Science Advances
AAAS

Supplementary Materials for

Toward machine-assisted tuning avoiding the underestimation of uncertainty in climate change projections

Frédéric Hourdin *et al.*

Corresponding author: Frédéric Hourdin, frederic.hourdin@lmd.ipsl.fr

Sci. Adv. **9**, eadf2758 (2023)
DOI: 10.1126/sciadv.adf2758

This PDF file includes:

SI1 to 4
Figs. S1 to S20
Tables S1 to S3

April 22, 2023

SI1: Tuning

We here give additional information on the tuning. The various tuning metrics are computed over different areas, and Fig. S1 shows the masks used to compute the various tuning metrics. The red area, when alone, corresponds to the region were the variable is taken, on average. Metrics associated with maps including both a red and a blue area are computed as the difference between the average on the red minus the blue area (`circAa` stands for circum Antarctica anomaly and `etoa` for East Tropical Ocean anomaly). The radiative metrics consists in yearly averaged values with the given mask and the name of the metrics is the concatenation `mask.variable`. The three rainfall metrics correspond to the global mean rainfall associated with daily rainfall larger than 50 mm (`PR>50`), the mean annual rainfall over a box over West-Africa used for the African Monsoon Mulit-disciplinary Analysis (`AMMA`) campaign, and finally an estimate of the intra-seasonal variability over the ocean in the region of the Madden-Julian-Oscillation (`MJO`), computed from the daily rainfall as the standard deviation of the 20-day running average minus the 120-day running average.

Tab. S1 shows both the target value and the tolerance to error, for each of the 14 metrics used for tuning.

Figs. 2, 3 and 4 show "implausibility matrices" obtained for the 3 tuning waves. Implausibility matrices constitute an attempt to visualize a n-parameter NROY space (here $n = 18$). The matrix itself is divided into 2D sub-matrices, each one being a restriction to two parameters, the names of which are given in the diagonal of the main matrix. For instance, the x-axis in the upper-right sub-matrix corresponds to DZTH and the y-axis to WDENSO. Each axis spans the initial $[min, max]$ range for the parameter considered. Each axis of the sub-matrix is divided into 15 sub-intervals (this number is adjustable within the tool), so that the matrix is made of 225 pixels. From a random sampling of (here) 3×10^5 parameter vectors, we compute the fraction of points with implausibility lower than the cutoff (see main text for the definition of the implausibility), when varying the $n-2$ other parameters. This fraction is displayed on the sub-matrices of the

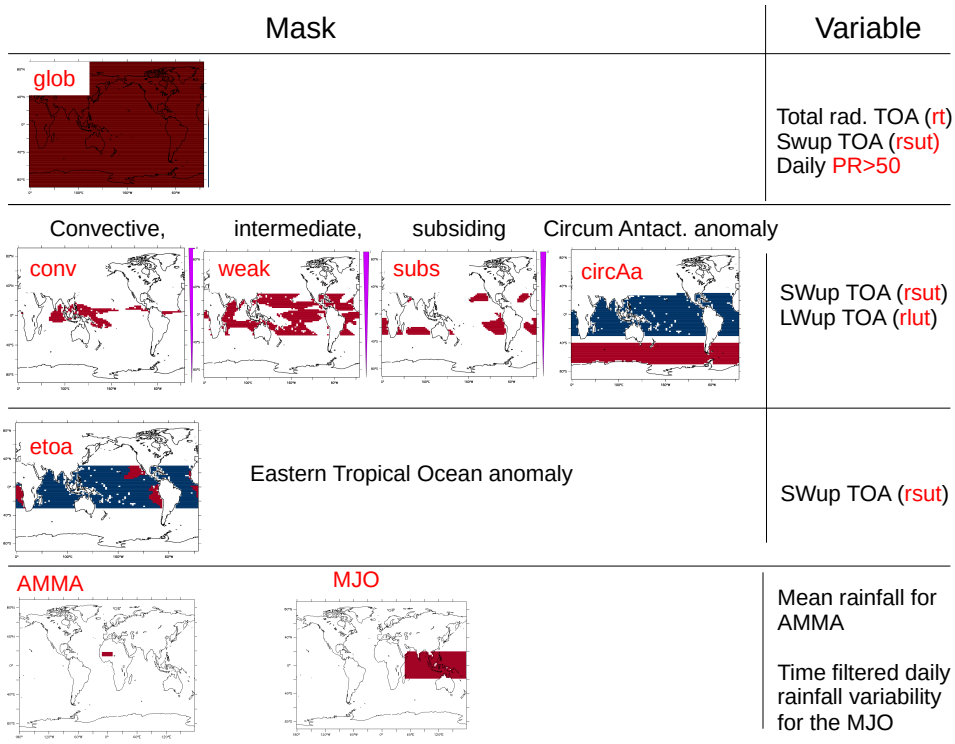

Figure S1: Mask and variables of the various metrics used for tuning.

Name	Unit	target	tolerance
glob.rt	W/m^2	2.7	0.7
glob.rlut	W/m^2	239.68	5
circAa.rsut	W/m^2	23.21	5
circAa.rlut	W/m^2	-44.10	5
subs.rsut	W/m^2	84.90	5
weak.rsut	W/m^2	81.78	5
conv.rsut	W/m^2	103.16	5
subs.rlut	W/m^2	274.64	5
weak.rlut	W/m^2	264.29	5
conv.rlut	W/m^2	235.76	5
etoa.crest	W/m^2	-10.67	5
MJO	mm/day	3.338	0.3
PR>50	mm/day	0.3394	0.17
AMMA	mm/day	1.164	0.63

Table S1: Name, unit, target value and tolerance to error for the 14 metrics used for tuning.

Name	code	range	6A	unit	comment
$wk_{dens,oce}$	WDENSO	$[10^{-11}, 10^{-8}]$	1.e-9	m^-2	(fixed) wake density over oceans
$\gamma_{\text{Alp,th}}$	ALPBLK	$[0.2, 0.8]$	0.5		$Alp = \gamma_{\text{Alp,th}} \times ALP_{\text{thermals}} + \gamma_{\text{Alp,wk}} \times ALP_{\text{wake}}$
$\gamma_{\text{Alp,wk}}$	ALPWKK	$[0.01, 0.5]$	0.25		
$\text{cld}_{l,conv}$	CLDLC	$[0.1, 1]$	0.65	g/kg	threshold for conversion of cloud liquid water to rainfall
$\gamma_{\text{cld,cv}}$	CLDCV	$[0.5, 10]$	1		scaling factor on the convective cloud water for computation of cloud cover
$\gamma_{\text{fall,v}}$	FALLV	$[0.3, 2.]$	0.8		scaling factor on the fall velocity of ice crystals
r_{P_0}	RQSPO	$[300, 600]$	450	hPa	width of the subgrid scale distribution of total water q_t is $\sigma_{q_t} = r \times q_t$
$r_{\Delta p}$	RQSDP	$[50, 300]$	100	hPa	where r varies between $r_{\min}$ (=0.002) and r_{max} with decreasing
$r_{\max}$	RQSTOP	$[0.05, 0.6]$	0.4		pressure p as $r = r_{min} + (r_{max} - r_{min})$ $\{\tanh[(r_{P_0} - P)/r_{\Delta p}] + 1\}/2$
b_0	AERIE	$[0.5, 2]$	1.3		droplet number concentration CDNC=$10^{b_0 + b_1 \log(q_{aer})}$, $b_1 = 0.02$
					q_{aer} being the concentration of soluble aerosols in μg/m^3
σ_{cvpr}	SIGDZ	$[0.001, 0.02]$	0.003		Grid cell fraction covered by unsaturated precipitating downdrafts
$w_{B,\text{srf}}$	WBSRF	$[0.05, 2.]$	0.5	m/s	Deep convection vertical velocity at cloud base (m/s)
$w_{B,\max}$	WBMAX	$[1.5, 6]$	2.8	m/s	$w_B = w_{B,\text{srf}} + w_{B,\max}/[1. + 500/(P_s - P_{LFC})]$, where P_s and P_{LFC}
					are the surface and "Level of Free Convection" pressure in hPa.
EP_{max}	1-OMEPMX	$[0.9, 0.9999]$	0.999		Rainfall efficiency for deep convection is
$q_{l,crit}$	ELCV	$[0.1, 2]$	0.3	g/kg	$EP = \min\{\max[1. - q_{l,crit} * (1.0 - T/T_{l,crit})/q_{\text{liq}}, 0], EP_{max}\}$
$T_{l,crit}$	TLCV	$[-65, -35]$	-55	$^{\text{o}}$C	where q_{liq} is the incloud liquid water (in g/kg).
γ_{ice}	REI	$[0.5, 1.3]$	1.		Scaling factor on an imposed vertical profile of ice particle size
λ_d	DZTH	$[0.05, 0.15]$	0.07		$z^* = z + \lambda_d z$ for plume detrainment

Table S2: List of parameters varied during the history matching procedure. For each parameter, the minimum and maximum values authorized during the tuning process are given, as well as the value used in the 6A configuration (not used in the tuning procedure). The authorized range arises from the expertise of the parametrization developers. The table is essentially reproduced from (37) and additional explanations on the meaning of the various parameters are given there. The second column gives the name of the parameter used in the tuning procedure. WDENSO CLDCV and OMEPMX are explored with a logarithmic rather than linear sampling.

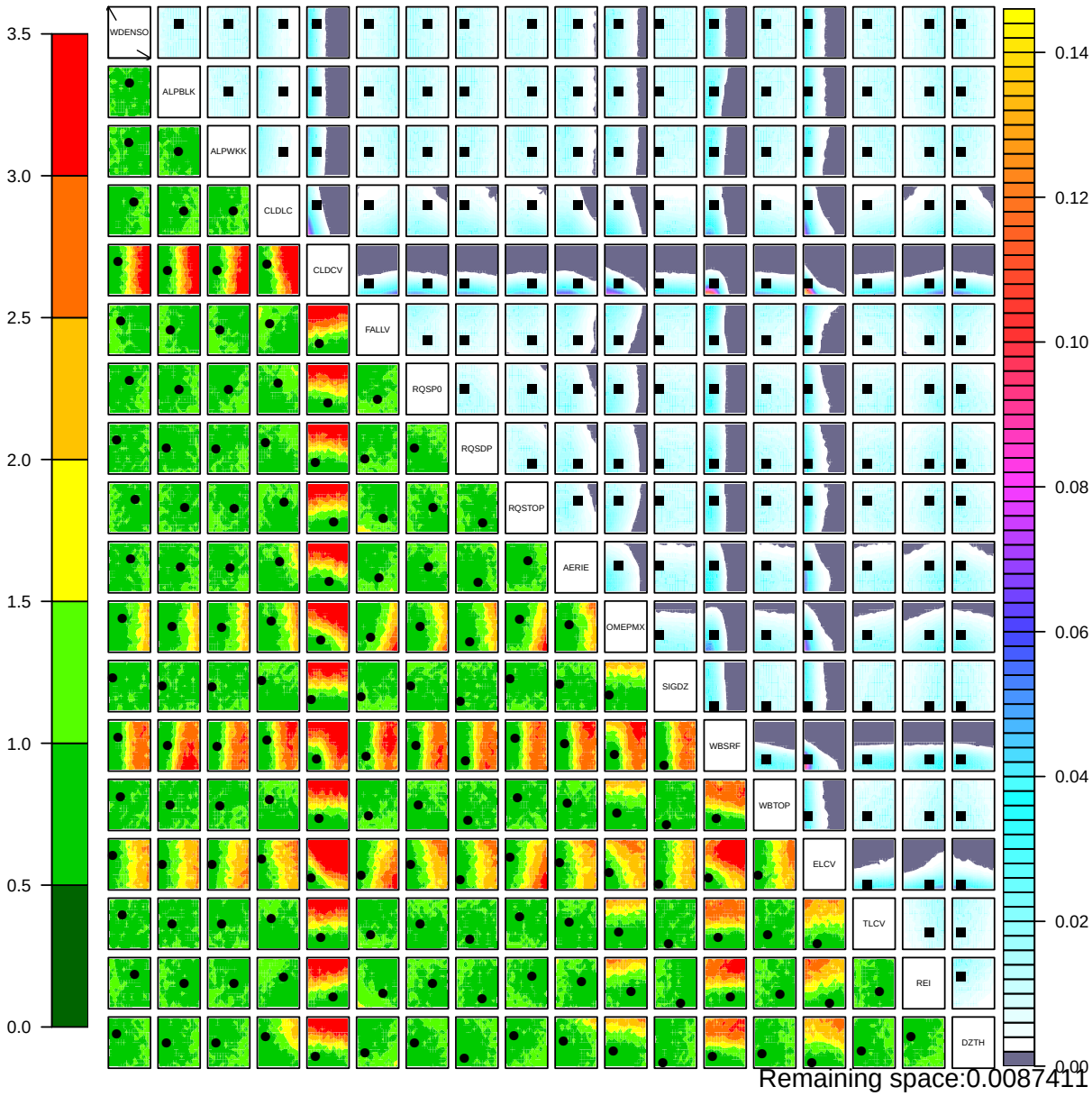

Figure S2: Implausibility matrix for wave 1 of the history matching exploration. The upper-right triangle is made of sub-matrices that display the fraction of points with implausibility lower than the chosen cutoff while the sub-matrices of the lower-left triangle show the minimum value of the implausibility when all the parameters are varied except those used as x- and y-axis, the name of which are given on the diagonal of the main matrix (additional details given in the text). The black squares and dots corresponds to the parameter values of the CTRL (or 6A) configuration.

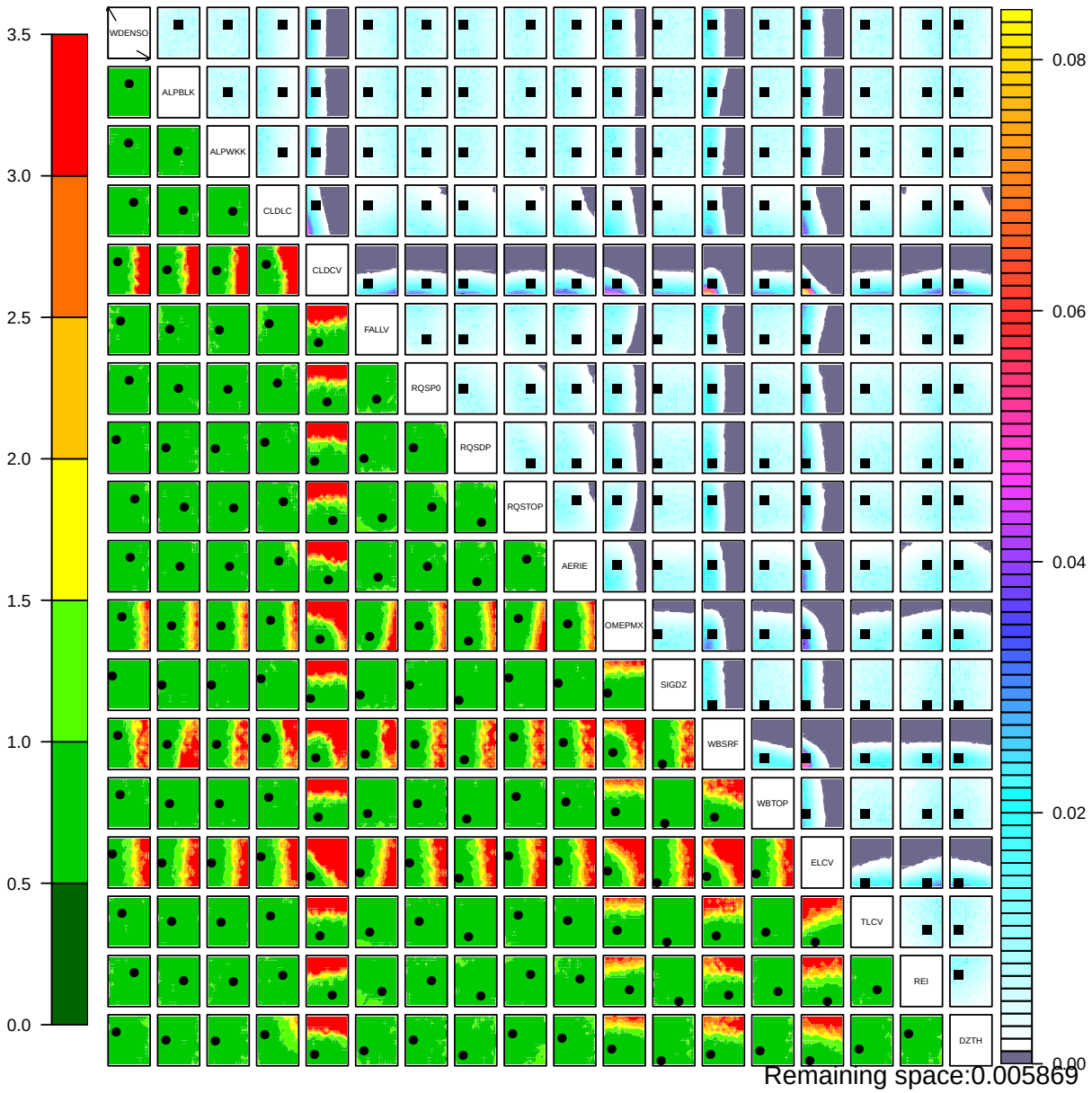

Figure S3: Same as Fig. S2 for wave 2

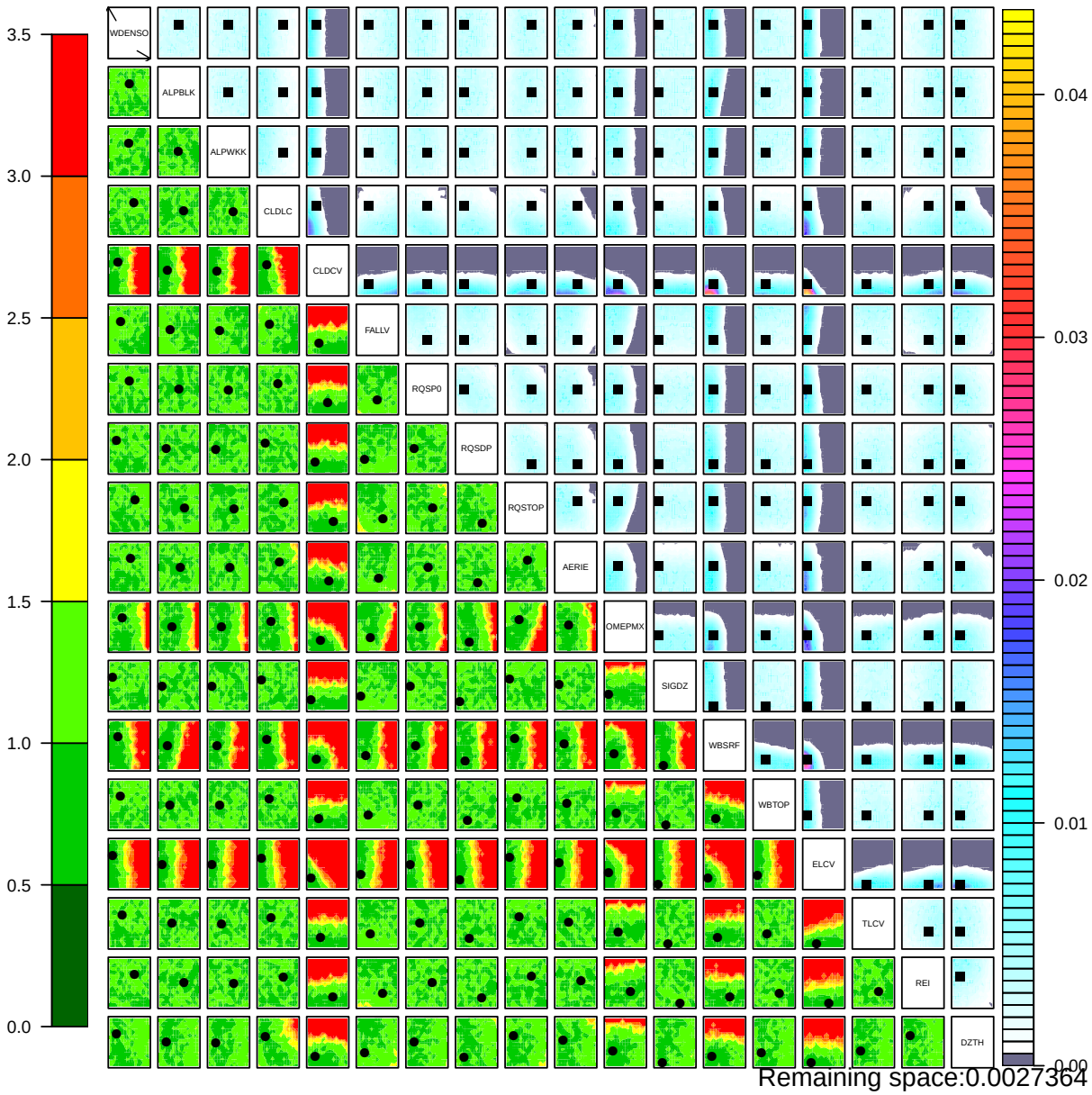

Figure S4: Same as Fig. S2 for wave 3

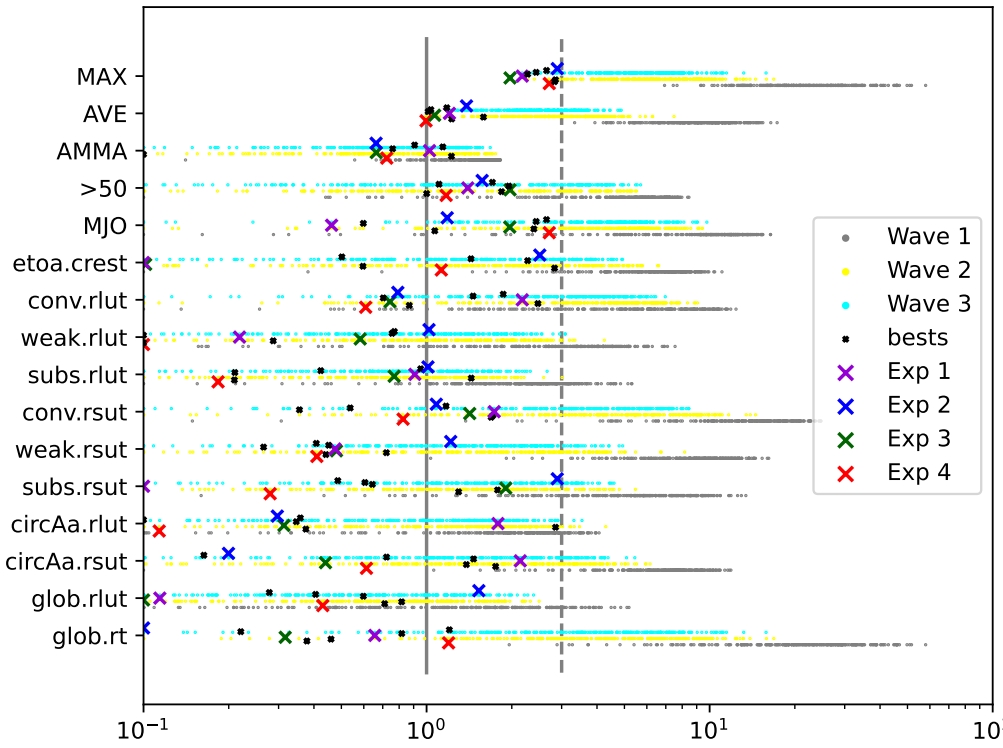

Figure S5: Values of error on each of the 14 metrics, normalized by the tolerance to error. The grey full (resp. dashed) line corresponds to an error/tolerance ratio of 1 (resp. 3). The values of the 14 normalized metrics are shown for each of the tuning simulations (wave 1 to 3), as well as the average (AVE) and maximum (MAX) normalized error across metrics. The MAX is used for the selection of the "best" simulations retained in the study including the Exp 1 to 4 configurations (see main text).

upper-right triangle. The total fraction of the volume of the NROY space relative to the initial n-dimension hypercube corresponding to the a priori $[min, max]$ values of the parameters is the average over the sub-matrix, which should be the same for all the sub-matrices of the upper-right triangle and which is also indicated in text below the figure. A dark grey color means that there is no way to fit the observations by varying the n-2 unfixed parameters while a value of 100% means that values of the two parameters in x and y axis can be retained whatever the values of these n-2 parameters.

The sub-matrices of the lower-left triangle are displaying for each pixel the minimum implausibility obtained when varying the $n-2$ other parameters. A minimum implausibility bellow 2 (to fix ideas), means that, for the bin associated with the two parameters fixed in the sub-matrix considered, there is at least one parameter vector for which the implausibility is smaller than 2 (equivalent to say that the distance of the emulator expectation to the observations is less than twice the root mean square of the tolerance to error and statistical error of the emulator). They are orientated the same way as those on the upper-right triangle, for easier visual comparison, so that the labeling of the axis should be inverted for this lower-left triangle, compared to the names given on the diagonal (i. e. DZTH corresponds to the x-axis and WDENSO to the y-axis for the lower-left sub-matrix as for the upper-right sub-matrix).

Fig. S5 shows the normalized (by tolerance) error for all the tuning metrics and Fig. S6 shows

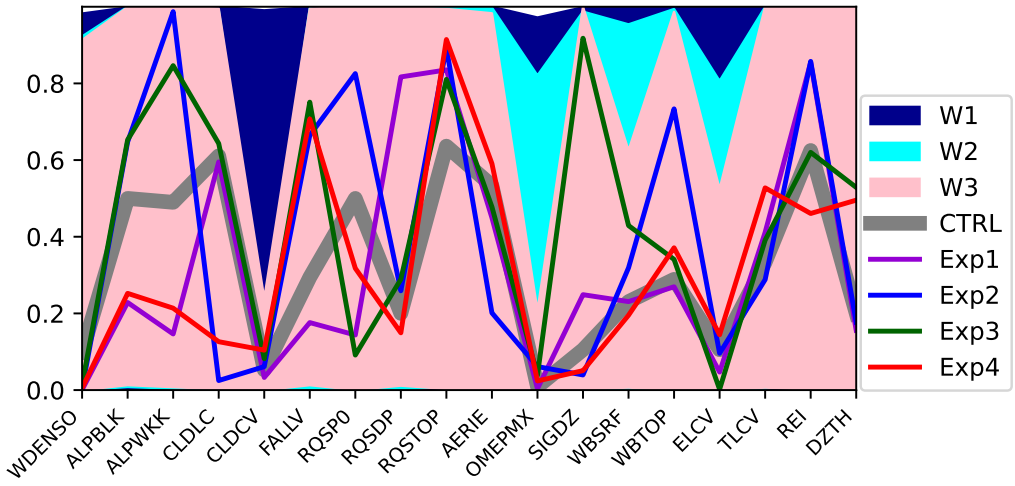

Figure S6: Normalized values of the tuning parameters for the 3 tuning waves (min/max envelopes) and CTRL and Exp1 to 4 simulations. On the y-axis, 0 corresponds to the minimum value and 1 to the maximum value allowed in the tuning procedure.

the range of parameters for the successive waves and for the Exp 1 to 4 simulations.

variable	configuration	6ALR	Exp1	Exp2	Exp3	Exp4
Net TOA (W/m^2)	stand-alone	2.70	2.33	2.90	2.74	1.87
Net TOA (W/m^2)	coupled	0.96	1.05	0.62	1.24	1.01
Global 2m temperature (K)	–	285.9	286.1	285.5	285.9	286.0
ECS (K)	–	4.64	3.85	4.24	4.53	5.35

Table S3: Net TOA radiative balance (in W/m^2) at in the stand-alone (averaged over 10 years) and coupled configurations (averaged over 50 years after a 100-year spin-up), and global mean temperature (K, same 50-year period) and ECS (K) for the control configuration 6ALR and Exp 1 to 4.

SI2: ECS and radiative feedback

In Fig. S7, we compare ways of computing the ECS and radiative feedback from coupled and atmosphere-alone simulations. Panels **a** and **b** illustrate Gregory's approach to compute the ECS and ERF from a pair of `piControl` and `abrupt4`×$\mathtt{CO_2}$ coupled experiments.

Panel **c** compares the coupled and atmosphere-alone estimates of the radiative feedback, and their decomposition into clear sky (CLR) and cloud radiative effect (CRE). From coupled simulations, the radiative feedback is computed as the slope of the regression line between flux imbalance ΔF and difference in global air surface temperature between the `abrupt4`×$\mathtt{CO_2}$ and `piControl` simulation (ΔT). For the stand-alone atmospheric simulations, $\gamma = \Delta F/\Delta T$ where Δ stands for the difference between the `clim+4K` and `clim` simulations. There are many reasons that explain that the coupled and atmosphere-alone estimates differ, which are discussed in dedicated papers (51). Noticeably, the agreement is better for the CRE than for the clear sky contributions.

Because of these differences, there are several ways of estimating the ECS from atmosphere-alone simulations. It can first be expressed directly as ECS=ERF/γ where gamma is the radiative feedback and ERF$\simeq$4 W/m^2. This "simple" estimate (in panel **c**) is the one used to select simulations with small and large ECS, and is displayed in Fig 2 of the paper (stars). In the case of our model, this direct estimate gives significantly lower values of the ECS than the final estimate from the pair of coupled simulations, however the range and ranking are quite similar.

Mauritsen and collaborators (42) proposed to partly palliate this by computing the ECS relatively to the reference configuration (for us the 6A configuration of the IPSL coupled model), ECS=ECS$_{6A}$ × γ_{6A}/γ, ECS$_{6A}$ being the value computed from coupled simulations with the 6A configuration, and γ_{6A} and γ being computed from `clim` and `clim+4K` simulations. This "Mauritsen" estimate in panel **d** gives the same ranking (by construction), a similar range but a better absolute value (by construction also). However, it makes no difference for selecting simulations with small and large ECS values. Therefore the "simple" approach (panel c) is used to select simulations with small and large ECS.

We show as well the estimates in which we take only one contribution to the change of radiative feedback, like the total CRE, in which case we compute the ECS as ECS=ECS$_{6A}$ × $\gamma_{6A}/(\gamma_{6A}+\delta\gamma)$, $\delta\gamma$ beeing here the difference in CRE contribution to the radiative feedback between the considered configuration and the reference 6A.

The proxy computed by considering only the change in CRE may give even closer results, and it is for a large part explained by the SW contribution to the CRE. Other ways of correcting feedbacks are proposed in the literature (51).

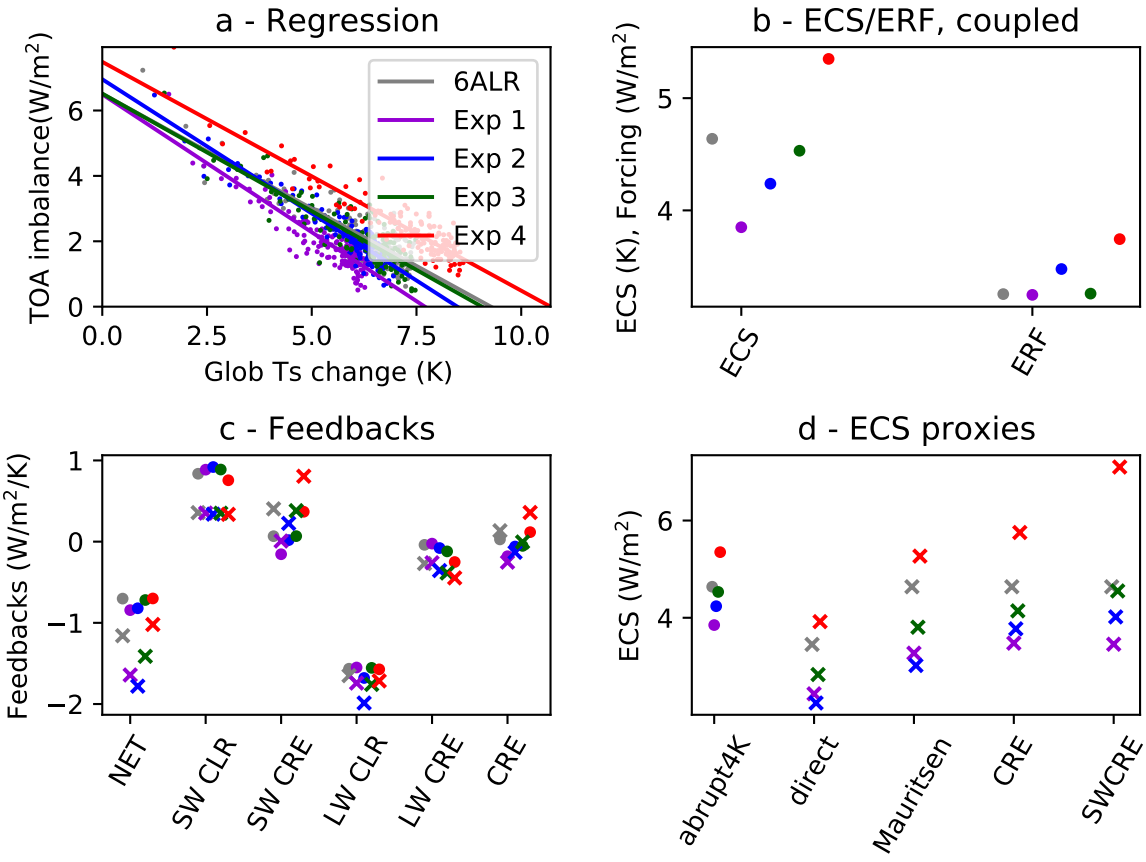

Figure S7: ECS and radiative feedback computed from pairs of coupled `piControl`/`abrupt4×CO2` simulations (•) and pairs of stand-alone forced atmospheric simulations (x) forced by climatological SST (`clim`) or SST increased by 4K (`clim+4K`). **a,b**: computation of the ECS and ERF with Gregory's approach, from the linear regression of the relationship between the global radiative imbalance at TOA and the global surface air temperature (yearly averaged). For each model configuration, the ECS (resp. ERF) shown in **b** is the intercept of the regression line with the x-axis (resp. y-axis). **c**: radiative feedback and its decomposition between clear sky and cloud radiative effect (CRE) for both long-wave (LW) and short-wave (SW) radiation. **d**: comparison with the `abrupt4×CO2` estimate of various proxies computed from the atmosphere-alone simulations (see text).

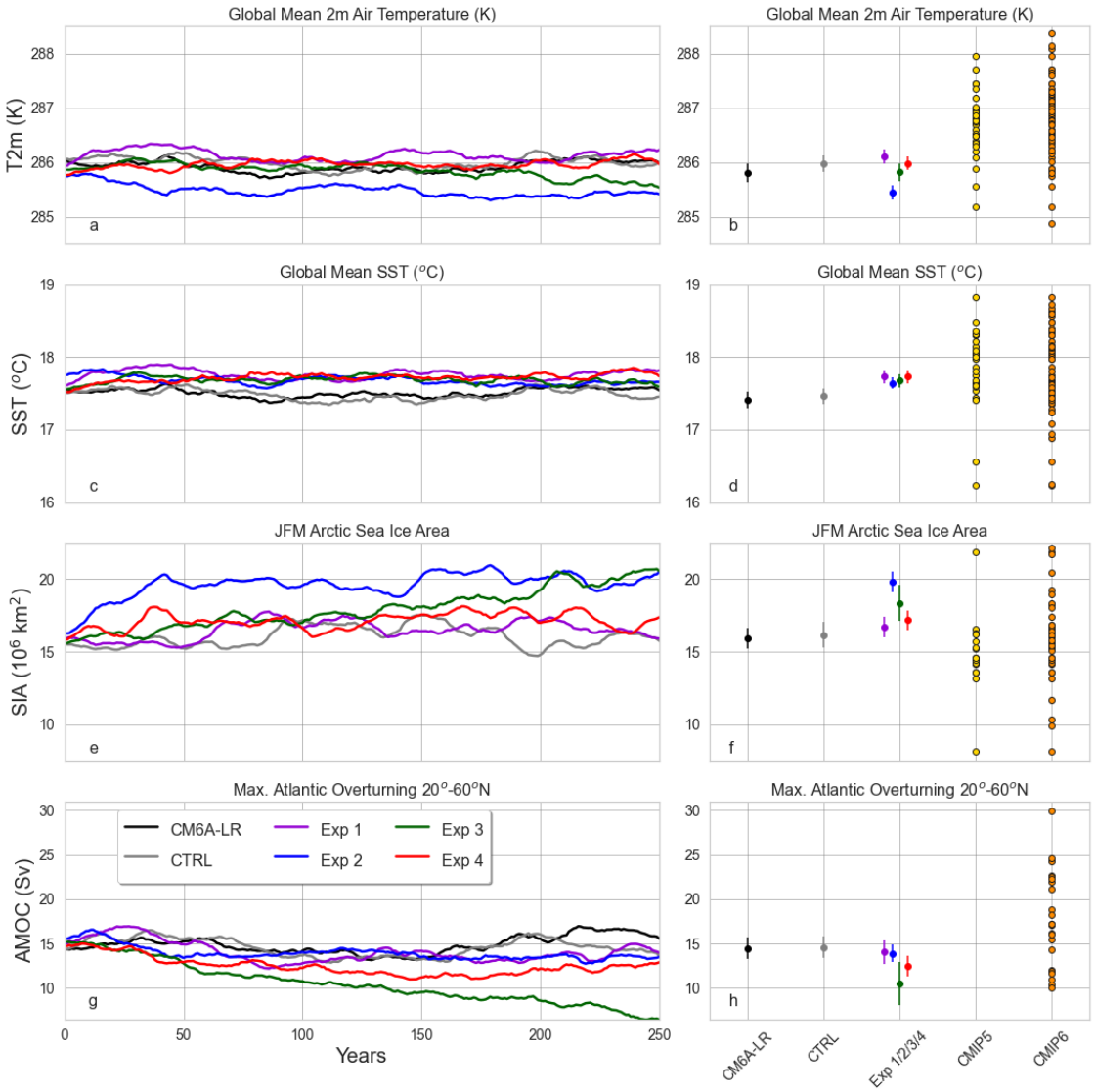

Figure S8: Time evolution of the global surface air temperature, global sea surface temperature, January-Frebruary-March Arctic sea ice cover and maximum intensity of the Atlantic meridional overturning circulation, for the IPSL-6A configuration and the CTRL et Exp 1 to 4 configurations. The dots on the right panels show the values averaged on the last 100 years, for those 6 configurations as well as for the CMIP5 and CMIP6 multi-model ensembles.

SI3: Evaluation of the coupled simulations

Here we present supplementary information on the performance of the Exp 1 to 4 coupled simulations. We first present in Fig. S8 time evolution of key variables which were monitored with a particular care during the tuning phase of the IPSL-6A model configuration.

In Tab. S3 we show the global imbalance at the atmospheric top for the stand-alone and coupled-to-ocean atmospheric simulations, as well as the global near surface (at 2m) air temperature. Note that the numbers shown fluctuate by typically 0.2 (W/m^2 or K) on centennial time-scales and that the simulations have not fully reached a steady-state regime. As explained in the main text, the global imbalance at the atmospheric top was a target of the tuning. More precisely, we targeted the value of the reference configuration 6ALR, since we new that an imbalance of 2.7 W/m^2 in the stand-alone configurations was associated with a steady-state climate in the coupled configuration. In this coupled configuration, the TOA imbalance of 0.9 W/m^2 correspond to an energy leak of the coupled model, attributed mostly to the atmospheric component. For the Exp 1 to 4 configuration, the TOA imbalance varies from 1.8 to 2.9 W/m^2 in the stand-alone configuration and from 0.6 to 1.2 W/m^2 in the coupled configuration. One assumption for the spread in the coupled mode is that it is related to a long-term drift in the global SSTs. In this case, the simulation Exp 2 (resp 3) may cool (resp warm) by a few tenth K. Anyhow, it is clear that with this methods, we succeeded to adjust SSTs in the coupled experiments to less that 1K by adjusting the TOA imbalanced in the stand-alone atmospheric simulations. Note that there is no relation between this tuning of the imbalance and the ECS (last line in the table).

In Fig. 9, 10 and 11, we show for each metrics considered two panels. The three upper panels are similar to those shown in Fig. 3 of the main text. They consist in root-mean-square error (RMSE) computed with respect to the observed mean seasonal cycle. In order to remove the global warming trend when considering temperature (Fig. 11), we show RMSC computed after removing the mean bias between the simulated and observed mean seasonal cycle. In order to compare models with different native resolutions, fields are first interpolated on a regular $3° \times 2°$ grid with a conservative regridding scheme before computing the global-mean RMSE against observations.

As explained in the main text, the CMIP5 and CMIP6 piControl simulations are displayed to contextualize our results. For the IPSL-6A configuration (one of the CMIP6 models), we show the same metrics computed for present-day conditions (1979-2005) on a 33-member ensemble of historical simulations described in (55).The various points for the IPSL-6A configuration in piControl conditions correspond to successive periods of 30 years. Comparing "IPSL-6A Ens" and "IPSL-6A pi" gives an idea of the contribution of the climate change trend to the considered metrics.

The lower panels are computed exactly as the upper ones except that observations are replaced by the IPSL-6A pi simulations for the 30-year period retained for the CMIP6 ensemble. By construction, both the "IPSL-6A pi" and "CMIP6" ensembles should show one dot on the 0 of the y axis.

The three last figures show more focused diagnostics that concern specific model biases some of which are rather systematic across models. On each figure, a map is shown with three highlighted regions. All the diagnostics consist in biases (BIAS) except the left panel of Fig. S13 which shows RMSC. Because biases are linear, the difference with "IPSL-6A pi" is readable directly from the figure. Fig. S12 focuses on tropical precipitation biases over the pacific ocean and Amazonia, with a particular focus on the double ITCZ pattern, and over rainfall over Amazonia. The last two figures show summer (June-July-August) temperature (Fig. S13) and precipitations biases (Fig. S14) over three continental regions of the northern hemisphere.

SI4: change in mean annual key variables

We show bellow the change in various key variables for IPSL-6A, Exp 1 and Exp 4, following Fig 4 of the main text, with a change after 30 years in the right column and at a given global temperature change of 5.5 K on the right column.

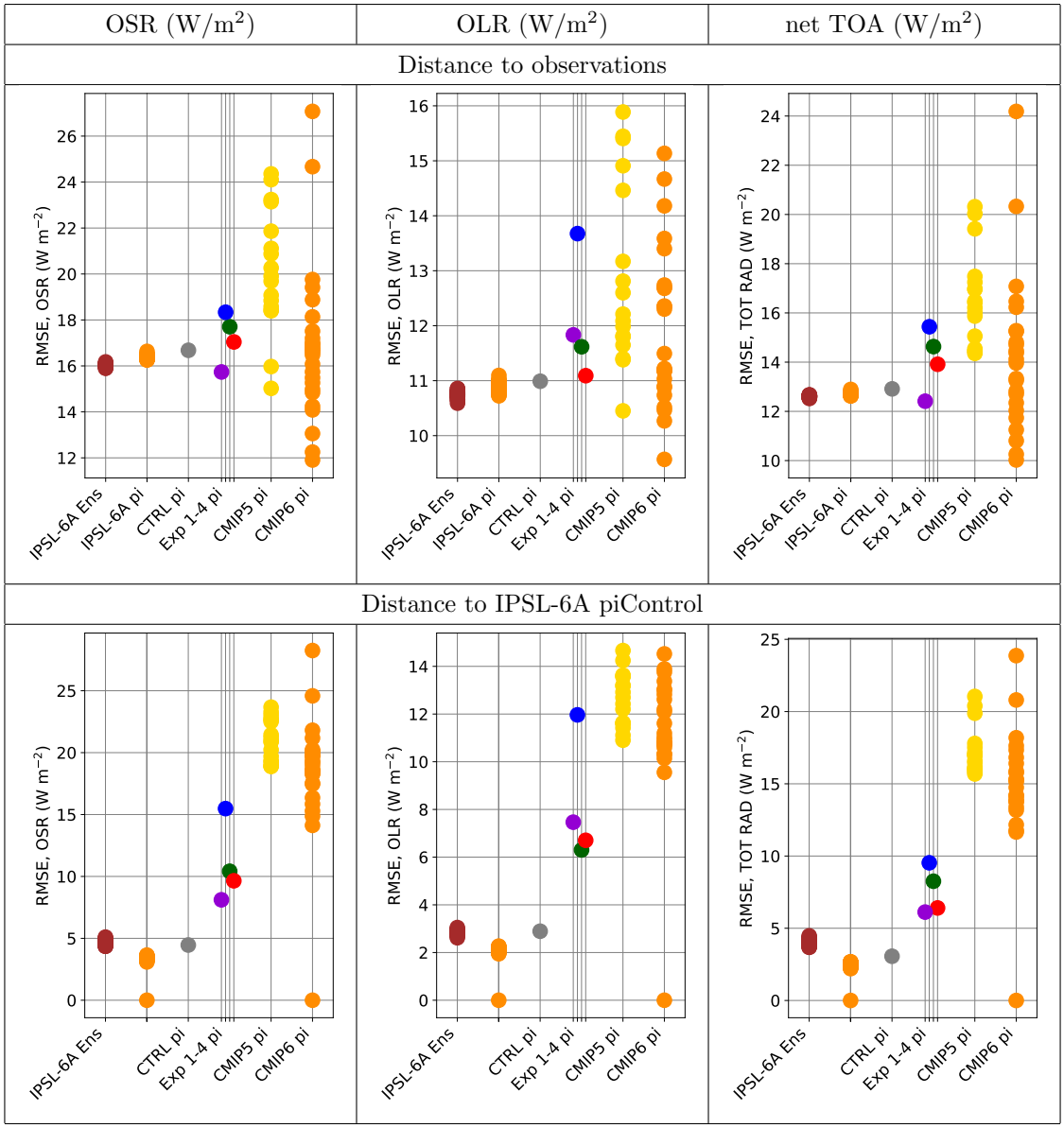

Figure S9: Root-mean-square error (RMSE in W m^{-2}) on the mean seasonal cycle of the outgoing shortwave radiation (OSR), outgoing longwave radiation (OLR) and total radiation (net TOA). For the three upper panels, the RMSE is computed with respect to the CERES EBAF climatology. For the lower panels, the RMSE is computed with respect to a 30-year average of the IPSL-6A piControl simulation. The explanations concerning the names of the simulations and ensembles displayed on the x-axis are given in the main text.

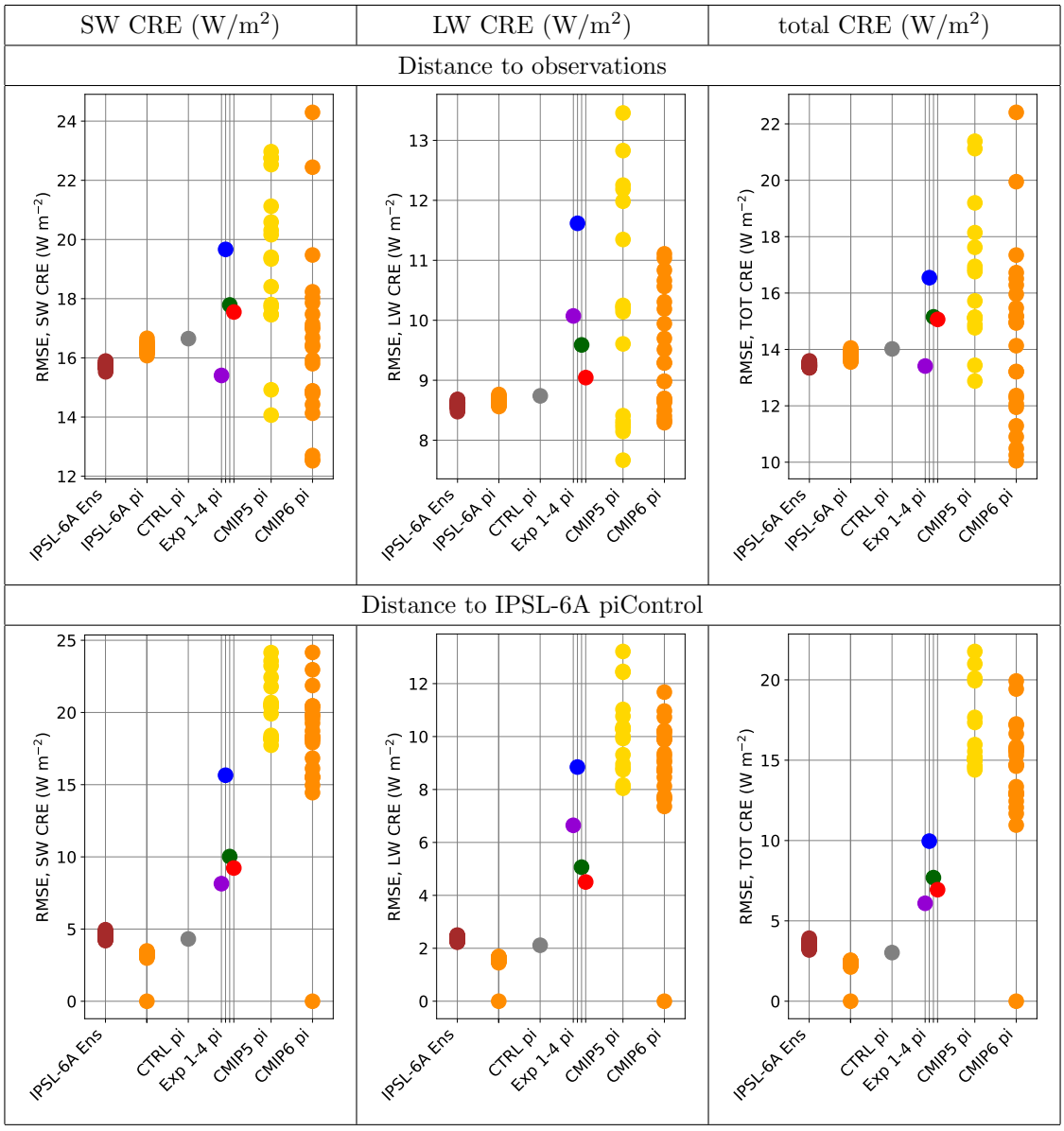

Figure S10: Same as Figure 9 but for the shortwave (SW), longwave (LW) and total cloud radiative effect (CRE).

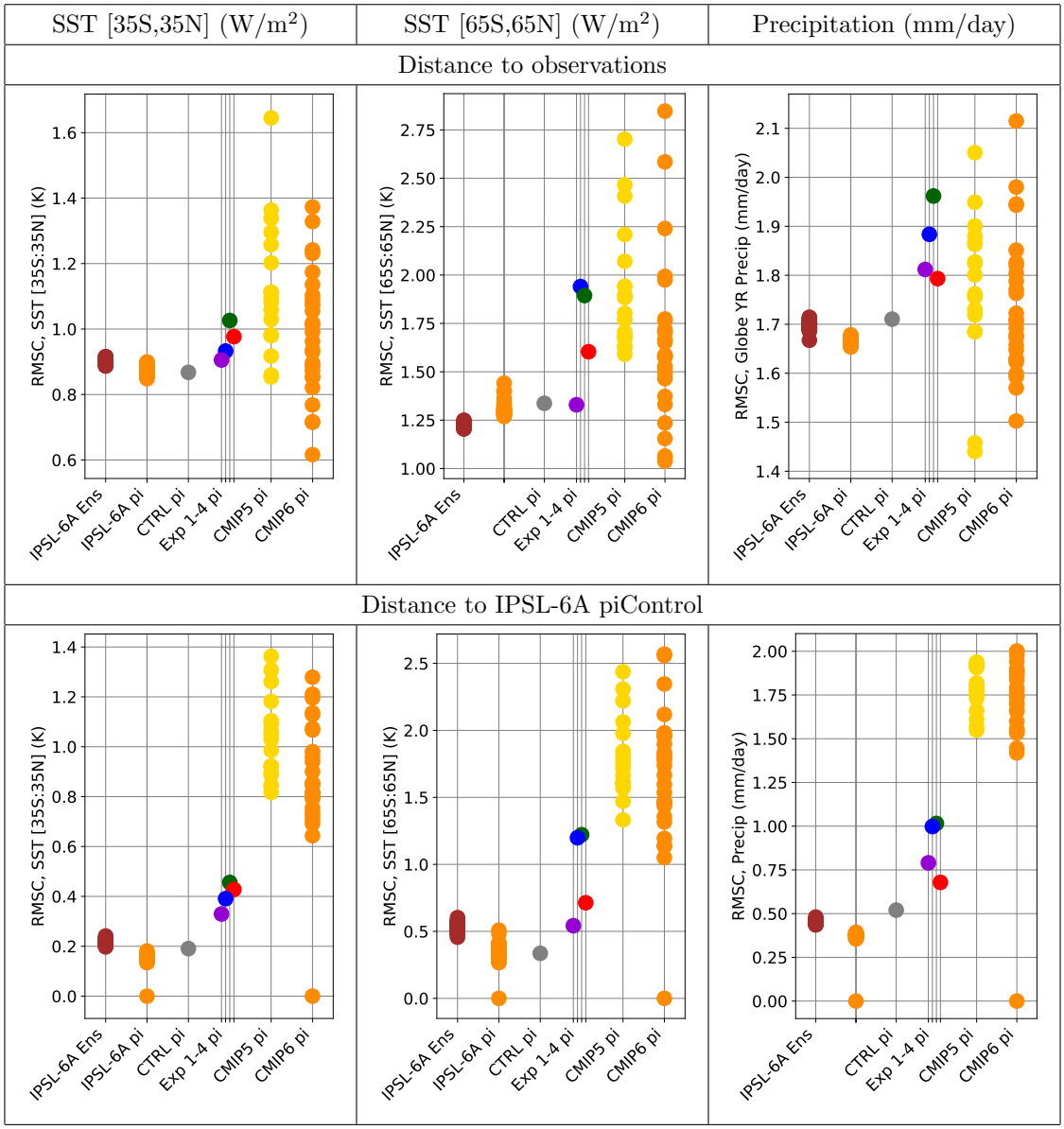

Figure S11: Root-mean-square error computed after removing the mean bias (RMSC) on the mean seasonal cycle of the sea-surface temperature (SST, K) in a tropical band (between 35S and 35N) for the first one and a more global band chosen to avoid regions with sea-ice (between 65S and 65N), and on the global mean seasonal cycle of rainfall (mm/day). For the three upper panels, the RMSC is computed with respect to observation : the Amip climatology for SST and GPCP for precipitation. For the lower panels, the RMSC is computed with respect to a 30-year average of the IPSL-6A piControl simulation. The explanations concerning the names of the simulations and ensembles displayed on the x-axis are given in the main text.

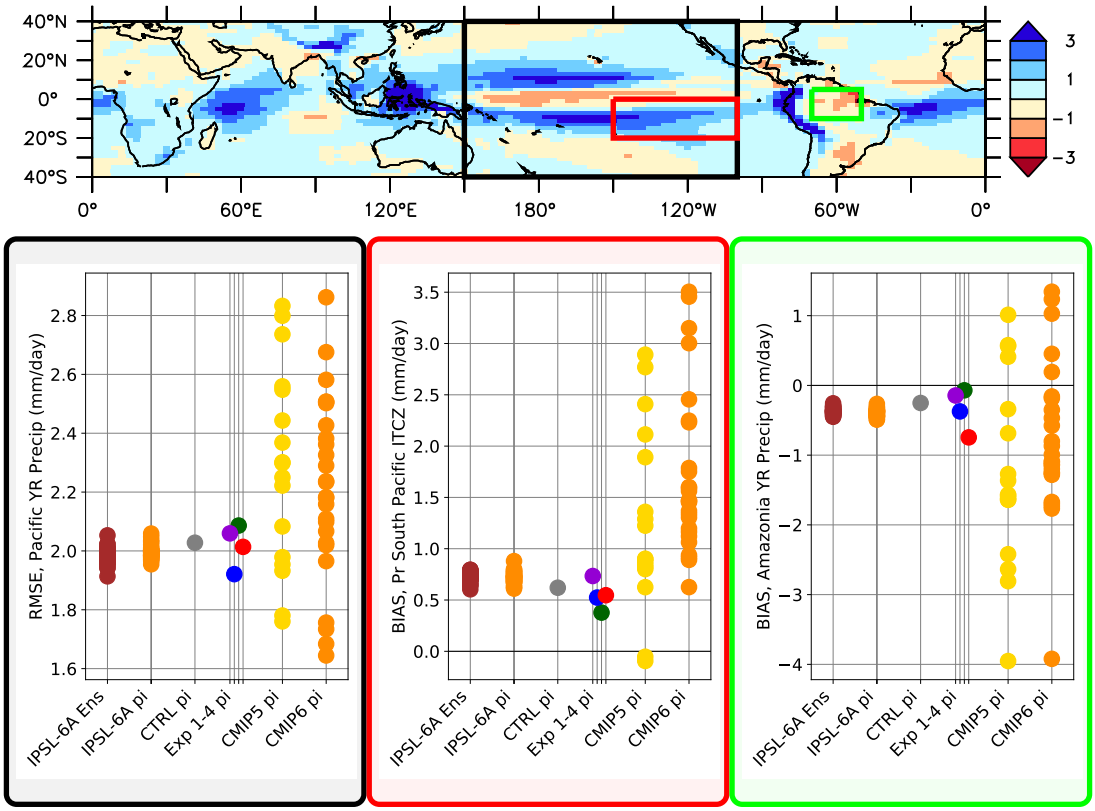

Figure S12: The upper panel shows the multi-model CMIP6 ensemble mean precipitation bias for historical simulations over the period 1979-2005, which guided the choice of the regions to compute the indices. The three bottom panels of the figure are associated with the three color boxes shown on the map. The left panel shows the RMSE of monthly rainfall (in mm day^{-1}) over the tropical pacific ocean [150°E–100°W, 40°S–40°N]. The middle panel shows the rainfall mean bias (in mm day^{-1}) over the south pacific ITCZ (Inter Tropical Convergence Zone), [150°W–100°W, 20°S–0°S]. The right panel shows the rainfall mean bias (in mm day^{-1}) over the Amazonian basin [70°W,50°W,20°S,0°S]. The explanations concerning the names of the simulations and ensembles displayed on the x-axis are given in the main text.

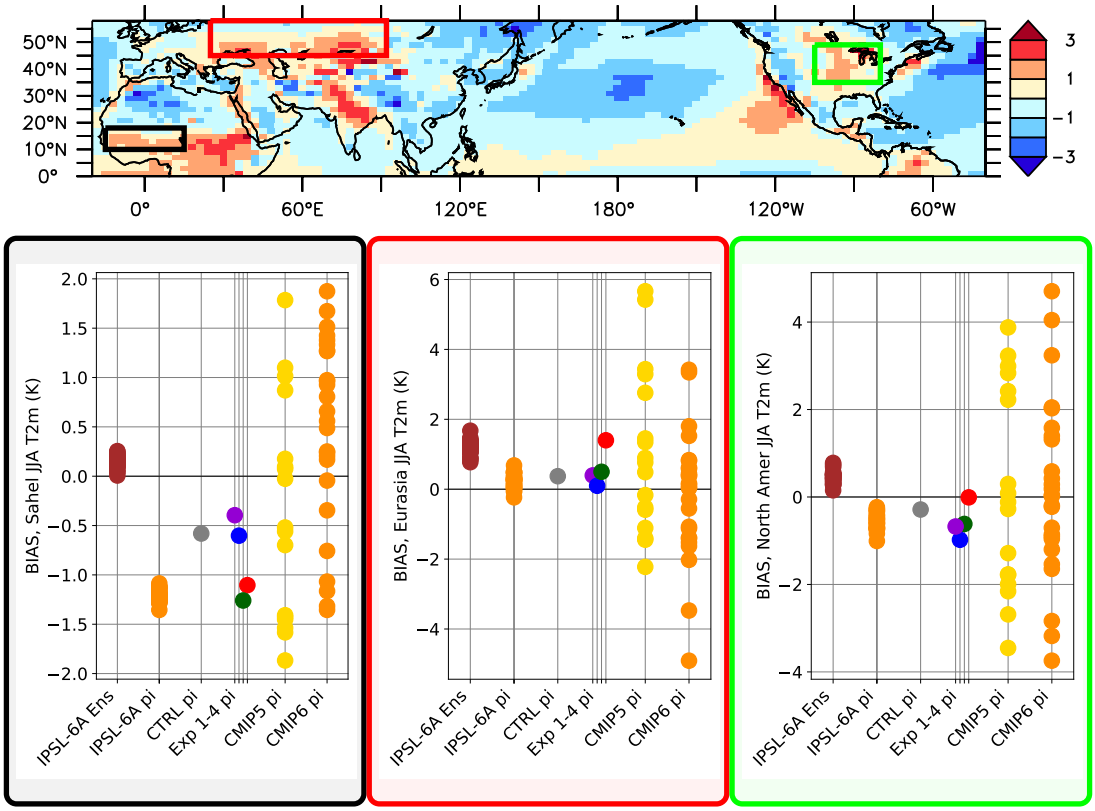

Figure S13: The upper panel shows the June-July-August (JJA) multi-model CMIP6 ensemble mean bias for the near surface air temperature over the period 1979-2005, which guided the choice of the regions to compute the indices. The lower panels show for the three colored rectangles of the upper panel, the mean JJA near surface air temperature bias (in K). The three rectangles are located respectively in Sahel [15°W–15°E,10°N–18°N], Eurasia [23°E–92°E, 45°N–58°N] and North American [105°W–80°W, 35°N–50°N]. The explanations concerning the names of the simulations and ensembles displayed on the x-axis are given in the main text.

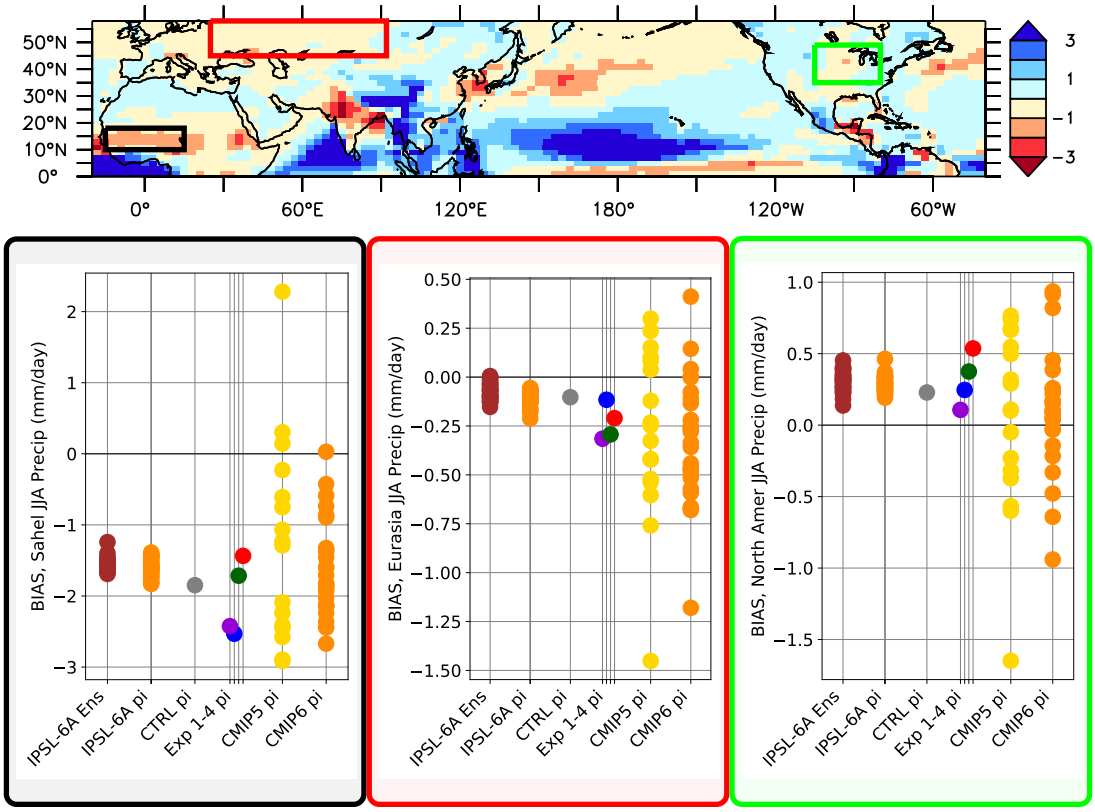

Figure S14: Same as Fig. S13 but for the JJA rainfall (in mm/day).

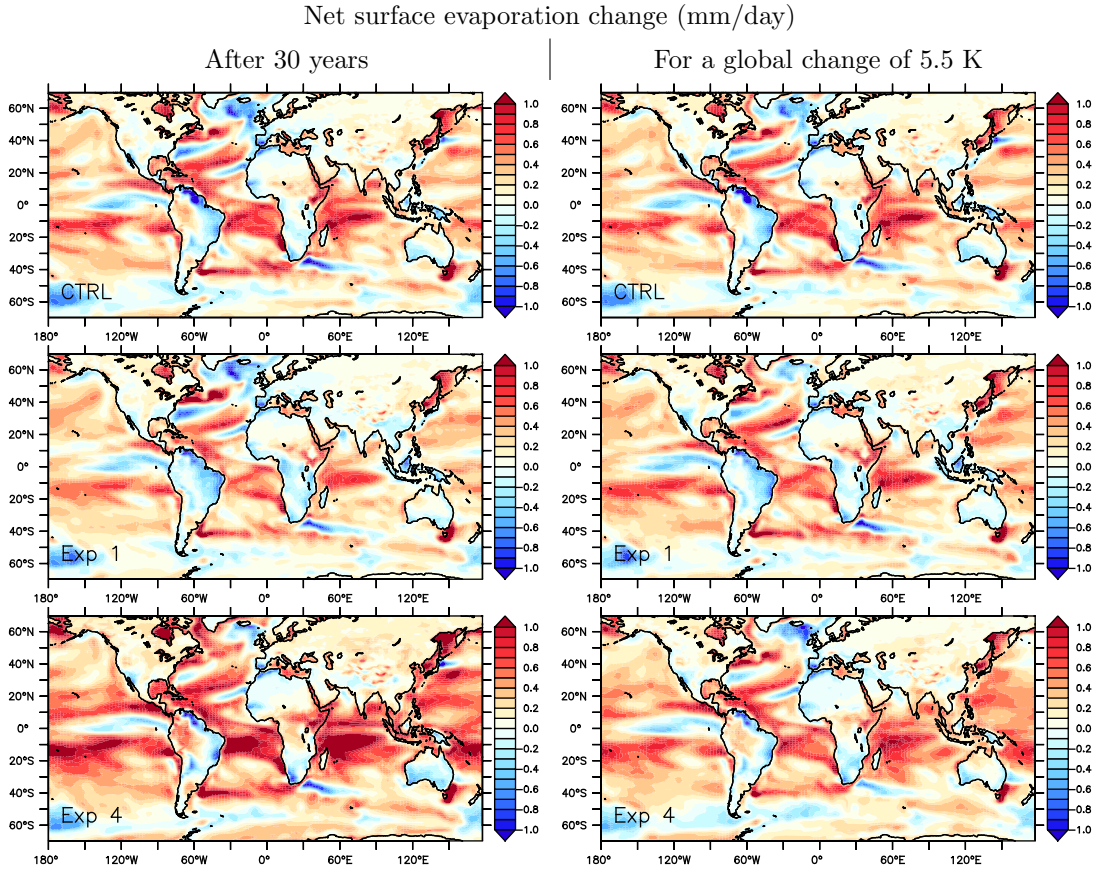

Figure S15: Net surface evaporation change (mm/day) change in coupled simulations CTRL (IPSL-6A configuration), Exp 1 and Exp 4. The maps correspond to differences between the `abrupt4`×CO_2 and `piControl` simulations, averaged over 21 consecutive years. For the left column, this time is centered at year 30 of the simulations. For the right column, it is centered at the time when the global 2m-temperature has increased by 5.5 K: *i. e.* at year 33 for CTRL, year 65 for Exp 1 and year 17 for Exp 4.

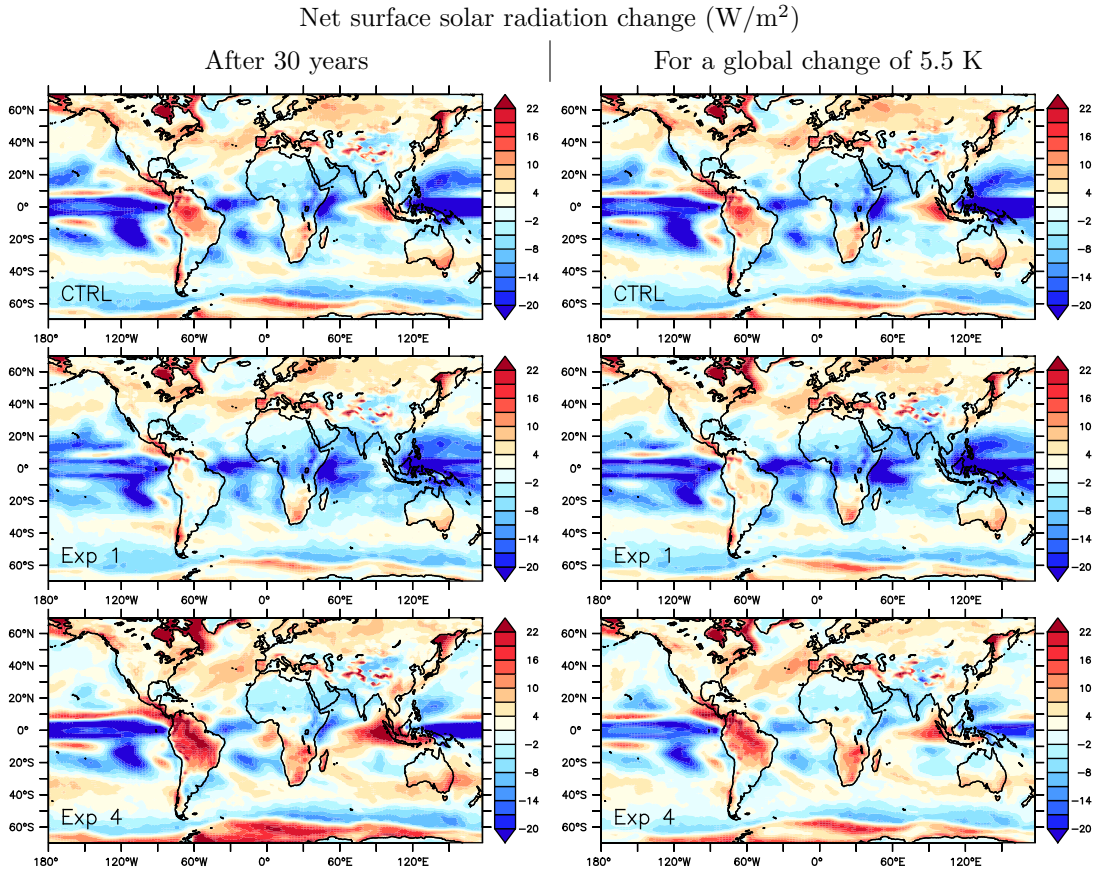

Figure S16: Net surface solar radiation change (W/m^2) change in coupled simulations CTRL (IPSL-6A configuration), Exp 1 and Exp 4. The maps correspond to differences between the `abrupt4×CO2` and `piControl` simulations, averaged over 21 consecutive years. For the left column, this time is centered at year 30 of the simulations. For the right column, it is centered at the time when the global 2m-temperature has increased by 5.5 K: *i. e.* at year 33 for CTRL, year 65 for Exp 1 and year 17 for Exp 4.

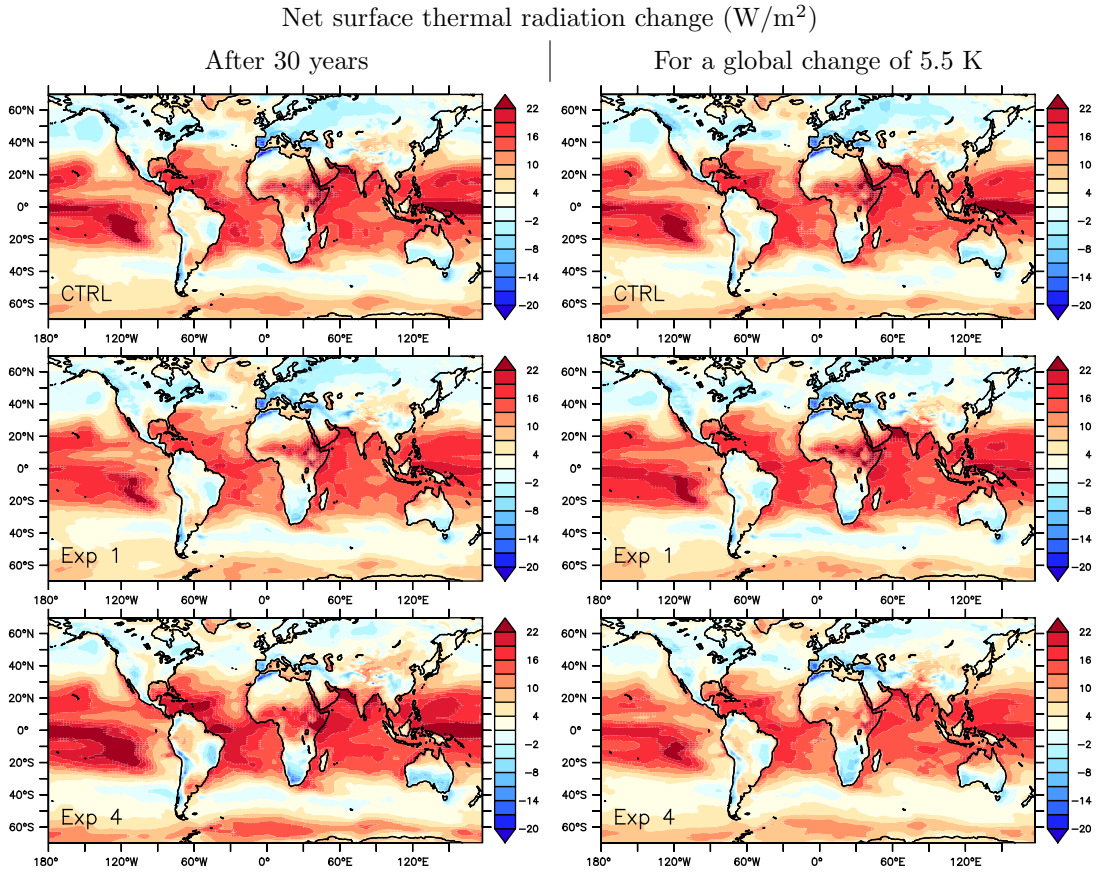

Figure S17: Net surface thermal radiation change (W/m^2) change in coupled simulations CTRL (IPSL-6A configuration), Exp 1 and Exp 4. The maps correspond to differences between the `abrupt4`×$\mathtt{CO_2}$ and `piControl` simulations, averaged over 21 consecutive years. For the left column, this time is centered at year 30 of the simulations. For the right column, it is centered at the time when the global 2m-temperature has increased by 5.5 K: *i. e.* at year 33 for CTRL, year 65 for Exp 1 and year 17 for Exp 4.

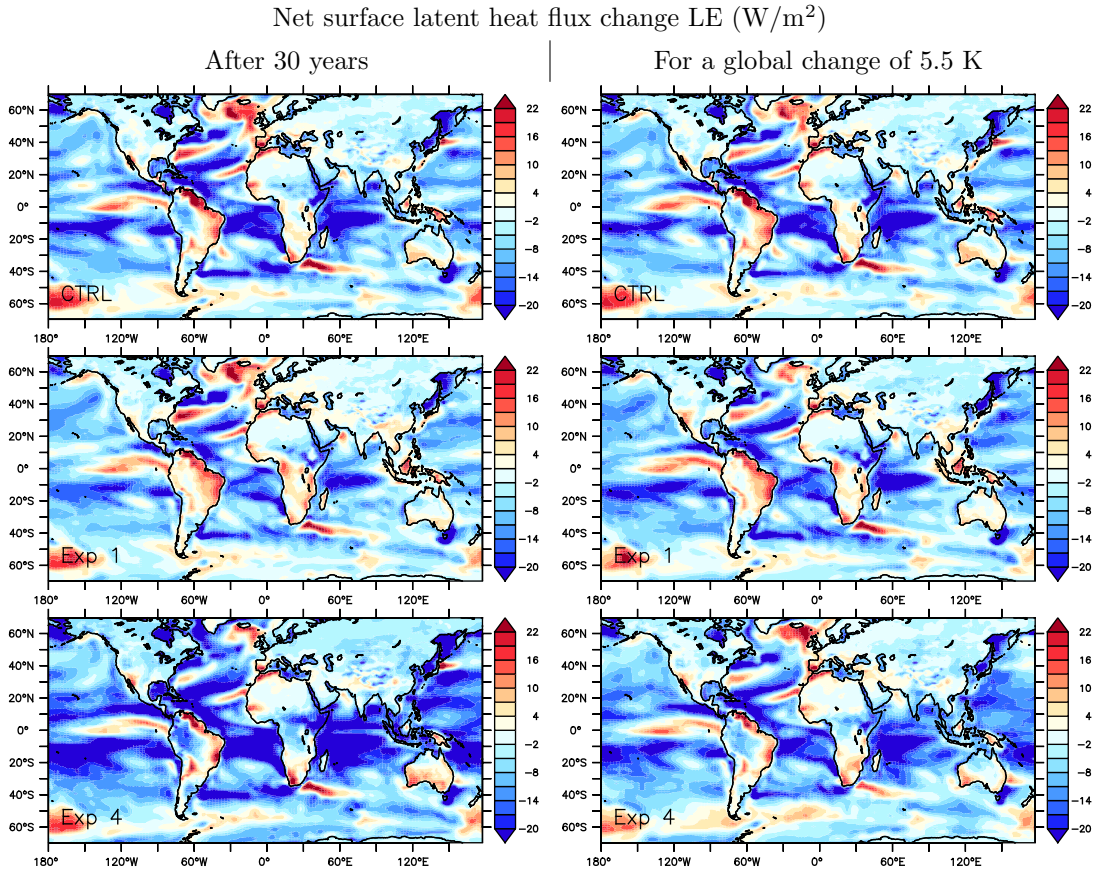

Figure S18: Net surface latent heat flux change LE (W/m^2) change in coupled simulations CTRL (IPSL-6A configuration), Exp 1 and Exp 4. The maps correspond to differences between the `abrupt4`×$\mathtt{CO_2}$ and `piControl` simulations, averaged over 21 consecutive years. For the left column, this time is centered at year 30 of the simulations. For the right column, it is centered at the time when the global 2m-temperature has increased by 5.5 K: *i. e.* at year 33 for CTRL, year 65 for Exp 1 and year 17 for Exp 4.

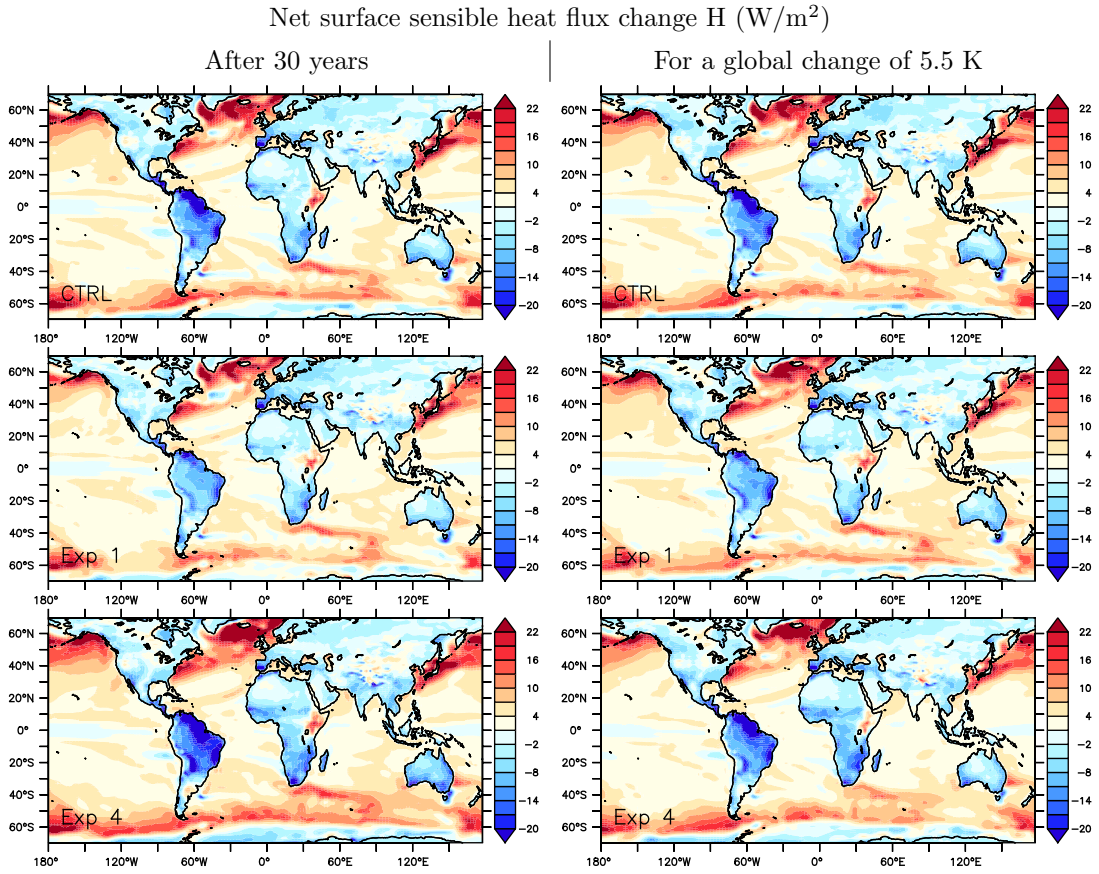

Figure S19: Net surface sensible heat flux change H (W/m^2) change in coupled simulations CTRL (IPSL-6A configuration), Exp 1 and Exp 4. The maps correspond to differences between the `abrupt4`×$\mathtt{CO_2}$ and `piControl` simulations, averaged over 21 consecutive years. For the left column, this time is centered at year 30 of the simulations. For the right column, it is centered at the time when the global 2m-temperature has increased by 5.5 K: *i. e.* at year 33 for CTRL, year 65 for Exp 1 and year 17 for Exp 4.

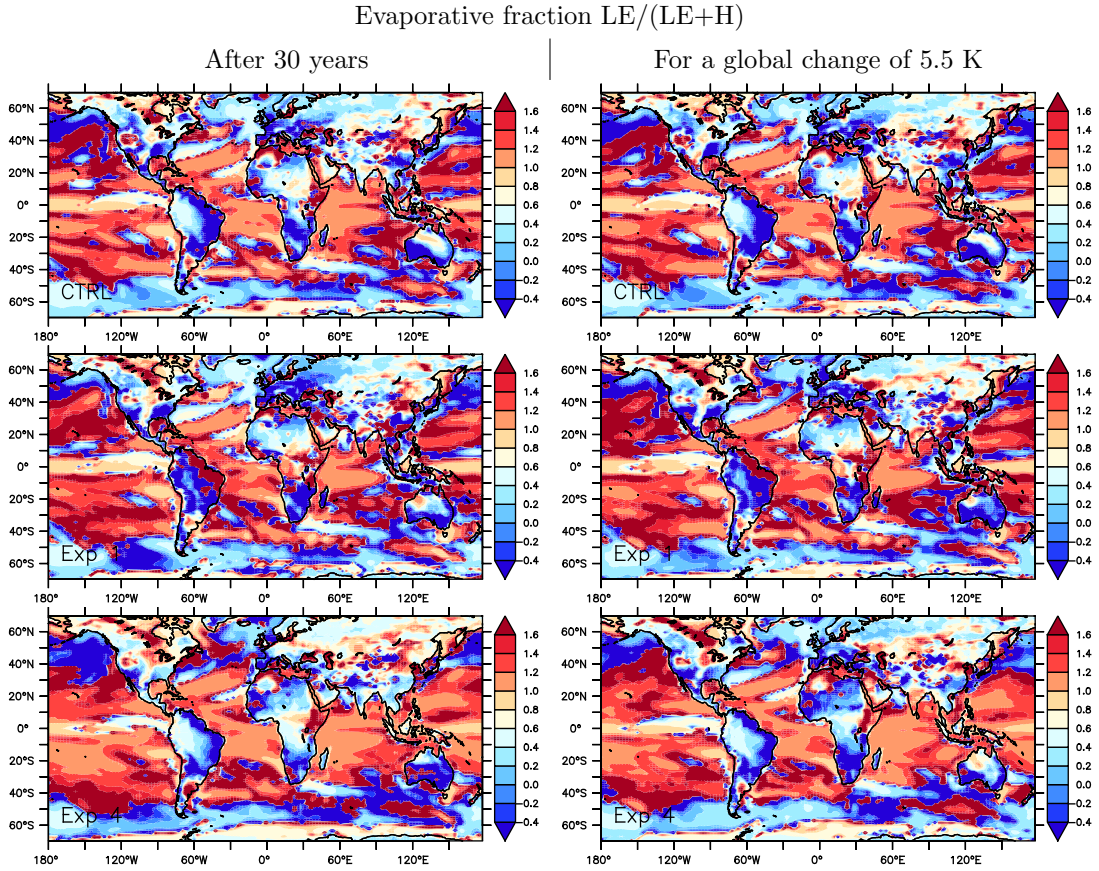

Figure S20: Evaporative fraction LE/(LE+H) change in coupled simulations CTRL (IPSL-6A configuration), Exp 1 and Exp 4. The maps correspond to differences between the `abrupt4`×CO_2 and `piControl` simulations, averaged over 21 consecutive years. For the left column, this time is centered at year 30 of the simulations. For the right column, it is centered at the time when the global 2m-temperature has increased by 5.5 K: *i. e.* at year 33 for CTRL, year 65 for Exp 1 and year 17 for Exp 4.