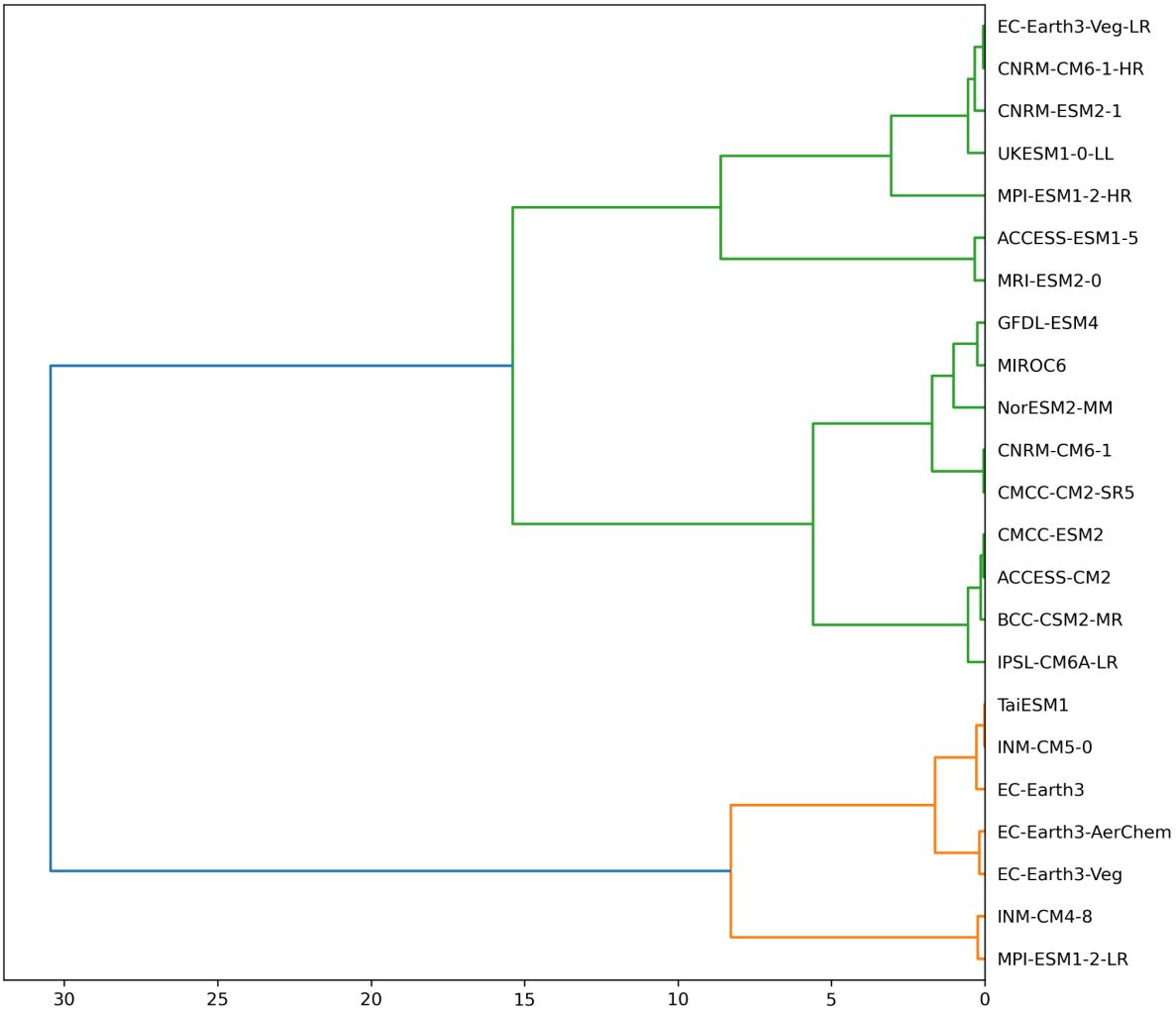

Figure s9. Dendrogram with hierarchical clustering of MJJASO regionally-averaged relative changes between the far-future (2070-2099) and the climatology (1961-1990) in seasonal total precipitation. Colours indicate the different clusters.

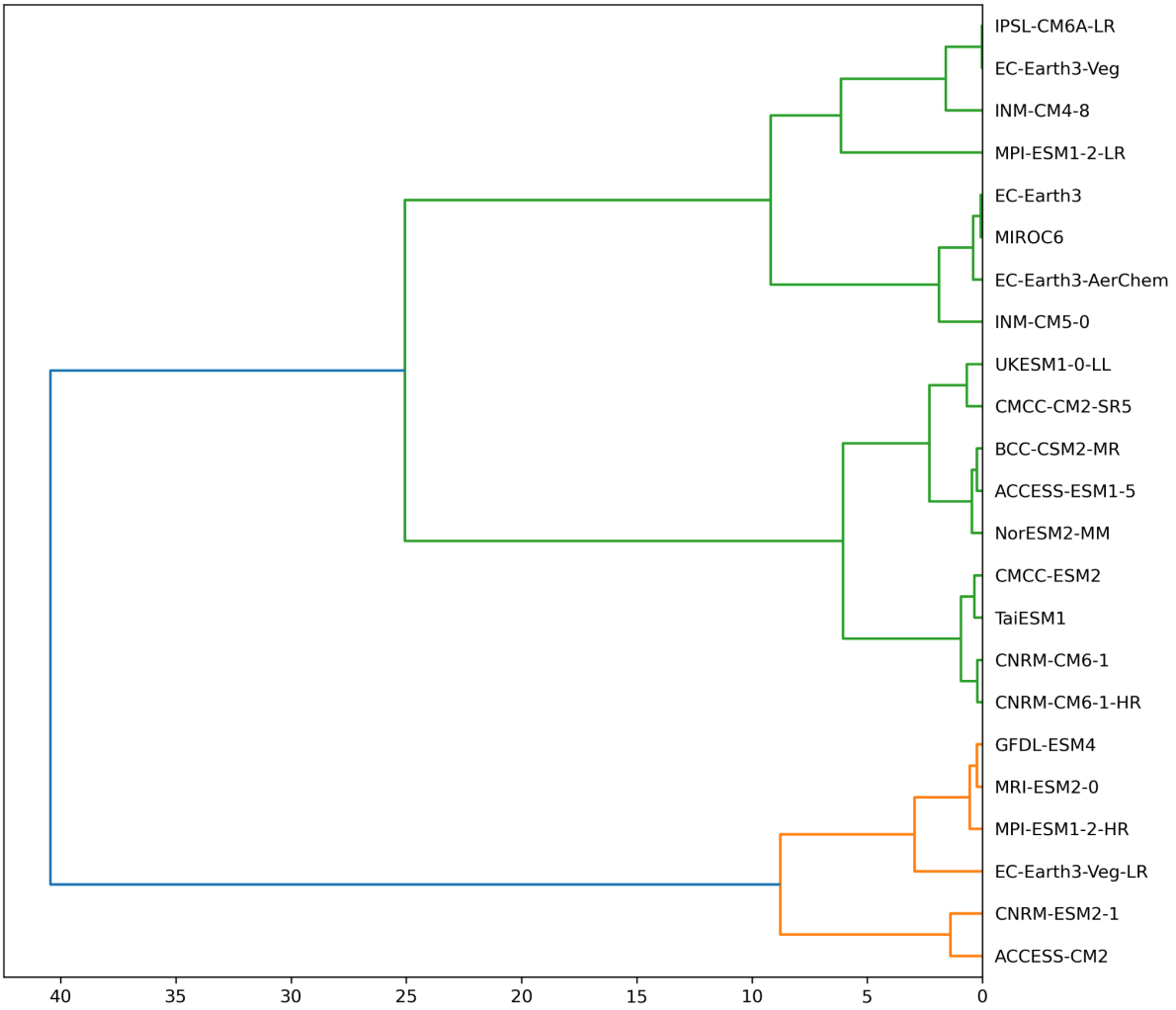

Figure s10. Similar to Figure s9 but for the dry season (November – April, NDJFMA).

Table s1. The climatological (1960-2014) monthly regionally averaged near-surface temperature over Southeast Asia are ranked from coldest to hottest for each CMIP6 simulation.

Simulations	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
APHRODITE	1	3	5	10	12	11	9	8	7	6	4	2
ACCESS-CM2	1	3	5	10	12	11	9	8	7	6	4	2
ACCESS-ESM1-5	1	3	6	11	12	10	9	8	7	5	4	2
BCC-CSM2-MR	1	3	5	8	12	11	10	9	7	6	4	2
CESM2	1	3	5	8	12	11	10	9	7	6	4	2
CMCC-CM2-HR4	1	3	5	11	12	10	9	8	7	6	4	2
CMCC-CM2-SR5	1	3	5	10	12	11	9	8	7	6	4	2
CMCC-ESM2	1	3	5	10	12	11	9	8	7	6	4	2
CNRM-CM6-1	1	3	6	11	12	10	9	8	7	5	4	2
CNRM-CM6-1-HR	1	3	7	11	12	10	9	8	6	5	4	2
CNRM-ESM2-1	1	3	6	11	12	10	9	8	7	5	4	2
E3SM-1-0	1	3	5	11	12	10	8	9	7	6	4	2
EC-Earth3	1	3	5	11	12	10	9	8	7	6	4	2
EC-Earth3-CC	1	3	5	11	12	10	8	9	7	6	4	2
EC-Earth3-Veg	1	3	5	11	12	10	9	8	7	6	4	2
EC-Earth3-Veg-LR	1	3	5	11	12	10	7	8	9	6	4	2
GFDL-CM4	1	3	6	11	12	10	9	8	7	5	4	2
GFDL-ESM4	1	3	6	11	12	10	8	9	7	5	4	2
HadGEM3-GC31-MM	1	3	6	11	12	10	9	8	7	5	4	2
INM-CM4-8	1	3	5	10	12	8	7	9	11	6	4	2
INM-CM5-0	1	3	5	11	12	8	7	9	10	6	4	2
IPSL-CM6A-LR	1	3	5	11	12	10	8	7	9	6	4	2
MIROC6	1	3	6	10	12	11	9	8	7	5	4	2
MPI-ESM1-2-HR	1	3	8	11	12	10	9	7	6	5	4	2
MPI-ESM1-2-LR	1	3	7	12	11	10	9	8	6	5	4	2
MRI-ESM2-0	1	3	5	10	12	11	9	8	7	6	4	2
NESM3	1	3	5	11	12	10	7	9	8	6	4	2
NorESM2-MM	1	3	5	10	12	11	8	9	7	6	4	2
SAM0-UNICON	1	3	5	7	12	11	9	10	8	6	4	2
TaiESM1	1	3	5	8	12	11	10	9	7	6	4	2
UKESM1-0-LL	1	3	5	11	12	10	9	8	7	6	4	2

Table s2. Summary of model performance against the MSMs for near-surface temperature. Models fail instantly the benchmarks are highlighted in red.

Simulations	MAPE	Scor	Seasonal cycle	Trend	Pass/4
ACCESS-CM2	+	+	+	+	**4**
ACCESS-ESM1-5	+	+	+	+	**4**
BCC-CSM2-MR	+	+	+	+	**4**
CESM2	+	+	+	+	**4**
CMCC-CM2-HR4	+	+	+	+	**4**
CMCC-CM2-SR5	+	+	+	+	**4**
CMCC-ESM2	+	+	+	+	**4**
CNRM-CM6-1	+	+	+	+	**4**
CNRM-CM6-1-HR	-	+	+	+	**3**
CNRM-ESM2-1	+	+	+	+	**4**
E3SM-1-0	+	+	+	+	**4**
EC-Earth3-AerChem	+	+	+	+	**4**
EC-Earth3-CC	+	+	+	+	**4**
EC-Earth3	+	+	+	+	**4**
EC-Earth3-Veg	+	+	+	+	**4**
EC-Earth3-Veg-LR	+	+	+	+	**4**
GFDL-CM4	+	+	+	+	**4**
GFDL-ESM4	+	+	+	+	**4**
HadGEM3-GC31-MM	+	+	+	+	**4**
INM-CM4-8	+	-	+	+	**3**
INM-CM5-0	+	-	+	+	**3**
IPSL-CM6A-LR	+	+	+	+	**4**
IPSL-CM6A-LR-INCA	+	+	+	+	**4**
MIROC6	+	+	+	+	**4**
MPI-ESM1-2-HR	+	+	+	+	**4**
MPI-ESM1-2-LR	+	+	+	+	**4**
MRI-ESM2-0	+	+	+	+	**4**
NESM3	-	+	+	+	**3**
NorESM2-MM	+	+	+	+	**4**
SAM0-UNICON	+	+	+	+	**4**
TaiESM1	+	+	+	+	**4**
UKESM1-0-LL	+	+	+	+	**4**

Table s3. Summary of model performance against the MSMs using multiple observational references for precipitation. The BMF is conducted over the common period of 1982-2014 among various references. Models fail instantly the benchmark are highlighted in red.

Simulations	Pass/7			
	APHRODITE	REGEN_ALL	GPDD_FDD	CHIRPSv2
ACCESS-CM2	7	7	7	7
ACCESS-ESM1-5	6	6	6	6
BCC-CSM2-MR	7	7	7	7
CESM2	7	7	7	7
CMCC-CM2-HR4	7	7	7	7
CMCC-CM2-SR5	7	7	7	7
CMCC-ESM2	7	7	7	7
CNRM-CM6-1	6	6	6	6
CNRM-CM6-1-HR	6	6	6	6
CNRM-ESM2-1	6	6	6	6
E3SM-1-0	7	7	7	7
EC-Earth3-AerChem	6	6	6	6
EC-Earth3-CC	7	7	7	7
EC-Earth3	7	7	7	7
EC-Earth3-Veg	7	7	7	7
EC-Earth3-Veg-LR	7	7	7	7
GFDL-CM4	7	7	7	7
GFDL-ESM4	7	7	7	7
HadGEM3-GC31-MM	7	7	7	7
INM-CM4-8	5	5	6	6
INM-CM5-0	5	5	5	5
IPSL-CM6A-LR	3	6	6	6
IPSL-CM6A-LR-INCA	3	6	6	6
MIROC6	7	7	7	7
MPI-ESM1-2-HR	5	5	5	5
MPI-ESM1-2-LR	7	7	6	7
MRI-ESM2-0	6	6	6	6
NESM3	5	5	5	5
NorESM2-MM	6	7	6	7
SAM0-UNICON	7	7	7	7
TaiESM1	7	7	7	7
UKESM1-0-LL	7	7	7	7

Table s4. Summary of model performance against the versatility metrics using multiple SST products. The second tier of BMF is conducted over the common period of 1981-2014 among various references. Models fail instantly the benchmark are highlighted in red.

Simulations	Pass/15		
	HadISS	OI SST	ERSST
ACCESS-CM2	15	15	15
BCC-CSM2-MR	13	12	13
CESM2	10	10	10
CMCC-CM2-HR4	11	10	11
CMCC-CM2-SR5	14	13	14
CMCC-ESM2	14	14	14
E3SM-1-0	15	15	15
EC-Earth3	15	15	15
EC-Earth3-Veg	15	15	15
EC-Earth3-Veg-LR	14	13	14
GFDL-CM4	14	13	14
GFDL-ESM4	15	15	15
HadGEM3-GC31-MM	15	15	15
MIROC6	13	11	13
MPI-ESM1-2-LR	14	13	14
SAM0-UNICON	15	15	15
TaiESM1	13	13	13
UKESM1-0-LL	15	15	15

Table s5. Summary of model performance against the MSMs using APHRODITE as a reference. The benchmarking is conducted at the native grid of CMIP6 GCMs. Models fail instantly the benchmark are highlighted in red.

Simulations	Wet season		Dry season		Seasonal cycle	Trend		Pass/7
	MAPE	Scor	MAPE	Scor		Wet	Dry	
ACCESS-CM2	+	+	+	+	+	+	+	7
ACCESS-ESM1-5	-	+	+	+	-	+	+	5
BCC-CSM2-MR	-	+	+	+	-	+	+	5
CESM2	+	+	+	+	+	+	+	7
CMCC-CM2-HR4	+	+	+	+	+	+	+	7
CMCC-CM2-SR5	+	+	+	+	+	+	+	7
CMCC-ESM2	+	+	+	+	+	+	+	7
CNRM-CM6-1	+	+	+	+	-	+	+	6
CNRM-CM6-1-HR	+	+	+	+	-	+	+	6
CNRM-ESM2-1	+	+	+	-	-	+	+	5
E3SM-1-0	+	+	+	+	+	+	+	7
EC-Earth3-AerChem	+	+	+	+	-	+	+	6
EC-Earth3-CC	+	+	+	+	-	+	+	6
EC-Earth3	+	+	+	+	+	+	+	7
EC-Earth3-Veg	+	+	+	+	+	+	+	7
EC-Earth3-Veg-LR	+	+	+	+	+	+	+	7
GFDL-CM4	+	+	+	+	+	+	+	7
GFDL-ESM4	+	+	+	+	+	+	+	7
HadGEM3-GC31-MM	+	+	+	+	+	+	+	7
INM-CM4-8	-	+	-	+	-	+	+	4
INM-CM5-0	-	+	-	+	+	+	+	5
IPSL-CM6A-LR	-	-	-	-	+	+	+	3
IPSL-CM6A-LR-INCA	-	-	-	-	-	+	+	2
MIROC6	+	+	+	+	+	+	+	7
MPI-ESM1-2-HR	+	+	+	+	+	+	-	6
MPI-ESM1-2-LR	+	+	+	+	+	+	+	7
MRI-ESM2-0	+	-	+	+	+	+	-	5
NESM3	+	+	+	-	-	+	+	5
NorESM2-MM	-	+	+	+	+	+	+	6
SAM0-UNICON	+	+	+	+	+	+	+	7
TaiESM1	+	+	+	+	+	+	+	7
UKESM1-0-LL	+	+	+	+	+	+	+	7