Supporting information for

Land-atmosphere feedbacks exacerbate concurrent soil drought and atmospheric aridity

Sha Zhou[1,2,3*], A. Park Williams[1], Alexis M. Berg[4], Benjamin I. Cook[1,5], Yao Zhang[3], Stefan Hagemann[6], Ruth Lorenz[7], Sonia I. Seneviratne[7], Pierre Gentine[2,3]

[1]Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

[2]Earth Institute, Columbia University, New York, New York 10027, USA

[3]Department of Earth and Environmental Engineering, Columbia University, New York, New York 10027, USA

[4]Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey 08544, USA

[5]NASA Goddard Institute for Space Studies, New York, New York 10027, USA

[6]Helmholtz-Zentrum Geesthacht, Institute of Coastal Research, 21502 Geesthacht, Germany

[7]Institute for Atmospheric and Climate Science, ETH Zurich, 8092 Zurich, Switzerland

*Correspondence to: sz2766@columbia.edu

Materials

GLEAM dataset. The Global Land Evaporation Amsterdam Model (GLEAM) (1) version 3.1a provides observationally constrained global daily surface and root-zone SM spanning the period from 1980 to 2017. The GLEAM surface SM was produced by employing an improved SM data assimilation system using three independent SM datasets, including two satellite-based SM products from the European Space Agency (ESA) Climate Change Initiative (CCI) and Soil Moisture Ocean Salinity (SMOS) and the surface SM from the Noah model in the Global Land Data Assimilation System (GLDAS). The root-zone SM was modelled from a three-layer water-balance module with input of precipitation infiltration and outputs of evapotranspiration and drainage. The forcing data of the water-balance module include Multi-Source Weighted Ensemble Precipitation (MSWEP) precipitation, ERA-Interim radiation and air temperature, CCI-LPRM vegetation optical depth, ESA CCI and GLDAS-Noah SM, etc. The GLEAM SM datasets perform well against SM measurements from eddy covariance towers and soil moisture sensors across a broad range of ecosystems (1). The daily SM data were aggregated into monthly values, and the SM data at a high spatial resolution of $0.25° \times 0.25$ were bilinearly interpolated to $1° \times 1°$ for analysis.

MERRA-2 reanalysis. We used monthly climatic data from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) to calculate monthly VPD. MERRA-2 is the latest atmospheric reanalysis produced by the NASA Global Modeling and Assimilation Office (2). In MERRA-2, global temperature and humidity data are largely determined from the direct assimilation of satellite radiances (see https://gmao.gsfc.nasa.gov/pubs/docs/McCarty885.pdf). We used monthly near-surface (2-m) temperature and dew-point temperature data for 1980-2018, with a spatial resolution of $1° \times 1°$. Monthly VPD was calculated as the difference between saturated water vapor pressure, determined by near-surface temperature, and actual water vapor pressure, determined by dew-point temperature. We only used data in the warm season, which was defined as the three-month period with highest mean temperature in each grid cell during the study period (Fig. S1), as drought and aridity events occur mainly in summer. In addition, focus on the warm season largely removes the impact of seasonal cycles of SM and VPD on the SM-VPD coupling. This definition for the warm season was also used for the simulations in GLACE-CMIP5 and CMIP5.

GLACE-CMIP5 experiment. We used three models (i.e., ACCESS, ECHAM6 and GFDL) participating in the GLACE-CMIP5 experiment, which was designed to assess the impact of land-atmosphere interactions, and specifically SM-climate feedbacks, in CMIP5 simulations (3). There are additional models in GLACE-CMIP5, but they did not provide atmospheric humidity (either specific or relative humidity) in the necessary format, so they were not used in this study. In each model, we used two twin simulations, i.e., the reference simulation (REF) and the experimental simulation (expB), covering the period from 1950 to 2100 (3). These two simulations share the same forcing, including sea surface temperature, sea ice, land use and CO_2 concentrations, from

the corresponding CMIP5 simulations (the historical simulations during 1950-2005, and the representative concentration pathway 8.5 scenario during 2006-2100). The difference between the two simulations is that SM in expB is prescribed as the mean seasonal cycle over a 30-year running mean monthly SM (with the exception of the first and last 15 years using the climatology of 1950-1979 and 2070-2100, respectively) from the REF simulation (Fig. S3). In addition to the seasonal cycle, the expB simulation preserves the long-term trend of REF SM. Therefore, the differences of the simulated climatic variables between the two simulations are attributed to the sub-seasonal and inter-annual variability of SM due to SM-atmosphere feedbacks.

To perform comparative analyses, we used monthly near-surface air temperature, relative humidity, total soil moisture content, and total precipitation in the warm season from the REF and expB simulations. VPD was calculated from temperature and relative humidity in the two simulations. The global analyses were bilinearly interpolated to 2°×2° spatial resolution to obtain the multi-model mean results.

CMIP5 model simulations. We used 26 CMIP5 models (61 pairs of simulations listed in Table S1) covering the historical (1861-2005) and future (2006-2100, the representative concentration pathway 8.5 scenario) periods. These models were selected because they make available the monthly means of near-surface air temperature, relative humidity, and total soil moisture content required for our analyses. VPD was calculated from temperature and relative humidity. The historical simulations were appended to the beginning of the future simulations to obtain the 30-year running mean (centered) monthly values of the variables (i.e., SM and VPD) covering the period 1881-2080. The total period was divided into two centennial periods: historical (1881-1980) and future (1981-2080). We calculated inter-annual variations for these variables as the difference between their monthly values and the 30-year running means for each month. Based on the 30-year running mean monthly values of the variables, we calculated the mean values in the warm season of each year to track their long-term trends, and the remaining components represent the seasonal cycles. All model results were bilinearly interpolated to 2°×2° spatial resolution, and the multi-model mean results were shown in the figures.

Attribution method for SM-VPD correlation. We used the Pearson's correlation coefficient to measure the correlation between SM and VPD. Here we show that this correlation coefficient can be decomposed to derive the contributions from the atmospheric circulation dynamics and LA feedbacks (i.e., the sub-seasonal and inter-annual variations in the variables), and the long-term trends and seasonality of the variables.

Time series of monthly SM and VPD in the warm season can be decomposed into two components. One is the long-term trends and seasonality of SM (SM_t) and VPD (VPD_t), calculated as the 30-year running mean monthly values. The other is the sub-seasonal and inter-annual variations in monthly SM (SM_v) and VPD (VPD_v):

$$SM = SM_t + SM_v \quad [S1]$$

$$VPD = VPD_t + VPD_v \quad [S2]$$

Pearson's correlation coefficient between SM and VPD ($r(SM, VPD)$) is defined as

$$r(SM, VPD) = \frac{cov(SM, VPD)}{\sigma_{SM} \cdot \sigma_{VPD}} \quad [S3]$$

where $cov()$ represents the covariance between two variables and σ is the standard deviation. Here we substitute Eqs. S1-S2 into Eq. S3, and obtain

$$r(SM, VPD) = \frac{cov(SM_t, VPD_t)}{\sigma_{SM} \cdot \sigma_{VPD}} + \frac{cov(SM_v, VPD_v)}{\sigma_{SM} \cdot \sigma_{VPD}} + \frac{cov(SM_t, VPD_v)}{\sigma_{SM} \cdot \sigma_{VPD}} + \frac{cov(SM_v, VPD_t)}{\sigma_{SM} \cdot \sigma_{VPD}} \quad [S4]$$

or

$$r(SM, VPD) = R(SM_t, VPD_t) + R(SM_v, VPD_v) + R(SM_t, VPD_v) + R(SM_v, VPD_t) \quad [S5]$$

where $R()$ is the normalized correlation coefficient, for example $R(SM_t, VPD_t) = r(SM_t, VPD_t) \cdot \frac{\sigma_{SM_t} \cdot \sigma_{VPD_t}}{\sigma_{SM} \cdot \sigma_{VPD}}$. The cross-correlation coefficients ($R(SM_t, VPD_v)$ and $R(SM_v, VPD_t)$) on the right side of Eq. S5 are very small (within ±0.05, indicating an independent relationship between SM_t and VPD_v, and between SM_v and VPD_t) and can be omitted (Fig. S12). Then we can get

$$r(SM, VPD) \approx R(SM_t, VPD_t) + R(SM_v, VPD_v) \quad [S6]$$

$R(SM_t, VPD_t)$ can be further decomposed to assess the contributions from the long-term trends and from the seasonal cycles of SM and VPD, so we can better understand how $R(SM_t, VPD_t)$ changes between historical and future simulations (see results in Fig. S9).

Attribution method for compound extreme events. There is a near-linear relationship between the natural logarithm of PMF and r(SM,VPD) across all grid cells (Fig. 3*E* and *F*, Table S2), that is

$$log(PMF) \approx a + b \times r(SM, VPD) \quad [S7]$$

where a and b represent intercept and slope parameters (see Table S2 for their values). The slope b represents the sensitivity of $log(PMF)$ to $r(SM, VPD)$, while the intercept a measures the impact of tail (in)dependence on PMF. As the Pearson's correlation coefficient $r(SM, VPD)$ only detects the linear dependency between SM and VPD, it may not well capture the co-occurrence frequency of extreme events if there is strong tail (in)dependence, which could be modeled by copulas (4). A positive intercept a indicates that the PMF is larger than unity when SM and VPD are linearly independent, because of the tail dependence in extreme deviations; on the contrary, a negative intercept a is induced by tail independence in extreme deviations. We observed positive values of the intercept a in both periods (Table S2), indicating the existence of tail dependence between SM and VPD in the simulations. The tail dependence is attributed to large-scale atmospheric dynamics and LA feedbacks that lead to SM and VPD extremes, as shown in GLACE-CMIP5 simulations.

We further separate PMF into PMF_t induced by the long-term trends and seasonality of SM and VPD, and PMF_v by large-scale atmospheric dynamics and SM-VPD feedback. Substituting Eq. S6 into Eq. S7, we can get

$$log(PMF_t \times PMF_v) = a + b \times [R(SM_t, VPD_t) + R(SM_v, VPD_v)] \quad [S8]$$

and

$$PMF_t = e^{b \times R(SM_t, VPD_t)} \quad [S9]$$

$$PMF_v = e^{a + b \times R(SM_v, VPD_v)} \quad [S10]$$

The dependence structure of SM and VPD varies, resulting in varying levels of tail dependence, and therefore varying parameters a and b, across different grid cells. The calibrated parameters a and b in Eq. S8 may lead to uncertainties of PMF_t and PMF_v in some grid cells, but weakly impact their relative magnitudes, which depend mainly on $R(SM_t, VPD_t)$ and $R(SM_v, VPD_v)$ (a is one order of magnitude smaller than b, see Table S2). Therefore, we used this method for the attribution of compound extreme events.

Table S1. Model ensembles from the CMIP5 experiments (historical and RCP8.5 simulations) used in this study.

Model Name	Ensemble	Institute ID	Modeling Center
ACCESS1-0	r1i1p1	CSIRO-BOM	Commonwealth Scientific and Industrial Research Organization (CSIRO) and Bureau of Meteorology (BOM), Australia
ACCESS1-3	r1i1p1		
CanESM2	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1	CCCMA	Canadian Center for Climate Modelling and Analysis
CCSM4	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1, r6i1p1	NCAR	National Center for Atmospheric Research
CSIRO-Mk3-6-0	r1i1p1, r2i1p1, r3i1p1, r4i1p1, r5i1p1, r6i1p1, r7i1p1, r8i1p1, r9i1p1, r10i1p1	CSIRO-QCCCE	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence
FGOALS-s2	r2i1p1, r3i1p1	LASG-IAP	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences
GFDL-ESM2G	r1i1p1	NOAA GFDL	NOAA Geophysical Fluid Dynamics Laboratory
GFDL-ESM2M	r1i1p1		
GISS-E2-H	r1i1p1, r1i1p2, r1i1p3, r2i1p1, r2i1p3	NASA GISS	NASA Goddard Institute for Space Studies
GISS-E2-H-CC	r1i1p1		
GISS-E2-R	r1i1p1, r1i1p2, r1i1p3, r2i1p1, r2i1p3		
GISS-E2-R-CC	r1i1p1		
HadGEM2-AO	r1i1p1	NIMR/KMA	National Institute of Meteorological Research/Korea Meteorological Administration
HadGEM2-CC	r1i1p1	MOHC (additional realizations by INPE)	Met Office Hadley Centre (additional HadGEM2-ES realizations contributed by Instituto Nacional de Pesquisas Espaciais)

HadGEM2-ES	r1i1p1, r2i1p1, r3i1p1, r4i1p1		
inmcm4	r1i1p1	INM	Institute for Numerical Mathematics
IPSL-CM5A-LR	r1i1p1, r2i1p1, r3i1p1, r4i1p1	IPSL	Institut Pierre Simon Laplace, France
IPSL-CM5A-MR	r1i1p1		
IPSL-CM5B-LR	r1i1p1		
MIROC5	r1i1p1, r2i1p1, r3i1p1	MIROC	Atmosphere and Ocean Research Institute (The University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology
MIROC-ESM	r1i1p1	MIROC	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies
MIROC-ESM-CHEM	r1i1p1		
MRI-CGCM3	r1i1p1	MRI	Meteorological Research Institute
MRI-ESM1	r1i1p1		
NorESM1-M	r1i1p1	NCC	Norwegian Climate Centre
NorESM1-ME	r1i1p1		

Table S2. The established functions to predict probability multiplication factor (PMF) using correlation coefficient between soil moisture (SM) and vapor pressure deficit (VPD).

	Historical period	Future period
Function	Log(PMF)=0.4557-2.2057× r(SM,VPD)	Log(PMF)=0.6440-1.9204× r(SM,VPD)
Predictive performance	R^2=0.77	R^2=0.79

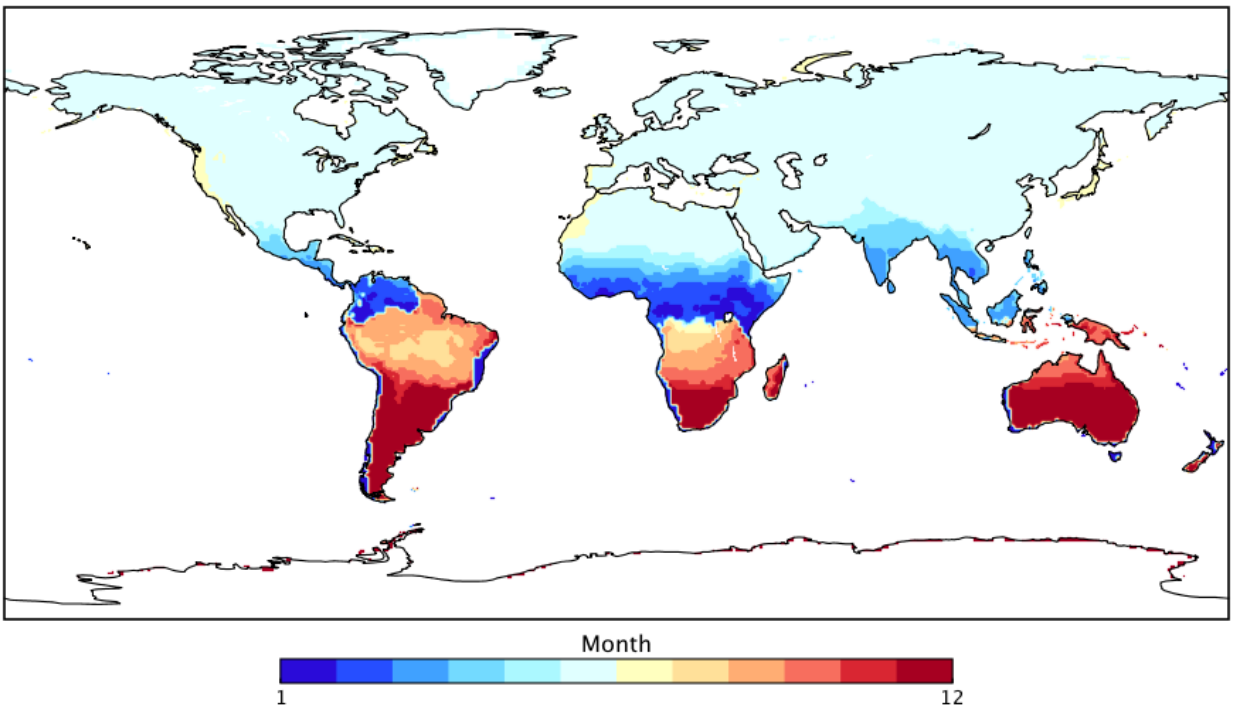

Fig. S1. Warm season in defined based on MERRA-2 temperature (1980-2017). The starting month of the three-month period (warm season) with highest mean temperature in each grid cell is shown.

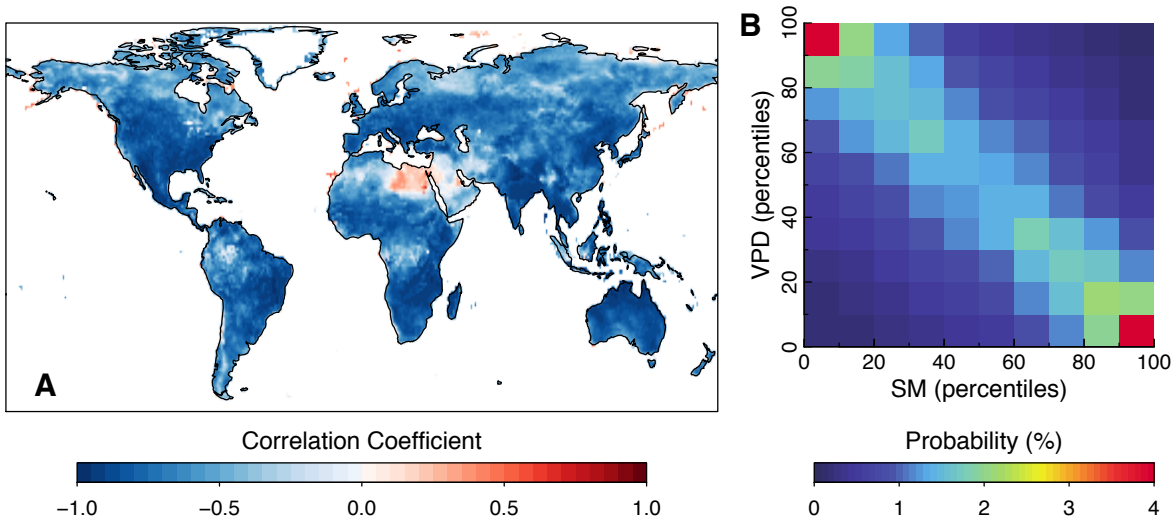

Fig. S2. Relationship between GLEAM surface SM and MERRA-2 VPD. (*A*) Correlation coefficient between surface SM and VPD during the period 1980-2017 (warm season in Fig. S1). (*B*) Mean probability of each percentile bin of surface SM and VPD across all land grid cells.

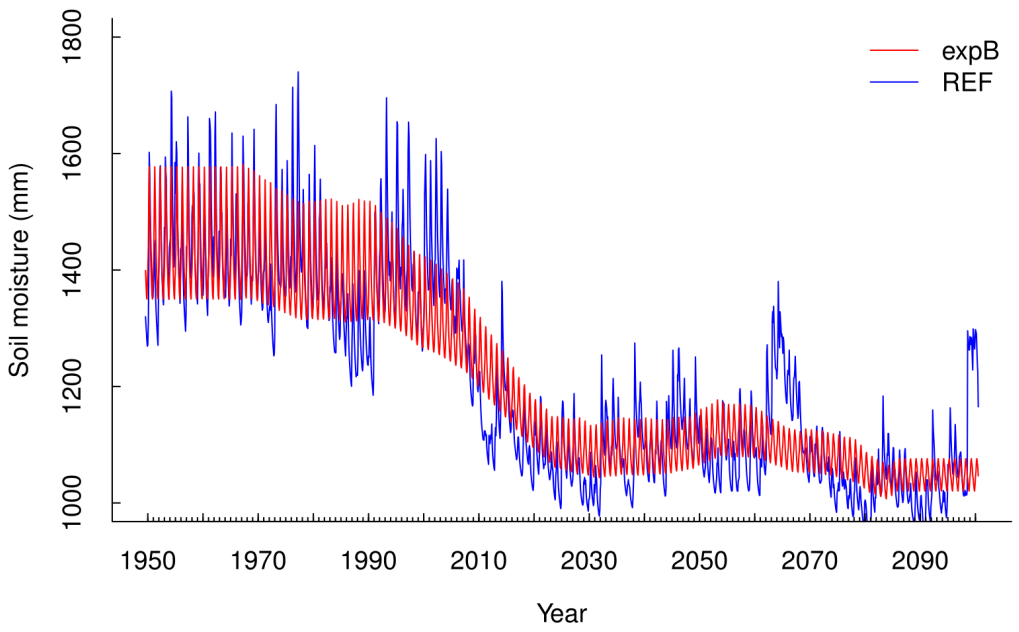

Fig. S3. Illustration of total column soil moisture in the simulations REF and expB in GLACE-CMIP5. Soil moisture data shown in the figure are obtained from a grid cell (13.75°N, 35.625°E) in the ACCESS model. The total soil column depth is 3164 mm.

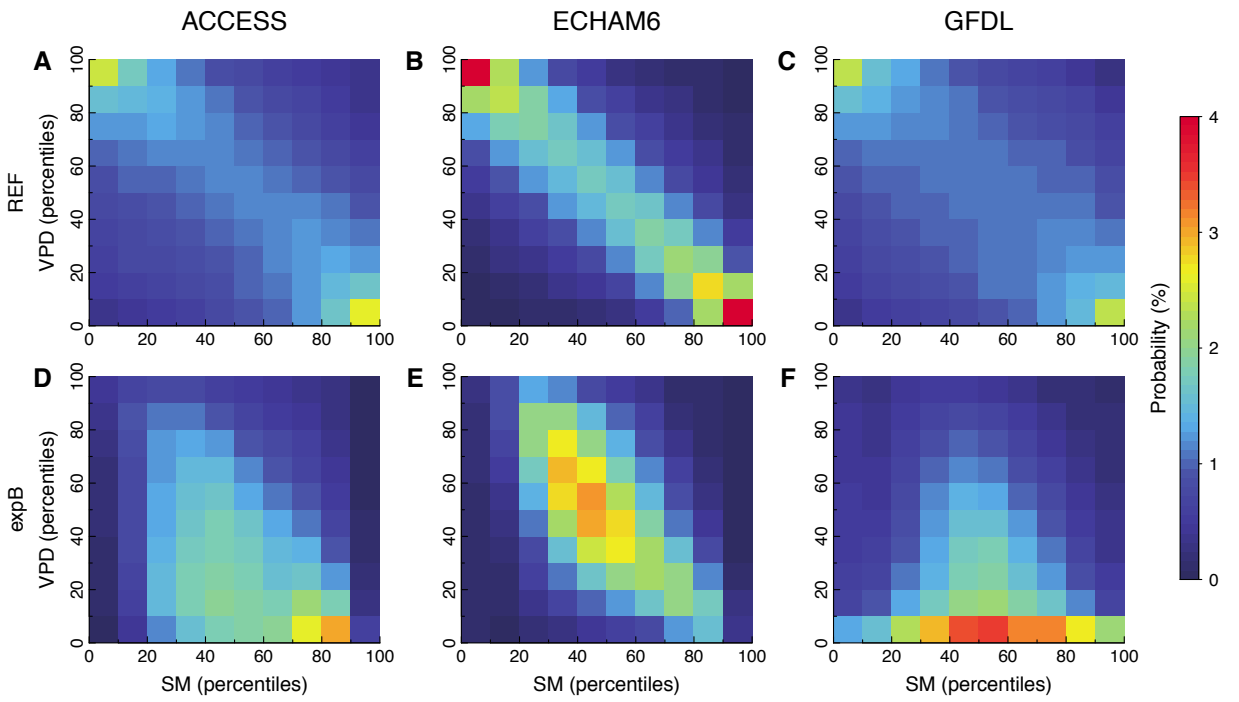

Fig. S4. Distributions of SM and VPD in GLACE-CMIP5 expB and REF simulations. Mean probability of each percentile bin of VPD and SM across all grid cells in (*A*-*C*) REF and (*D*-*F*) expB. (*A* and *D*) ACCESS; (*B* and *E*) ECHAM6; (*C* and *F*) GFDL.

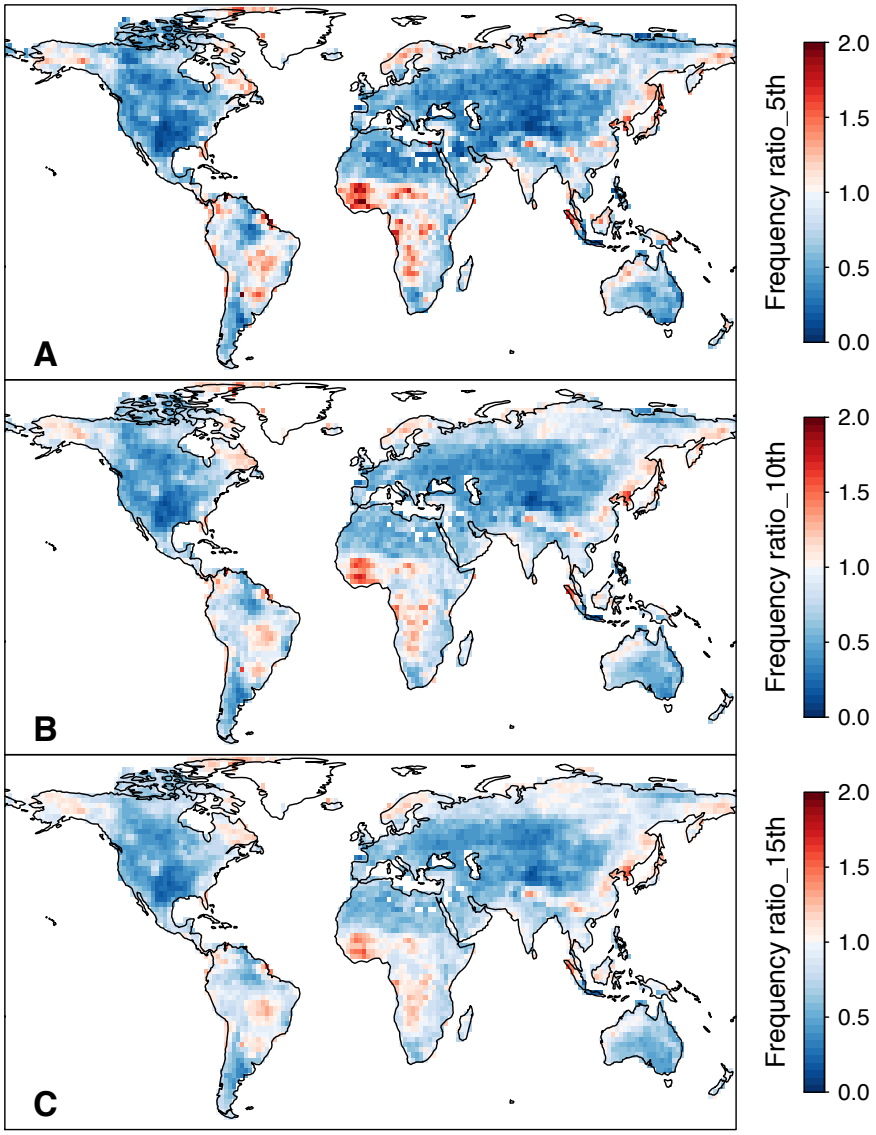

Fig. S5. Frequency ratio of extreme low precipitation between expB and REF simulations in GLACE-CMIP5. (*A*) Ratio of the frequency of extreme low precipitation (below 5th percentile precipitation in REF) between expB and REF (expB over REF). (*B* and *C*) The same as (*A*), but for (*B*) 10th percentile precipitation and (*C*) 15th percentile precipitation.

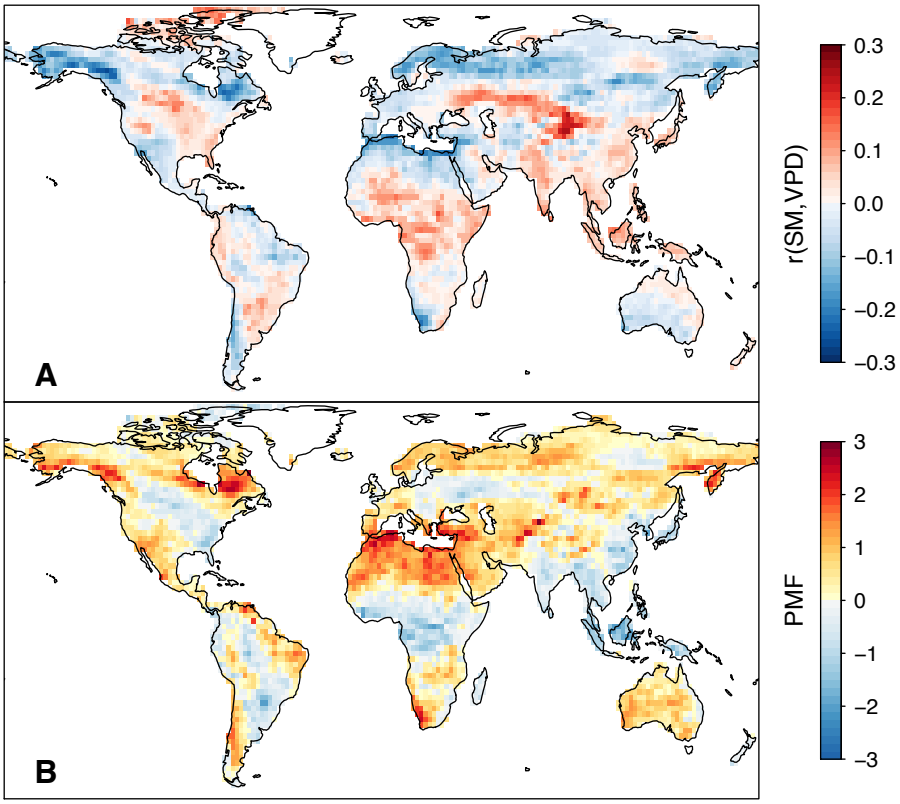

Fig. S6. Future changes in the SM-VPD correlation and probability multiplication factor (PMF) in CMIP5. (*A*) Changes in mean correlation coefficient (r(SM,VPD)) between historical and future simulations (future minus historical values). (*B*) Changes in model mean PMF of concurrent extreme VPD above its 95th percentile and extreme SM below its 5th percentile between the two simulations.

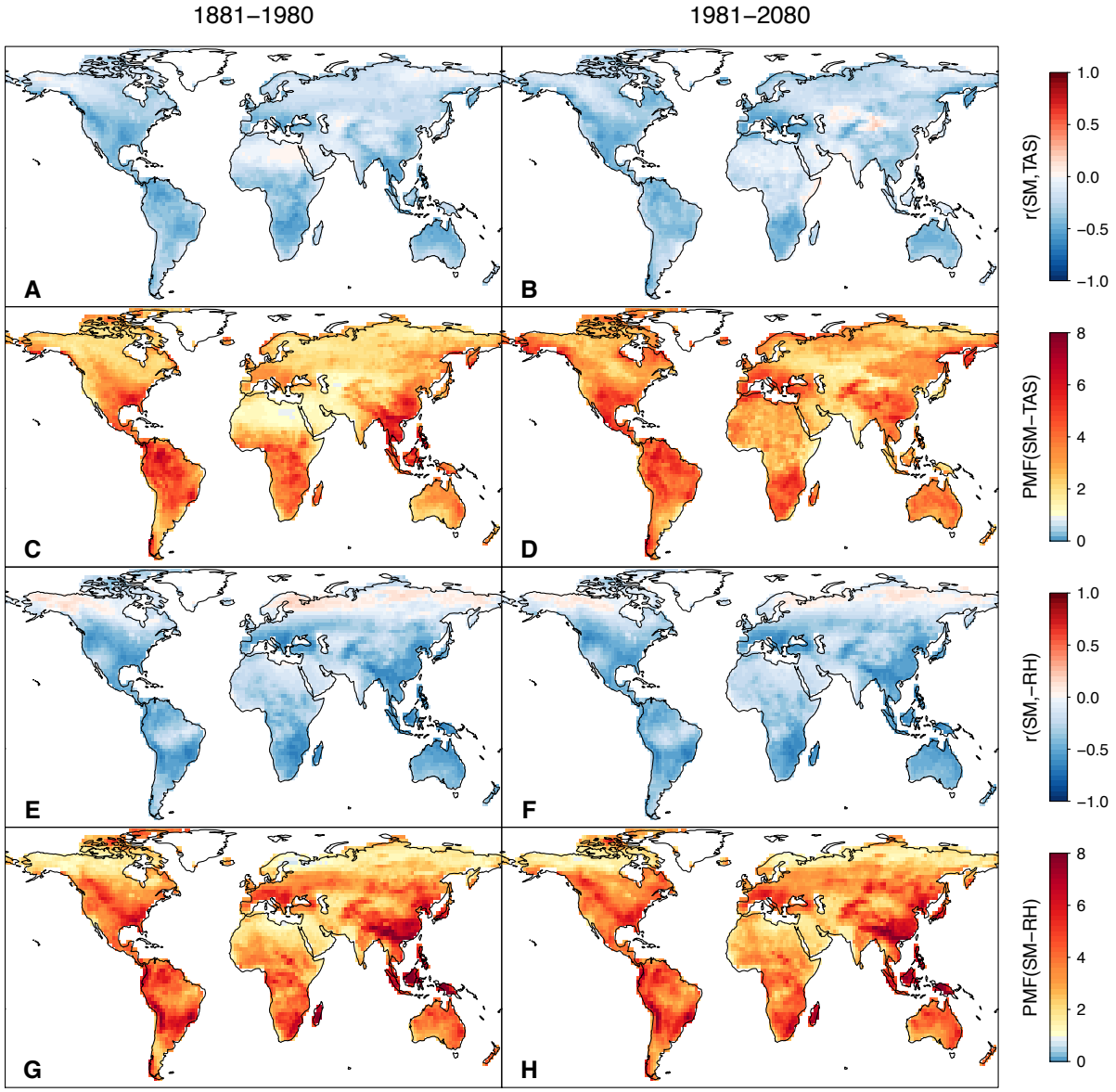

Fig. S7. SM-TAS and SM-RH correlations and probability multiplication factor (PMF) in CMIP5. (*A* and *B*) Mean correlation coefficient (r(SM,TAS)) between SM and TAS across 61 pairs of historical and future simulations. (*C* and *D*) Model mean PMF of concurrent extreme TAS above its 95th percentile and extreme SM below its 5th percentile. (*E* and *F*) Mean correlation coefficient (r(SM,-RH)) between SM and negative RH. (*G* and *H*) Model mean PMF of concurrent extreme RH and extreme SM below their 5th percentiles.

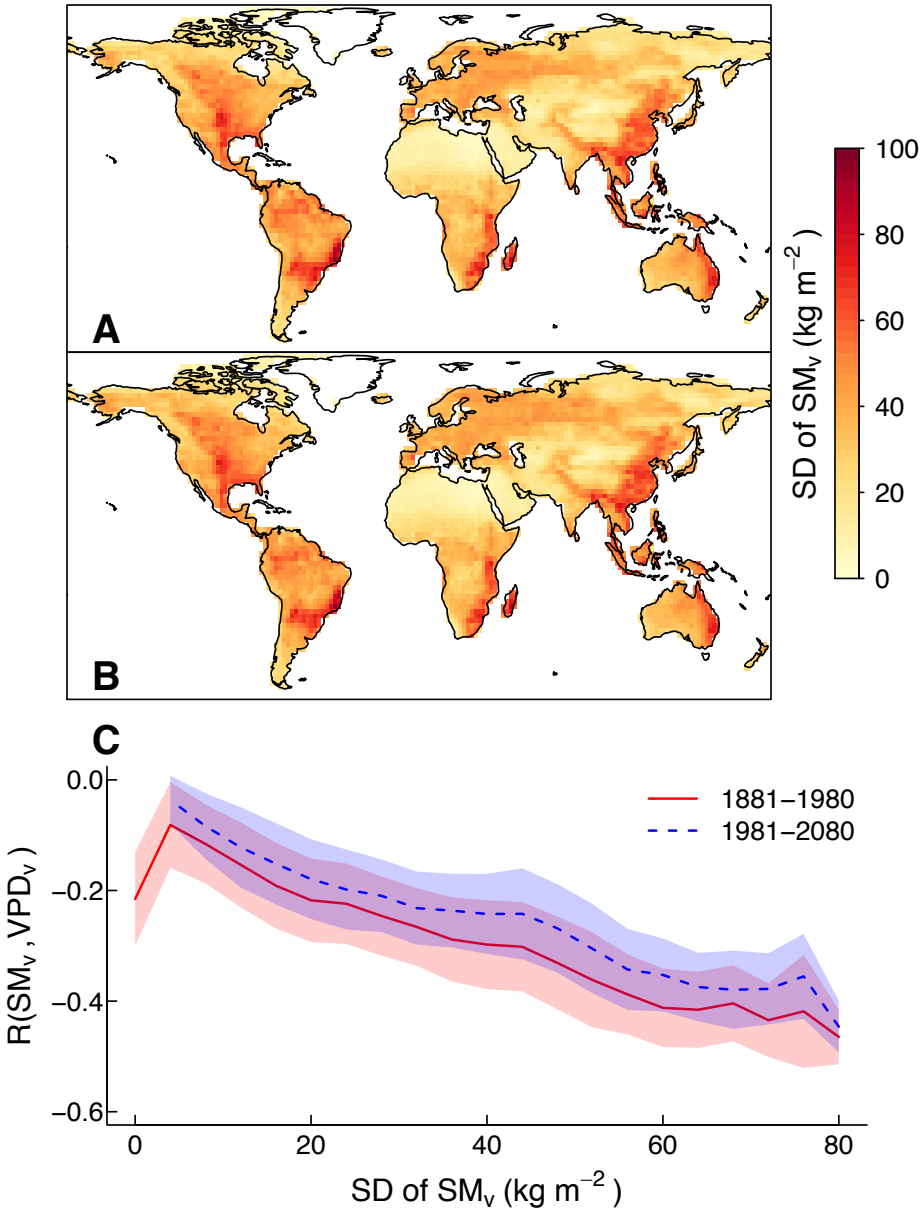

Fig. S8. Dependence of the SM-VPD correlation on SM variability in CMIP5. (*A* and *B*) The standard deviation (SD) of SM_v in (*A*) historical (1881-1980) and (*B*) future (1981-2080) simulations. SM_v was obtained by removing the long-term trends and seasonal cycles of SM. (*C*) Relationship between $R(SM_v,VPD_v)$ and the SD of SM_v.

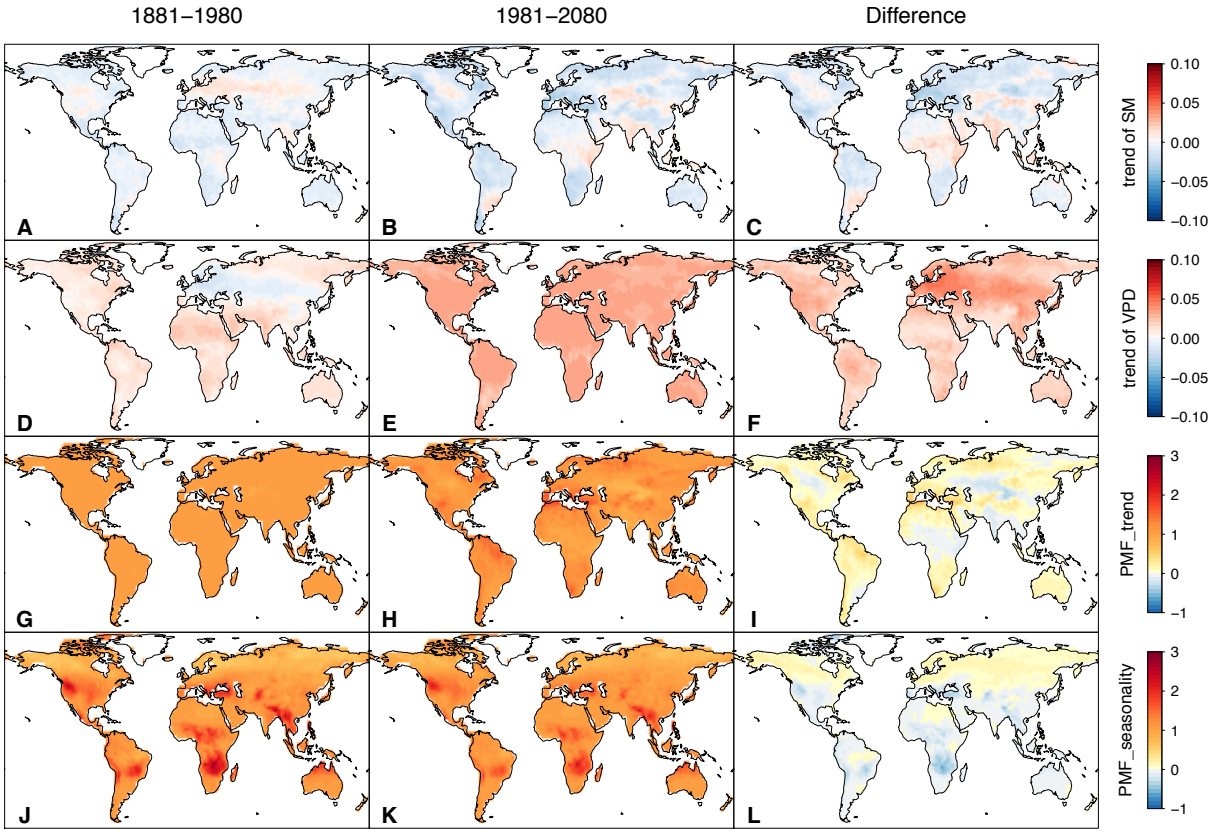

Fig. S9. Long-term trends in SM and VPD and the contributions of trends and seasonality to probability multiplication factor (PMF) in CMIP5. (*A*-*F*) Long-term linear trends in warm season SM and VPD normalized by their standard deviation in each grid cell. (*G*-*L*) Separated contributions of the (*G*-*I*) long-term trends and (*J*-*L*) the seasonal cycles of SM and VPD. The long-term trends of SM and VPD for attribution were tracked according to the 30-year running mean values in the warm season, which are different from the linear trends shown in *A*-*F*. The last column shows the difference between historical (1881-1980) and future (1981-2080) periods (future minus historical values).

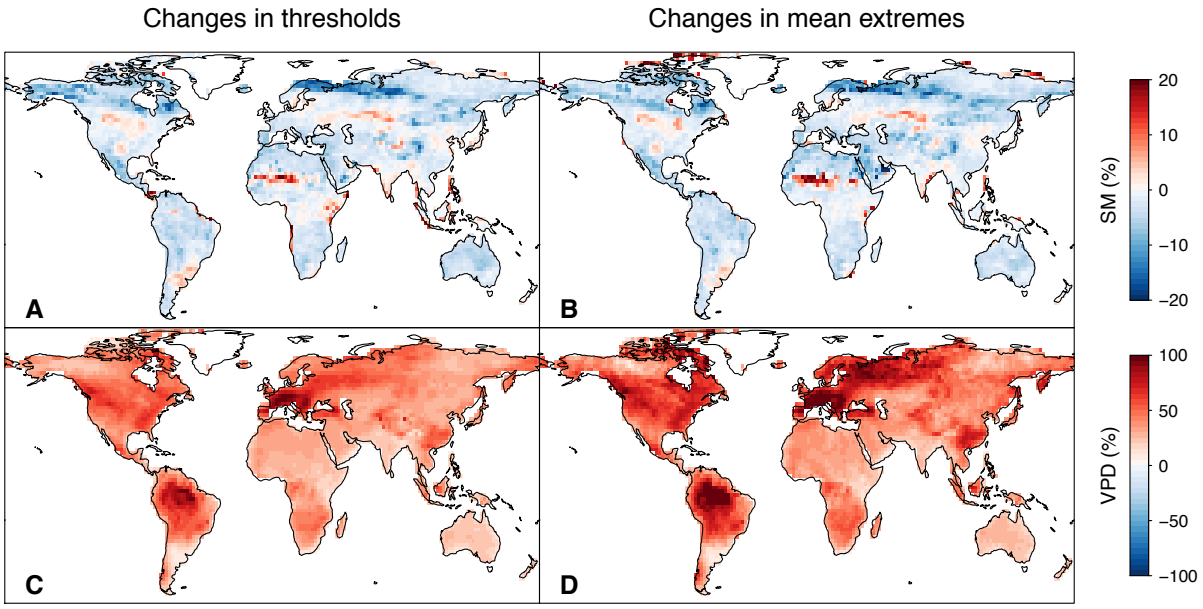

Fig. S10. Changes in the thresholds and mean values for compound extreme events from historical to future periods in CMIP5. (*A* and *C*) Changes (future minus historical values) in thresholds of extreme low SM and extreme high VPD. (*B* and *D*) Changes in mean SM and mean VPD for compound extreme events. The changes in the thresholds and mean values of SM and VPD for compound extreme events are normalized by mean SM and VPD in historical simulations, respectively.

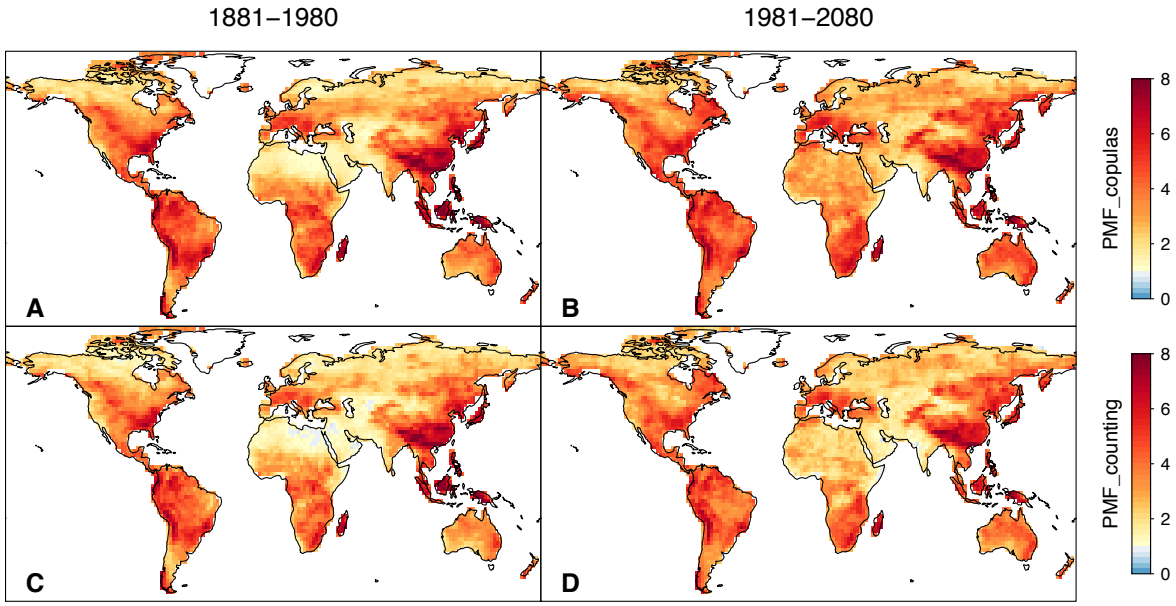

Fig. S11. Comparison of the probability multiplication factor (PMF) in CMIP5. PMF for compound SM and VPD extremes is derived (*A* and *B*) with copulas and (*C* and *D*) by counting in historical (1881-1980) and future (1981-2080) simulations.

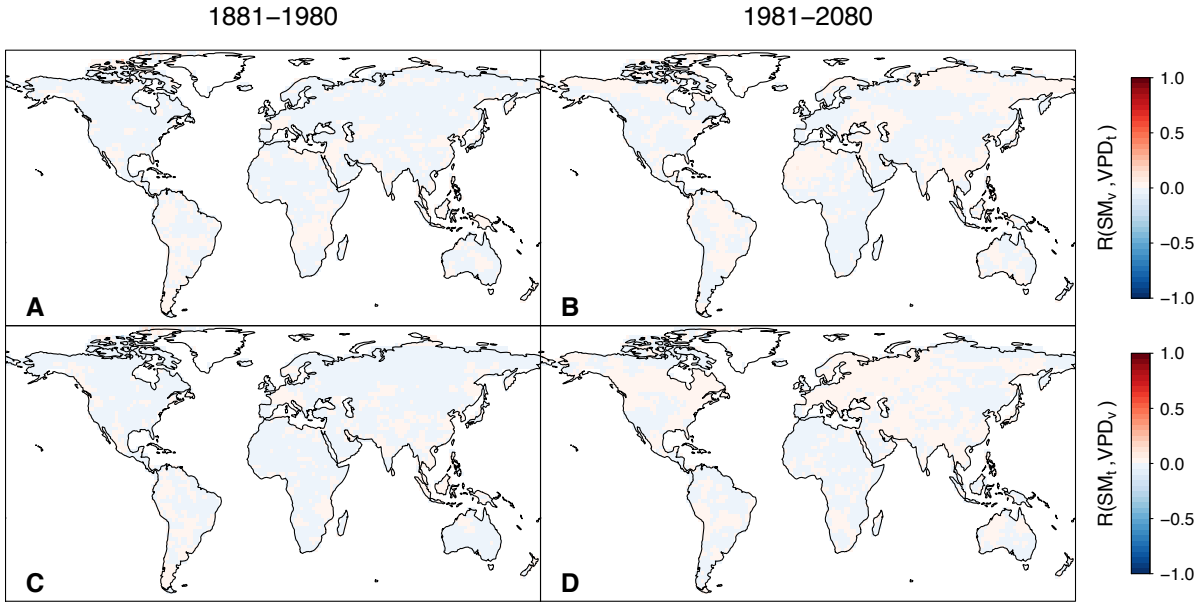

Fig. S12. Cross-correlation coefficients in CMIP5 during the two periods. (*A* and *C*) Historical simulations (1881-1980). (*B* and *D*) Future simulations (1981-2080).

References

1. Martens B*, et al.* (2017) GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development* 10(5):1903-1925.
2. Gelaro R*, et al.* (2017) The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate* 30(14):5419-5454.
3. Seneviratne SI*, et al.* (2013) Impact of soil moisture-climate feedbacks on CMIP5 projections: First results from the GLACE-CMIP5 experiment. *Geophysical Research Letters* 40(19):5212-5217.
4. Joe H, Li H, & Nikoloulopoulos AK (2010) Tail dependence functions and vine copulas. *Journal of Multivariate Analysis* 101(1):252-270.