

Depth-dependent propagation of precipitation deficits through the root zone and cropland exposure to soil-moisture drought across China

Jun Zhang

junzh2026@163.com

Chongqing Water Resources and Electric Engineering College

Research Article

Keywords: Agricultural drought, Drought propagation, Root-zone soil moisture, Soil-moisture memory, Spatiotemporal drought events, Cropland exposure

Posted Date: September 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10794446/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Precipitation deficits reach different soil depths after different delays, but grid-cell summaries do not show whether soil-moisture drought forms spatially connected events or how much cropland those events affect. We quantified these patterns across China at 0.1° for 2000–2022 using the standardized precipitation index (SPI) at 1-, 3-, and 6-month accumulation scales; the standardized soil moisture index (SSI) for the upper (0–40 cm), lower (40–100 cm), and full (0–100 cm) root zones; directional event matching; and connected-event tracking. Among five precipitation products compared with 840 gauges, CHM_PRE V2 showed the closest agreement with the gauge record and was selected for the remaining analyses. For SPI-3, the area-weighted mean of grid-cell maximum correlations was higher in the upper than in the lower root zone, and the corresponding area-weighted mean optimal lag was shorter. Correlations in the lower zone declined more slowly with increasing lag. SPI-3 gave the highest critical success index (CSI) and F₁ score, although SPI–SSI event correspondence remained modest [probability of detection (POD) ≤ 0.33; false-alarm ratio (FAR) ≥ 0.68]. Across the three profile-specific catalogs, 229,003 connected SSI droughts were identified; the largest 1% accounted for 71–78% of cropland area–time exposure. Exposure was concentrated in the eastern agricultural plains, while western arid and high-elevation regions showed weaker coupling and longer lags. Meteorological XGBoost models explained 5.1–15.9% of held-out SSI variance, and SSI showed strong month-to-month persistence ($r = 0.66–0.70$). Precipitation indices alone did not capture the depth-dependent persistence or spatial concentration of root-zone drought.

1 Introduction

Agricultural drought develops as atmospheric water deficits are transmitted through the soil profile rather than as a direct copy of meteorological drought. Infiltration, evapotranspiration, runoff, and antecedent storage filter the precipitation signal before it appears as anomalously low root-zone moisture (Koster et al. 2004, Seneviratne et al. 2010). Soil-moisture drought may consequently begin later, persist longer, and occupy a different spatial footprint from the corresponding precipitation deficit (Sheffield and Wood 2008). Precipitation and soil-moisture indices therefore describe different stages of the drought process (Apuv et al. 2017).

The gridded precipitation product used to calculate SPI is another source of uncertainty. Satellite, reanalysis, gauge-interpolated, and merged products differ in sampling, bias correction, and gauge dependence, especially across China's monsoon gradients and complex terrain. Most evaluations emphasize precipitation totals and conventional error statistics, but agreement in monthly rainfall does not necessarily imply agreement in standardized drought anomalies or event timing (Beck et al. 2017, Sun et al. 2018). Product screening for drought applications should therefore consider both precipitation fields and the SPI derived from them (Stagge et al. 2015).

Soil depth also affects drought propagation. Moisture in the upper soil profile responds rapidly to infiltration and atmospheric demand, whereas deeper storage changes more slowly and can retain anomalies after the surface has begun to recover (Rahmati et al. 2024). Observational studies further show that the precipitation accumulation period most closely associated with soil-moisture drought tends to increase with depth and that SPI alone reproduces only part of the observed soil-water signal. A single shallow layer or bulk 0–100 cm average can therefore obscure differences in response time and persistence within the root zone (Dai et al. 2022, Zhu et al. 2021).

Most drought-propagation studies rely on grid-cell correlations or local event series. These approaches quantify coupling and lag, but they do not retain the spatial connectivity, expansion, merging, and persistence of drought patches. Tracking connected space–time drought events retains information on duration, footprint, movement, and area–time exposure (Gu et al. 2020, Herrera-Estrada et al. 2017). Intersecting these events with annual cropland maps provides a direct measure of where and for how long root-zone drought coincides with cultivated land (Gaupp et al. 2020, Lesk et al. 2016).

This study examines how precipitation deficits propagate through the upper and lower root zones and how connected SSI droughts intersect with cropland across China. We first compare five gridded precipitation products with 840 gauges and select the precipitation forcing for the remaining analyses. We then evaluate SPI–SSI coupling and event timing for SPI-1, SPI-3, and SPI-6 and for upper (0–40 cm), lower (40–100 cm), and full (0–100 cm) root-zone SSI. We track connected SSI events nationally and within nine agricultural regions to quantify their duration, spatial extent, and cropland exposure. Lagged meteorological associations and retrospective XGBoost models test how much of the SSI signal can be reconstructed from meteorological information. We address three questions: How do soil depth and SPI accumulation scale affect coupling and event correspondence? How strongly is cropland exposure concentrated within the largest connected SSI events? How much of the SSI signal can be recovered from meteorological forcing without antecedent SSI?

2 Data and methods

2.1 Study domain and datasets

The analysis covers mainland China on a common 0.1° grid comprising 95,785 valid land cells ($9.446 \times 10^6 \text{ km}^2$; Fig. 1; Table 1). Grid-cell centers span 18.1–54.0°N and 73.0–135.9°E. Daily volumetric soil moisture at ten nominal depths (10–100 cm) was obtained from the 0.1° distribution of SMC11.0 for 2000–2022. SMC11.0 was distributed at both 1 km and 0.1° resolution (Li et al. 2022); the 0.1° product was used directly rather than aggregated in this study. Values of –999 were treated as missing, and packed values were converted to $\text{m}^3 \text{ m}^{-3}$ using the documented scale factor of 0.001 before monthly averaging. The 0–100 cm interval was used as a common root-zone representation. The upper profile comprised the 10–40 cm levels, the lower profile the 50–100 cm levels, and the full profile all ten levels. Monthly precipitation from Multi-Source Weighted-Ensemble Precipitation (MSWEP) (Beck et al. 2019), the Climate Prediction Center morphing technique (CMORPH) (Joyce et al. 2004), Integrated Multi-satellite Retrievals for Global Precipitation Measurement Final (IMERG Final) (Skofronick-Jackson et al. 2017), CN05.1, and CHM_PRE V2 was evaluated against observations from 840 China Meteorological Administration (CMA) V3 gauges (Han et al. 2023).

Annual cropland area was obtained from the 30 m China Annual Cropland Dataset (CACD) for 2000–2021 and conservatively aggregated to 0.1° (Tu et al. 2024); cropland-dependent values for 2022 were left missing. Regional summaries used nine agricultural-region polygons, labeled 1–9. Eight variables from the China Meteorological Forcing Dataset (CMFD) V02 series—precipitation, air temperature, downward shortwave and longwave radiation, surface pressure, relative humidity, specific humidity, and wind speed—covered 1951–2024 on a source 0.1° grid (He et al. 2020). The 0.05° center offset between the source and soil grids was resolved by conservatively harmonizing CMFD to the analysis grid. Spherical grid-cell areas were used for national weighting, and fractional polygon–cell overlaps were used for regional aggregation.

Table 1

Datasets, spatial support, product-specific gauge-evaluation periods, analysis periods, and roles. CHM_PRE V2 Gamma SPI was fitted over 1960–2024; the 2000–2022 subset was used for SPI–SSI analyses.

Dataset	Category	Spatial support	Product-specific station-evaluation period	Analysis period	Role
CMA V3.0 station precipitation	gauge reference	point (840 stations)	1951-01 to 2020-07	reference only	precipitation-product evaluation reference
MSWEP	merged precipitation	native grid	1979-02 to 2020-07	gauge comparison only	candidate product used only in gauge comparison
CMORPH	satellite precipitation	native grid	1998-01 to 2020-07	gauge comparison only	candidate product used only in gauge comparison
IMERG Final	satellite precipitation	native grid	2000-06 to 2020-07	gauge comparison only	candidate product used only in gauge comparison
CN05.1	gauge-interpolated precipitation	native grid	1961-01 to 2020-07	gauge comparison only	candidate product used only in gauge comparison
CHM_PRE V2	gauge-based precipitation	native grid harmonized to 0.1°	1960-01 to 2020-07	2000–2022 SPI–SSI subset; 1960–2024 annual SPI-area summaries	selected precipitation forcing; Gamma SPI fitted on the common 0.1° grid over 1960–2024
SMC11.0 multilayer soil moisture	soil moisture	published 1 km and 0.1°; 0.1° product used	—	2000-01 to 2022-12	volumetric water content at ten nominal depths; upper/lower/full root-zone SSI
CACD annual cropland	land cover	30 m aggregated to 0.1°	—	2000-01 to 2021-12	cropland exposure; 2022 unavailable and not gap-filled
CMFD V02	meteorological forcing	source 0.1° harmonized to the analysis grid	—	2000-01 to 2022-12 (1951–1999 baseline)	lagged meteorological associations and reconstruction features
Nine agricultural-region dataset	regional zoning	source Albers equal-area polygons intersected with the 0.1° grid	—	static	nine reporting regions using fractional grid-cell overlap

2.2 Precipitation evaluation and drought-index construction

Each precipitation product was evaluated on its native grid by extracting the nearest valid grid cell at each gauge location. Only months with paired product and gauge observations were used; out-of-domain stations and masked nearest cells remained missing. Metrics were calculated at each station and then averaged with equal station weights. They included Pearson and Spearman correlation, mean absolute error, root-mean-square error (RMSE), bias, relative bias, unbiased root-mean-square error, and the 2009 Kling–Gupta efficiency (KGE-2009) (Gupta et al. 2009), together with event agreement between product- and gauge-derived SPI. Gauge and product SPI were fitted separately over the same finite overlap record at each station. Events were contiguous runs of $SPI \leq -1$ lasting at least 2 months, and missing months broke an event. Candidate pairs had onset dates no more than 6 months apart and were matched one-to-one by Hungarian assignment using onset difference, temporal overlap, and duration difference. With H matched pairs, M missed gauge events, and F false product events, station-scale $F_1 = \frac{2H}{2H+M+F}$. Because the products covered different periods, rankings refer to their respective gauge-overlap periods.

CHM_PRE V2 monthly precipitation was conservatively aggregated to the common 0.1° grid, after which Gamma SPI-1, SPI-3, and SPI-6 were fitted over 1960–2024 with the climate-indices implementation. SPI–SSI analyses used the 2000–2022 subset aligned with SSI, whereas the annual meteorological-drought area summary used the full 1960–2024 series (Stagge et al. 2015). Both summaries therefore used the same fitted SPI record. Fits that lacked the required valid record remained missing.

Monthly soil moisture was averaged across the constituent depth levels before each profile was transformed, separately by calendar month, to SSI. Empirical plotting positions, $p = \frac{rank}{n+1}$, were converted to standard-normal quantiles, with average ranks assigned to ties (Hao and AghaKouchak 2014). The 40 cm level belonged only to the upper root zone. The full-profile SSI is a composite of the upper and lower intervals.

2.3 Lagged coupling and local event correspondence

For each valid grid cell, SPI scale, and soil profile, Pearson and Spearman correlations were calculated between $SPI(t - \tau)$ and $SSI(t)$ for lags $\tau = 0$ –6 months; positive lag indicates that SPI precedes SSI (Zhou et al. 2021). We also calculated the conditional occurrence fraction, $P[SSI(t) \leq -1 \mid SPI(t - \tau) \leq -1]$. All seven lags used the same jointly finite monthly sample, with at least 120 paired months required. Each cell was assigned its maximum correlation and the earliest lag at which that maximum occurred. National and regional summaries were calculated as area-weighted means of the grid-cell values.

Local drought events were identified independently at each grid cell as at least two consecutive months with a standardized index ≤ -1 ; missing months terminated an event. SPI events were retained only when the following 0–6-month SSI response window was complete, and SSI events only when the preceding attribution window was complete. Directional one-to-one matching used the Hungarian algorithm and allowed SSI onset from 0 to 6 months after SPI onset (Huang et al. 2017). Among admissible pairs, the assignment minimized the sum of normalized onset lag, one minus temporal intersection-over-union, and the absolute duration difference divided by the longer event. Exact ties were resolved deterministically. With H matched pairs, M unmatched SSI events, and F unmatched SPI events, $POD = \frac{H}{H+M}$, $FAR = \frac{F}{H+F}$, $CSI = \frac{H}{H+M+F}$, and $F_1 = \frac{2H}{2H+M+F}$. The matched-SPI-event fraction was $\frac{H}{H+F}$, equivalent to $1 - FAR$. Onset and duration differences were defined as SSI event minus SPI event; right-censored pairs were excluded from duration summaries.

2.4 Connected-event tracking, cropland exposure, and regional aggregation

Connected SSI droughts were defined independently of cropland. Within each month, cells with $SSI \leq -1$ were grouped using 8-neighbor connectivity. Patches in consecutive months were compared after allowing a displacement of up to one grid cell in the eight neighboring directions. Link strength was defined as the maximum physical overlap area divided by the area of the smaller patch, and links with a score below 0.1 were discarded. Conflicting candidates were resolved one-to-one by maximizing match count and then total score; a matched child continued the parent event, unmatched children started new events, and unmatched parents terminated. Events lasting at least two months were retained (Basara et al. 2019).

Event boundaries were defined from SSI alone, after which annual cropland fractions were applied to quantify exposure (Tu et al. 2024). For each event, we calculated duration, peak concurrent cropland area, unique affected cropland area, cropland area–time exposure, mean and maximum SSI intensity, cell-area-weighted cumulative severity, and intensity-weighted cropland exposure. Unique area counted each grid cell once using the maximum applicable annual cropland area encountered during the event. Cropland exposure was calculated for events occurring entirely within 2000–2021. Events extending into 2022 retained their SSI-based characteristics but had missing cropland metrics because CACD ends in 2021.

The nine agricultural-region polygons were transformed to World Geodetic System 1984 and intersected with the 0.1° grid using fractional cell overlap measured in the source Albers equal-area coordinate system. These fractions aggregated grid-cell propagation fields and annual drought-area statistics, whereas gauge stations were assigned by point-in-polygon and connected events by drought centroid. For annual area summaries, mild drought was $-1.0 < \text{index} \leq -0.5$, moderate drought was $-1.5 < \text{index} \leq -1.0$, and severe drought was $\text{index} \leq -1.5$. These classes differ from the $\text{SSI} \leq -1$ threshold used to identify connected events. Annual SPI area fractions cover 1960–2024, SSI area fractions 2000–2022, and cropland exposure 2000–2021.

2.5 Meteorological associations and retrospective reconstruction

SPI-1, SPI-3, SPI-6, and eight CMFD variables were related to each SSI profile at lags of 0–6 months. CMFD V02 spans 1951–2024, allowing precipitation, air temperature, downward shortwave and longwave radiation, surface pressure, relative humidity, specific humidity, and wind speed to be standardized by calendar month against a 1951–1999 climatology before analysis of 2000–2022. For each grid cell and variable, we retained the largest absolute Pearson correlation and its associated lag. The variables were evaluated individually, so the resulting coefficients describe pairwise lagged associations.

XGBoost was used to reconstruct held-out SSI from meteorological information. SPI-only, climate-only, and combined models included contemporaneous and 1–6-month-lagged meteorological variables, seasonal terms, and coordinates (Chen et al. 2016). Models were trained on 2000–2016, tuned on 2017–2019, and evaluated on 2020–2022. A reproducible coordinate hash withheld 20% of grid cells within each period. Months with $\text{SSI} \leq -1$ were up-sampled during model fitting. Antecedent SSI was not included as a predictor. Because SSI was standardized over the full 2000–2022 record, the experiment represents temporal hold-out reconstruction with spatial interpolation rather than a prospective forecast. SHapley Additive exPlanations (SHAP) values summarized model dependence on the supplied features (Lundberg and Lee 2017).

Lag-1 Pearson correlation was calculated separately as a descriptive measure of SSI persistence. A second XGBoost model reconstructed each region's monthly cropland-weighted fraction of $\text{SSI} \leq -1$ from cropland-weighted SPI and CMFD distribution summaries with 1-, 3-, 6-, and 12-month rolling features. Predicted fractions were multiplied by the evaluated cropland area to obtain affected cropland area (km²) for regional evaluation and aggregation (Shen 2018). This model was tuned on 2017–2019 and evaluated on 2020–2021.

2.6 Statistical summaries

National and regional summaries were area weighted. Event distributions were described using medians, interquartile ranges, empirical cumulative distributions, upper quantiles, and the share of total exposure contributed by the largest events. Cross-depth and cross-region contrasts are presented descriptively.

3 Results

3.1 Agreement of gridded precipitation and SPI with gauge observations

Agreement with gauge observations differed substantially among the five precipitation products (Fig. 2). Within each product's available overlap period, CHM_PRE V2 had the highest station-equal-weight agreement for monthly precipitation ($r = 0.997$, RMSE = 4.81 mm/month, bias = -0.30 mm/month, and KGE-2009 = 0.987) and the highest station-scale SPI-3 event F_1 (0.965). CN05.1 and IMERG Final generally ranked next, followed by MSWEP and CMORPH. These rankings should be interpreted in light of their overlap periods and gauge dependence (Table 1). CHM_PRE V2 was used as the precipitation forcing for the subsequent SPI–SSI analysis.

3.2 Depth- and scale-dependent SPI–SSI coupling and event correspondence

Across all three SPI scales, upper-zone SSI was more strongly and more rapidly coupled to precipitation deficits than lower-zone SSI (Fig. 3; Table 2). For SPI-3, the area-weighted mean of grid-cell maximum correlations was 0.346 in the upper root zone and 0.270 in the lower root zone, while the corresponding area-weighted mean optimal lags were 0.61 and 1.37 months. The full-profile response fell between the two. The same vertical ordering occurred for SPI-1 and SPI-6. Differences between the upper and lower root zones diminished with increasing lag; at a 6-month lag under SPI-6, the lower-zone correlation (0.087) exceeded the upper-zone correlation (0.070), indicating slower correlation decay at depth.

Event correspondence also varied with SPI scale (Fig. 4; Table 2). SPI-1 produced the largest matched-SPI-event fraction but had low POD, whereas SPI-3 gave the highest CSI and F_1 across the three profiles. Matched SSI events began 1.40–1.84 months after their SPI counterparts, with a shorter delay in the upper root zone. At each SPI scale, the duration difference was larger in the lower than in the upper root zone. SPI-1 differences were positive in all profiles. Under SPI-3, the upper-zone difference was close to zero (-0.07 months), whereas

the lower and full profiles had positive differences. All profiles showed negative differences for SPI-6 because the matched SPI-6 events lasted longer than the SSI events. Overall event correspondence remained modest ($POD \leq 0.33$; $FAR \geq 0.68$).

Table 2

National area-weighted SPI–SSI coupling and local event metrics. Mean maximum r is the area-weighted mean of grid-cell maximum correlations; mean optimal lag is the area-weighted mean of grid-cell optimal lags. Onset and duration differences are SSI minus SPI. Matched-SPI-event fraction = $1 - FAR$.

SPI scale	Root-zone profile	Mean maximum r	Area-weighted mean optimal lag (months)	POD	FAR	CSI	F_1	Onset difference (months)	Duration difference (months)
SPI-1	Upper (0–40 cm)	0.281	0.85	0.187	0.679	0.13	0.213	1.4	0.71
SPI-1	Lower (40–100 cm)	0.212	1.57	0.177	0.714	0.118	0.194	1.84	0.99
SPI-1	Full (0–100 cm)	0.253	1.1	0.184	0.695	0.126	0.205	1.53	0.93
SPI-3	Upper (0–40 cm)	0.346	0.61	0.321	0.716	0.178	0.285	1.42	–0.07
SPI-3	Lower (40–100 cm)	0.27	1.37	0.321	0.736	0.169	0.272	1.71	0.36
SPI-3	Full (0–100 cm)	0.319	0.83	0.325	0.725	0.176	0.282	1.47	0.24
SPI-6	Upper (0–40 cm)	0.335	0.59	0.255	0.734	0.15	0.244	1.56	–1.36
SPI-6	Lower (40–100 cm)	0.281	1.3	0.262	0.748	0.147	0.238	1.73	–0.95
SPI-6	Full (0–100 cm)	0.321	0.79	0.26	0.741	0.149	0.242	1.56	–1.06

3.3 Spatiotemporal SSI events and regional cropland exposure

Across the three profile-specific catalogs, the tracking algorithm identified 229,003 connected SSI droughts: 64,815 in the upper root zone, 97,635 in the lower root zone, and 66,553 in the full root-zone profile (Fig. 5). Lower-zone events were more numerous and had a median duration of 3 months, compared with 2 months in the upper and full profiles. Event exposure was strongly right-skewed. Median unique cropland area ranged from 1.96 to 5.69 km², but the largest 1% of events with complete cropland records accounted for 71.2–78.3% of total cropland area–time exposure; the largest 5% accounted for 86.3–91.7%.

The largest connected events persisted across seasons and covered broad agricultural areas (Table 3; Fig. 6). The largest upper-zone event lasted from December 2010 to September 2011 and intersected 706,203 km² of unique cropland. The largest full-profile event extended from February 2018 to October 2019 and intersected 627,841 km². Annual cropland exposure was highest in 2011, when upper-, lower-, and full-profile droughts affected 77.3%, 68.1%, and 70.5% of cropland within the nine-region union.

Table 3
Largest connected SSI droughts by unique affected cropland area under the one-to-one overlap tracking rule.

Event ID	Start month	End month	Duration (months)	Peak affected cropland area (km ²)	Unique affected cropland area (km ²)	Cropland area–time exposure (km ² ·month)	Cumulative severity (km ² ·month)	Intensity-weighted cropland exposure (km ² ·month)	Mean drought intensity	Profile
33381	2010-12	2011-09	10	552,445	706,203	830,479	3,589,281	1,308,920	1.555	upper root zone (0–40 cm)
143741	2018-07	2019-10	16	403,663	558,877	711,571	1,861,170	1,056,482	1.48	lower root zone (40–100 cm)
215302	2018-02	2019-10	21	429,621	627,841	1,196,491	3,149,334	1,771,106	1.473	full root zone (0–100 cm)

Regional results showed that hydrologic coupling and cultivated-land distribution followed different spatial patterns (Fig. 7; Table 4). Upper-zone SPI-3 coupling was strongest and fastest in the eastern and southern regions and weakest and most delayed in the arid and high-elevation west. Regions with stronger coupling generally had shorter area-weighted mean optimal lags ($r = -0.90$ across the nine regional means). Raw event counts were largest in the geographically extensive western regions, whereas cropland area–time exposure was concentrated in the Middle-lower Yangtze Plain, Northeast China Plain, and Huang-Huai-Hai Plain. Event frequency and cropland exposure therefore produced different regional rankings. The annual records peaked in different years: SPI-3/6 drought area in 1965, SSI drought area around 2009, and cropland exposure in 2011 (Fig. 8).

Table 4
Descriptive regional SPI-3 upper-zone coupling, raw connected-event counts, and cropland exposure across the nine agricultural regions.

Region	Agricultural region	Mean maximum r	Area-weighted mean optimal lag (months)	POD	F_1	Number of connected upper-zone events	Median duration (months)	Cropland area–time (10 ⁶ km ² ·month)
1	Northeast China Plain	0.44	0.22	0.472	0.402	4589	2	10.67
2	Northern arid and semiarid region	0.261	1.03	0.273	0.242	25452	3	8.01
3	Huang-Huai-Hai Plain	0.484	0.02	0.396	0.329	2402	2	9.53
4	Loess Plateau	0.465	0.01	0.375	0.32	1459	2	3.8
5	Qinghai Tibet Plateau	0.236	1.09	0.219	0.223	19055	3	0.34
6	Sichuan Basin and surrounding regions	0.34	0.18	0.285	0.254	4385	2	3.48
7	Middle-lower Yangtze Plain	0.522	0	0.395	0.355	2820	2	12.54
8	Yunnan-Guizhou Plateau	0.439	0	0.406	0.352	3388	2	3.88
9	Southern China	0.546	0	0.478	0.398	1265	2	1.38

3.4 Meteorological associations and SSI reconstruction

The lagged meteorological analysis reproduced the depth ordering found for SPI–SSI coupling (Fig. 9; Table 5). The strongest positive associations with upper-zone SSI occurred for SPI-3, SPI-6, relative humidity, and CMFD precipitation. Downward shortwave radiation and air temperature had the largest negative correlations. For the leading variables, correlation magnitudes were generally larger and selected lags shorter in the upper than in the lower root zone.

Table 5
National area-weighted pairwise lagged associations between eleven meteorological predictors and upper-root-zone SSI (2000–2022).

Driver	Max r	Signed r at optimal lag	Optimal lag (months)
SPI-3	0.364	0.326	0.59
SPI-6	0.358	0.314	0.61
Relative humidity	0.323	0.27	0.74
SPI-1	0.293	0.264	0.79
Precipitation	0.273	0.246	0.91
Specific humidity	0.245	0.199	0.97
Downward shortwave	0.217	−0.186	1.01
Air temperature	0.181	−0.104	1.47
Downward longwave	0.174	0.137	1.54
Wind speed	0.169	−0.005	2.62
Surface pressure	0.136	0.005	2.55

SSI also showed substantial month-to-month persistence: lag-1 correlations ranged from 0.657 in the upper root zone to 0.697 for the full profile (Fig. 10; Table 6). The combined meteorological XGBoost models explained 5.1–15.9% of held-out SSI variance, with the lowest R^2 in the lower root zone. Longitude, contemporaneous SPI-6, relative humidity, and recent precipitation were the leading SHAP features of the upper-zone model. Longitude ranked among the leading features, showing that spatial climatology contributed substantially to the reconstruction; event-to-event variability remained difficult to reproduce. After conversion of the predicted regional fractions to area, the regional model reconstructed monthly affected cropland area for the nine-region aggregate with correlations of 0.469–0.777 and aggregate bias within $\pm 9\%$; individual-region bias reached +83.7%.

Table 6

Retrospective SSI reconstruction, descriptive lag-1 SSI correlation, and regional drought-area reconstruction. SSI-model statistics use the 2020–2022 temporal hold-out. The regional model was evaluated in 2020–2021 after predicted drought fractions were converted to affected cropland area; Lag-1 SSI r denotes observed persistence, whereas the remaining statistics quantify reconstruction performance.

Root-zone profile	SPI-only r	Climate-only r	Combined r	Combined R^2	Lag-1 SSI r	Regional area-model monthly r	Regional drought-area bias (%)
Upper (0–40 cm)	0.451	0.48	0.494	0.159	0.657	0.469	5.8
Lower (40–100 cm)	0.384	0.387	0.41	0.051	0.689	0.777	−8.8
Full (0–100 cm)	0.423	0.447	0.468	0.122	0.697	0.522	−4.2

4 Discussion

4.1 Depth-dependent propagation and soil-moisture memory

The upper and lower root zones filtered precipitation deficits on different time scales. Moisture in the upper 0–40 cm is replenished directly by infiltration and depleted rapidly by evaporation and root uptake, which is consistent with its stronger correlation and shorter lag. Deeper storage changes more slowly. Its delayed response and slower correlation decay indicate that anomalies are retained after the upper profile has begun to adjust. Similar depth dependence has been reported for soil-moisture memory across a range of climates and land-surface conditions. The intermediate behavior of full-profile SSI follows from its integration of both depth intervals (Seneviratne et al. 2006).

The selected SPI accumulation scale changed the apparent agreement with SSI(Afshar et al. 2022). SPI-1 was sensitive to short precipitation deficits, many of which were not followed by an SSI event. SPI-6 described longer anomalies but produced lower event F_1 and, in matched cases, often lasted longer than the SSI event itself. SPI-3 gave the strongest overall event agreement, although its POD and FAR still indicated substantial mismatch. Several processes can produce this mismatch: antecedent storage and atmospheric demand alter the effect of a given precipitation deficit, while snow, soil properties, vegetation, groundwater, and water management introduce additional spatial variability(Wu et al. 2022). SPI represents atmospheric water-supply anomalies, whereas SSI reflects the soil state after these processes act.

4.2 Connected drought events and the geography of cropland exposure

Grid-cell frequency maps did not show how strongly exposure was concentrated: a small number of spatially extensive, persistent droughts dominated national cropland exposure(Tigchelaar et al. 2018). Because the distribution was strongly right-skewed, the mean alone did not describe a typical event; both the median and upper tail were informative. Frequent local droughts and rare large connected events have different implications for monitoring—the former describe recurring field-scale deficits, while the latter dominate aggregate cropland area–time exposure.

Regional propagation and regional exposure did not follow the same spatial pattern. Stronger, faster SPI–SSI coupling occurred in the eastern and southern regions, whereas the largest raw event counts occurred in the much larger western regions. Cropland weighting shifted the exposure maximum to the eastern agricultural plains, especially the Middle-lower Yangtze, Northeast China, and Huang-Huai-Hai plains(Leng and Hall 2019). The 2011 event illustrates this separation between hydrologic footprint and land use: a spatially connected upper-zone drought coincided with the largest annual cropland exposure in the available record. Event counts and exposure concentration depend on the rule used to connect monthly drought patches. The one-to-one overlap rule provides a reproducible event catalog, but a split–merge tracker would allocate some patch divisions and mergers differently.

4.3 Meteorological information, soil-state persistence, and model limitations

The meteorological associations partly explain the limited event agreement between SPI and SSI. Positive associations with precipitation, SPI, and humidity and negative associations with shortwave radiation and temperature are consistent with the joint effects of water supply and atmospheric demand(Qing et al. 2022). These variables covary, however, and the reported coefficients are maxima selected from seven lags; the coefficients therefore describe association patterns and should not be interpreted as independent effects.

The reconstruction models captured broad spatial and seasonal structure better than local SSI variability, as indicated by the modest held-out R^2 and the high SHAP rank of longitude. SSI, however, remained strongly correlated from one month to the next, especially in the lower and full profiles. Antecedent soil state therefore carries information that is not represented directly by the meteorological feature set. The full-record standardization and interspersed spatial hold-out mean that the XGBoost results describe retrospective reconstruction within China rather than prospective forecasting(Fang et al. 2017).

The product comparison is conditional on each product's overlap period and gauge dependence, and the 23-year soil-moisture record limits inference about multidecadal change. Event totals and concentration are also sensitive to the SSI threshold, minimum duration, and patch-tracking convention. The present cropland analysis resolves the spatial coincidence of root-zone drought and cultivated land; crop type, phenology, irrigation, soil properties, and yield observations would be needed to relate that exposure to crop-specific responses(Leng and Hall 2019, Zhao et al. 2017).

5 Conclusions

Across China, upper-zone SSI responded earlier and more strongly to precipitation deficits than lower-zone SSI. Correlations decayed more slowly with depth, and lower-zone events were generally more persistent. Among the tested accumulation scales, SPI-3 produced the highest CSI and F_1 , yet event correspondence remained modest ($POD \leq 0.33$; $FAR \geq 0.68$). Precipitation indices alone therefore missed part of the timing and occurrence of root-zone soil-moisture drought.

Cropland exposure was highly concentrated among connected events. Across the three profile-specific catalogs, the largest 1% of events accounted for 71–78% of cropland area–time exposure. Coupling was generally weaker and slower in western arid and high-elevation regions, whereas cropland exposure was concentrated in the eastern agricultural plains. Meteorological features reconstructed only a limited share of held-out SSI variance, while SSI showed strong month-to-month persistence. Drought monitoring should therefore combine precipitation with depth-resolved soil moisture and track connected drought footprints to represent both the timing and spatial concentration of agricultural soil-water deficits.

Declarations

CRedit authorship contribution statement

Jun Zhang: Conceptualization; Methodology; Software; Validation; Formal analysis; Data curation; Writing – original draft; Writing – review & editing; Visualization; Supervision; Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-K202303801 and KJZD-K202503802).

Data Availability

The gridded precipitation, SMCI1.0 soil-moisture, CACD cropland, and CMFD datasets are available from the providers and persistent identifiers cited in Section 2.1 and Table 1. The CMA V3.0 daily surface-climate observations (SURF_CLI_CHN_MUL_DAY V3.0) were obtained through the China Meteorological Data Service Centre, where access is available to registered users for education and research.

References

1. Afshar MH, Bulut B, Duzenli E, Amjad M, Yilmaz MT (2022) Global spatiotemporal consistency between meteorological and soil moisture drought indices. *Agric For Meteorol* 316
2. Apurv T, Sivapalan M, Cai XM (2017) Understanding the Role of Climate Characteristics in Drought Propagation. *Water Resour Res* 53(11):9304–9329
3. Basara JB, Christian JI, Wakefield RA, Otkin JA, Hunt EH, Brown DP (2019) The evolution, propagation, and spread of flash drought in the Central United States during 2012. *Environ Res Lett* 14(8)
4. Beck HE, Vergopolan N, Pan M, Levizzani V, van Dijk A, Weedon GP, Brocca L, Pappenberger F, Huffman GJ, Wood EF (2017) Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol Earth Syst Sci* 21(12):6201–6217
5. Beck HE, Wood EF, Pan M, Fisher CK, Miralles DG, van Dijk A, McVicar TR, Adler RF (2019) MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. *Bull Am Meteorol Soc* 100(3):473–502
6. Chen TQ, Guestrin C, Assoc Comp M (2016) XGBoost: A Scalable Tree Boosting System, pp. 785–794, San Francisco, CA
7. Dai M, Huang SZ, Huang Q, Zheng XD, Su XL, Leng GY, Li ZY, Guo Y, Fang W, Liu YJ (2022) Propagation characteristics and mechanism from meteorological to agricultural drought in various seasons. *J Hydrol* 610
8. Fang K, Shen CP, Kifer D, Yang X (2017) Prolongation of SMAP to Spatiotemporally Seamless Coverage of Continental US Using a Deep Learning Neural Network. *Geophys Res Lett* 44(21):11030–11039
9. Gaupp F, Hall J, Hochrainer-Stigler S, Dadson S (2020) Changing risks of simultaneous global breadbasket failure. *Nat Clim Change* 10(1):54–
10. Gu L, Chen J, Yin JB, Xu CY, Chen H (2020) Drought hazard transferability from meteorological to hydrological propagation. *J Hydrol* 585
11. Gupta HV, Kling H, Yilmaz KK, Martinez GF (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J Hydrol* 377(1–2):80–91
12. Han JY, Miao CY, Gou JJ, Zheng HY, Zhang Q, Guo XY (2023) A new daily gridded precipitation dataset for the Chinese mainland based on gauge observations. *Earth Syst Sci Data* 15(7):3147–3161
13. Hao ZC, AghaKouchak A (2014) A Nonparametric Multivariate Multi-Index Drought Monitoring Framework. *J Hydrometeorol* 15(1):89–101

14. He J, Yang K, Tang WJ, Lu H, Qin J, Chen YY, Li X (2020) The first high-resolution meteorological forcing dataset for land process studies over China. *Sci Data* 7(1)
15. Herrera-Estrada JE, Satoh Y, Sheffield J (2017) Spatiotemporal dynamics of global drought. *Geophys Res Lett* 44(5):2254–2263
16. Huang SZ, Li P, Huang Q, Leng GY, Hou BB, Ma L (2017) The propagation from meteorological to hydrological drought and its potential influence factors. *J Hydrol* 547:184–195
17. Joyce RJ, Janowiak JE, Arkin PA, Xie PP (2004) CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J Hydrometeorol* 5(3):487–503
18. Koster RD, Dirmeyer PA, Guo ZC, Bonan G, Chan E, Cox P, Gordon CT, Kanae S, Kowalczyk E, Lawrence D, Liu P, Lu CH, Malyshev S, McAvaney B, Mitchell K, Mocko D, Oki T, Oleson K, Pitman A, Sud YC, Taylor CM, Verseghy D, Vasic R, Xue YK, Yamada T, Team G (2004) Regions of strong coupling between soil moisture and precipitation. *Science* 305(5687):1138–1140
19. Leng GY, Hall J (2019) Crop yield sensitivity of global major agricultural countries to droughts and the projected changes in the future. *Sci Total Environ* 654:811–821
20. Lesk C, Rowhani P, Ramankutty N (2016) Influence of extreme weather disasters on global crop production. *Nature* 529(7584):84–
21. Li QL, Shi GS, Shanguan W, Nourani V, Li JD, Li L, Huang FN, Zhang Y, Wang CY, Wang DG, Qiu JX, Lu XJ, Dai YJ (2022) A 1 km daily soil moisture dataset over China using in situ measurement and machine learning. *Earth Syst Sci Data* 14(12):5267–5286
22. Lundberg SM, Lee SI (2017) *A Unified Approach to Interpreting Model Predictions*, Long Beach, CA
23. Qing YM, Wang S, Ancell BC, Yang ZL (2022) Accelerating flash droughts induced by the joint influence of soil moisture depletion and atmospheric aridity. *Nat Commun* 13(1)
24. Rahmati M, Amelung W, Brogi C, Dari J, Flammini A, Bogen H, Brocca L, Chen H, Groh J, Koster RD, McColl KA, Montzka C, Moradi S, Rahi A, S, F.S. and, Vereecken H (2024) Soil Moisture Memory: State-Of-The-Art and the Way Forward. *Rev Geophys* 62(2)
25. Seneviratne SI, Corti T, Davin EL, Hirschi M, Jaeger EB, Lehner I, Orlowsky B, Teuling AJ (2010) Investigating soil moisture-climate interactions in a changing climate: A review. *Earth Sci Rev* 99(3–4):125–161
26. Seneviratne SI, Koster RD, Guo ZC, Dirmeyer PA, Kowalczyk E, Lawrence D, Liu P, Lu CH, Mocko D, Oleson KW, Verseghy D (2006) Soil moisture memory in AGCM simulations: Analysis of global land-atmosphere coupling experiment (GLACE) data. *J Hydrometeorol* 7(5):1090–1112
27. Sheffield J, Wood EF (2008) Global trends and variability in soil moisture and drought characteristics, 1950–2000, from observation-driven Simulations of the terrestrial hydrologic cycle. *J Clim* 21(3):432–458
28. Shen CP (2018) A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists. *Water Resour Res* 54(11):8558–8593
29. Skofronick-Jackson G, Petersen WA, Berg W, Kidd C, Stocker EF, Kirschbaum DB, Kakar R, Braun SA, Huffman GJ, Iguchi T, Kirstetter PE, Kummerow C, Meneghini R, Oki R, Olson WS, Takayabu YN, Furukawa K, Wilheit T (2017) THE GLOBAL PRECIPITATION MEASUREMENT (GPM) MISSION FOR SCIENCE AND SOCIETY. *Bull Am Meteorol Soc* 98(8):1679–1695
30. Stagge JH, Tallaksen LM, Gudmundsson L, Van Loon AF, Stahl K (2015) Candidate Distributions for Climatological Drought Indices (SPI and SPEI). *Int J Climatol* 35(13):4027–4040
31. Sun QH, Miao CY, Duan QY, Ashouri H, Sorooshian S, Hsu KL (2018) A Review of Global Precipitation Data Sets: Data Sources, Estimation, and Intercomparisons. *Rev Geophys* 56(1):79–107
32. Tigchelaar M, Battisti DS, Naylor RL, Ray DK (2018) Future warming increases probability of globally synchronized maize production shocks. *Proc Natl Acad Sci USA* 115(26):6644–6649
33. Tu Y, Wu SB, Chen B, Weng QH, Bai YQ, Yang J, Yu L, Xu B (2024) A 30 m annual cropland dataset of China from 1986 to 2021. *Earth Syst Sci Data* 16(5):2297–2316
34. Wu GY, Chen J, Shi XY, Kim JS, Xia J, Zhang LP (2022) Impacts of Global Climate Warming on Meteorological and Hydrological Droughts and Their Propagations. *Earths Future* 10(3)
35. Zhao C, Liu B, Piao SL, Wang XH, Lobell DB, Huang Y, Huang MT, Yao YT, Bassu S, Ciais P, Durand JL, Elliott J, Ewert F, Janssens IA, Li T, Lin E, Liu Q, Martre P, Müller C, Peng SS, Peñuelas J, Ruane AC, Wallach D, Wang T, Wu DH, Liu Z, Zhu Y, Zhu ZC, Asseng S (2017) Temperature increase reduces global yields of major crops in four independent estimates. *Proc Natl Acad Sci USA* 114(35):9326–9331
36. Zhou ZQ, Shi HY, Fu Q, Ding YB, Li TX, Liu SN (2021) Investigating the Propagation From Meteorological to Hydrological Drought by Introducing the Nonlinear Dependence With Directed Information Transfer Index. *Water Resour Res* 57(8)
37. Zhu Y, Liu Y, Wang W, Singh VP, Ren LL (2021) A global perspective on the probability of propagation of drought: From meteorological to soil moisture. *J Hydrol* 603

Figures

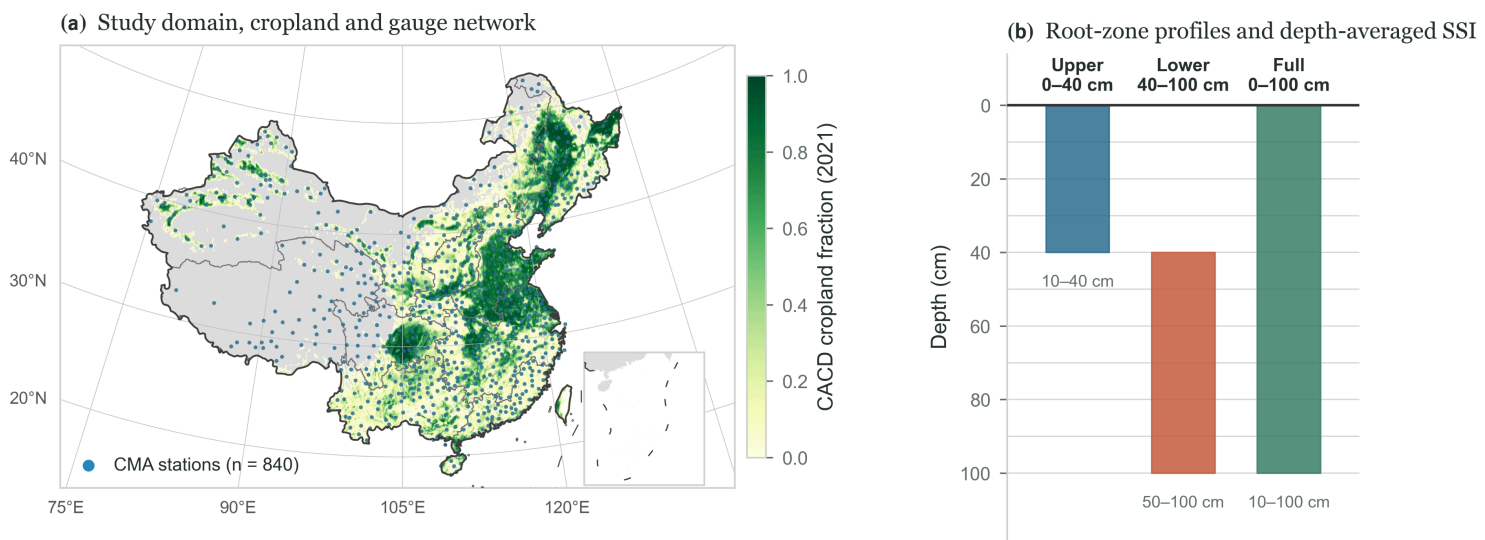


Figure 1

Study domain and root-zone profile definitions. (a) Analysis domain (0.1° , 95,785 valid cells) with 2021 CACD cropland fraction and 840 CMA V3 gauge stations. (b) Upper (0–40 cm), lower (40–100 cm), and full (0–100 cm) root-zone profiles. Soil moisture was available at nominal 10 cm intervals; the upper profile used the 10–40 cm levels, the lower profile the 50–100 cm levels, and the full profile all ten levels.

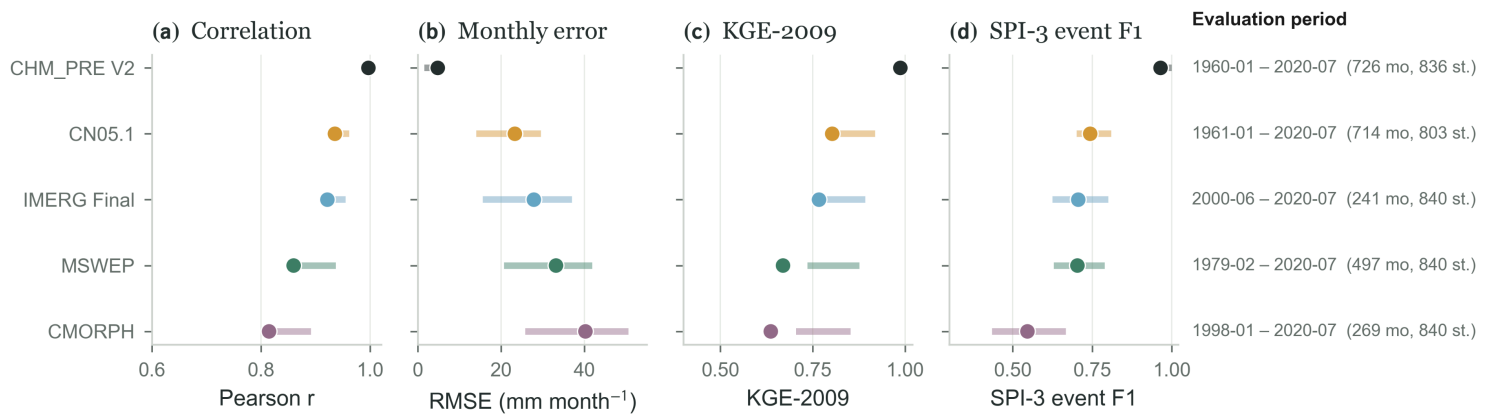


Figure 2

Gauge-based comparison of five gridded precipitation products at 840 CMA V3 stations. Station-equal-weight mean (dot) and 25th–75th percentile interval (bar) of (a) Pearson r , (b) RMSE, (c) KGE-2009, and (d) station SPI-3 drought-event F_1 . Each product's evaluation period and month/station count are annotated. Products were evaluated over their available gauge-overlap periods; CHM_PRE V2 incorporates gauge information. Error direction is product minus gauge.

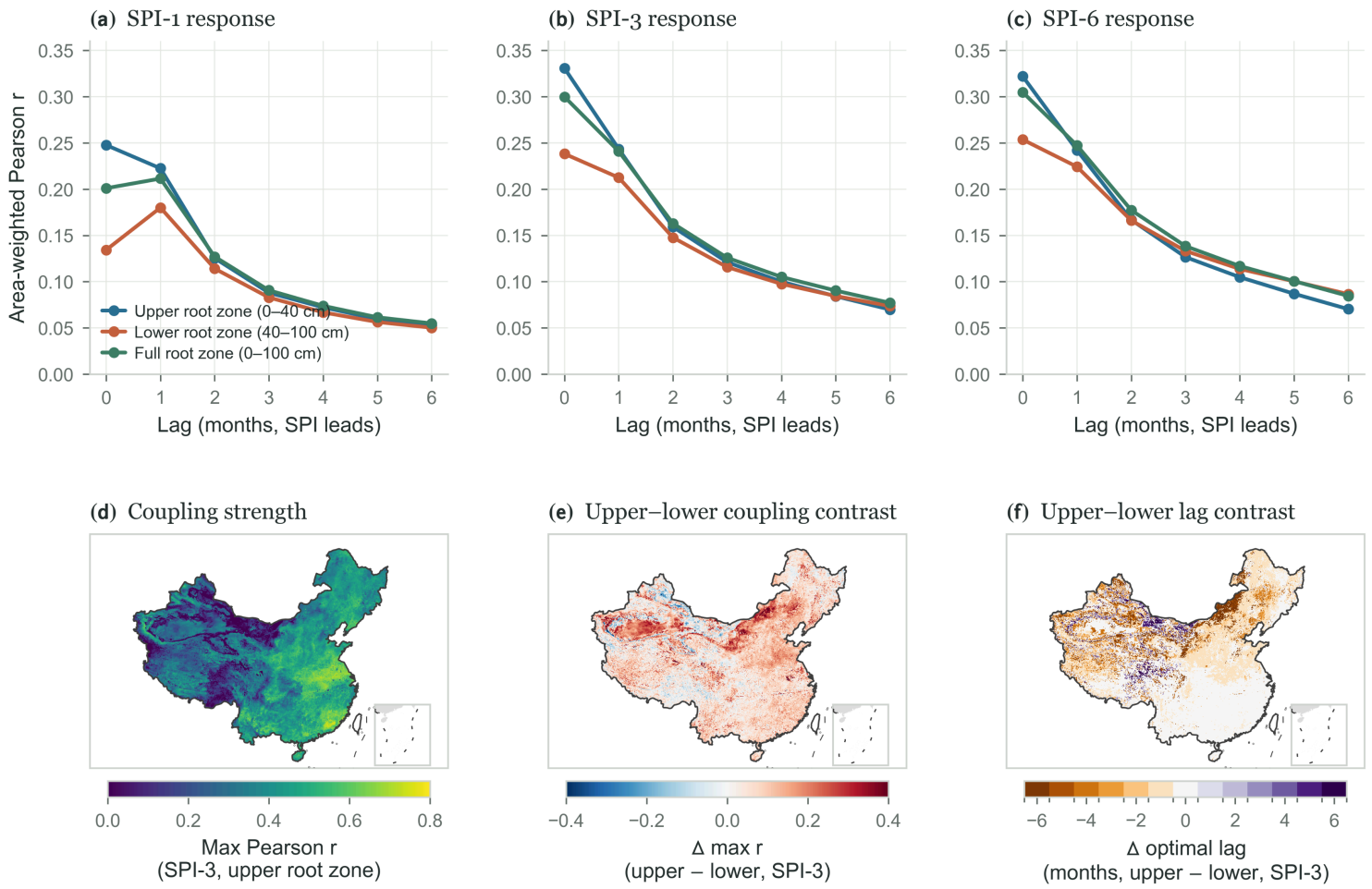


Figure 3

Accumulation-scale and depth dependence of SPI–SSI coupling. (a–c) National area-weighted lagged Pearson correlation for SPI-1/3/6 and the three root-zone profiles (positive lag indicates that SPI precedes SSI). (d) Grid-cell maximum SPI-3 correlation for the upper root zone. (e) Upper-minus-lower grid-cell maximum correlation for SPI-3. (f) Upper-minus-lower grid-cell optimal lag for SPI-3; negative values indicate an earlier upper-zone response. Values reported nationally or regionally are area-weighted means of grid-cell maxima and the area-weighted mean of grid-cell optimal lags.

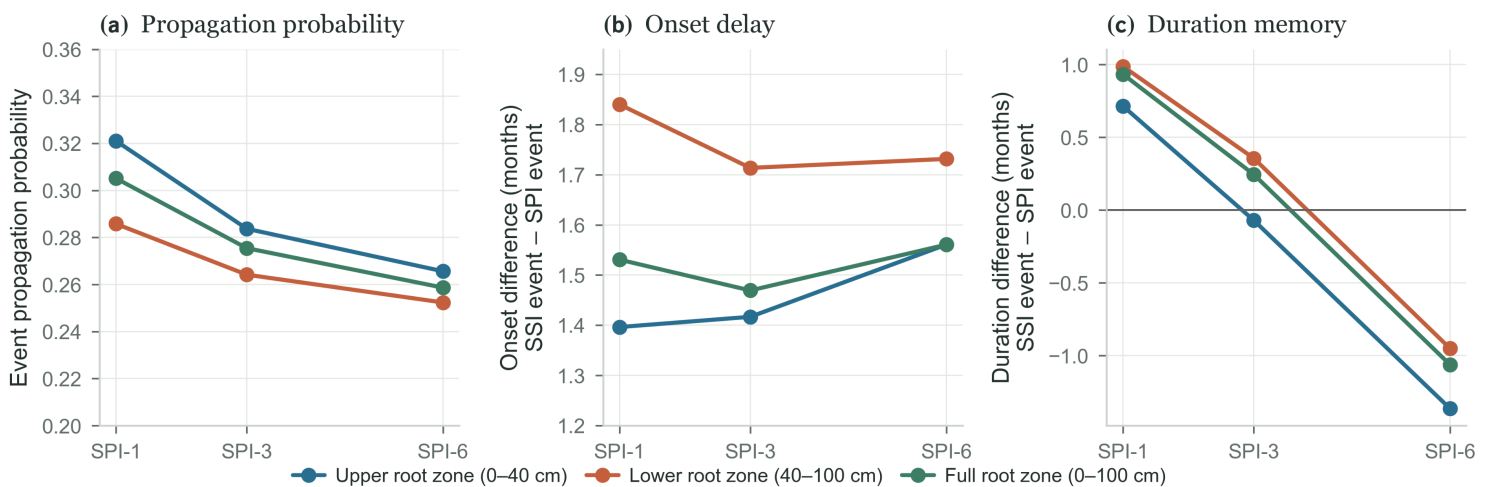


Figure 4

Event-scale SPI–SSI correspondence and timing (national area-weighted; directional one-to-one matching within a 0–6-month response window). (a) Matched-SPI-event fraction by SPI scale and root-zone profile. (b) Onset difference (SSI minus SPI). (c) Duration difference (SSI minus SPI). Lower-zone duration differences were consistently larger than upper-zone differences, although all profile differences were negative for SPI-6 because the matched SPI-6 events lasted longer. The matched-SPI-event fraction equals $1 - \text{FAR}$; overall correspondence remained modest ($\text{POD} \leq 0.33$; $\text{FAR} \geq 0.68$).

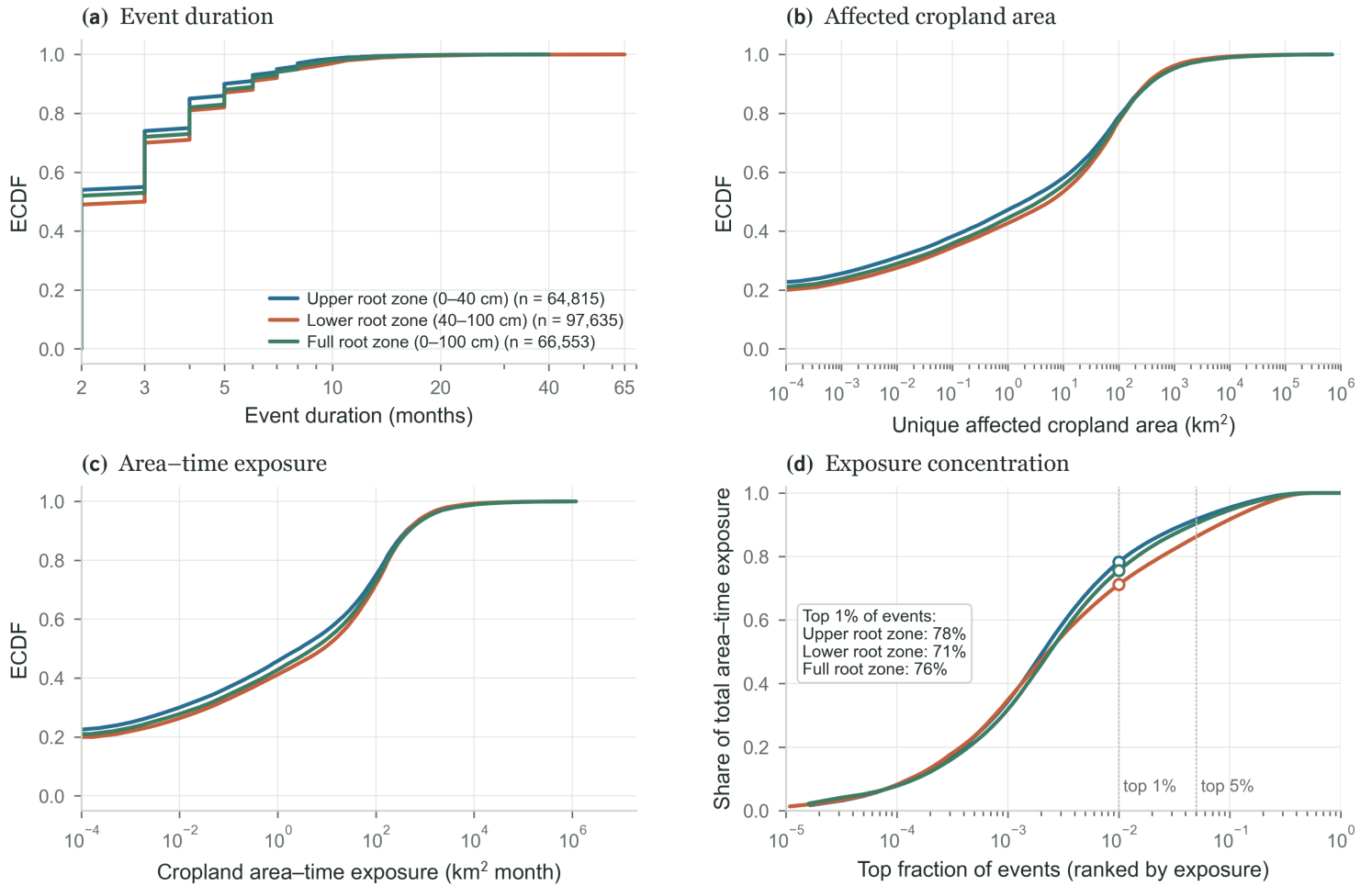


Figure 5

Distributions of connected SSI events and cropland exposure. Empirical cumulative distributions (logarithmic x axis) of (a) event duration, (b) unique affected cropland area, and (c) cropland area–time exposure by profile. (d) Cumulative share of total area–time exposure against the largest-event fraction under the one-to-one tracking rule. Panels (b–d) use events with complete cropland records; the largest 1% accounted for 71.2–78.3% of exposure.

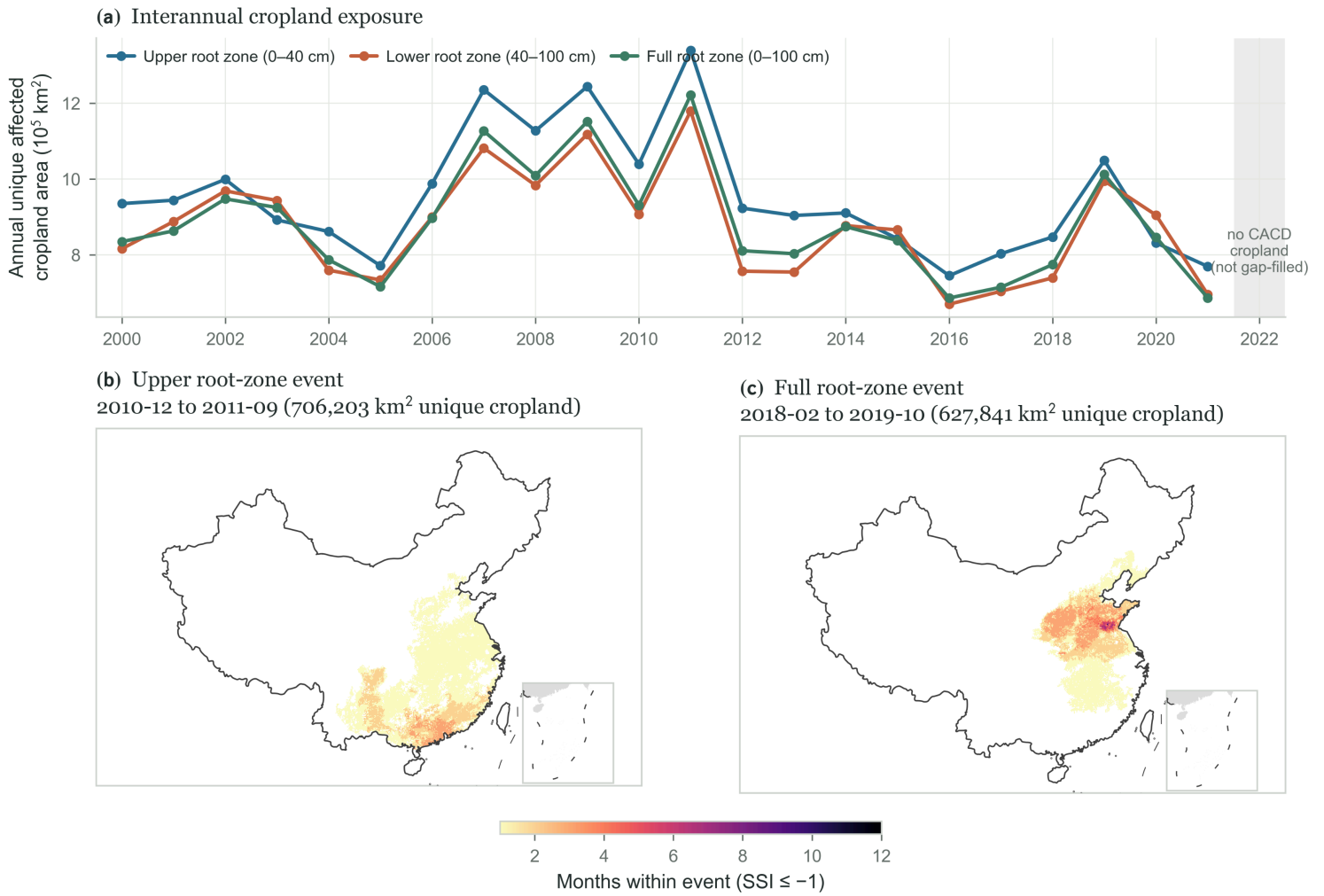
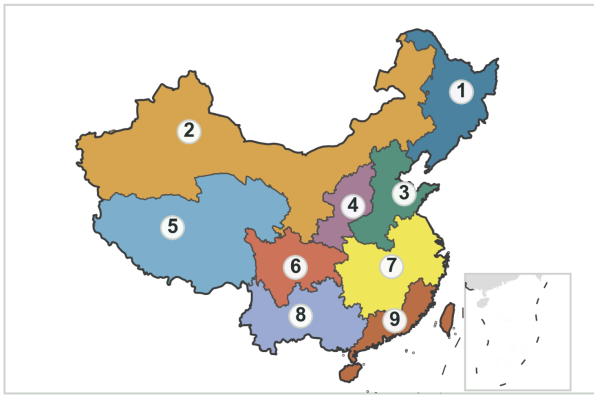


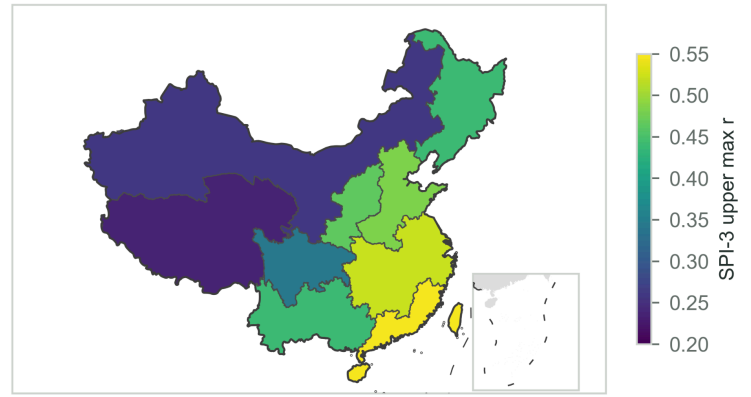
Figure 6

Interannual cropland exposure to connected soil-moisture droughts and the two largest events. (a) Annual unique affected cropland area (10^5 km^2) by profile; 2022 is blank because CACD ends in 2021. (b) Monthly footprint of the largest upper-zone event (December 2010–September 2011). (c) Monthly footprint of the largest full-profile event (February 2018–October 2019). The cases are the events with the largest unique cropland area in their respective catalogs.

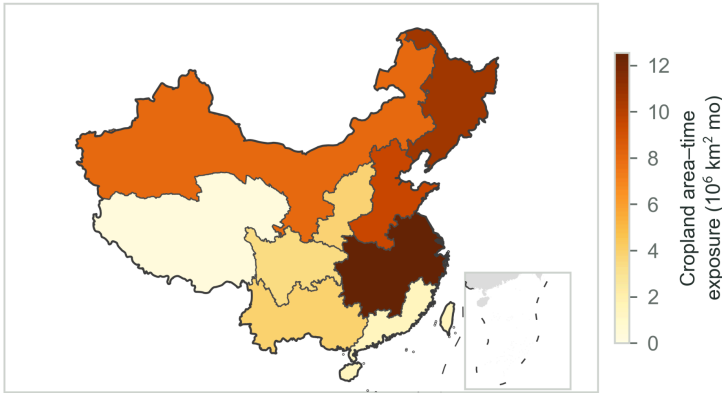
(a) Agricultural regions



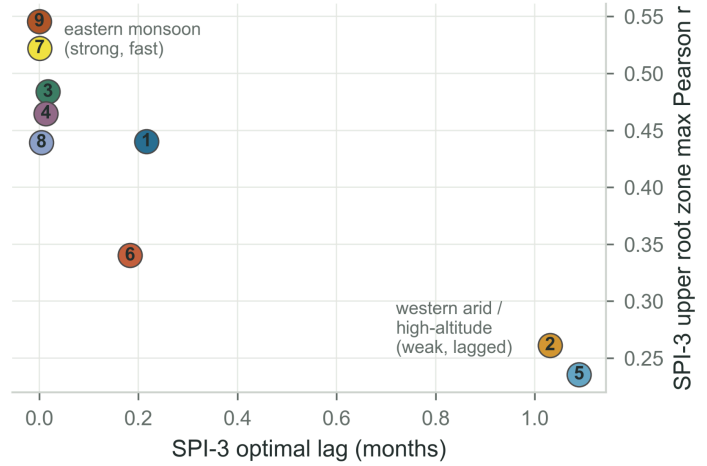
(b) Regional coupling strength



(c) Cropland exposure (upper root zone)



(d) Coupling-lag gradient



- | | | |
|-------------------------------------|---|------------------------------|
| 1 Northeast China Plain | 4 Loess Plateau | 7 Middle-lower Yangtze Plain |
| 2 Northern arid and semiarid region | 5 Qinghai Tibet Plateau | 8 Yunnan-Guizhou Plateau |
| 3 Huang-Huai-Hai Plain | 6 Sichuan Basin and surrounding regions | 9 Southern China |

Figure 7

Propagation and cropland exposure across nine agricultural regions. (a) Region map. (b) Area-weighted mean of grid-cell maximum SPI-3 correlations for the upper root zone. (c) Upper-zone cropland area-time exposure. (d) Regional relationship between SPI-SSI coupling strength and area-weighted mean optimal lag (descriptive $r = -0.90$; $n = 9$). Numbers identify the regions listed beneath panel (a).

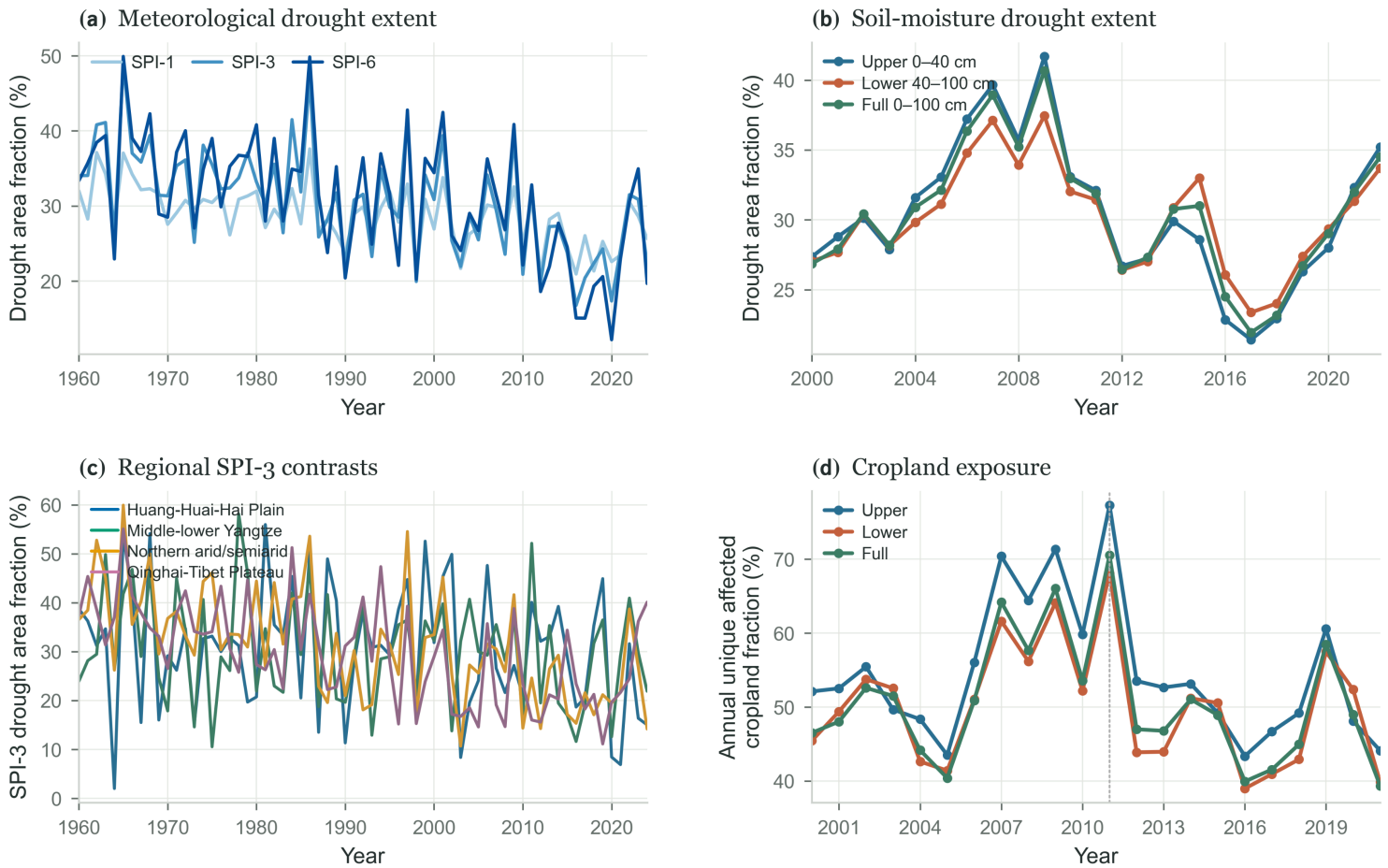


Figure 8

Annual meteorological drought, soil-moisture drought, and cropland-exposure records. (a) Annual SPI-1/3/6 drought area fraction for the nine-region union, 1960–2024. (b) Annual SSI drought area fraction by profile, 2000–2022. (c) SPI-3 drought area fraction for four contrasting regions, 1960–2024. (d) Annual unique affected cropland fraction by profile, 2000–2021; the dotted line marks the 2011 maximum.

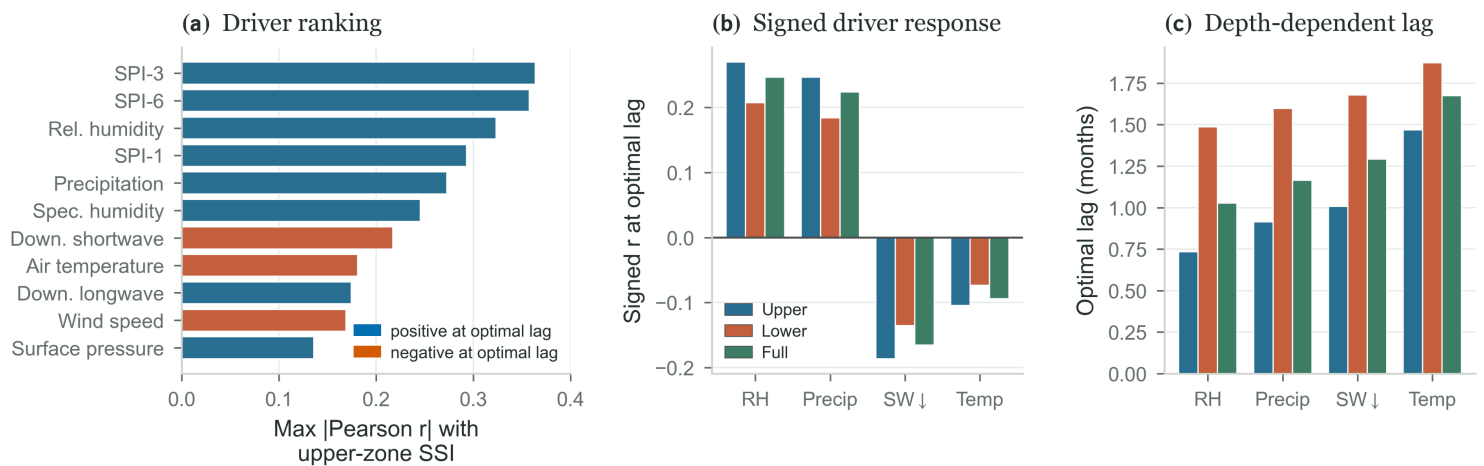


Figure 9

Pairwise lagged meteorological associations with upper-root-zone SSI (national area-weighted, 2000–2022). (a) Maximum absolute correlation of eleven predictors with SSI, coloured by the sign at the selected lag. (b) Signed correlation of the four leading predictors by root-zone profile. (c) Selected lag of those predictors by profile. Variables covary, and the maximum was selected from seven candidate lags; the coefficients do not partition independent effects.

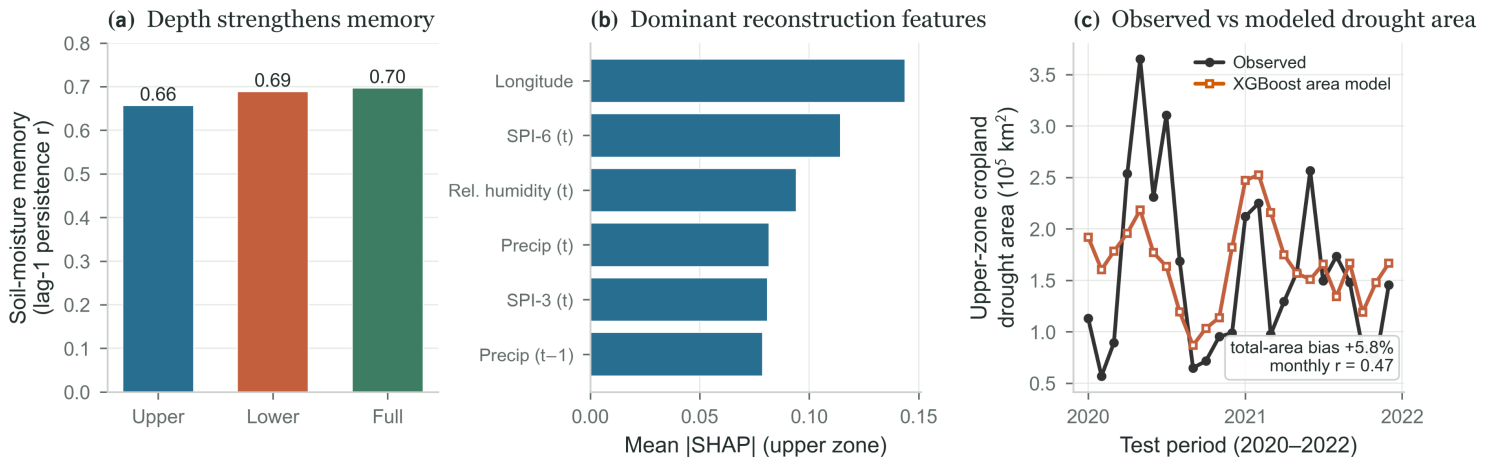


Figure 10

SSI persistence and retrospective cropland-drought reconstruction. (a) Observed lag-1 SSI correlation (Pearson r) by profile. Lag-1 SSI correlation was lowest in the upper profile and higher in the lower and full profiles. (b) Leading SHAP features of the combined upper-zone reconstruction. (c) Observed and reconstructed national monthly cropland-drought area for the upper root zone in the 2020–2021 hold-out. Aggregate bias was below 10%, while individual-region bias reached +83.7%.