Supplement of

Uncertainty in Regional Monsoon Responses to Solar Radiation Modification Is Often Greater Across Models Than Across Strategies

Cindy Wang[1], Daniele Visioni[1], Brendan Clark[1], Zachary M. Labe[2], Bor-Ting Jong[3,4], Matthew Henry[5], Walker Lee[6], and Haruki Hirasawa[7]

[1]Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, USA

[2]Climate Central, Inc., Princeton, NJ, United States of America

[3]Department of Earth Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[4]Institute for Environmental Studies, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[5]Department of Mathematics, University of Exeter, UK

[6]Climate & Global Dynamics Division, National Science Foundation's National Center for Atmospheric Research (NSF NCAR), Boulder, CO, USA

[7]Department of Atmospheric and Climate Science, University of Washington, Seattle, WA, USA

Contents of this file

1. Text S1
2. Tables S1 to S5
3. Figures S1 to S6

S1 Model-versus-strategy comparison

For each region and metric, let $\widehat{\Delta}_{m,s}$ be the ensemble-mean change of model m under strategy s, relative to the 2020–2039 baseline of its control simulation.

With $N_M = 4$ models and N_S strategies ($N_S = 3$, or 2 when only the two SAI strategies are compared), the mean absolute pairwise differences across models and across strategies are

$$M = \frac{1}{N_S} \sum_s \left(\frac{N_M}{2}\right)^{-1} \sum_{a<c} \left|\widehat{\Delta}_{a,s} - \widehat{\Delta}_{c,s}\right|, \qquad S = \frac{1}{N_M} \sum_m \left(\frac{N_S}{2}\right)^{-1} \sum_{s<t} \left|\widehat{\Delta}_{m,s} - \widehat{\Delta}_{m,t}\right| \tag{1}$$

and Fig. 6a shows their ratio M/S. The sign-agreement measure of Fig. 6b is defined analogously, as the fraction of the same pairs whose changes have the same sign.

Sampling noise does not enter M and S equally. When two strategies are compared against the same control run, the sampling error of that control cancels in their difference, whereas a difference between two models contains the sampling errors of both models' controls; and the absolute value of a noisy difference is biased upward. We therefore also compare noise-corrected squared differences,

$$\widetilde{M}^2 = \frac{1}{N_S} \sum_s \left(\frac{N_M}{2}\right)^{-1} \sum_{a<c} \left[\left(\widehat{\Delta}_{a,s} - \widehat{\Delta}_{c,s}\right)^2 - \widehat{\mathrm{Var}}\left(\widehat{\Delta}_{a,s} - \widehat{\Delta}_{c,s}\right)\right] \tag{2}$$

and $\widetilde{S}^2$ defined correspondingly over pairs of strategies. Because the expected square of an unbiased estimate exceeds the square of its true value by the estimate's variance, $\widetilde{M}^2$ and $\widetilde{S}^2$ are unbiased estimates of the mean squared underlying differences, without any assumption about the shape of the sampling distribution. The variance of each difference is estimated in two ways: from the hierarchical bootstrap (Sect. 2.5), in which all scenarios that share a control run share its resamples, so that a common control cancels exactly (for MIROC the same SSP2-4.5 run serves as the control of all three strategies, whereas for CESM2, UKESM1 and E3SM G6-1.5K-MCB has its own control); and from the spread across ensemble members, as the sum of s^2/n over the independent runs entering the difference. With three members the member-spread estimate is itself uncertain, whereas the two-stage bootstrap tends to overstate the variance, so the two estimates bracket the correction. A region–metric combination is counted as model-dominated when $\widetilde{M}^2 > \widetilde{S}^2$ and strategy-dominated when $\widetilde{S}^2 > \widetilde{M}^2$, provided at least one of the two is positive, that is, exceeds its sampling noise.

We further recompute the count of combinations with $M/S > 1$ for each of 2000 bootstrap replicates and repeat the comparison (i) leaving out one model at a time, (ii) using only UKESM1 and MIROC, whose configurations are

identical across the SAI and MCB experiments, and (iii) with the alternative thresholds and harmonics of Sect. 2.4 (Table S4). Because onset, retreat and duration are related, as are the precipitation metrics, the 36 region–metric combinations are not independent, and counts of combinations are descriptive summaries rather than independent tests.

Table S1: Multi-model mean change in active-season precipitation (%) between 2020–2039 and 2065–2084 for each monsoon region and scenario. Each experiment is compared with the 2020–2039 mean of its own control simulation (Sect. 2.1). Values are the mean of the four models (CESM2, UKESM1, E3SM and MIROC) ± one inter-model standard deviation. Daggers (†) mark changes for which fewer than three of the four models agree on the sign of the multi-model mean.

Region	**SSP2-4.5**	**G6-1.5K-SAI**	**G6-1.5K-HiLLA**	**G6-1.5K-MCB**
NAmerM	$-1.8 \pm 22.0^\dagger$	-15.9 ± 18.2	-16.4 ± 15.7	$+1.8 \pm 5.9$
WAfriM	$+7.0 \pm 12.6$	-14.1 ± 14.3	-18.9 ± 11.2	$+8.4 \pm 19.2^\dagger$
SAsiaM	$+5.9 \pm 2.3$	-2.9 ± 3.4	$-1.3 \pm 2.8^\dagger$	$+4.1 \pm 3.2$
EAsiaM	$+9.7 \pm 2.5$	-2.9 ± 7.4	$+0.5 \pm 6.7^\dagger$	$+4.4 \pm 7.1$
SAmerM	-1.3 ± 0.7	-0.2 ± 2.5	$+0.4 \pm 0.6^\dagger$	$+0.7 \pm 2.4^\dagger$
AusMCM	$+17.4 \pm 18.4$	$+2.7 \pm 5.8$	$+3.3 \pm 16.4^\dagger$	$+3.4 \pm 2.2$

Table S2: How closely each SRM strategy returns the monsoon metrics to the baseline climatology, for each model and the multi-model mean, summarized over the 36 region–metric combinations (six regions × six metrics) of Fig. 4 of the main text. For every combination, Δ is the 2065–2084 change relative to the 2020–2039 baseline of each experiment's own control simulation, and $\Delta_n = \Delta / \max|\Delta|$ is the same change normalized as in Fig. 4. $\overline{|\Delta_n|}$: mean of $|\Delta_n|$ over the 36 combinations (lower = closer to the baseline). N_{close}: the number of combinations (out of 36) in which the strategy has the smallest $|\Delta|$ of the three SRM strategies; a tie between two (three) strategies gives each 1/2 (1/3), so the three N_{close} values of a model add up to 36 (up to rounding) and can be fractional. N_{red}: the number of combinations (out of 36) in which the strategy's $|\Delta|$ is smaller than the $|\Delta|$ under SSP2-4.5, i.e. in which the strategy reduces the departure from the baseline relative to no intervention; it is counted separately for each strategy, so the three values need not add up to 36. Bold marks the best strategy for each model by $\overline{|\Delta_n|}$ and by N_{close}. For reference, $\overline{|\Delta_n|}$ under SSP2-4.5 is 0.32 (CESM2), 0.24 (UKESM1), 0.22 (E3SM), 0.18 (MIROC) and 0.16 (multi-model mean). For the multi-model mean, Δ_n is first averaged over the four models (keeping its sign) and the absolute value is taken afterwards.

	G6-1.5K-SAI			**G6-1.5K-HiLLA**			**G6-1.5K-MCB**		
Model	$\overline{\lvert\Delta_n\rvert}$	N_{close}	N_{red}	$\overline{\lvert\Delta_n\rvert}$	N_{close}	N_{red}	$\overline{\lvert\Delta_n\rvert}$	N_{close}	N_{red}
CESM2	0.29	8	14	0.30	8	19	**0.18**	**20**	22
UKESM1	0.18	6.3	17	0.23	11.3	19	**0.17**	**18.3**	19
E3SM	**0.11**	**16.3**	27	0.19	13.8	21	0.20	5.8	19
MIROC	**0.09**	12	27	0.09	**16**	27	0.13	8	23
Multi-model mean	0.12	11	19	0.13	12	20	**0.09**	**13**	26

Table S3: Historical model biases compared with projected changes, for the 36 region–metric combinations of Fig. 4 of the main text, in physical units: days for onset, retreat, duration and peak timing; percent for active-season precipitation and wet-day intensity (biases relative to the IMERG value, changes relative to each model's own 2020–2039 baseline). Bias: range of the four models' historical (1998–2014; 1999–2014 for SAmerM and AusMCM) ensemble-mean biases relative to IMERG. SRM change: range of the 2065–2084 changes across the four models and three SRM strategies. SSP2-4.5: largest absolute SSP2-4.5 change across models. $r_{\max}$: largest absolute bias divided by largest absolute SRM change; r_{range}: bias range divided by the range of SRM changes. Ratios above 1 indicate that the historical bias is larger; ratios of 1 or less are in bold (3 and 3 of 36, respectively).

Region	**Metric**	**Bias**	**SRM change**	**SSP2-4.5**	$r_{\max}$	r_{range}
NAmerM	Onset (days)	0 to +30	−5 to +18	15	1.64	1.29
	Retreat (days)	−8 to +38	−12 to +5	10	3.29	2.80
	Duration (days)	−25 to +38	−30 to +4	25	1.28	1.86
	Peak timing (days)	+12 to +50	−17 to +5	7	3.00	1.79
	Active-season precipitation (%)	−31 to +75	−42 to +6	35	1.79	2.20
	Wet-day intensity (%)	+3 to +30	−13 to +5	4	2.20	1.46
WAfriM	Onset (days)	−2 to +25	−5 to +11	7	2.20	1.63
	Retreat (days)	−30 to +12	−12 to +8	7	2.57	2.13
	Duration (days)	−55 to +2	−20 to +13	13	2.75	1.73
	Peak timing (days)	−7 to +5	−5 to +3	4	1.33	1.40
	Active-season precipitation (%)	−62 to +15	−29 to +27	24	2.13	1.38
	Wet-day intensity (%)	−10 to +12	−8 to +6	4	1.55	1.57
SAsiaM	Onset (days)	+2 to +17	−3 to +7	5	2.50	1.42
	Retreat (days)	−12 to +10	−3 to +5	5	2.50	2.70
	Duration (days)	−22 to −2	−10 to +8	6	2.17	1.09
	Peak timing (days)	−10 to +10	−7 to +10	10	**1.00**	1.15
	Active-season precipitation (%)	−19 to +9	−6 to +9	8	2.26	1.96
	Wet-day intensity (%)	−6 to +11	−2 to +2	6	4.31	3.78
EAsiaM	Onset (days)	−30 to +15	−9 to +8	8	3.33	2.60
	Retreat (days)	−2 to +18	−3 to +6	7	2.93	2.09
	Duration (days)	−17 to +35	−10 to +13	10	2.65	2.22
	Peak timing (days)	0 to +22	−12 to +13	10	1.62	**0.87**
	Active-season precipitation (%)	−21 to +56	−10 to +12	12	4.51	3.39
	Wet-day intensity (%)	0 to +13	−2 to +4	4	3.32	2.33
SAmerM	Onset (days)	−2 to +17	−8 to 0	5	2.00	2.30
	Retreat (days)	+2 to +13	−3 to +2	3	4.00	2.29
	Duration (days)	−15 to +10	−2 to +7	5	2.25	2.88
	Peak timing (days)	−52 to −23	−12 to +18	17	2.86	**0.97**
	Active-season precipitation (%)	+2 to +31	−2 to +4	2	7.40	4.57
	Wet-day intensity (%)	+12 to +24	−4 to +2	3	5.96	1.96
AusMCM	Onset (days)	−5 to +13	−18 to +5	20	**0.73**	**0.79**
	Retreat (days)	−22 to +18	−12 to +13	28	1.69	1.63
	Duration (days)	−30 to +23	−17 to +32	48	**0.95**	1.10
	Peak timing (days)	+8 to +45	−15 to +17	20	2.65	1.15
	Active-season precipitation (%)	−31 to +26	−16 to +24	39	1.31	1.44
	Wet-day intensity (%)	−4 to +3	−3 to +1	3	1.53	1.80

Table S4: Change in intermodel spread relative to SSP2-4.5, by region. For each region, the entries count the six metrics (onset, retreat, duration, peak timing, active-season precipitation and wet-day intensity) for which the range of the four models' normalized changes under each SRM strategy is larger than ($>$), smaller than ($<$) or within 5% of ($\approx$) the corresponding SSP2-4.5 range. These correspond to the symbols to the right of each panel in Fig. 4 of the main text.

Region	**G6-1.5K-SAI** $>$ / $<$ / $\approx$	**G6-1.5K-HiLLA** $>$ / $<$ / $\approx$	**G6-1.5K-MCB** $>$ / $<$ / $\approx$
NAmerM	3 / 2 / 1	2 / 4 / 0	1 / 5 / 0
WAfriM	4 / 1 / 1	2 / 3 / 1	6 / 0 / 0
SAsiaM	2 / 2 / 2	3 / 3 / 0	3 / 2 / 1
EAsiaM	4 / 2 / 0	4 / 2 / 0	4 / 2 / 0
SAmerM	1 / 4 / 1	1 / 4 / 1	1 / 3 / 2
AusMCM	0 / 6 / 0	1 / 4 / 1	2 / 4 / 0
Total (of 36)	14 / 17 / 5	13 / 20 / 3	17 / 16 / 3

Table S5: Robustness of the model-versus-strategy comparison (Fig. 6 and Fig. S3 of the main text). Entries give the number of the 36 region–metric combinations in which differences across models exceed differences across strategies, for all three SRM strategies (Fig. 6a) and for the two SAI strategies only (Fig. S3a), and the number of the six SAsiaM metrics in which strategy differences exceed model differences (all three strategies). Unless noted, the comparison uses the ratio of mean absolute pairwise differences (Sect. 2.6). Bootstrap: 2000 hierarchical member–year resamples (Sect. 2.5) in which scenarios sharing a control run share its resamples. [a]Model configuration identical across the SAI and MCB experiments. [b]The AusMCM season is not detected in most MIROC members at the raised threshold (9 mm day^{-1}). [c]Mean squared pairwise differences minus their estimated sampling variances (Eq. 2), with the variance estimated from the bootstrap or from the spread across ensemble members; in brackets, the number of model-dominated combinations among those in which either difference exceeds its sampling noise. [d]Not a model-versus-strategy count: number of the 36 combinations in which the model with the largest absolute historical (1998–2014) bias relative to IMERG is also the model with the largest absolute projected change, averaged over the SRM strategies in that column (6 of 36 for SSP2-4.5).

Test	**All strategies**	**SAI only**	**SAsiaM, strategy-dominated**
Reference (four models)	28	31	4
Bootstrap replicates, median [5–95%]	24 [18–29]	28 [23–33]	–
Noise-corrected squared contrasts (bootstrap variance)[c]	19 (9 of 18)	20 (14 of 17)	6
Noise-corrected squared contrasts (member-spread variance)[c]	20 (13 of 26)	22 (18 of 24)	6
Without CESM2	20	29	4
Without UKESM1	31	32	1
Without E3SM	28	31	5
Without MIROC	21	27	4
UKESM1 and MIROC only[a]	26	27	5
Thresholds $\times 0.8$	23	30	3
Thresholds $\times 1.2$[b]	18	32	2
8 Fourier harmonics	23	32	6
16 Fourier harmonics	24	29	6
Most-biased model also has the largest projected change[d]	8	5	–

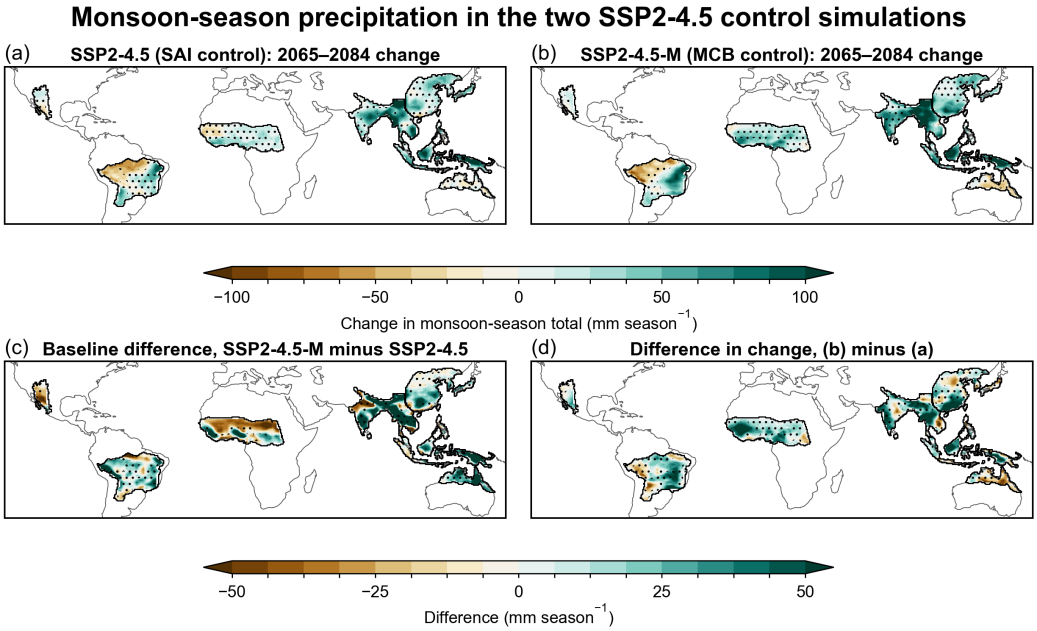


Figure S1: Comparison of the two SSP2-4.5 control simulations: SSP2-4.5, the control for G6-1.5K-SAI and G6-1.5K-HiLLA, and SSP2-4.5-M, the control run alongside G6-1.5K-MCB. Maps show the multi-model mean monsoon-season total precipitation (mm season^{-1}) within the six monsoon regions, using the core monsoon months of each region as in Fig. 1 of the main text. (a) Change between 2020–2039 and 2065–2084 under SSP2-4.5. (b) As in (a), but for SSP2-4.5-M. (c) Difference between the 2020–2039 baselines (SSP2-4.5-M minus SSP2-4.5). (d) Difference in the projected change ((b) minus (a)). The multi-model means weight CESM2, UKESM1 and E3SM equally; MIROC is omitted because its SSP2-4.5-M and SSP2-4.5 output are identical. The SSP2-4.5 ensembles have 3 members per model; the SSP2-4.5-M ensembles have 16 members for CESM2 and 3 for UKESM1 and E3SM. For CESM2 and E3SM the two controls use different model versions (CESM2(WACCM6) and CESM2.1-CAM6; E3SMv3 and E3SMv2; Table 1), so the differences in (c) and (d) combine internal variability with model-version differences. Stippling indicates differences that are not statistically significant at the 95% level. Note the different color scales in (a, b) and (c, d).

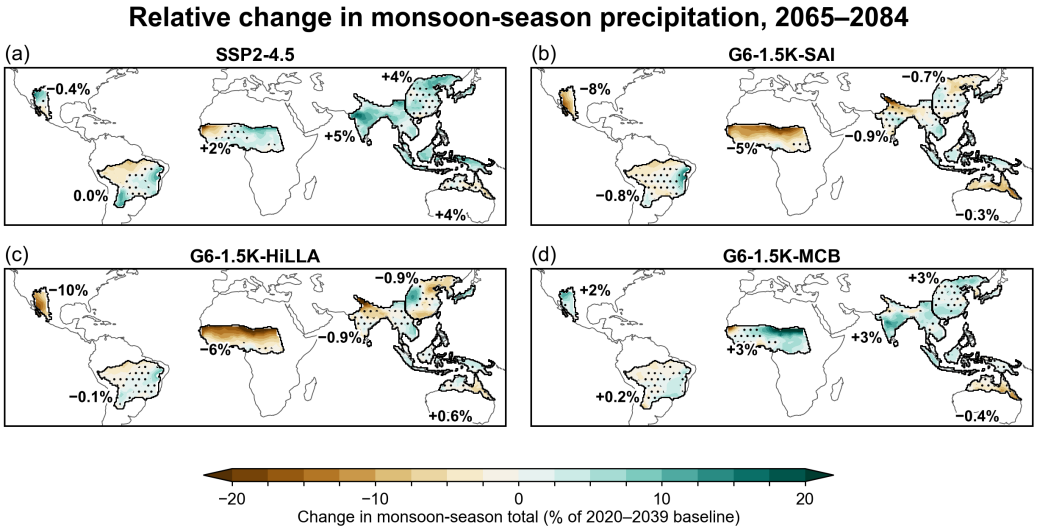


Figure S2: As Fig. 1b–e of the main text, but showing the 2065–2084 change in monsoon-season total precipitation as a percentage of the 2020–2039 baseline of each experiment's own control simulation: SSP2-4.5 for (a) SSP2-4.5 and (b) G6-1.5K-SAI; SSP2-4.5 (CESM2, UKESM1) or SSP2-4.5-H (E3SM, MIROC) for (c) G6-1.5K-HiLLA; and SSP2-4.5-M for (d) G6-1.5K-MCB. Monsoon-season totals use the core monsoon months of each region, as in Fig. 1. For each model, the percentage change is computed from the ensemble means at each grid cell after conservative remapping of each member's 20-year mean to the UKESM1 grid; the maps show the mean of the four models' percentage changes, with equal weight per model. Numbers give the change in the region-total monsoon-season precipitation (cosine-latitude weighted) for each region, averaged over models. Stippling indicates changes that are not statistically significant at the 95% level.

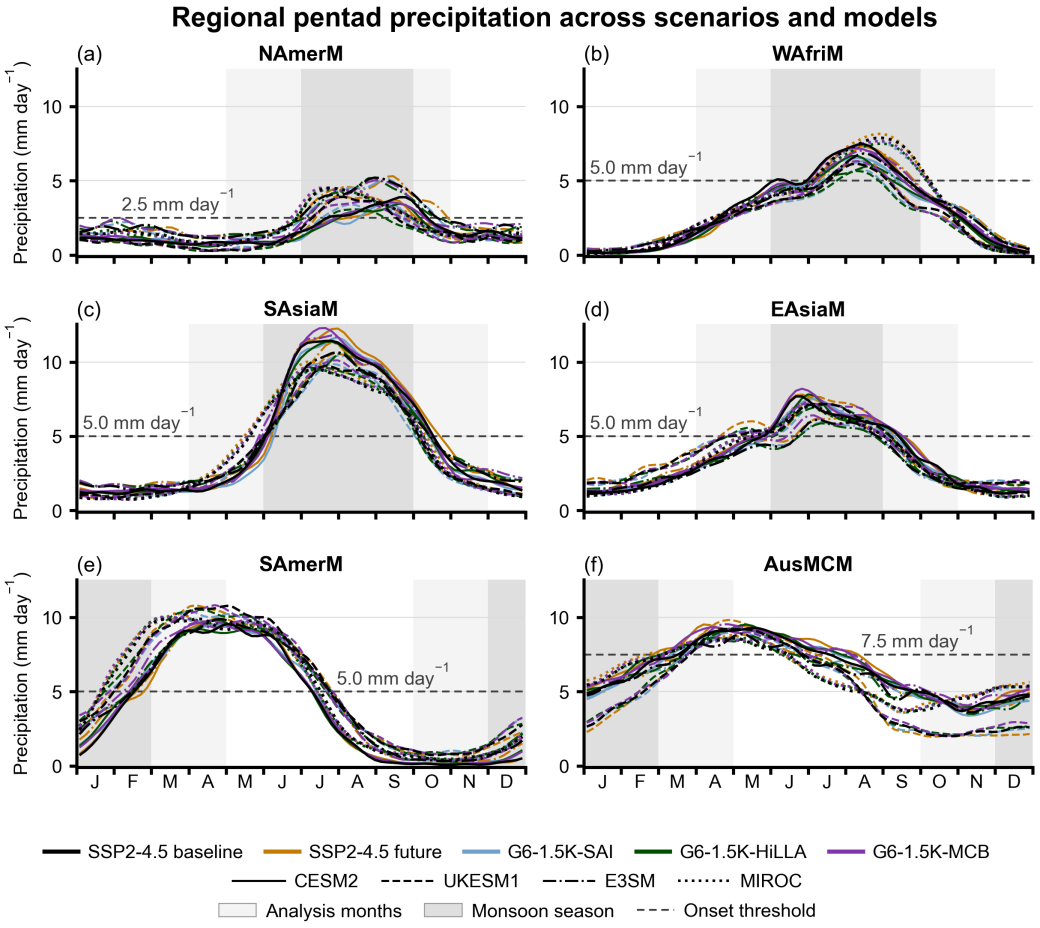


Figure S3: Regional pentad precipitation climatologies for six monsoon regions under SSP2-4.5 and SRM scenarios across CESM2, UKESM1, E3SM, and MIROC: (a) North American monsoon (NAmerM), (b) West African monsoon (WAfriM), (c) South Asian monsoon (SAsiaM), (d) East Asian monsoon (EAsiaM), (e) South American monsoon (SAmerM), and (f) Australian-Maritime Continent monsoon (AusMCM). Curves show the ensemble-mean climatological pentad precipitation reconstructed using the mean seasonal cycle plus the first 12 Fourier harmonics for the SSP2-4.5 baseline (black), SSP2-4.5 future (orange), G6-1.5K-SAI (light blue), G6-1.5K-HiLLA (dark green), and G6-1.5K-MCB (purple). Line style indicates model (solid: CESM2; dashed: UKESM1; dash-dotted: E3SM; dotted: MIROC). Dark gray shading marks the core monsoon season, light gray shading the analysis window (Table 2 of the main text), and the dashed line the onset threshold.

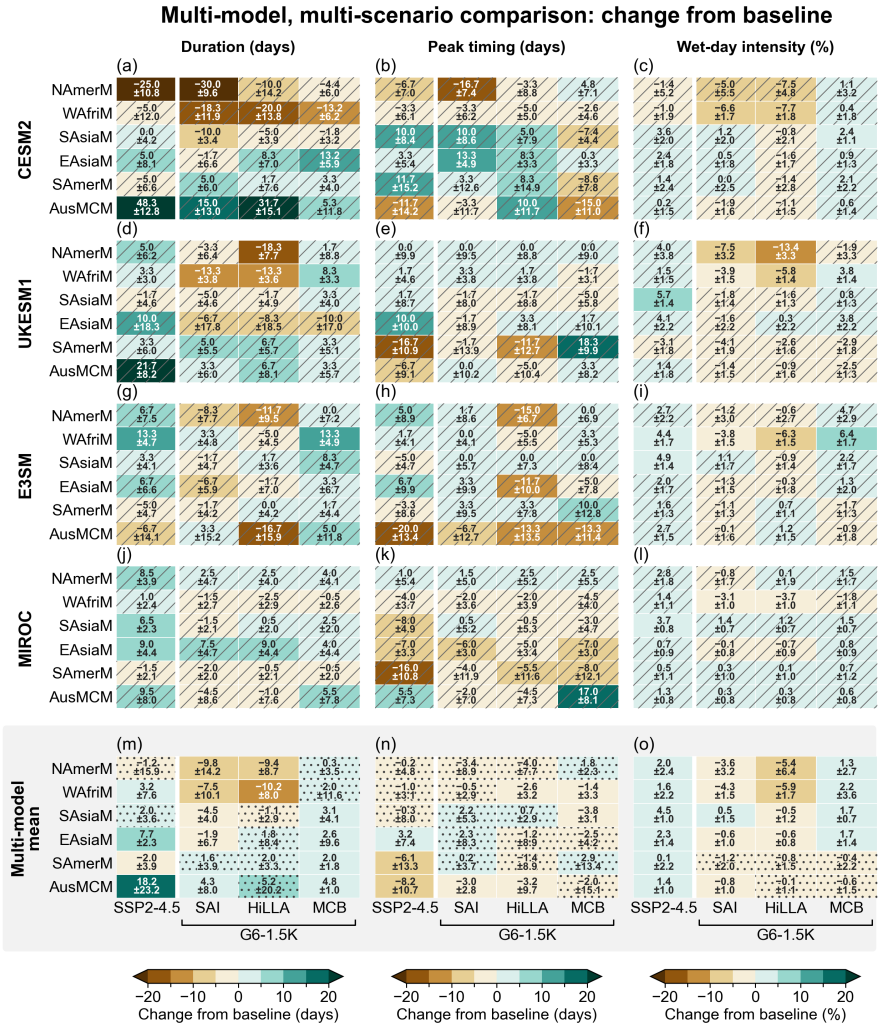


Figure S4: Multi-model, multi-scenario comparison of future monsoon changes across CESM2, UKESM1, E3SM, and MIROC. Panel rows correspond to models and columns correspond to monsoon metrics: duration change (days), peak timing change (days) and wet-day intensity change (%). Within each panel, rows show monsoon regions and columns show scenarios: SSP2-4.5, G6-1.5K-SAI, G6-1.5K-HiLLA, and G6-1.5K-MCB. Text annotations show the ensemble-mean change and one standard deviation derived from bootstrapping (inter-model standard deviation for the multi-model mean). Hatching indicates changes that are not statistically significant (95% hierarchical bootstrap interval includes zero) or, for the multi-model mean, fewer than three of four models agreeing on the sign.

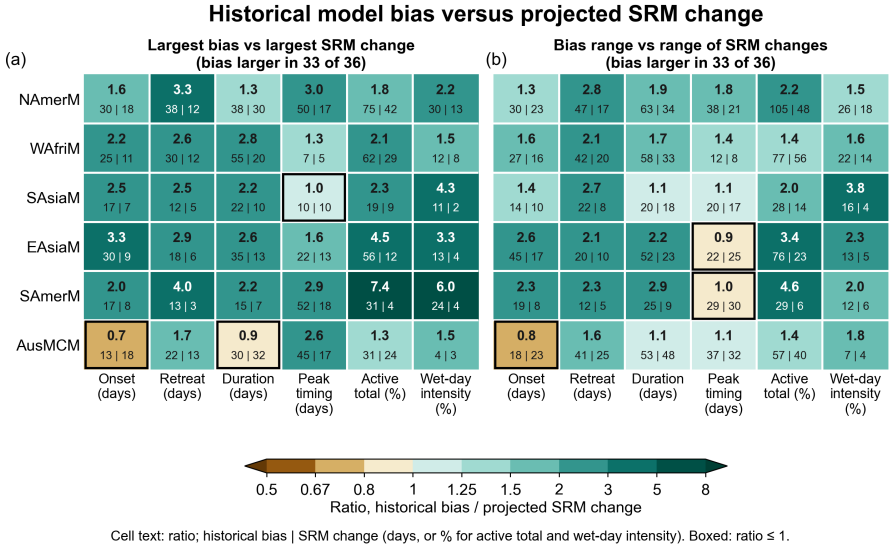


Figure S5: Historical model biases compared with projected SRM changes for the 36 region–metric combinations shown in Fig. 4 of the main text. Both are expressed in physical units: days for onset, retreat, duration and peak timing, and percent for active-season precipitation and wet-day intensity (biases relative to the IMERG value and changes relative to each model's own 2020–2039 baseline). Historical biases are ensemble-mean model values minus IMERG over 1998–2014 (1999–2014 for SAmerM and AusMCM); SRM changes are 2065–2084 changes under G6-1.5K-SAI, G6-1.5K-HiLLA and G6-1.5K-MCB in all four models. (a) Ratio of the largest absolute historical bias across models to the largest absolute SRM change across models and strategies. (b) Ratio of the range of historical biases across models to the range of SRM changes across models and strategies. Values above 1 (teal) indicate that the historical bias is larger than the projected SRM change, and values below 1 (brown) that the projected change is larger. In each cell, the upper number is the ratio and the lower numbers give the historical bias and the SRM change. Boxed cells have ratios of 1 or less (3 of 36 in each panel).

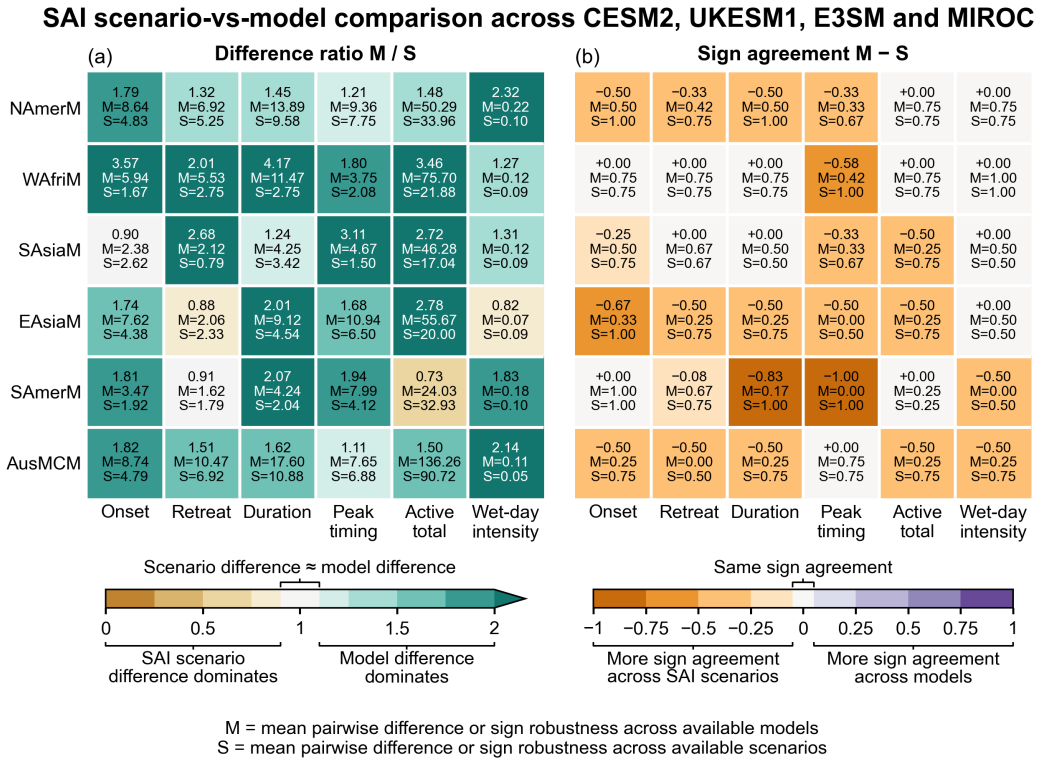


Figure S6: Comparison of model and SAI scenario dependence in projected monsoon changes across CESM2, UKESM1, E3SM, and MIROC for onset, retreat, duration, peak timing, active-season precipitation, and wet-day intensity. (a) Left: Ratio of mean pairwise difference across models to the mean pairwise difference across scenarios for each monsoon region and metric. Values greater than 1 (teal) indicate that differences among models are larger than differences among SAI scenarios, while values less than 1 (brown) indicate that scenario differences dominate. (b) Right: the difference between sign robustness across models and sign robustness across scenarios, where sign agreement is defined as the fraction of pairwise comparisons with the same sign of change. Positive values (purple) indicate greater sign consistency across models, while negative values (orange) indicate greater sign consistency across scenarios.