Supplement of Earth Syst. Sci. Data, 13, 4385–4405, 2021
https://doi.org/10.5194/essd-13-4385-2021-supplement

Open Access Earth System Science Data

Supplement of

Development of observation-based global multilayer soil moisture products for 1970 to 2016

Yaoping Wang et al.

Correspondence to: Jiafu Mao (maoj@ornl.gov)

Table S1: List of satellite soil moisture products used in the merging.

Dataset	Original resolution (°)	Original depth	Used depth (cm)	Time period	Time period used	Reference
ESA CCI v4.5 (combined passive and active)	0.25×0.25	Top few centimeters	0–10	1978–2018	1981–2016, 1981–2010	(Dorigo et al., 2017)

Table S2: List of reanalysis soil moisture products used in the merging.

Dataset	Original resolution	Original depth (cm)	Interpolated depth (cm)	Time period	Time period used	Reference
GLEAM v3.3a	0.25°×0.25°	0–10	0–10	1980–2018	1981–2016	(Martens et al., 2017)
CERA20C	~125 km	0–7, 7–28, 28–100, 100–289	0–10, 10–30	1900–2010	1981–2010, 1970–2010	(Laloyaux et al., 2017)
ERA20C	~125 km	0–7, 7–28, 28–100, 100–255	0–10, 10–30	1900–2010	1981–2010, 1970–2010	(Poli et al., 2016)
ERA-Interim	~80 km	0–7, 7–28, 28–100, 100–289	0–10, 10–30	1979–2018	1981–2016, 1981–2010	(Dee et al., 2011)
ERA5	~30 km	0–7, 7–28, 28–100, 100–289	0–10, 10–30	1979–2020	1981–2016, 1981–2010	(Hersbach et al., 2020)

Table S3: List of offline LSM-simulated SM datasets used in the merging. The BG1 simulation of MsTMIP has time-varying climate, land use, atmospheric CO_2, and nitrogen deposition forcings, whereas the SG3 forcing has constant nitrogen deposition. The S3 simulation of TRENDY v7 has time-varying climate, land use, atmospheric CO_2, and, if the model has a nitrogen cycle, nitrogen deposition forcings.

Dataset	Simulation	Model	Original resolution (°)	Original depth (cm, to the precision of 3 or 4 digits)	Interpolated depth (cm)	Time period	Time period used	Reference
GLDAS NOAH025 M2.0†	GLDAS	Noah-MP	0.25×0.25	0–10, 10–40, 40–100, 100–200	0–10, 10–30, 30–50, 50–100	1948–2010	1970–2010, 1981–2010	(Beaudoing et al., 2019)
ERA-Intrim/Land†	—	HTESSEL	0.75×0.75	0–7, 7–28, 28–100, 100–289	0–10, 10–30	1979–2010	1981–2010	(Balsamo et al., 2015)
MsTMIP	BG1	CLASS-CTEM-N	0.5×0.5	0–10, 10–35, 35–400	0–10, 30–50	1901–2010	1970–2010	(Huntzinger et al., 2018)
	BG1	CLM4	0.5×0.5	0–0.18, 0.18–0.45, 0.45–0.91, 0.91–1.66, 1.66–2.89, 2.89–4.93, 4.93–8.29, 8.29–13.8, 13.8–23.0, 23.0–38.0, 38.0–62.8, 62.8–104, 104–171, 171–283, 283–421	0–10, 10–30			
	BG1	CLM4VIC	0.5×0.5	0–1.75, 1.75–4.51, 4.51–9.06, 9.06–16.6, 16.6–28.9, 28.9–49.3, 49.3–82.9, 139–230, 230–380	0–10, 10–30, 30–50, 50–100			

Dataset	Simulation	Model	Original resolution (°)	Original depth (cm, to the precision of 3 or 4 digits)	Interpolated depth (cm)	Time period	Time period used	Reference
	SG3	SiBCASA	0.5×0.5	0.0–2.0, 2.0–4.5, 4.5–7.5, 7.5–11.2, 11.2–15.8, 15.8–21.4, 21.4–28.3, 28.3–36.7, 36.7–47.1, 47.1–59.9, 59.9–75.7, 75.7–95.0, 95.0–119, 119–148, 148–184, 184–228, 228–282, 282–349, 349–430, 430–531, 531–655, 655–807, 807–993, 993–1223, 1223–1504	0–10, 10–30, 30–50, 50–100			
TRENDY v7	S3	CLM5.0	~0.94×1.25	0–2, 2–6, 6–12, 12–20, 20–32, 32–48, 48–68, 68–92, 92–120, 120–152, 152–188, 188–228, 228–272, 272–326, 326–390, 390–464, 464–548, 548–642, 642–746, 746–860	0–10, 10–30, 30–50, 50–100	1901–2017	1970–2016, 1970–2010, 1981–2010, 1981–2016	(Sitch et al., 2008)
		ISAM	0.5×0.5	0.00–1.75, 1.75–4.51, 4.51–9.06, 9.06–16.6, 16.6–28.9, 28.9–49.3, 49.3–82.9, 82.9–138, 138–230, 230–343	0–10, 10–30, 30–50, 50–100			
		JSBACH	~1.86×1.875	0–6.5, 6.5–31.9, 31.9–123, 123–413, 413–983	0–10, 30–50			
		JULES	1.25×1.875	0–10, 10–35, 35–100, 100–300	0–10, 10–30, 50–100			
		LPX-Bern	0.5×0.5	0–10, 10–20, 20–30, 30–50, 50–70, 70–100, 100–150, 150–200	0–10, 10–30, 30–50, 50–100			
		ORCHIDEE	2×2	0–0.1, 0.1–0.4, 0.4–1, 1–2.2, 2.2–4.5, 4.5–9.2, 9.2–18.6, 18.6–37.4, 37.4–74.9, 74.9–150, 150–200	0–10, 10–30, 30–50, 50–100			
		ORCHIDEE-CNP	2×2	0–0.1, 0.1–0.4, 0.4–1, 1–2.2, 2.2–4.5, 4.5–9.2, 9.2–18.6, 18.6–37.4, 37.4–74.9, 74.9–150, 150–200	0–10, 10–30, 30–50, 50–100			

† Although the GLDAS is a data assimilation system, the "2.0" simulations do not use any data assimilation (Rui and Beaudoing, 2020). The ERA-Interim/Land is an offline run without data assimilation and forced with bias-corrected meteorological inputs from ERA-Interim (Balsamo et al., 2015).

Table S4: List of coupled ESM-simulated SM datasets used in the merging. Only the first ensemble member (CMIP5 r1i1p1, CMIP6 r1i1p1f1) was used for all the models. All the models were used for 1970–2016 and the shorter periods (1970–2010, 1981–2016, 1981–2010) contained within. The CMIP5 historical experiment was used for 1970–2005, and the CMIP5 RCP8.5 for 2006–2016. The CMIP6 historical experiment was used for 1970–2014, and the CMIP6 SSP5-85 experiment for 2015–2016.

Dataset	Model	Original resolution (°)	Original depth (cm, to the precision of 3 or 4 digits)	Interpolated depth (cm)	Reference
CMIP5	ACCESS1-0	1.25×1.875	0–10, 10–35, 35–100, 100–300	0–10, 30–50	(Taylor et al., 2012)
	ACCESS1-3	1.25×1.875	0–2.2, 2.2–8, 8–23.4, 23.4–64.3, 64.3–173, 173–460	0–10, 10–30, 50–100	
	CanESM2	~2.79×2.81	0–10, 10–35, 35–41	0–10, 30–50	
	CNRM-CM5	~1.40×1.40	0–1, 1–20, 20–50, 50–100, 100–150, 150–200, 200–300, 300-800	0–10, 10–30, 50–100	
	FGOALS-s2	~2.79×2.81	0–1.75, 1.75–4.51, 4.51–9.06, 9.06–16.6, 16.6–28.9, 28.9–49.3, 49.3–82.9, 82.9–138, 138–230, 230–343	0–10, 10–30, 30–50, 50–100	
	GFDL-CM3	2×2.5	0–2, 2–6, 6–10, 10–15, 15–20, 20–30, 30–40, 40–60, 60–80, 80–100, 100–140, 140–180, 180–220, 220–260, 260–300, 300–400, 400–500, 500–600, 600–750, 750–1000	0–10, 10–30, 30–50, 50–100	
	GFDL-ESM2G	~2×2			
	GFDL-ESM2M	~2×2.5			
	GISS-E2-H-CC	2×2.5	0–10, 10–27.3, 27.3–57.0, 57.0–108, 108–197, 197–350	0–10, 10–30, 50–100	
	GISS-E2-H	2×2.5			
	GISS-E2-R-CC	2×2.5			
	GISS-E2-R	2×2.5			
	HadGEM2-ES	1.25×1.875	0–10, 10–35, 35–67.5, 67.5–300	0–10, 30–50, 50–100	
	HadGEM2-CC	1.25×1.875			
	inmcm4	1.5×2	0.5–1.5, 1.5–3, 3–6, 6–11.5, 11.5–20, 20–30, 30–40, 40–50, 50–60, 60–70, 70–80, 80–90, 90–100, 100–110, 110–120, 120–130, 130–140, 140–150, 150–178, 178–250, 250–400, 400–750, 750–1500	0–10, 10–30, 30–50, 50–100	
	MIROC5	1.40×1.40	0–5, 5–25, 25–100, 100–200, 200–400, 400–1400	0–10, 10–30	
	MIROC-ESM	~2.79×2.81			
	MIROC-ESM-CHEM	~2.79×2.81			
	NorESM1-ME	~1.89×2.5	0–1.75, 1.75–4.51, 4.51–9.06, 9.06–16.6, 16.6–28.9, 28.9–49.3, 49.3–82.9, 82.9–138.3, 138.3–230, 230–380, 380–628, 628–1038, 1038–1712, 1712–2825, 2825–4210	0–10, 10–30, 30–50, 50–100	
CMIP6	BCC-CSM2-MR	~1.12×1.125	0–1.75, 1.75–4.51, 4.51–9.06, 9.06–16.5, 16.5–28.9, 28.9–49.3, 49.3–82.9, 82.9–138, 138–230, 230–343	0–10, 10–30, 30–50, 50–100	(Eyring et al., 2016)
	CanESM5	~2.79×2.81	0–10, 10–35, 35–410	0–10, 30–50	
	CESM2	~0.94×1.25	0–2.5, 2.5–6.5, 6.5–12.5, 12.5–21, 21–33, 33–49, 49–69, 69–93, 93–121, 121–153, 153–189, 189–229, 229–274, 274 –328, 328–392, 392–466, 466–550, 550–644, 644–748, 748–858	0–10, 10–30, 30–50, 50–100	
	CESM2-WACCAM	~0.94×1.25			
	EC-Earth3	~0.70×0.70	0–7, 7–28, 28–100, 100–289	0–10, 10–30	
	EC-Earth3-Veg	~0.70×0.70			

Dataset	Model	Original resolution (°)	Original depth (cm, to the precision of 3 or 4 digits)	Interpolated depth (cm)	Reference
	IPSL-CM6A-LR	~1.27×2.5	0–0.1, 0.1–0.4, 0.4–1, 1–2.2, 2.2–4.5, 4.5–9.2, 9.2–18.6, 18.6–37.4, 37.4–74.9, 74.9–150, 150–200	0–10, 10–30, 30–50, 50–100	
	MIROC6	~1.4×1.4	0–5, 5–25, 25–100, 100–200, 200–400, 400–1400	0–10, 10–30	
	MPI-ESM1-2-HR	~0.94×0.94	0–6.5, 6.5–31.9, 31.9–123, 123–413, 413–983	0–10, 30–50	

30

Table S5: List of the corresponding temperature and precipitation forcings to the ORS SM datasets.

SM dataset	Temperature and precipitation dataset	Reference
ESA CCI v4.5 (combined passive and active)	CRU TS v4.03	(Harris et al., 2014)
GLEAM v3.3a		
CERA20C	CERA20C	(Laloyaux et al., 2017)
ERA20C	ERA20C	(Poli et al., 2016)
ERA-Interim	ERA-Interim	(Dee et al., 2011)
ERA5	ERA5	(Hersbach et al., 2020)
ERA-Land	ERA-Interim	(Balsamo et al., 2015)
GLDAS NOAH025_M2.0	GLDAS NOAH025_M2.0	(Beaudoing et al., 2019)
MsTMIP	CRU TS v3.20	(Harris et al., 2014)
TRENDY v7	CRU TS v3.26	(Harris et al., 2014)

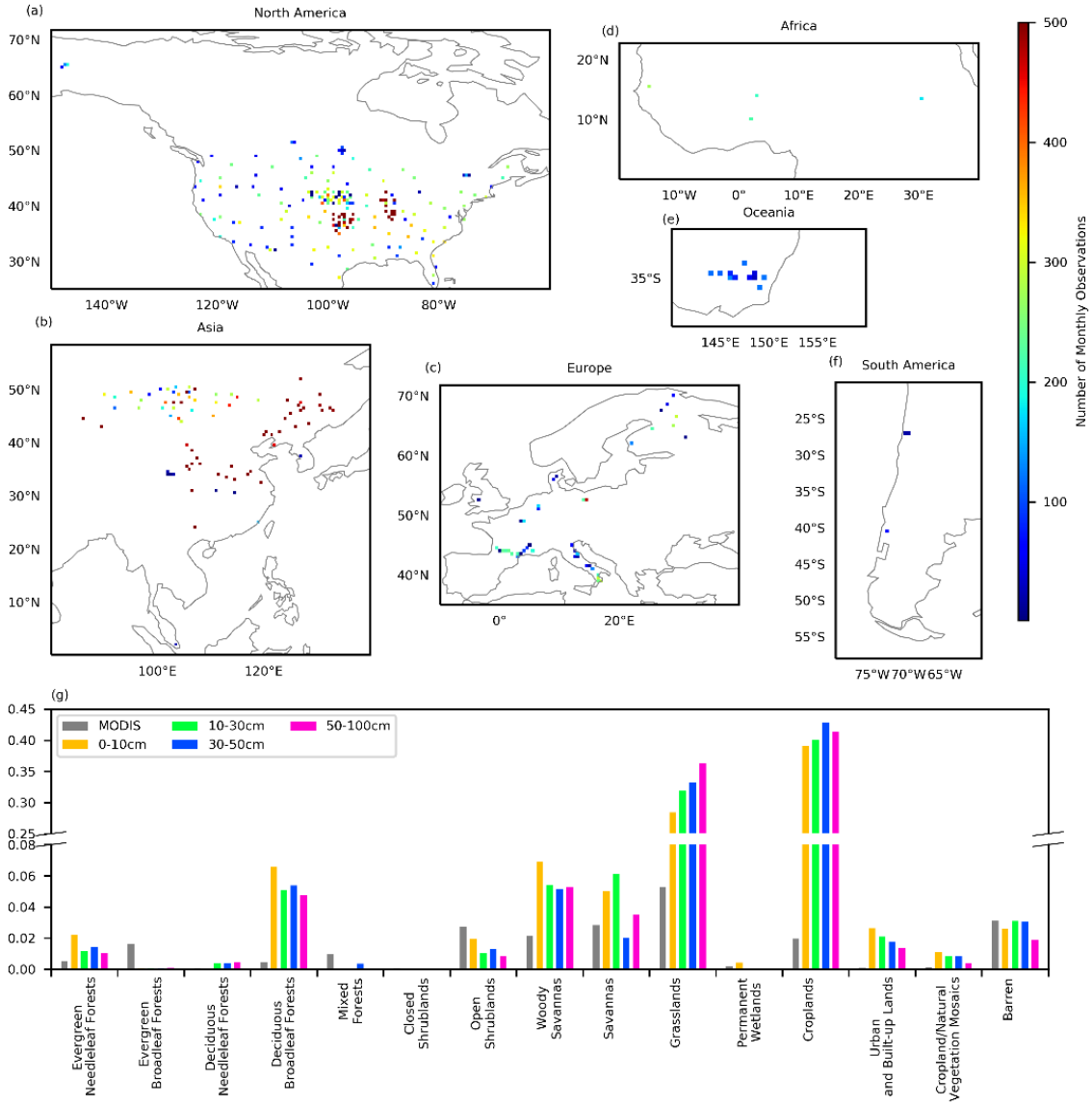


Figure S1: (a–f) Spatial distribution of the SM observations from the ISMN. The colors indicate the number of months that have valid SM observations within each 0.5° grid. When observations exist for more than one of the four target soil layers (0–10, 10–30, 30–50, and 50–100 cm), the number of months for each soil layer was summed up. (g) Fraction of station-months in each land use type according to MODIS IGBP classification (Friedl and Sulla-Menashe, 2015). The fractions of grid cells in each land use type from MODIS are shown for comparison.

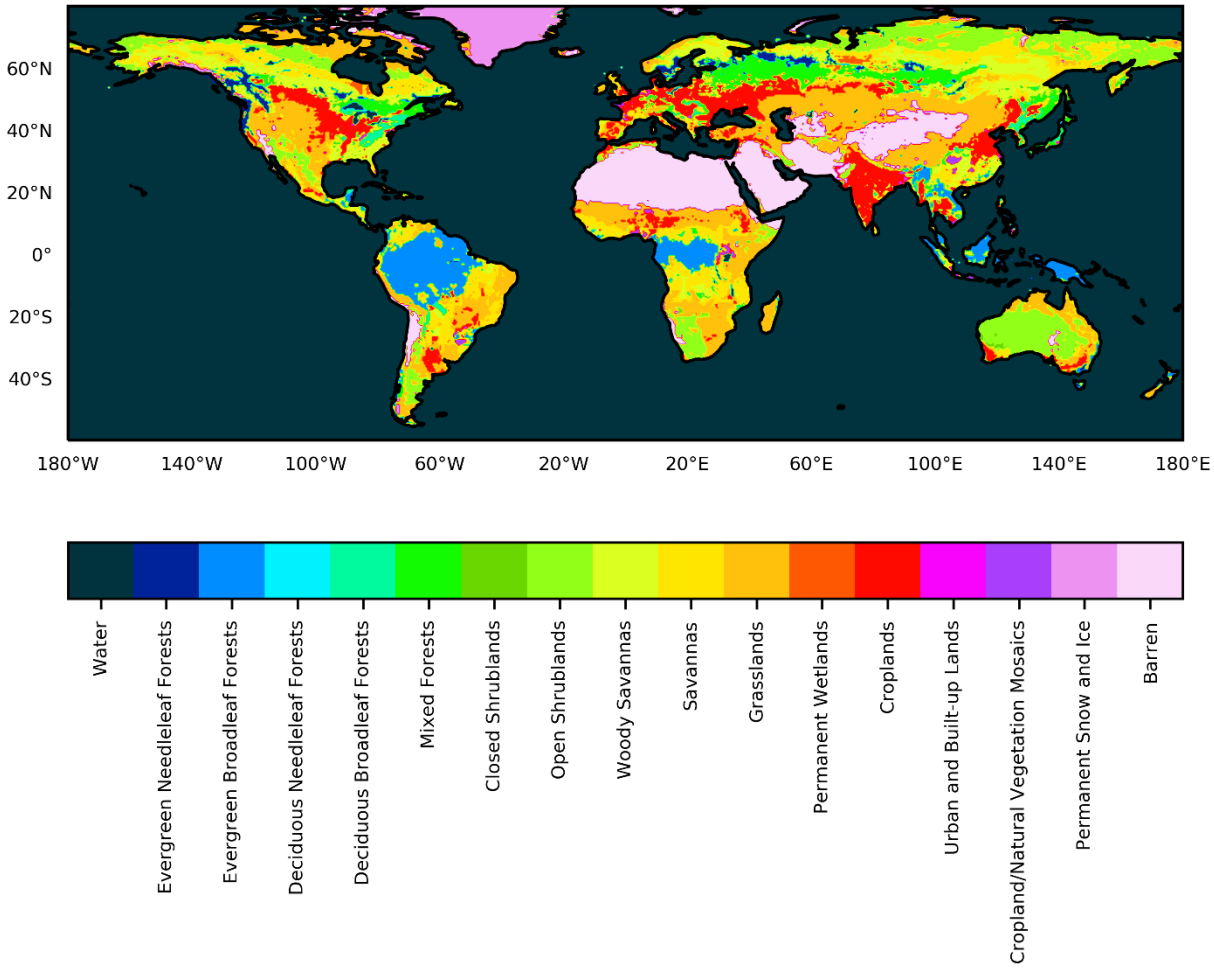


Figure S2: The spatial distribution of IGBP land cover types in the MODIS MCD12C1 product (Friedl and Sulla-Menashe, 2015).

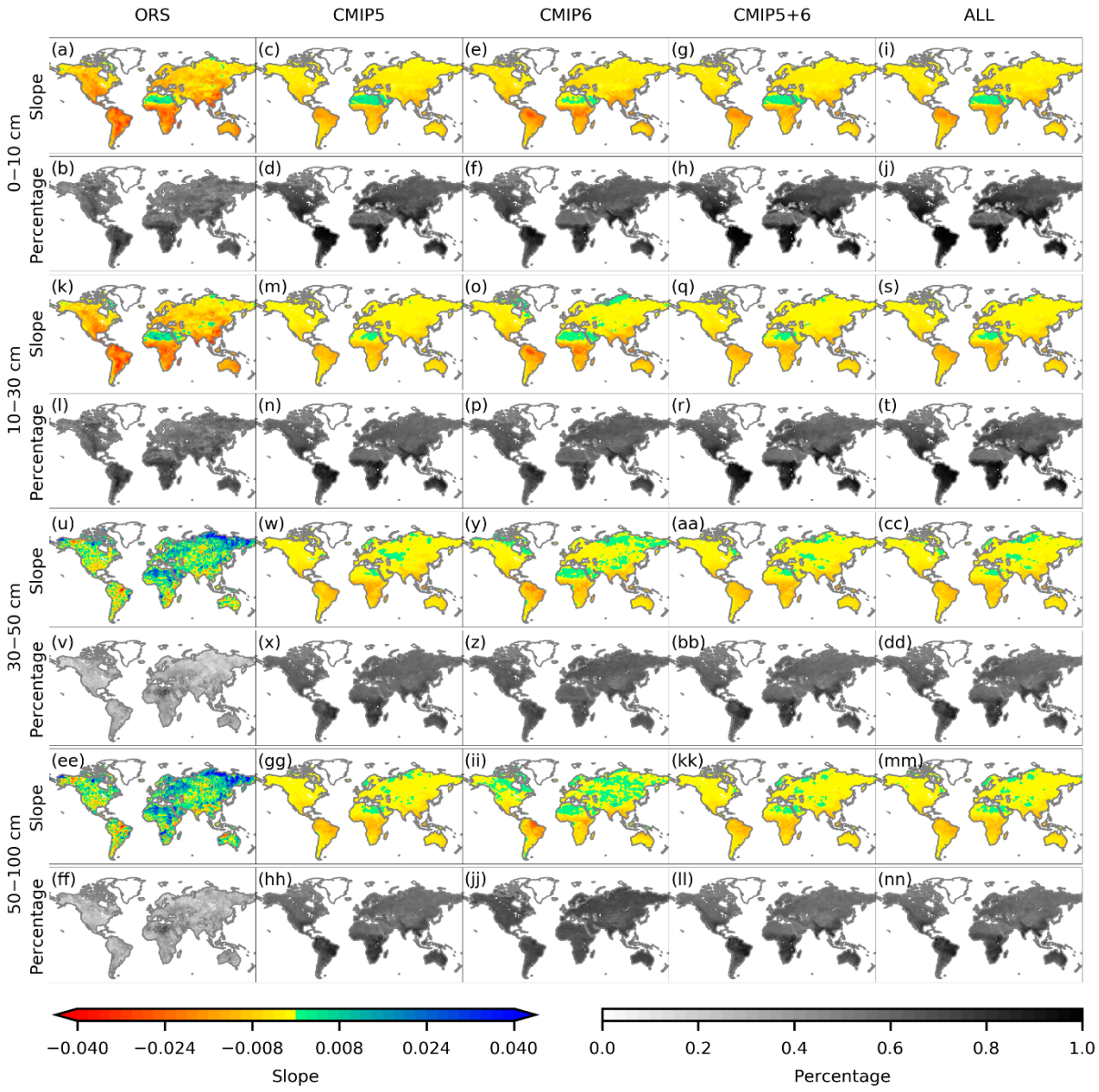


Figure S3: Averaging regression coefficients (significant-only) for the temperature part of the EC relationships, and the percentage of months when temperature was used for the EC.

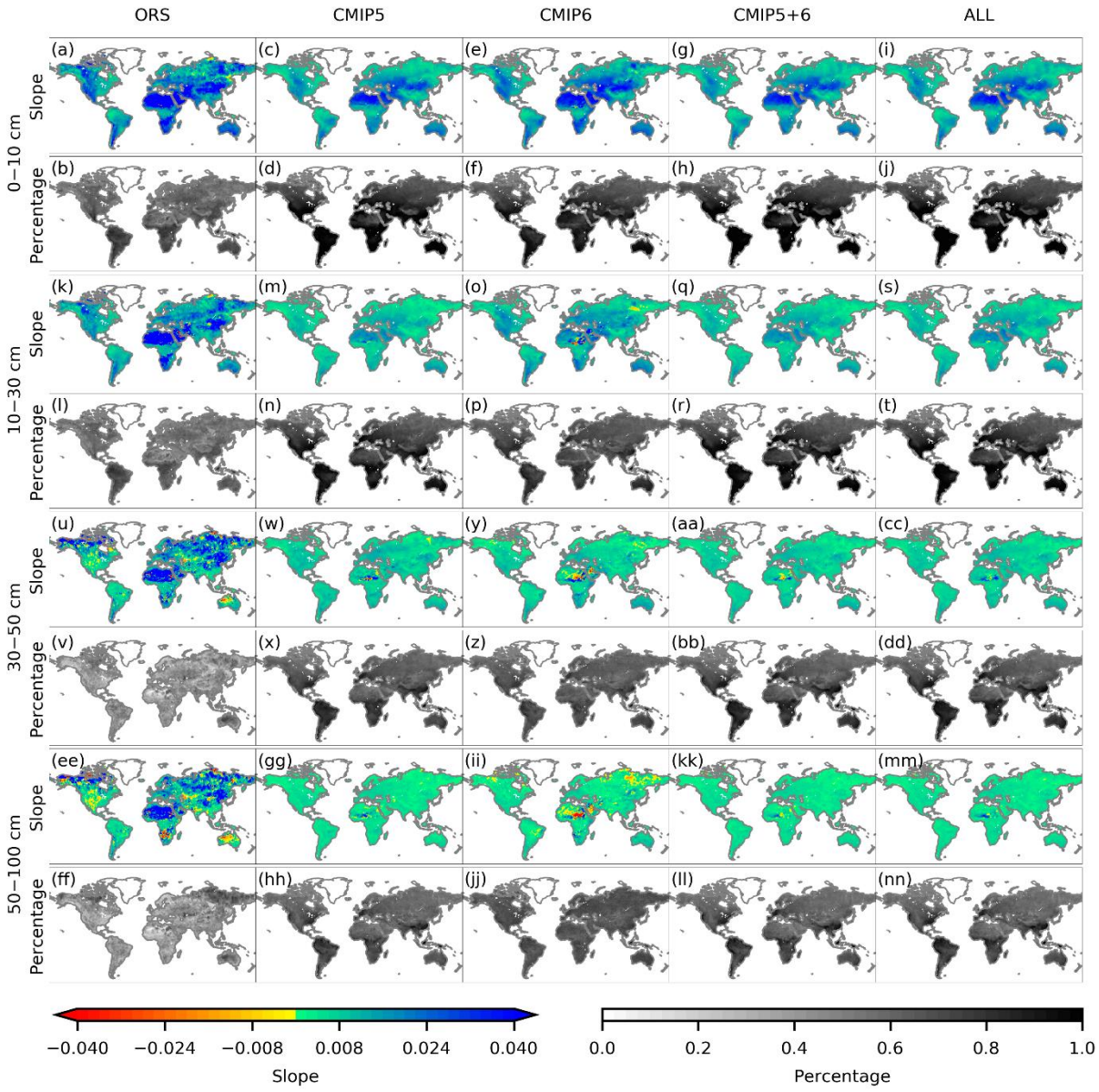


Figure S4: Averaging regression coefficients (significant-only) for the precipitation part of the EC relationships, and the percentage of months when precipitation was used for the EC.

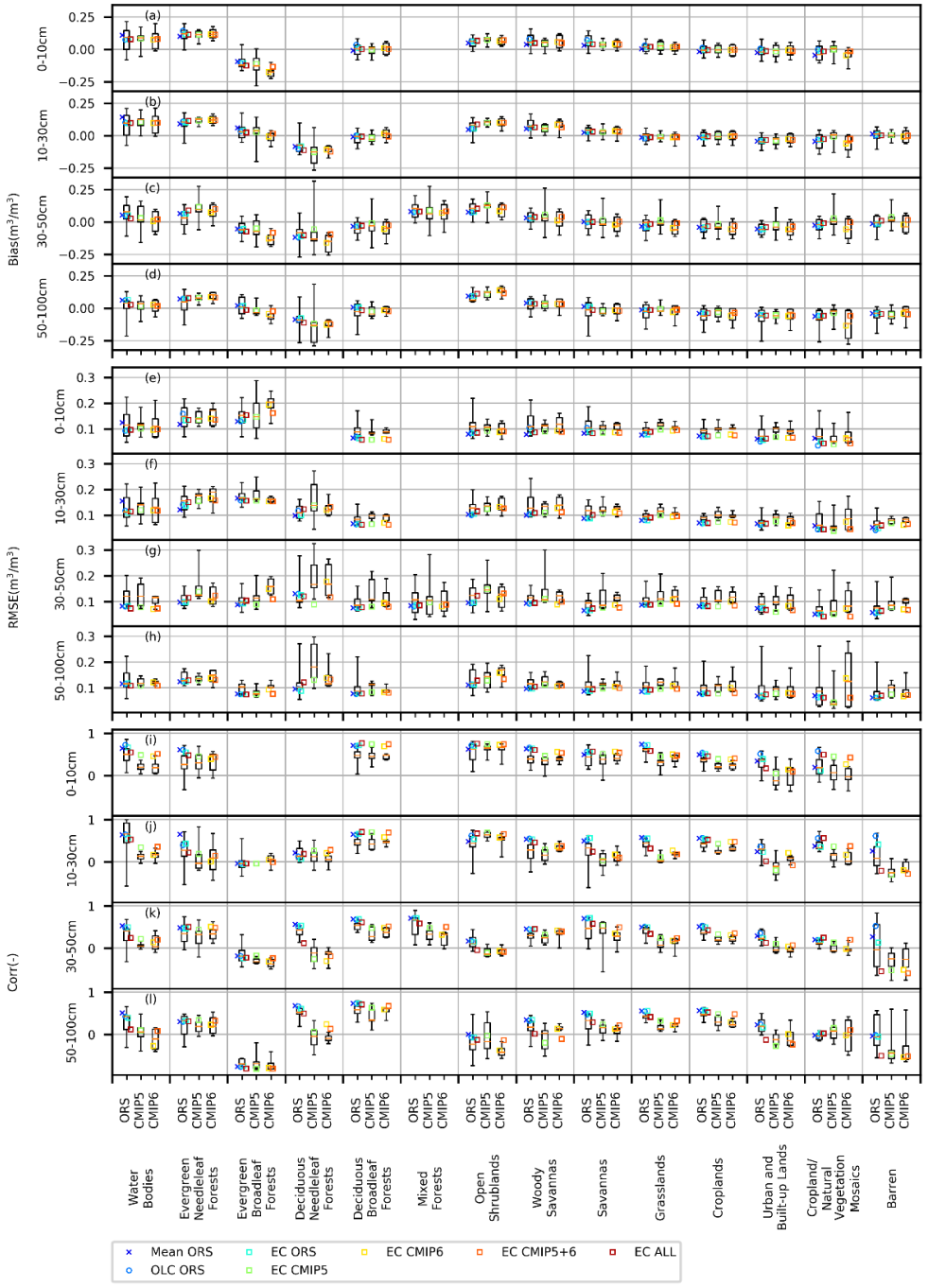


Figure S5: By land-cover type performance of the original ORS, CMIP5, and CMIP6 datasets (boxplots) and the merged products (scatter plots) on the validation set of observations. The boxplots show (from top to bottom) maximum, 75th percentile, median, 25th percentile, and minimum. The observations were classified into the most prevalent land cover types in their occupying 0.5° grids based on the MODIS MCD12C1 product (Friedl and Sulla-Menashe, 2015). Some land cover types and depths are not displayed because too few data points were available for the calculation of the validation metrics (at least 2 are needed for Bias and RMSE, at least 3 for Corr).

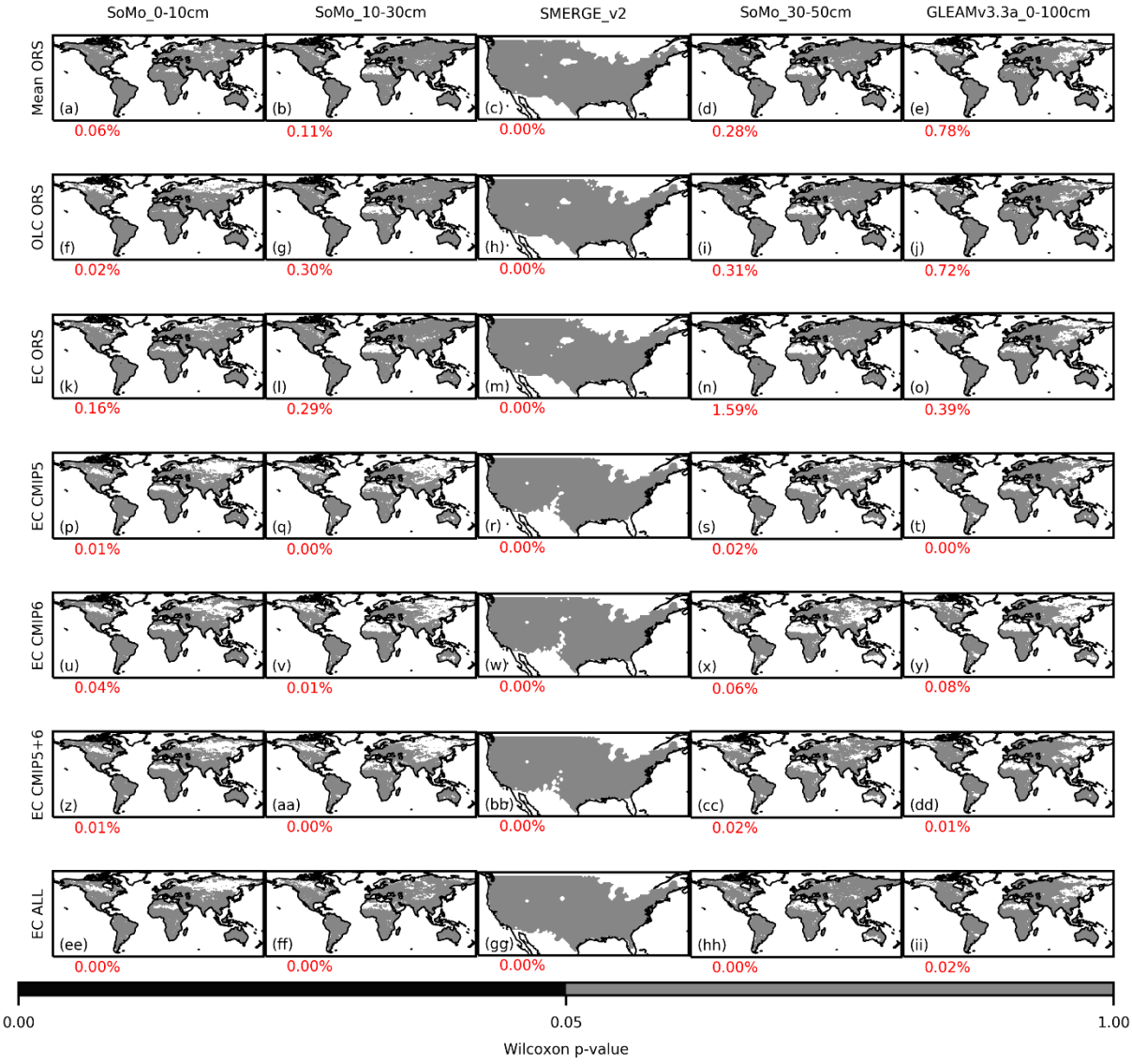


Figure S6: The *p*-values of homogeneity test on the discontinuity in mean (Wilcoxon rank-sum test) between the time periods 1981−2010 and 2010−2016 in the merged products.

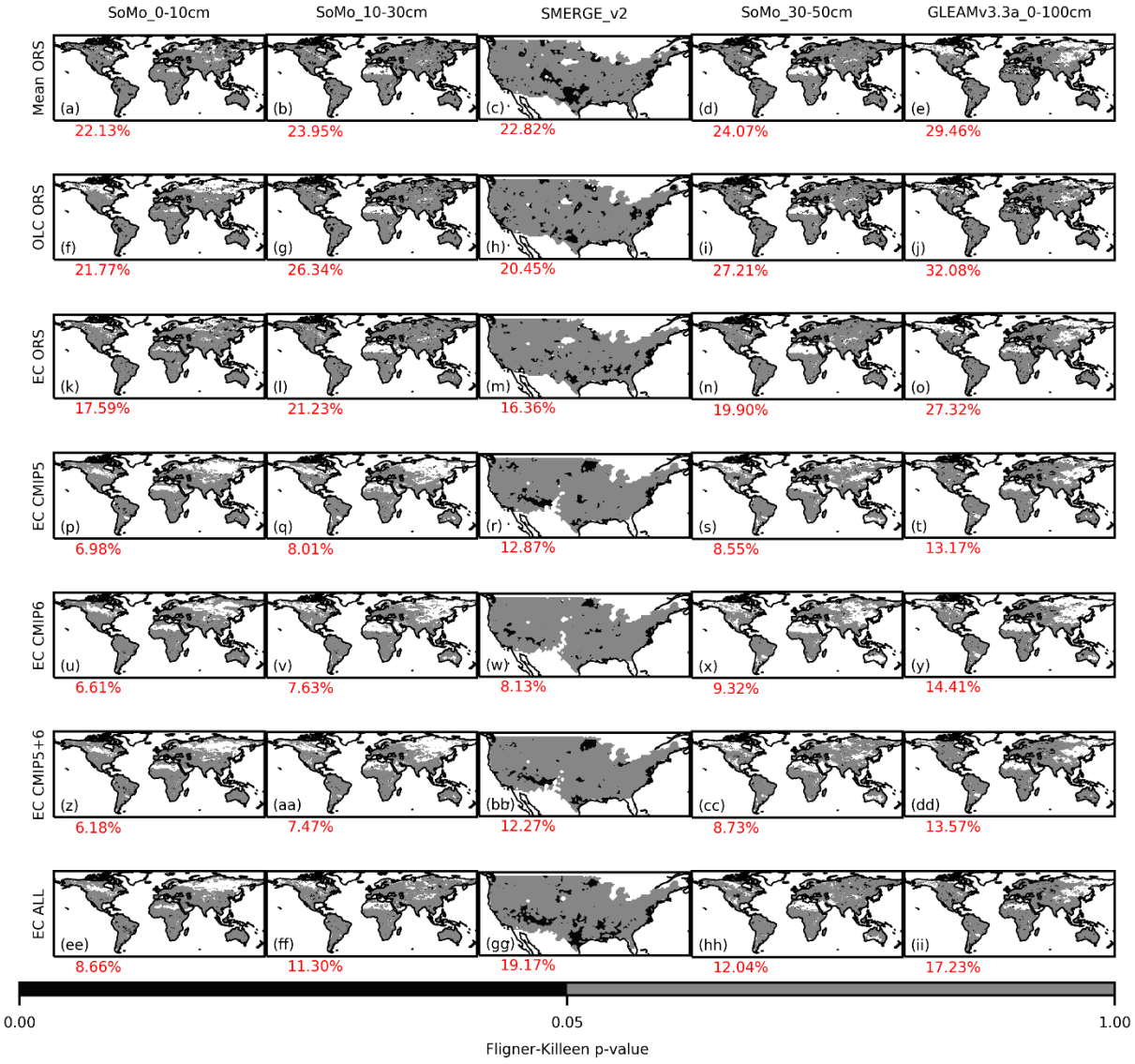


Figure S7: The *p*-values of homogeneity test on the discontinuity in variance (Fligner-Killeen test) between the time periods 1981−2010 and 2010−2016 in the merged products.

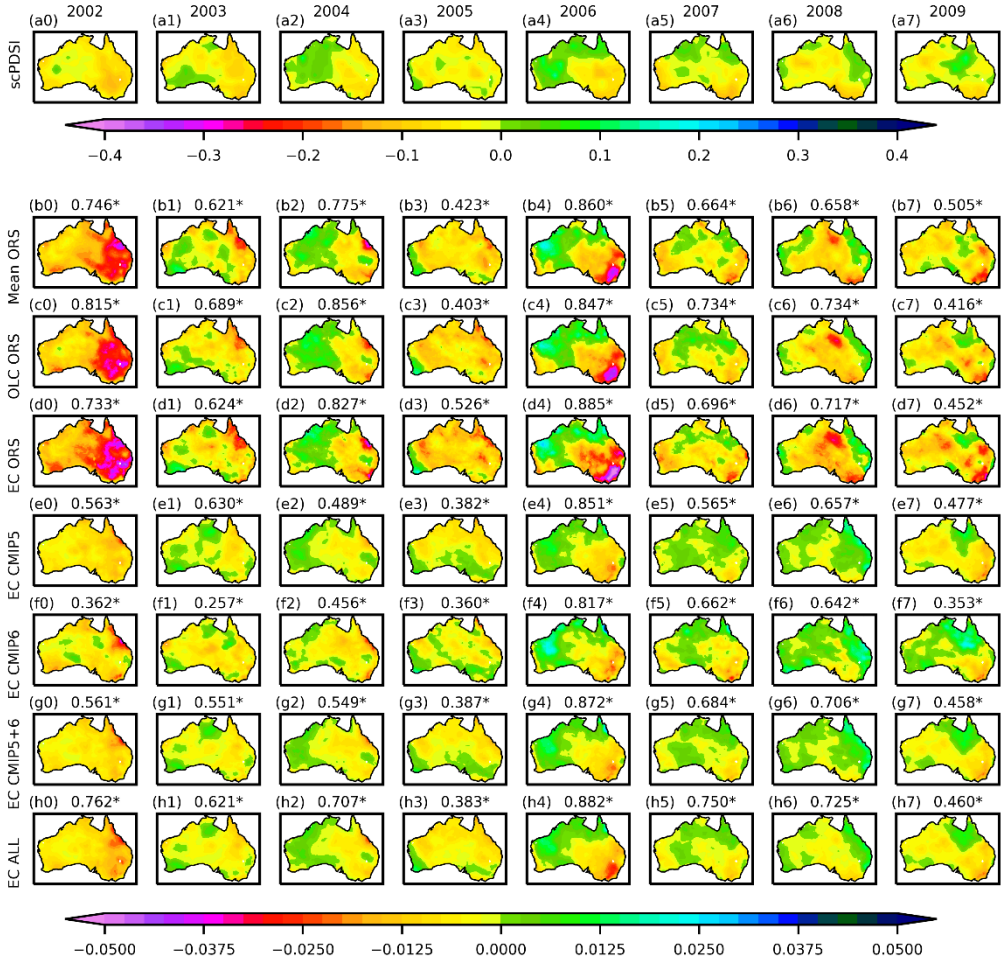


Figure S8: The annual mean scPDSI anomalies (no unit) and annual mean SM anomalies (m^3/m^3) during the Australian millennium drought 2002–2009. The numbers on top of plots are the Spearman correlation between the anomalies of the merged product and scPDSI, and the asterisk (*) indicates that the correlation is significant at $p = 0.05$. The anomalies were calculated relative to the climatology of 1970–2016. The SM anomalies are for 0–10 cm.

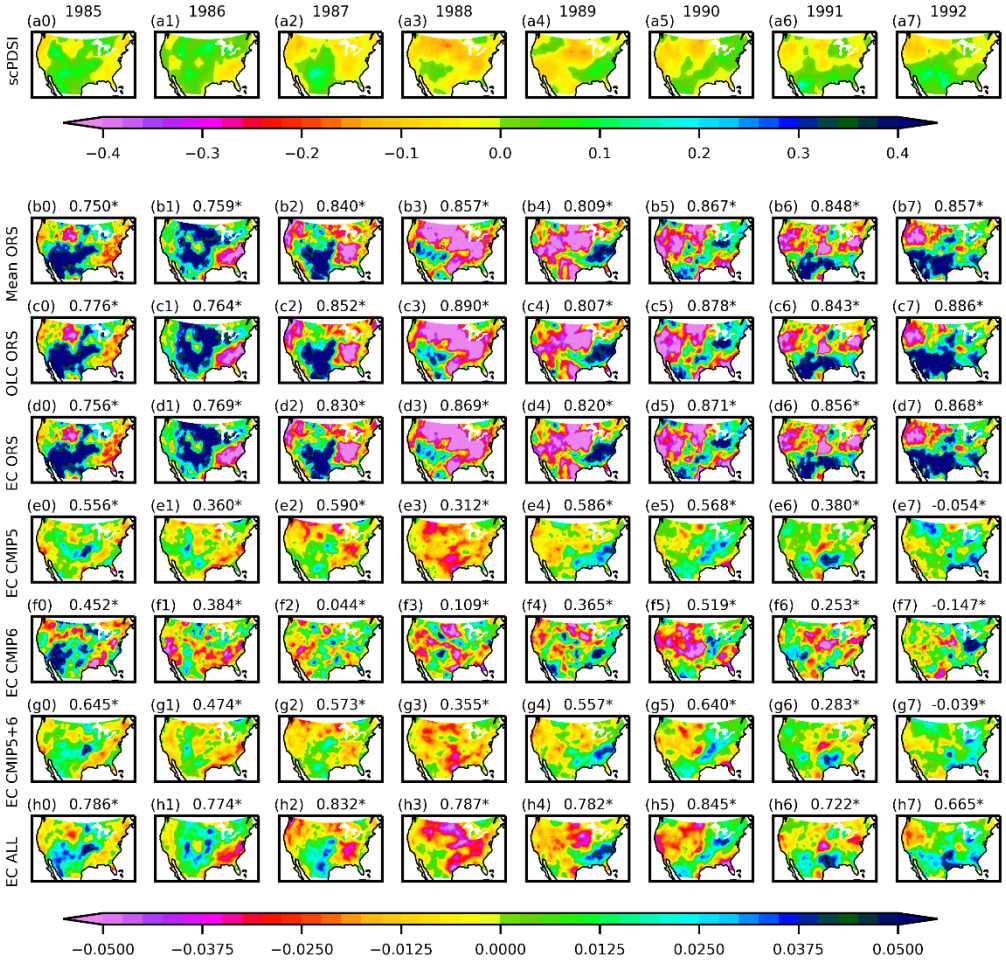


Figure S9: The annual mean scPDSI anomalies (no unit) and annual mean SM anomalies (m^3/m^3) during the US drought 1985–1992. The numbers on top of plots are the Spearman correlation between the anomalies of the merged product and scPDSI, and the asterisk (*) indicates that the correlation is significant at $p = 0.05$. The anomalies were calculated relative to the climatology of 1970–2016. The SM anomalies are for 0–100 cm.

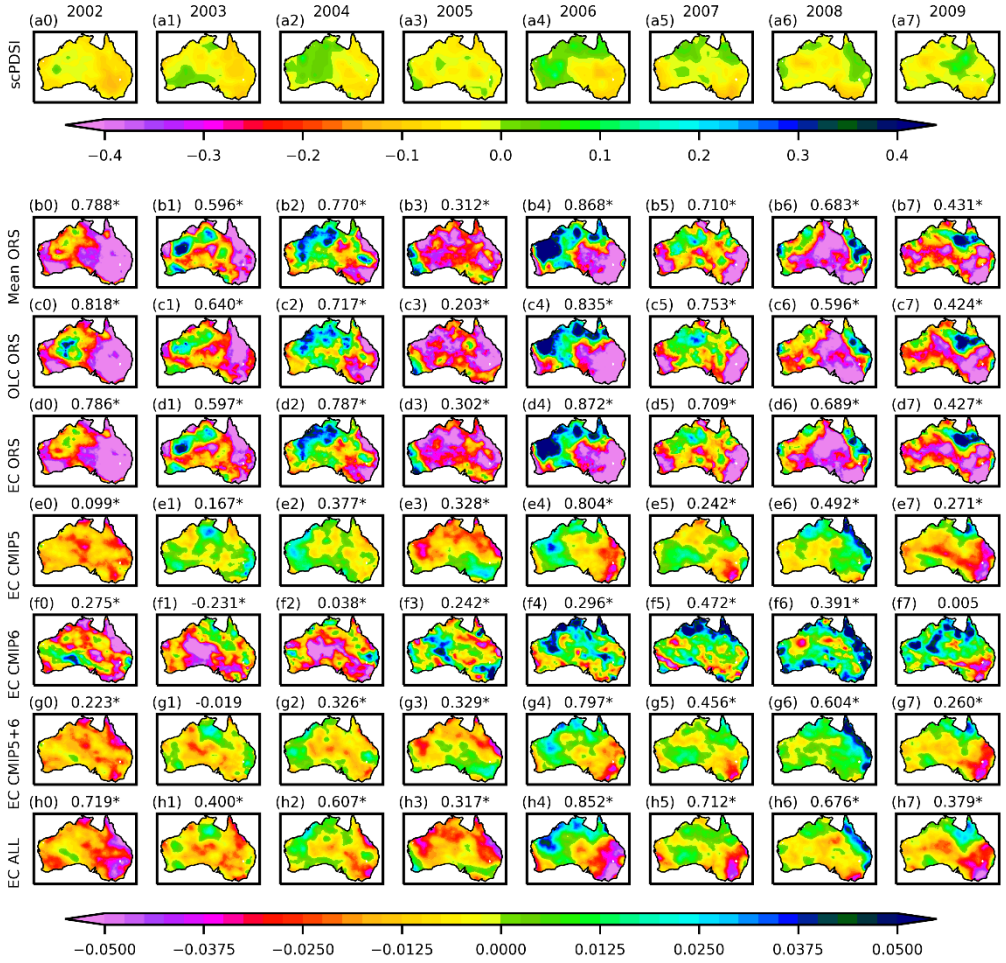


Figure S10: The annual mean scPDSI anomalies (no unit) and annual mean SM anomalies (m^3/m^3) during the Australian millennium drought 2002–2009. The numbers on top of plots are the Spearman correlation between the anomalies of the merged product and scPDSI, and the asterisk (*) indicates that the correlation is significant at $p = 0.05$. The anomalies were calculated relative to the climatology of 1970–2016. The SM anomalies are for 0–100 cm.

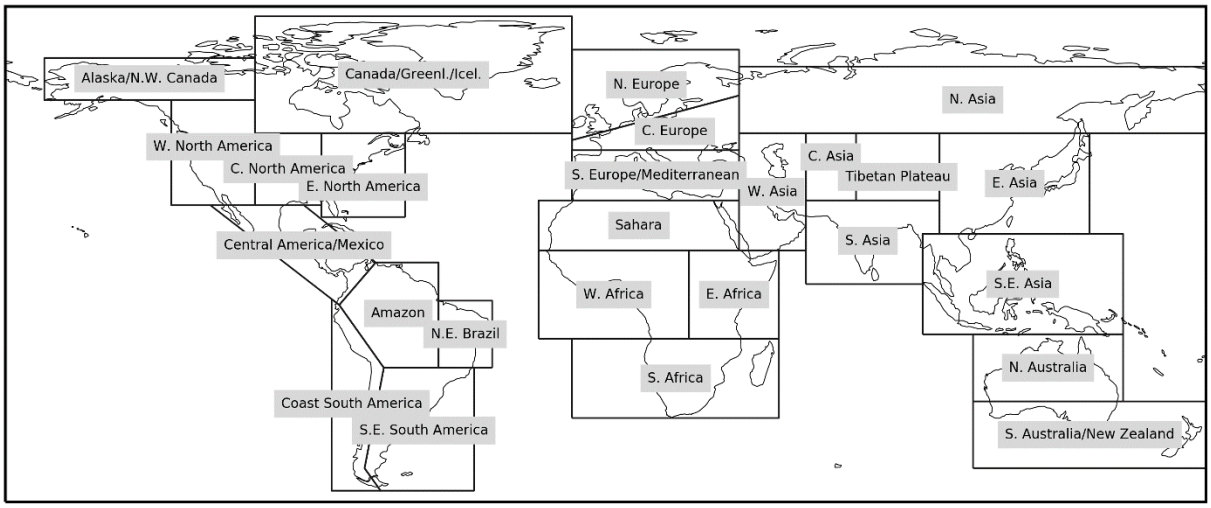
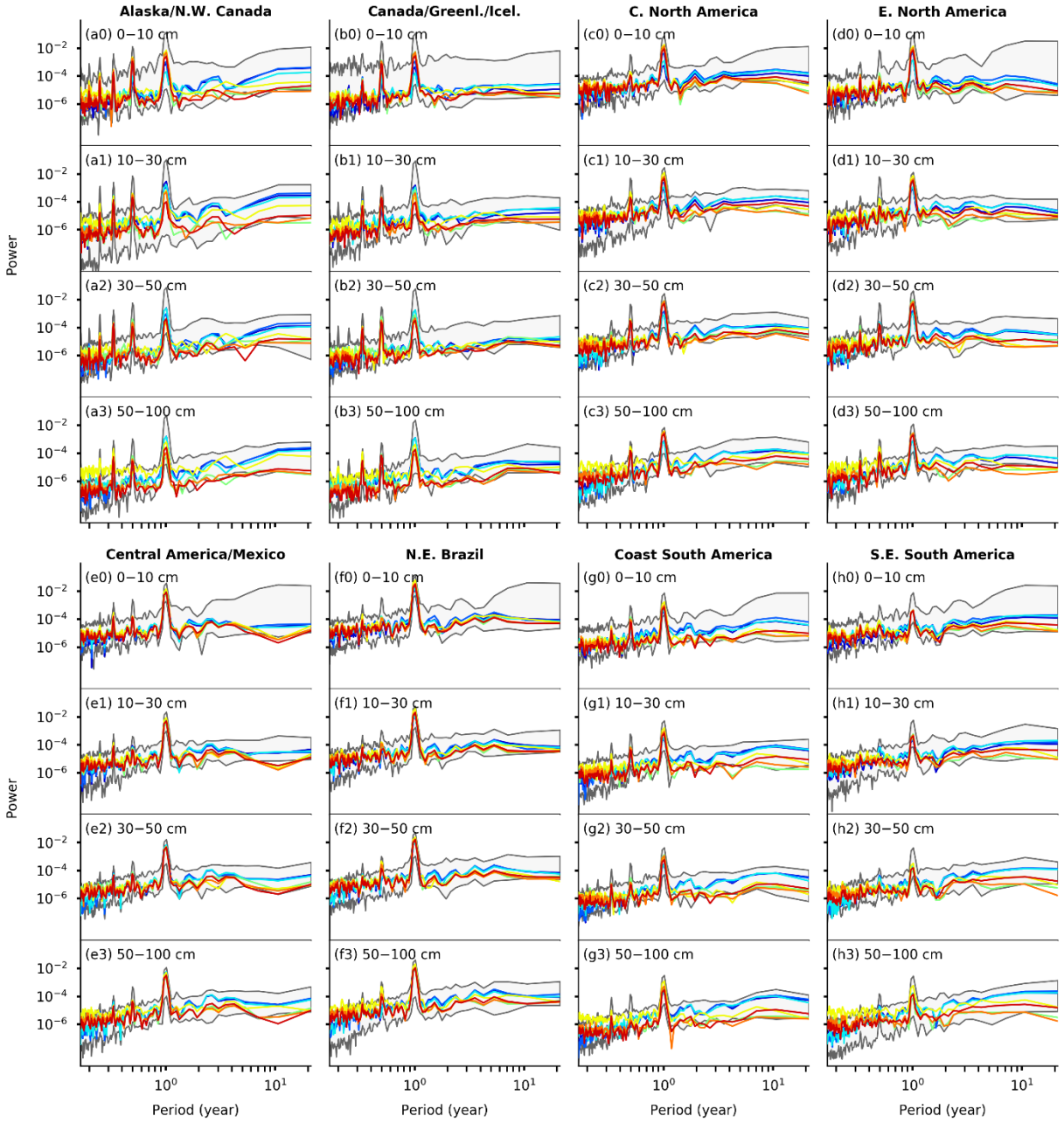
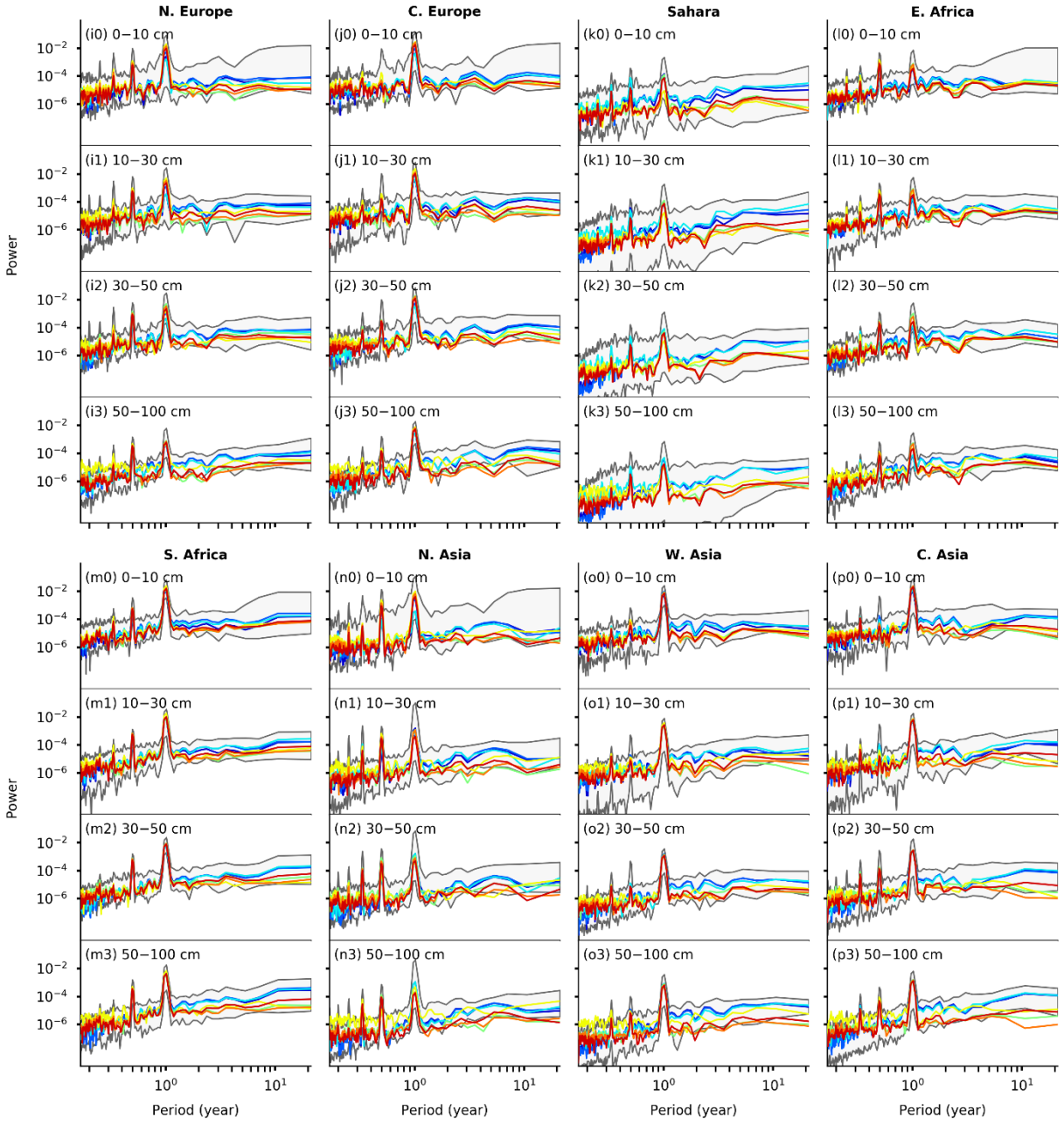


Figure S11: The locations of the IPCC SREX regions. The plot was drawn using the Python package regionmask v0.4.0 (Hauser, 2019).

85



(continued on next page)



(continued on next page)

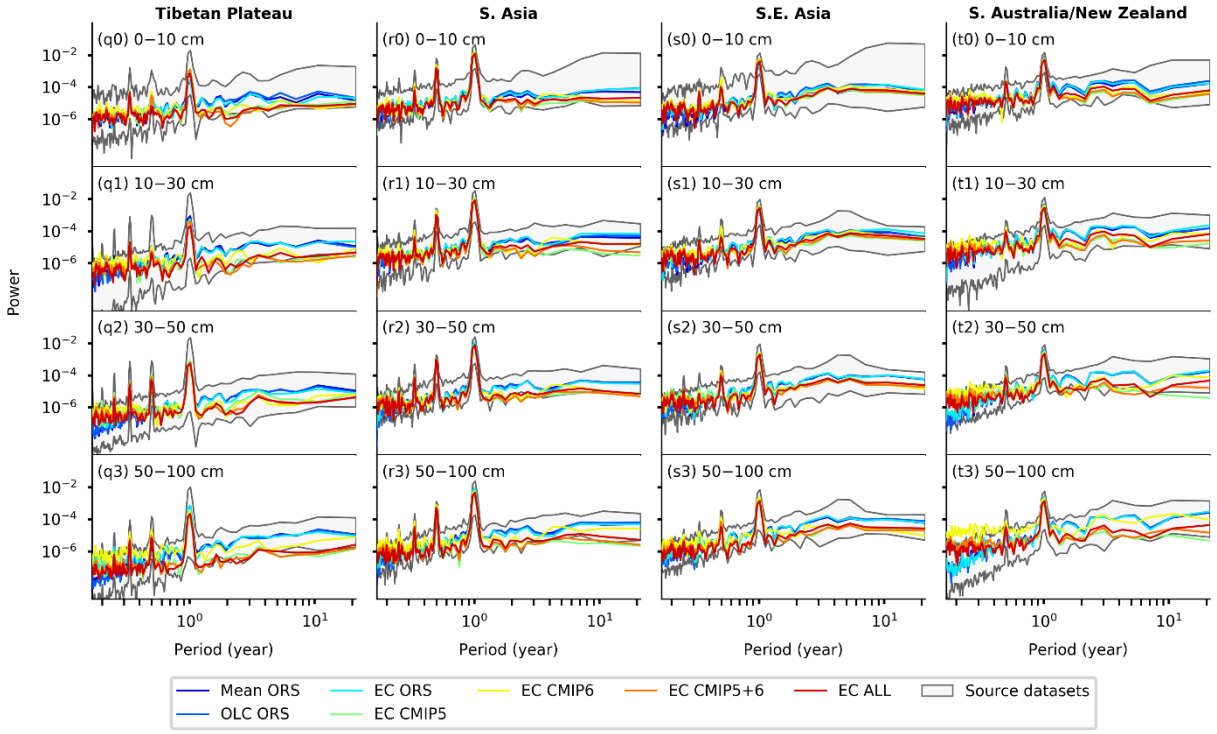


Figure S12: The power spectral density of the spatially averaged time series of monthly SM over selected IPCC SREX regions (Field et al., 2012). The envelopes of the source datasets were created from the maximum and minimum values across all the ORS, CMIP5, and CMIP6 datasets. Abbreviations: C. – central, W. – west, N. – north, S. – south, E. – east.

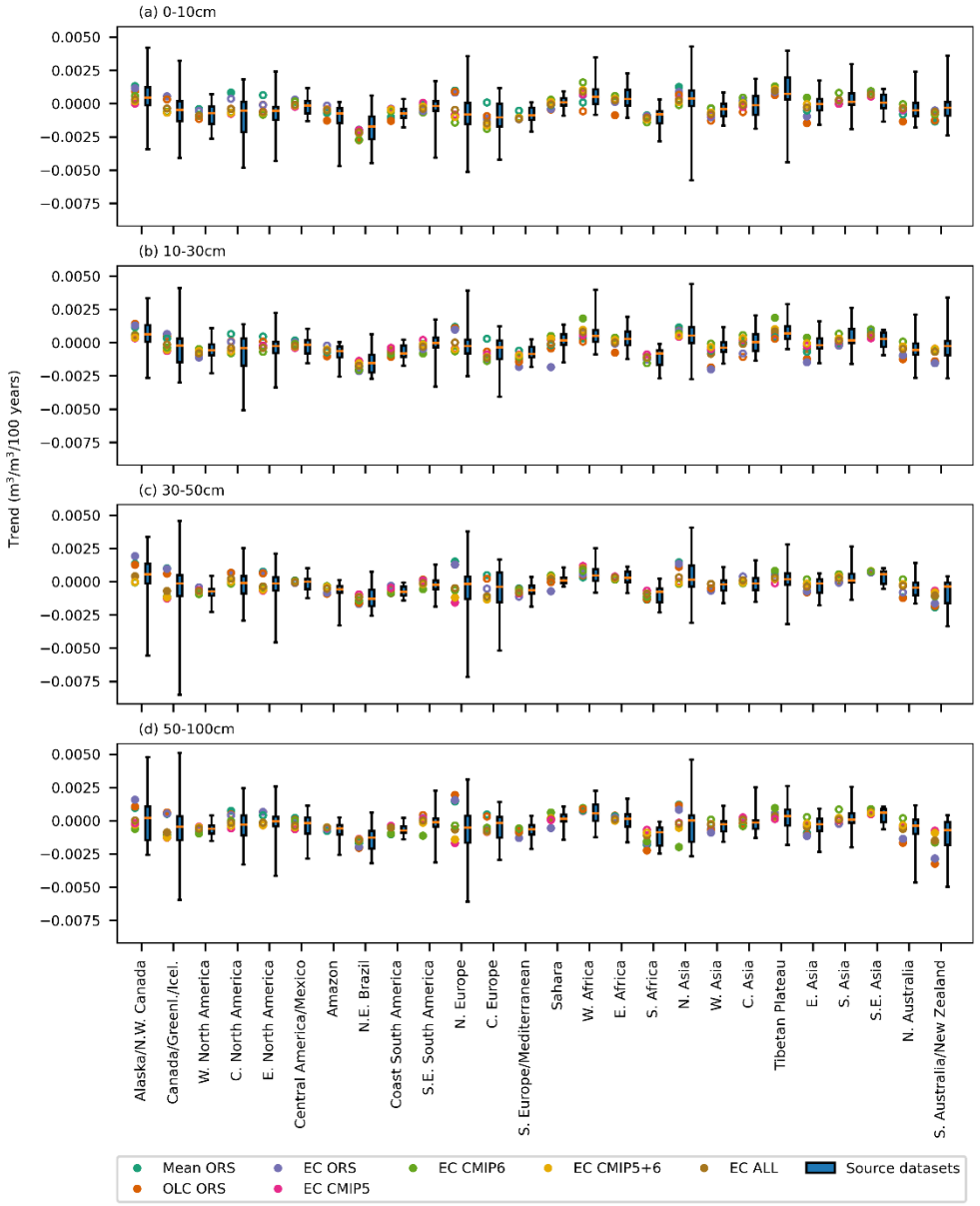


Figure S13: The least squares linear trend in regionally averaged SM between 1971 and 2016 in the merged products and the source datasets (ORS, CMIP5, and CMIP6). Solid dots indicate significant trends in the merged product at $p = 0.05$, and hollow dots indicate insignificant trends ($p > 0.05$). The boxplots show (from top to bottom), the maximum, 75th percentile, median, 25th percentile, and minimum trends across the source datasets. The boxplots only include the source datasets that span the entire 1971−2016 period, and the significant and insignificant trends were not separately presented to keep the plots readable. The regions are IPCC SREX regions (Field et al., 2012).

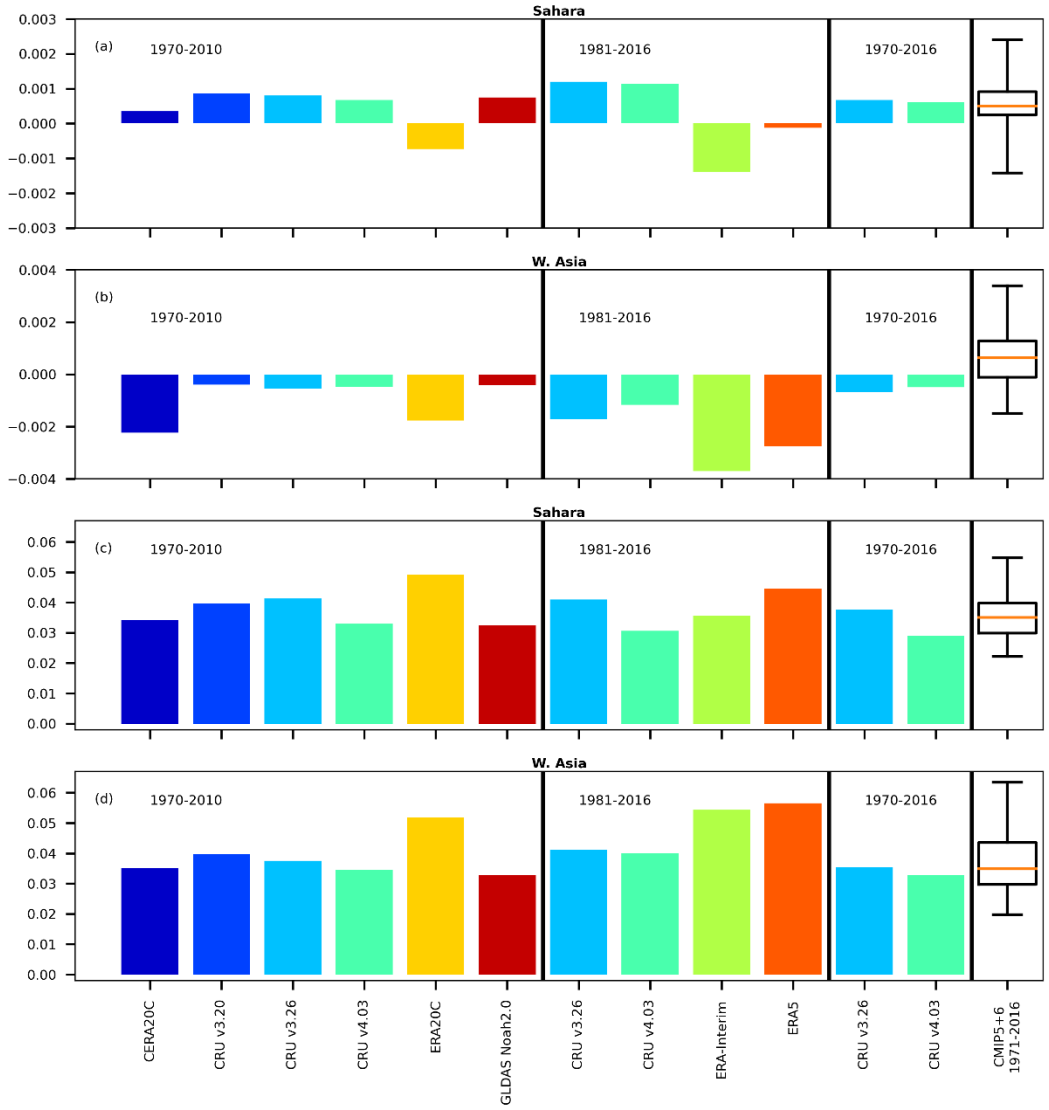
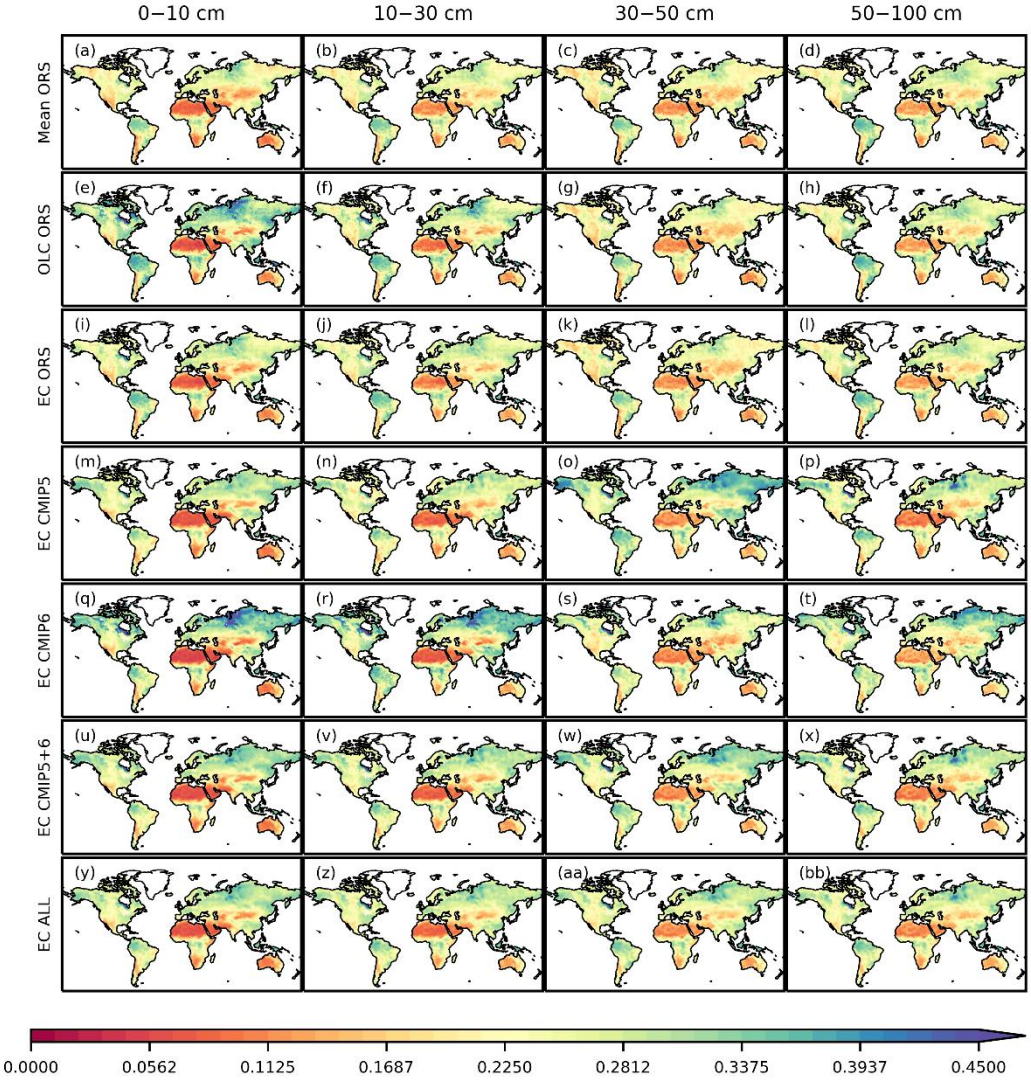


Figure S14: Historical least squares linear trends in the precipitation and temperature drivers of the source datasets (CERA20C, CRU v3.20, CRU v3.26, ERA20C, ERA-Interim, ERA5, GLDAS Noah2.0, CMIP5, and CMIP6), and the CRU TS v4.03 dataset, which was used to create the EC-based merged datasets. Some of source datasets do not cover the entire 1970−2016 period, and therefore were compared to the CRU TS v4.03 dataset over the overlapping periods (1970−2010 or 1981−2016). The trends were calculated after the SM values were averaged over the Sahara and the West (W.) Asia regions. The boxplots show (from top to bottom), the maximum, 75th percentile, median, 25th percentile, and minimum trends, across the CMIP5 and CMIP6 ESMs. Significance of the trends were not indicated and the boxplots include both significant and insignificant trends.



115 Figure S15: The annual mean climatology of SM (m^3/m^3) for 1971–2016 in the merged products.

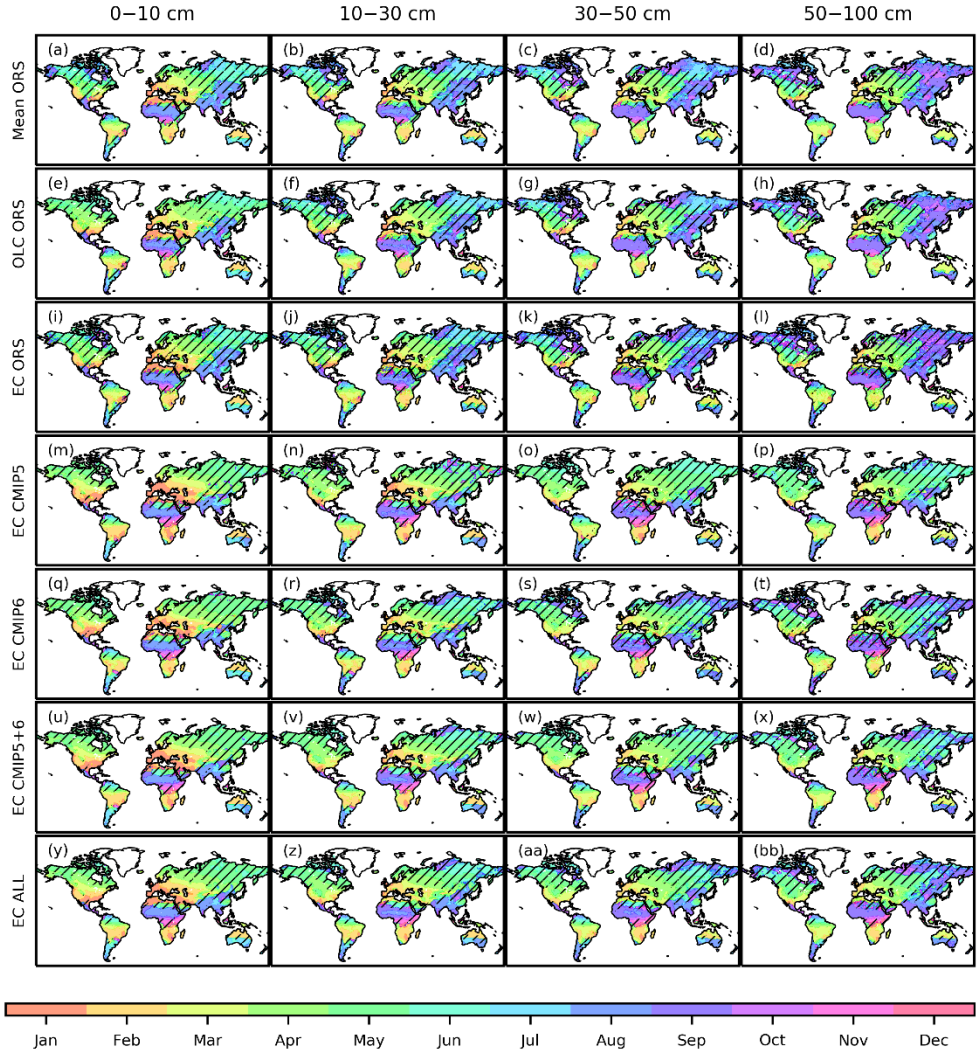


Figure S16: The month of maximum SM in the monthly climatology of 1971–2016 of the merged products. Hatching indicates that more than one local maximum exists in the monthly climatology.

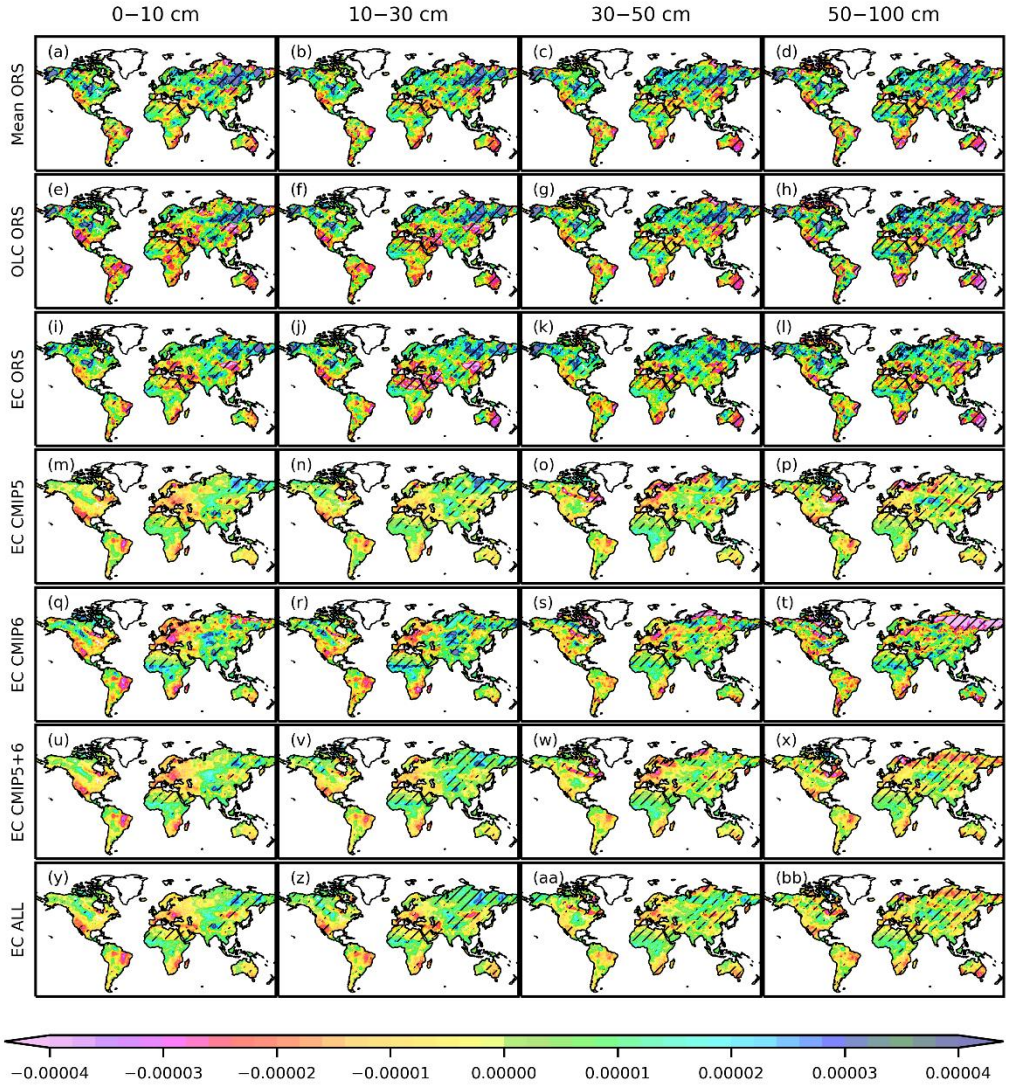


Figure S17: The least squares linear trend in SM (m^3/m^3 per year) between 1971 and 2016 in the merged products. Hatching indicates significance at $p = 0.05$. The SM values of all the months of the year were pooled in calculating the trend.

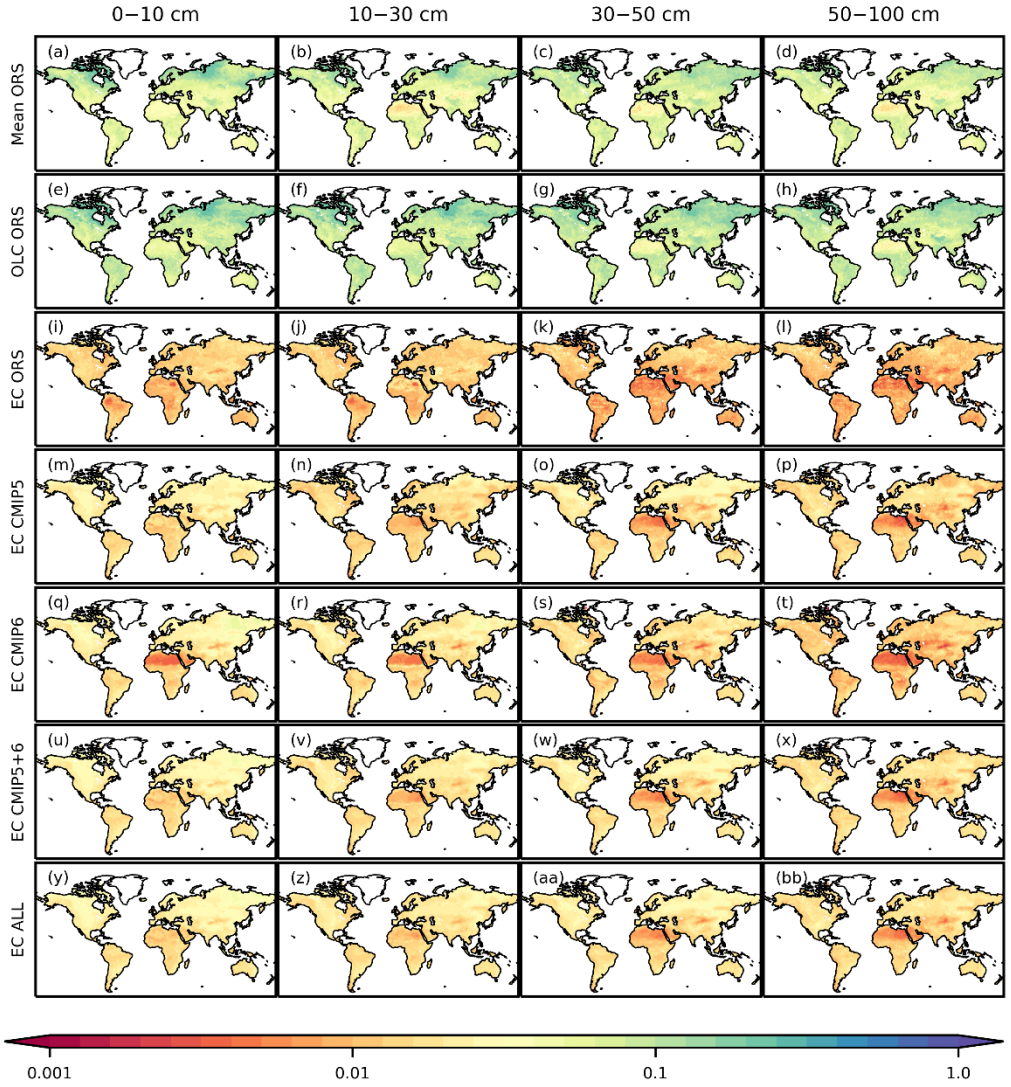


Figure S18: The uncertainty, as characterized by standard deviation, in the estimated SM (m^3/m^3) between 1971 and 2016 in the merged products. The uncertainty of the Mean ORS product was estimated by the standard deviation of the source datasets at each grid. The uncertainties of the OLC- and EC-based products were estimated following the methods in Sects. 2.7 and 2.8. All the uncertainty values were simply averaged over all time steps in the merged products.

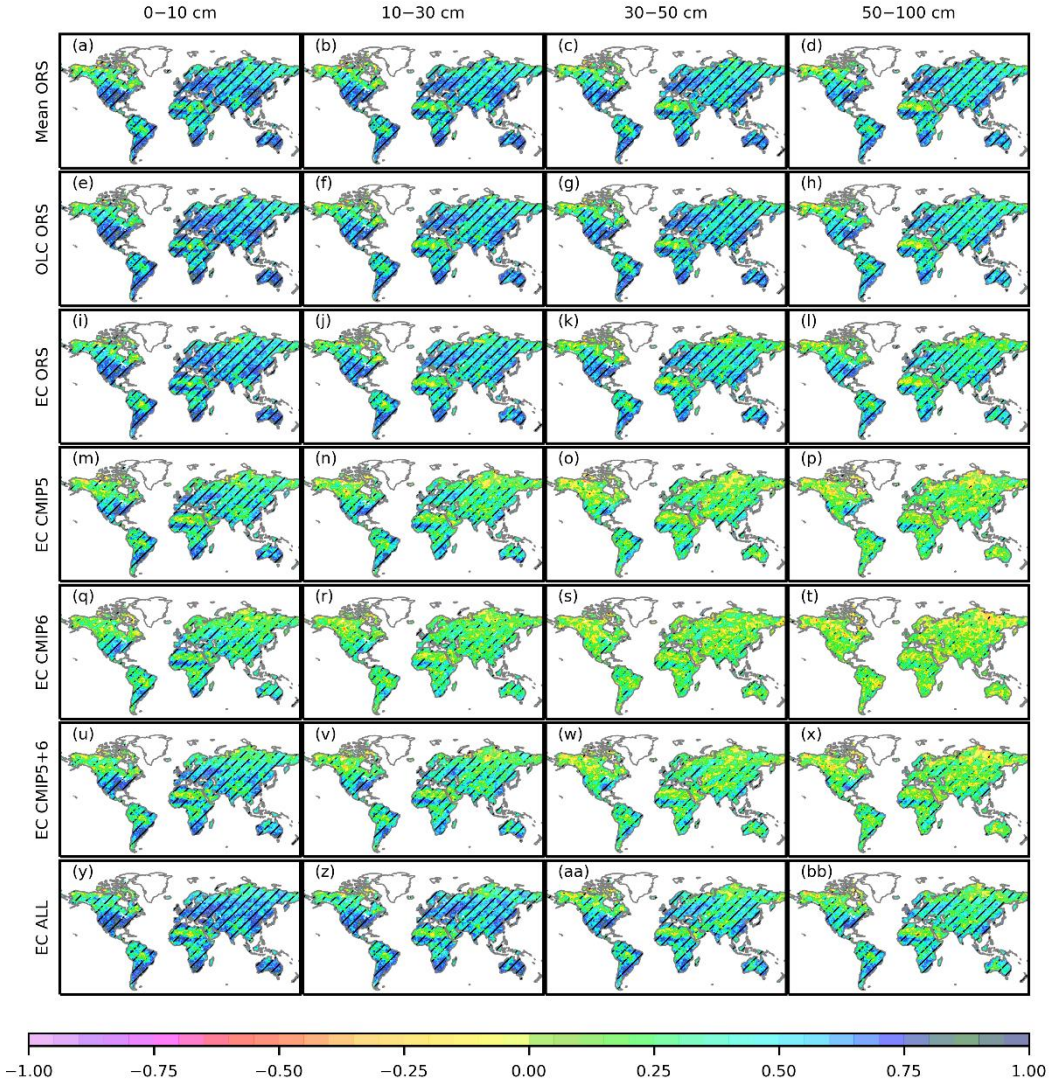


Figure S19: Partial correlation coefficients (no unit) between annual mean SM in the merged products and observed precipitation. Hatching indicates significance at $p = 0.05$.

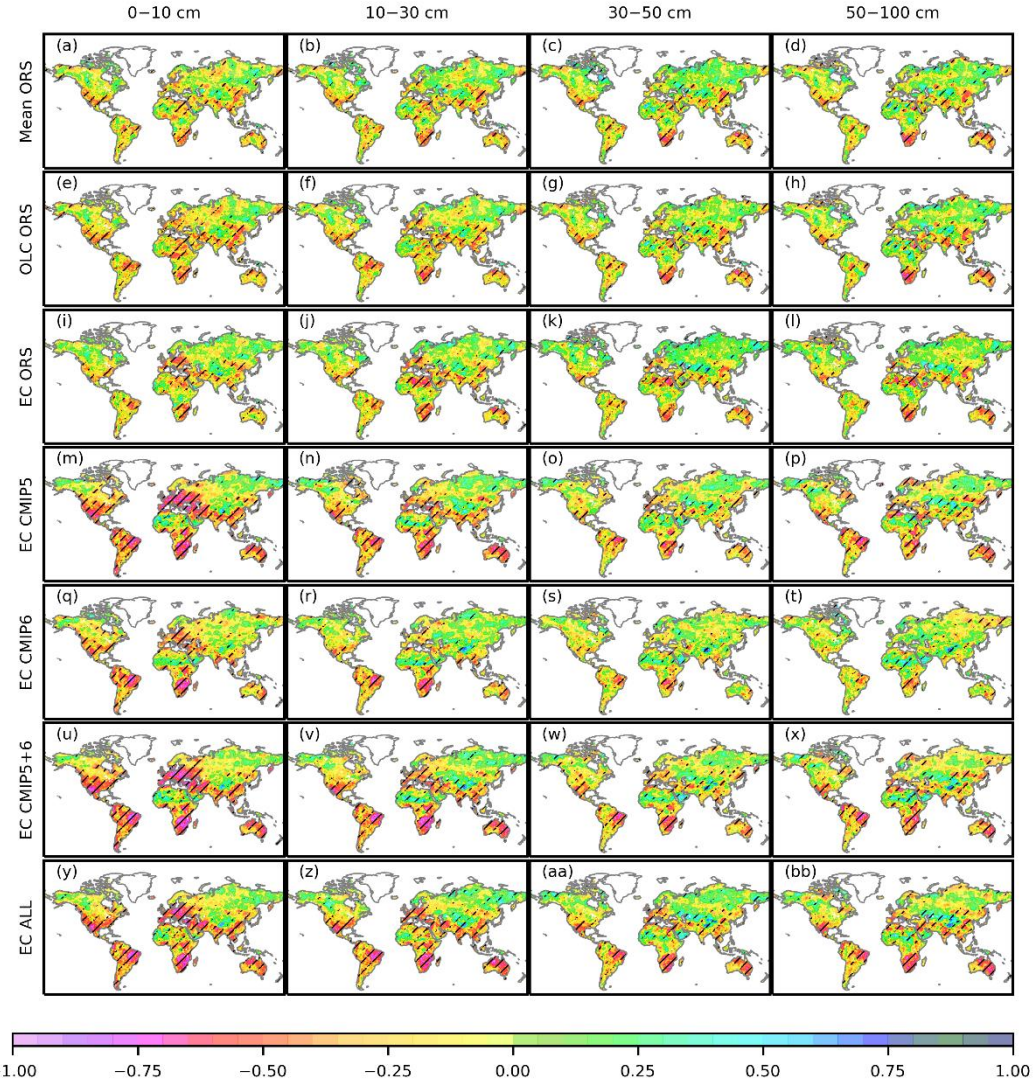


Figure S20: Partial correlation coefficients (no unit) between annual mean SM in the merged products and observed air temperature. Hatching indicates significance at $p = 0.05$.

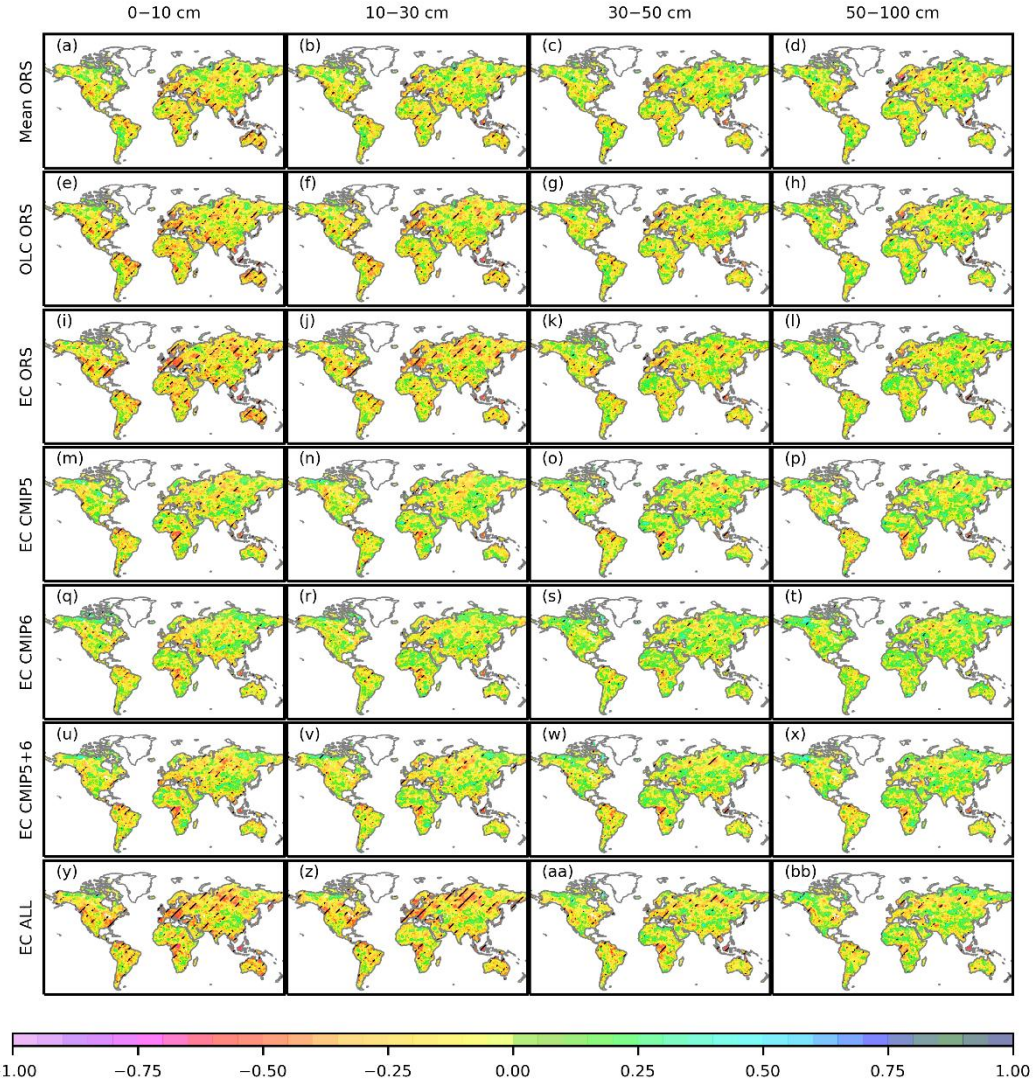


Figure S21: Partial correlation coefficients (no unit) between annual mean SM in the merged products and observed downwelling shortwave radiation. Hatching indicates significance at $p = 0.05$.

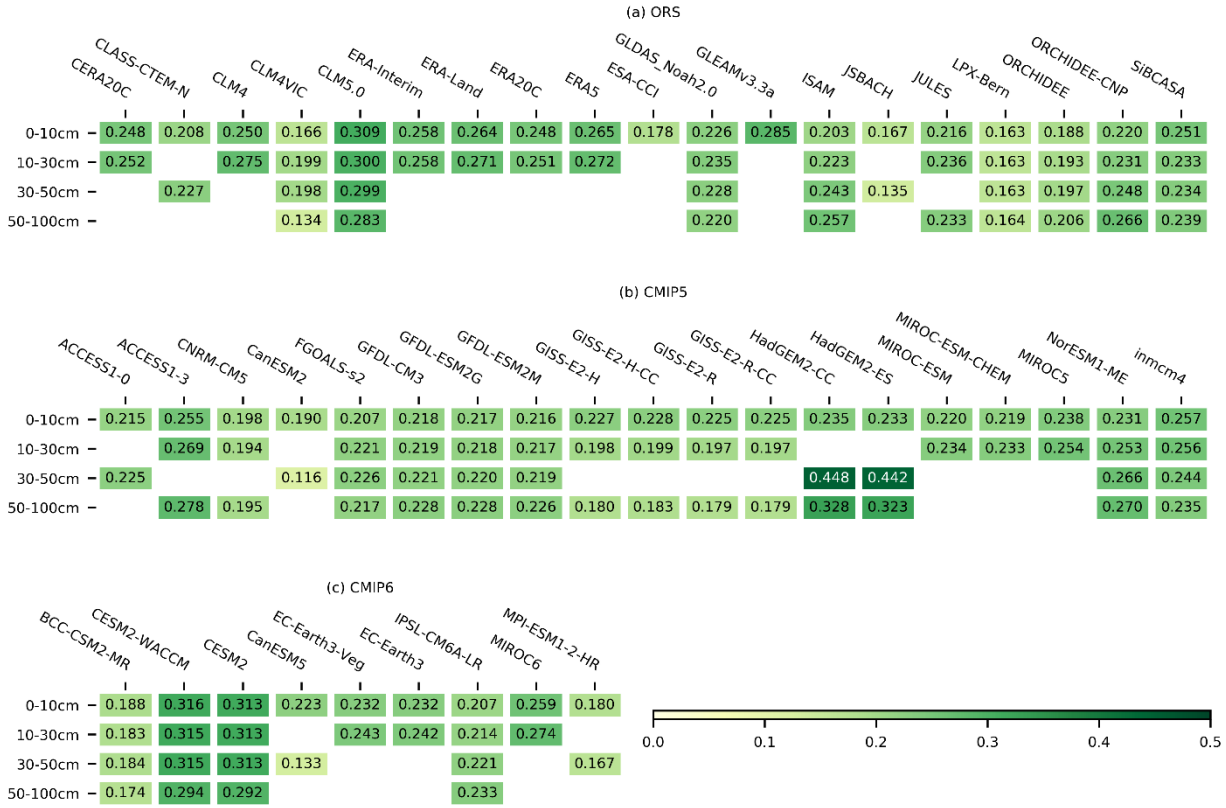


Figure S22: Global mean SM values (m^3/m^3) of the source datasets.