Supplementary Information

Deep soil moisture reveals hidden water stress in African rainfed maize systems

Study domain and data coverage

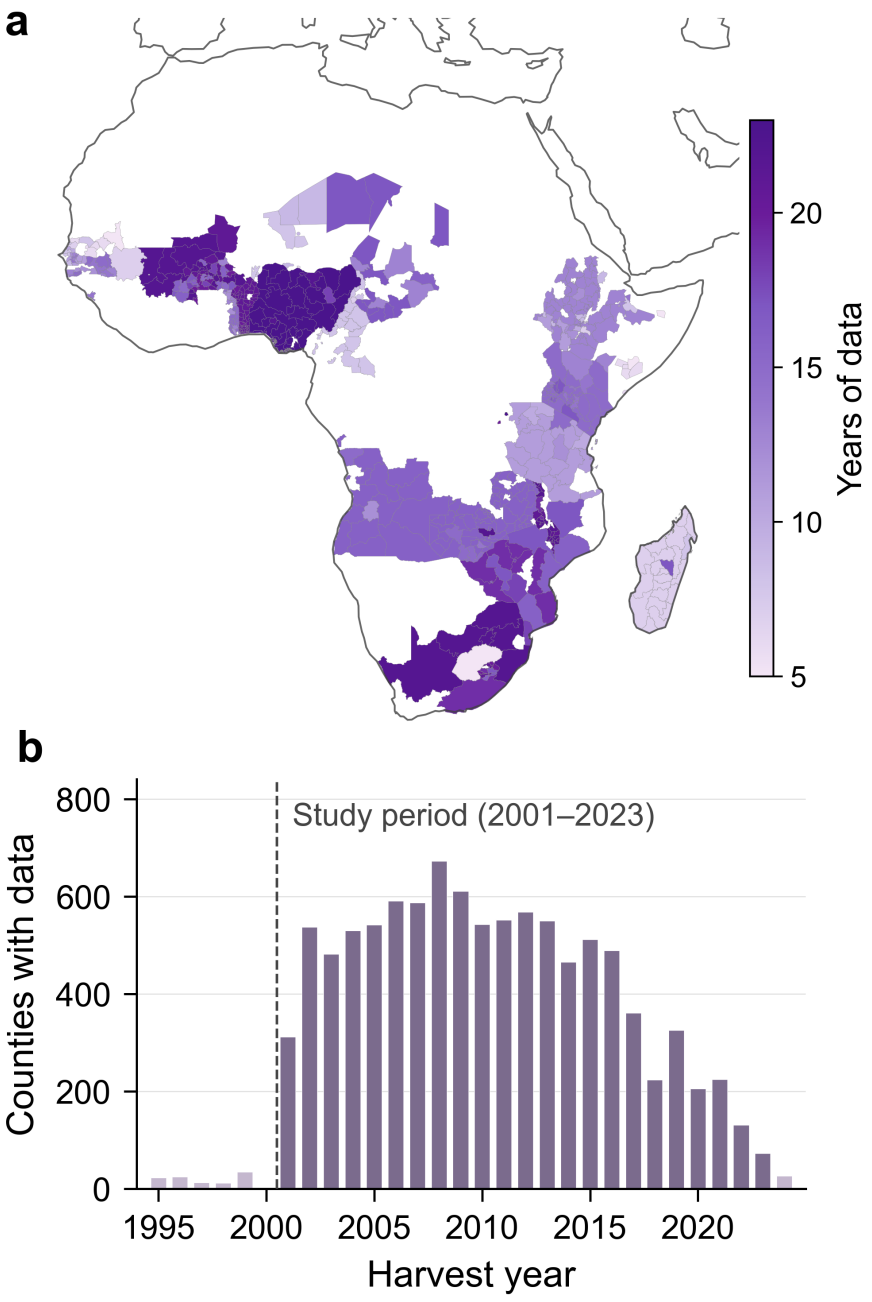

Figure S1: **Spatial and temporal coverage of subnational maize yield data.** (a) Number of years with valid yield observations per county after quality control filtering (2001–2023). The final dataset comprises 606 counties across 22 countries. (b) Total number of counties with available yield data per harvest year. The dashed line marks the start of the study period; data prior to 2001 were excluded owing to limited spatial coverage.

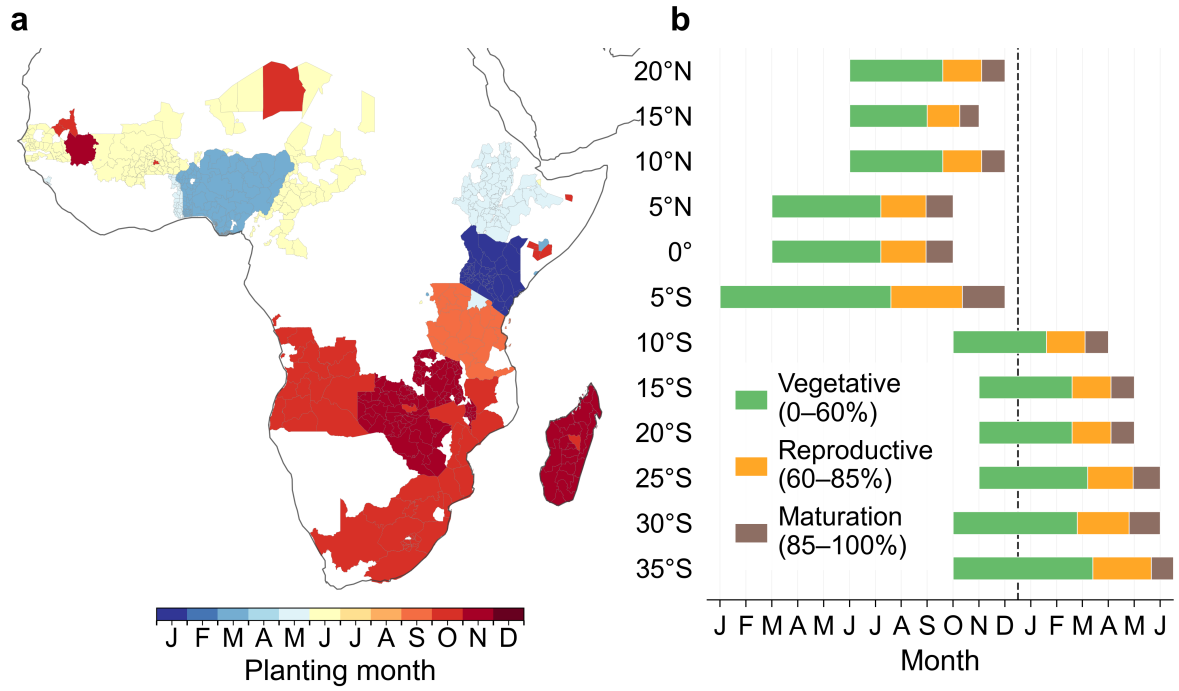

Figure S2: **Growing season timing and phenological phase partitioning across African maize regions.** (a) Median planting month per county from the HarvestStat Africa dataset. (b) Median growing season windows aggregated by 5° latitude bands. Each bar spans from planting to harvest, with shading indicating the vegetative (0-60% of season length), reproductive (60-85%), and maturation (85-100%) phases used in the phenological decomposition (Fig. 3c). The dashed line marks the calendar year boundary; southern-hemisphere growing seasons typically span two calendar years.

Robustness to alternative SM datasets

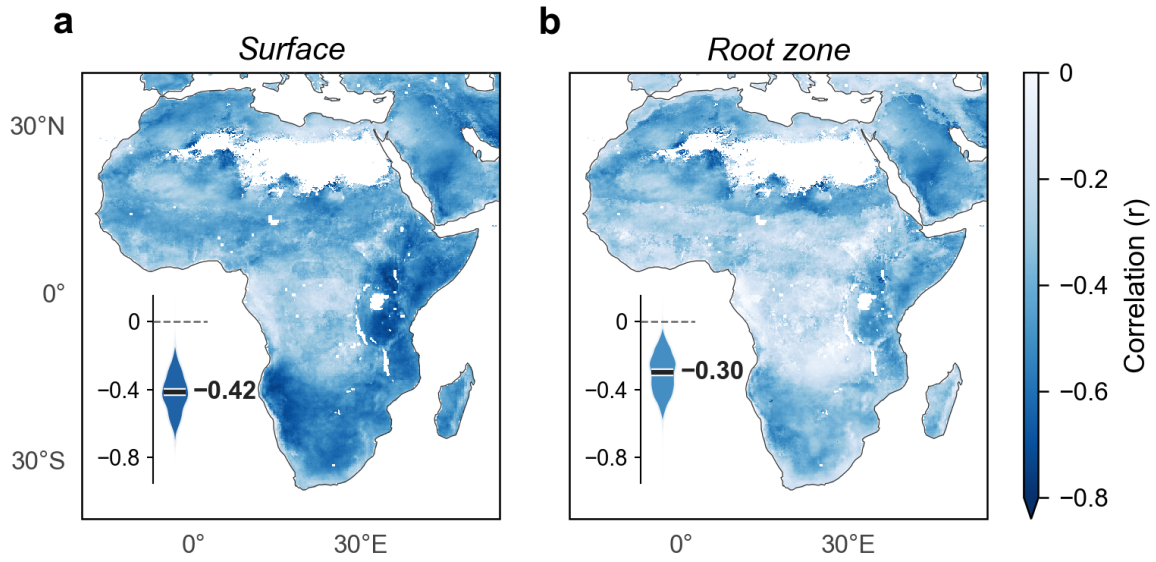

Figure S3: **Depth-dependent attenuation of SM–VPD coupling reproduced with GLEAM.** (a,b) Spatial patterns of daily correlation between VPD and SM anomalies for surface SM (SMs, top few centimetres) and root-zone SM (SMrz, integrated from the surface over the full rooting depth).

To assess the sensitivity of our findings to the choice of SM dataset, we repeated the SM–VPD coupling analysis using GLEAM v4.2b. The depth-dependent attenuation of coupling strength is reproduced, with surface SM showing stronger negative correlation with VPD (median r = -0.42) than root-zone SM (median r = -0.30) (Supplementary Fig. S3). Yield attribution analyses were not replicated with GLEAM because its root-zone product integrates the full soil column including the surface layer, preventing the vertically exclusive decomposition required by our framework.

To further assess robustness beyond the SM–VPD coupling analysis, we replicated all four main analyses using soil moisture from the global Land Data Assimilation System (GLDAS) Version 2.1 Noah land surface model at 0.25° resolution. GLDAS Noah provides 3-hourly soil moisture at four depths, of which we used the upper three layers: 0–10 cm (L1), 10–40 cm (L2), and 40–100 cm (L3). We aggregated the 3-hourly data to daily means to match the temporal resolution of the ERA5-Land analysis. These layer definitions differ slightly from ERA5-Land (0–7, 7–28, and 28–100 cm) owing to differences in the vertical discretization of the Noah and HTESSEL land surface models. Despite these structural differences, GLDAS provides an independent test of whether the depth-dependent patterns identified in the main analysis are robust to the choice of land surface model. For the yield-based analyses (Supplementary Figs. S4–S6), only soil moisture inputs were replaced with GLDAS; all meteorological predictors (VPD, precipitation, maximum temperature, radiation) and yield data remained identical to the main analysis. For the SIF-based machine learning attribution (Supplementary Fig. S7), all predictor variables were aggregated to a common 0.25° resolution before county-level aggregation to ensure a fair comparison of variable importance across groups.

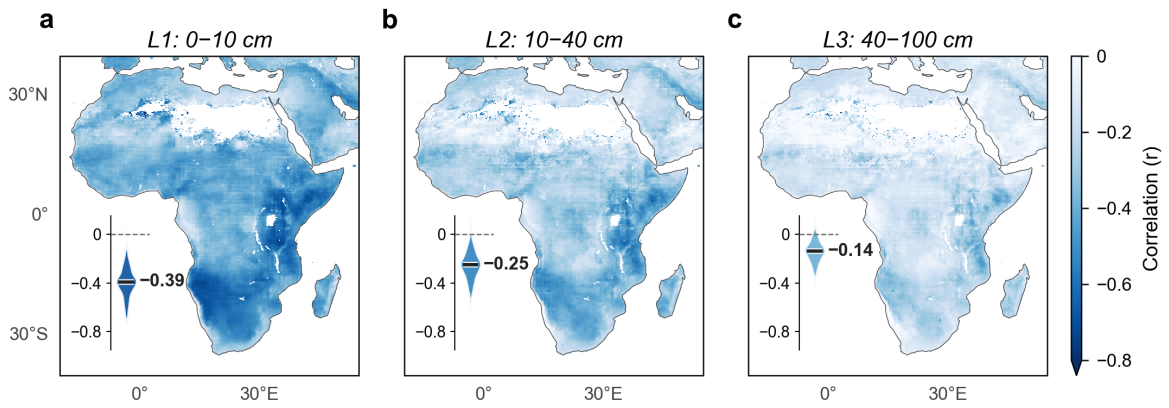

Figure S4: **Depth-dependent decoupling between daily SM and VPD anomalies using GLDAS Noah.** (a–c) Spatial patterns of correlation for L1 (0–10 cm), L2 (10–40 cm), and L3 (40–100 cm). Same as Fig. 1a–c but using GLDAS Noah soil moisture.

The depth-dependent attenuation of SM–VPD coupling is reproduced with GLDAS Noah (Supplementary Fig. S4). Surface SM shows stronger negative correlation with VPD (r = -0.39) than root-zone SM (r = -0.14), confirming that the progressive weakening of SM–VPD coupling with depth is a robust physical feature independent of SM dataset choice.

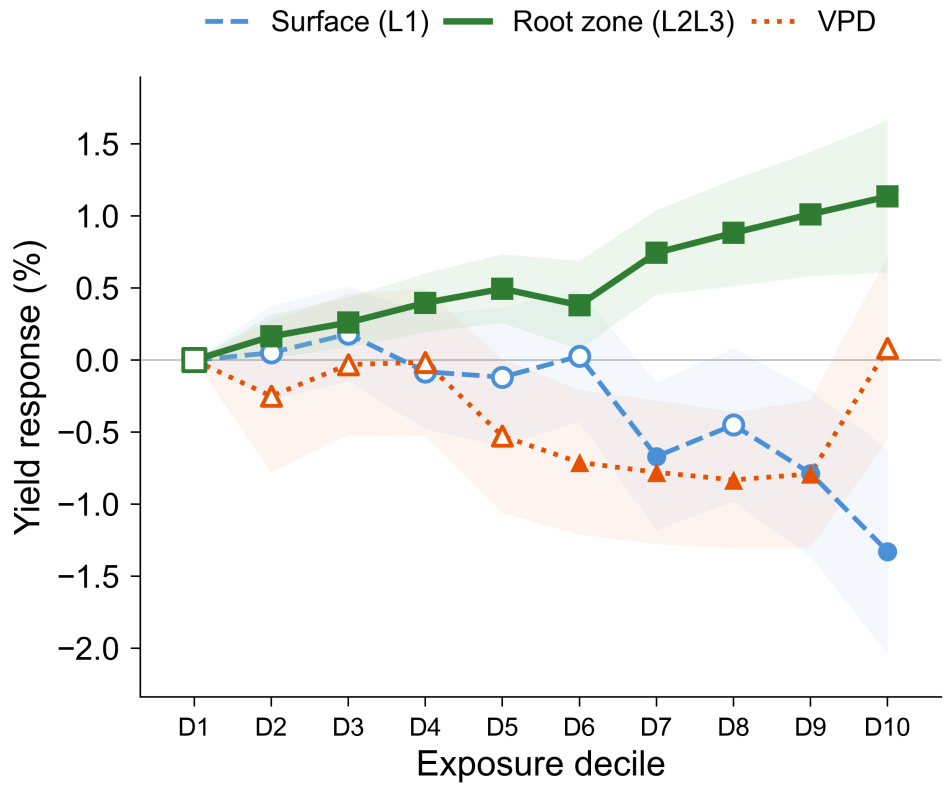

Figure S5: **Nonlinear yield response to GLDAS Noah soil moisture and VPD.** Marginal yield response to surface SM (L1, blue circles), root-zone SM (L2L3, green squares), and VPD (orange triangles). Filled markers indicate statistically significant coefficients ($p < 0.05$); open markers indicate non-significant coefficients. Shaded areas denote 95% confidence intervals. Same as Fig. 2a but using GLDAS Noah soil moisture.

The divergent yield response patterns are replicated with GLDAS Noah (Supplementary Fig. S5). Root-zone SM maintains a monotonically increasing relationship with yield through the wettest decile, consistent with the ERA5-Land results (Fig. 2a). Surface SM coefficients at the wettest deciles are negative in both datasets, with statistical significance in GLDAS. The key finding that root-zone SM provides consistent and positive yield information while surface SM does not is robust to the choice of SM datasets.

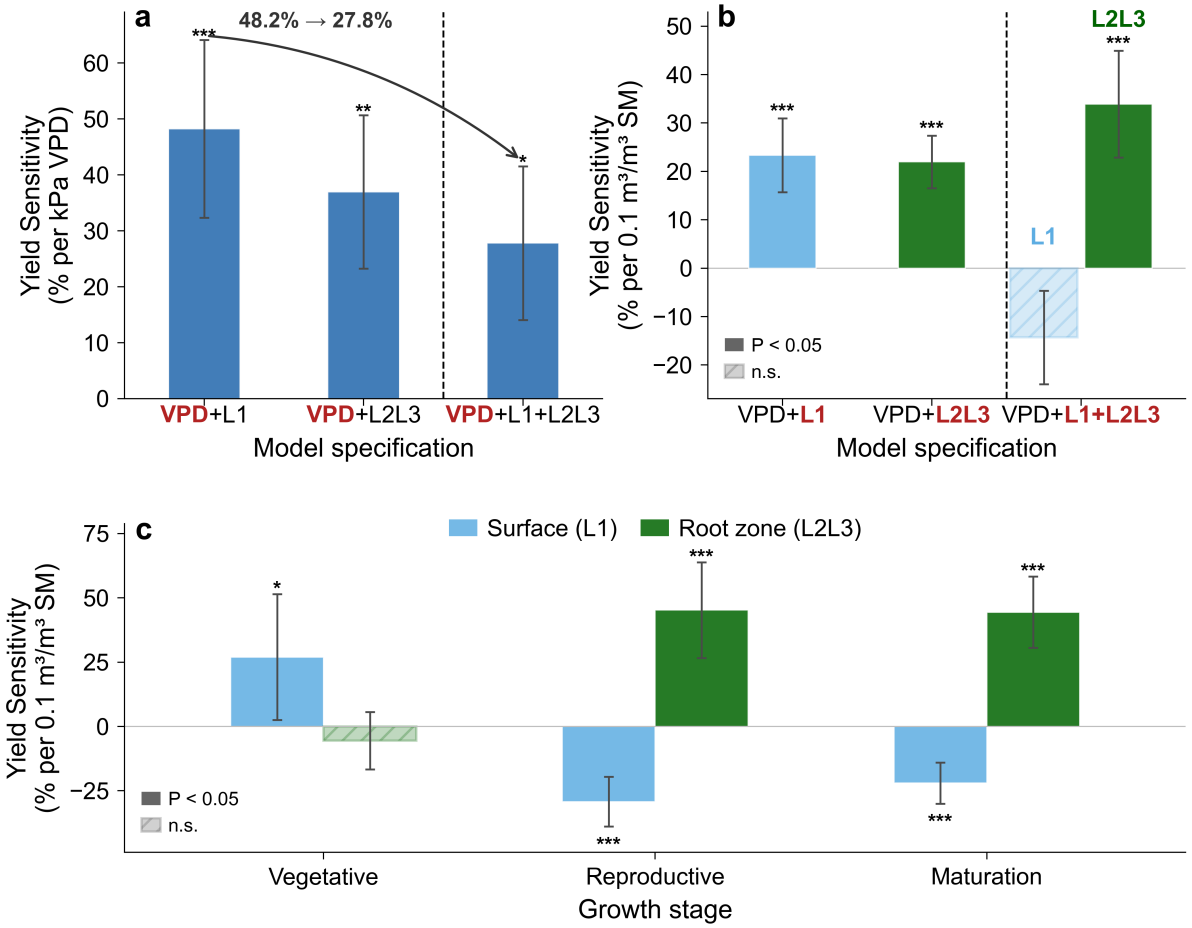

Figure S6: **VPD and SM yield sensitivity using GLDAS Noah.** (a) VPD yield sensitivity (% loss per kPa) across model specifications. (b) SM yield sensitivity (% per 0.1 m^3/m^3) for surface (L1) and root-zone (L2L3) layers across separate and combined models. (c) Phenological divergence. Yield response estimated across vegetative (0–60% of growing season), reproductive (60–85%), and maturation (85–100%) phases. Error bars denote 95% confidence intervals; significance levels: *P < 0.05, **P < 0.01, ***P < 0.001. Same as Fig. 3 but using GLDAS Noah soil moisture.

The seasonal regression results with GLDAS Noah confirm and strengthen the main findings (Supplementary Fig. S6). Estimated VPD yield sensitivity decreases from 48.2% to 27.8% per kPa when root-zone SM is included (Supplementary Fig. S6a). In the model that includes both layers (VPD+L1+L2L3), the surface SM coefficient becomes statistically non-significant while root-zone SM remains large and significant (Supplementary Fig. S6b), mirroring the ERA5-Land results (Fig. 3). The phenological decomposition reveals consistent depth-dependent patterns across growth stages (Supplementary Fig. S6c). Surface SM becomes significantly negative during reproductive and maturation phases, while root-zone SM dominates yield sensitivity during these critical periods, consistent with the ERA5-Land results (Fig. 3c).

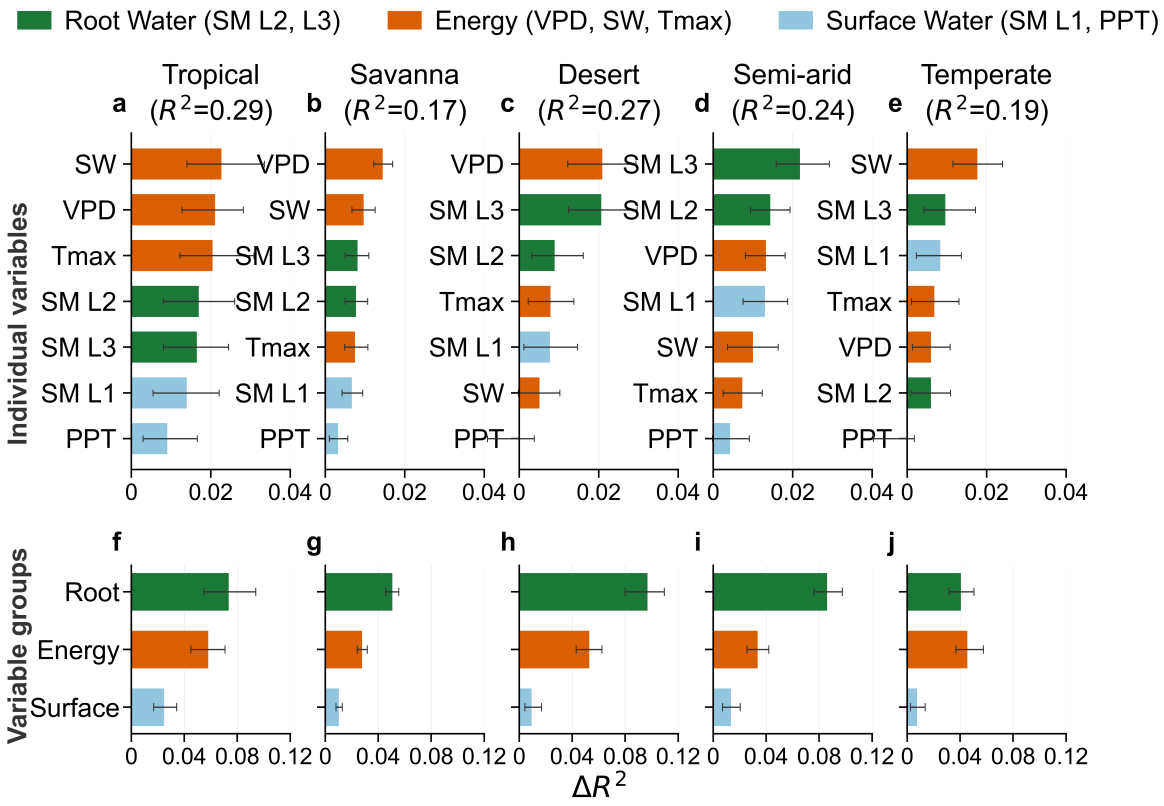

Figure S7: **SIF attribution using GLDAS Noah soil moisture.** (a-e) Individual-variable drop-column importance (ΔR^2) with bootstrap 95% CIs across five climate zones. (f-j) Group-level ΔR^2 (Root Water vs Energy vs Surface Water). Same as Fig. 4 but using GLDAS Noah soil moisture.

The SIF-based machine learning attribution was replicated using GLDAS Noah soil moisture (Supplementary Fig. S7). At the individual variable level, rankings differ somewhat from the main analysis, with energy-related predictors (SW, VPD) ranking higher in tropical zones (Supplementary Fig. S7a), reflecting differences in soil hydraulic parameterizations between the Noah and HTESSEL land surface models that affect how soil moisture variability is partitioned across layers. However, at the group level, Root Water still dominates in all five climate zones (Supplementary Fig. S7f–j), with only Temperate regions showing comparable contributions from Root Water and Energy. The conclusion that root-zone SM provides the most irreplaceable information for predicting vegetation productivity is robust to the choice of SM datasets.

Regression coefficients

Table S1: **Full regression coefficients for seasonal panel models.** Coefficients represent percentage yield change per unit increase: 1 kPa for VPD, 0.1 m^3/m^3 for SM, 100 mm for precipitation, 1°C for maximum temperature, and 1 W m^{-2} for radiation. All models include county and year fixed effects with standard errors clustered at the county level.

(a) Seasonal models

	VPD + SM_{L1}	VPD + SM_{L2L3}	VPD + SM_{L1} + SM_{L2L3}
VPD	+47.0***	+37.9***	+38.0**
SM_{L1}	+14.0**	—	−0.0
SM_{L2L3}	—	+18.9***	+18.9**
PPT	+1.2***	+1.2***	+1.2***
T_{max}	−7.4**	−4.7	−4.7
Radiation	−2.3	−2.4	−2.4

(b) Phenological models (VPD + SM_{L1} + SM_{L2L3})

	Vegetative	Reproductive	Maturation
VPD	+3.2***	−1.8***	−0.6
SM_{L1}	+12.8*	−2.5	−14.2***
SM_{L2L3}	−0.1	+7.1*	+27.8***
PPT	+1.4***	+1.3**	+0.9
T_{max}	+0.5	−4.3***	+1.0
Radiation	−6.5***	+3.3***	+0.9

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

SIF attribution: model performance

XGBoost models were trained on deseasonalized anomalies, which remove the climatological seasonal cycle and retain only interannual and intraseasonal variability. Test R^2 values therefore reflect the fraction of non-seasonal variance explained, and are expected to be lower than models fitted to raw values that include the seasonal cycle. Training and test sets were split using grouped stratified sampling by year (80/20), with all observations from a given county-year combination assigned entirely to either set to prevent information leakage from spatial and temporal autocorrelation. The optimal number of boosting iterations was determined via 5-fold cross-validation with early stopping. Model performance across climate zones is shown in Supplementary Fig. S8.

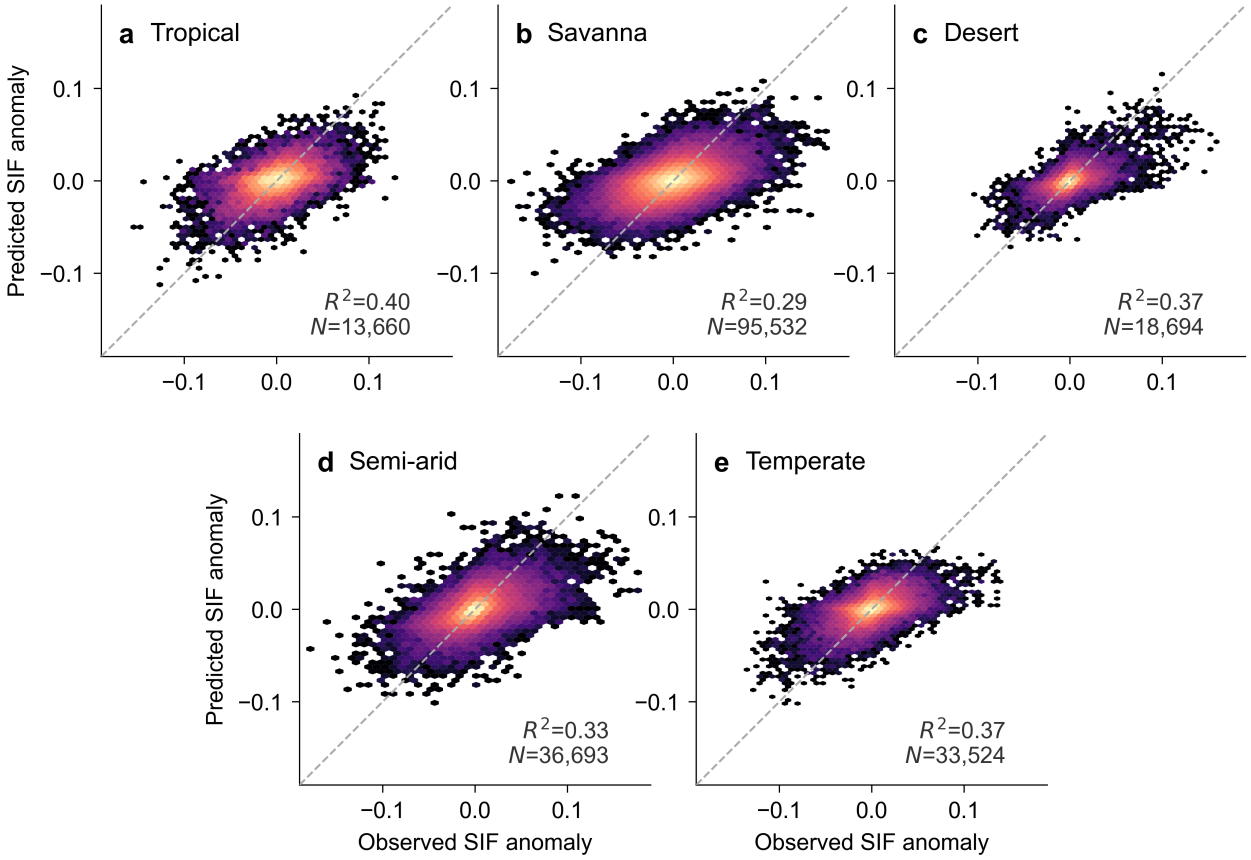

Figure S8: **XGBoost model performance for predicting SIF anomalies across climate zones.** (a-e) Predicted versus observed deseasonalized SIF anomalies on the held-out test set for each Köppen climate zone. Colour indicates point density. Dashed lines indicate the 1:1 relationship. R^2 and N (number of test observations) are shown in each panel.

SIF attribution: Shapley decomposition

The drop-column importance used in the main analysis (Fig. 4f-j) measures the unique, non-redundant contribution of each variable group by removing it and retraining the model. While this directly addresses whether a group provides irreplaceable information, it does not account for variance shared among groups. As a complementary approach, we applied Shapley R^2 decomposition from cooperative game theory, which distributes the total explained variance (R^2) into additive contributions that sum exactly to the full-model R^2 (the efficiency property), thereby accounting for both unique and shared variance.

We treated the three variable groups as players in a cooperative game: Root water (R: SM L2, SM L3), Energy (E: VPD, shortwave radiation, maximum temperature), and Surface water (S: SM L1, precipitation). The characteristic function $v(A)$ assigns to each coalition $A \subseteq \{E, S, R\}$ the weighted test-set R^2 of an XGBoost model trained using only the features in A, with $v(\emptyset) = 0$. For three players, this requires training $2^3 - 1 = 7$ coalition models. All models used identical hyperparameters and the same number of boosting rounds determined by cross-validation on the full feature set, ensuring direct comparability of R^2 values across coalitions.

The Shapley value of each group i is:

$$\varphi_i = \sum_{A \subseteq N \setminus \{i\}} \frac{|A|!\,(|N| - |A| - 1)!}{|N|!} \left[v(A \cup \{i\}) - v(A) \right] \tag{1}$$

which averages the marginal contribution of group i across all possible orderings in which groups can be added to the model. By construction, $\varphi_E + \varphi_S + \varphi_R = v(\{E, S, R\})$.

Bootstrap confidence intervals were obtained by resampling the test set with replacement (1,000 iterations) and recomputing all coalition values and Shapley values for each resample. Models were not retrained, so the intervals reflect uncertainty in the test-set evaluation.

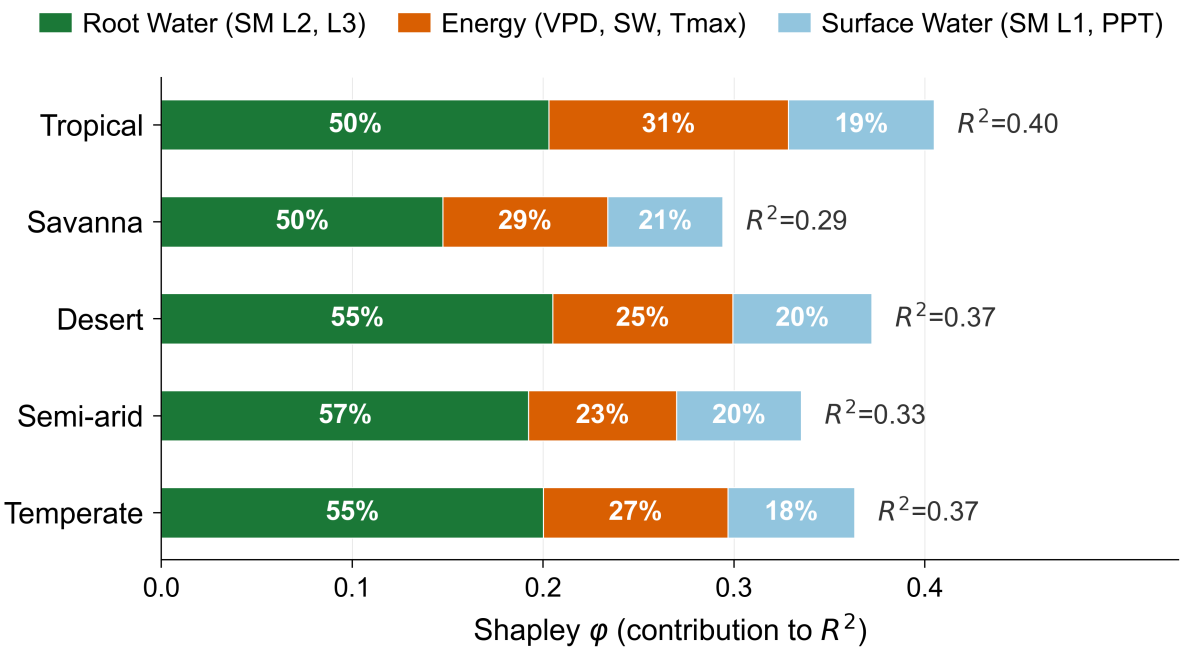

Figure S9: **Shapley R^2 decomposition of SIF anomaly variability across climate zones.** Stacked bars show the additive Shapley value contribution of each variable group to the total explained variance. Percentages denote each group's share of the total R^2; values to the right indicate the full-model R^2. Bootstrap 95% confidence intervals on relative contributions are within ±2 percentage points.

The results confirm the findings from the drop-column analysis (Supplementary Fig. S9). Root Water consistently explains the largest share of the total R^2 across all five climate zones, contributing 50% in tropical and savanna zones and increasing to 55-57% in desert, semi-arid, and temperate zones. Energy contributes 23-31%, while Surface water accounts for 18-21%. The ranking Root Water > Energy > Surface Water is preserved in every climate zone, and bootstrap 95% confidence intervals on relative contributions are within ±2 percentage points, confirming the stability of the decomposition. Notably, surface water's Shapley contribution (18-21%) substantially exceeds its drop-column importance, reflecting variance shared with root-zone water that is redistributed under the Shapley framework but does not constitute unique, irreplaceable information.

SIF attribution: hard energy filtering

To assess sensitivity to the energy filtering approach, we repeated the SIF attribution analysis using hard thresholds that exclude observations below energy-sufficient conditions (maximum temperature < 15°C, VPD < 0.5 kPa, or shortwave radiation < 200 W m^{-2}), following previous studies [**?**, **?**]. The results are consistent with the soft energy gating used in the main analysis (Fig. 4 and Fig. S10). At the individual variable level, the rankings show minor differences relative to the main analysis: for example, VPD ranks higher in savanna zones under hard filtering, likely because excluding low-energy observations increases the relative weight of high-VPD periods. These differences do not propagate to the group level, where Root Water still dominates in all five zones, confirming that the main findings are robust to the choice of energy filtering method.

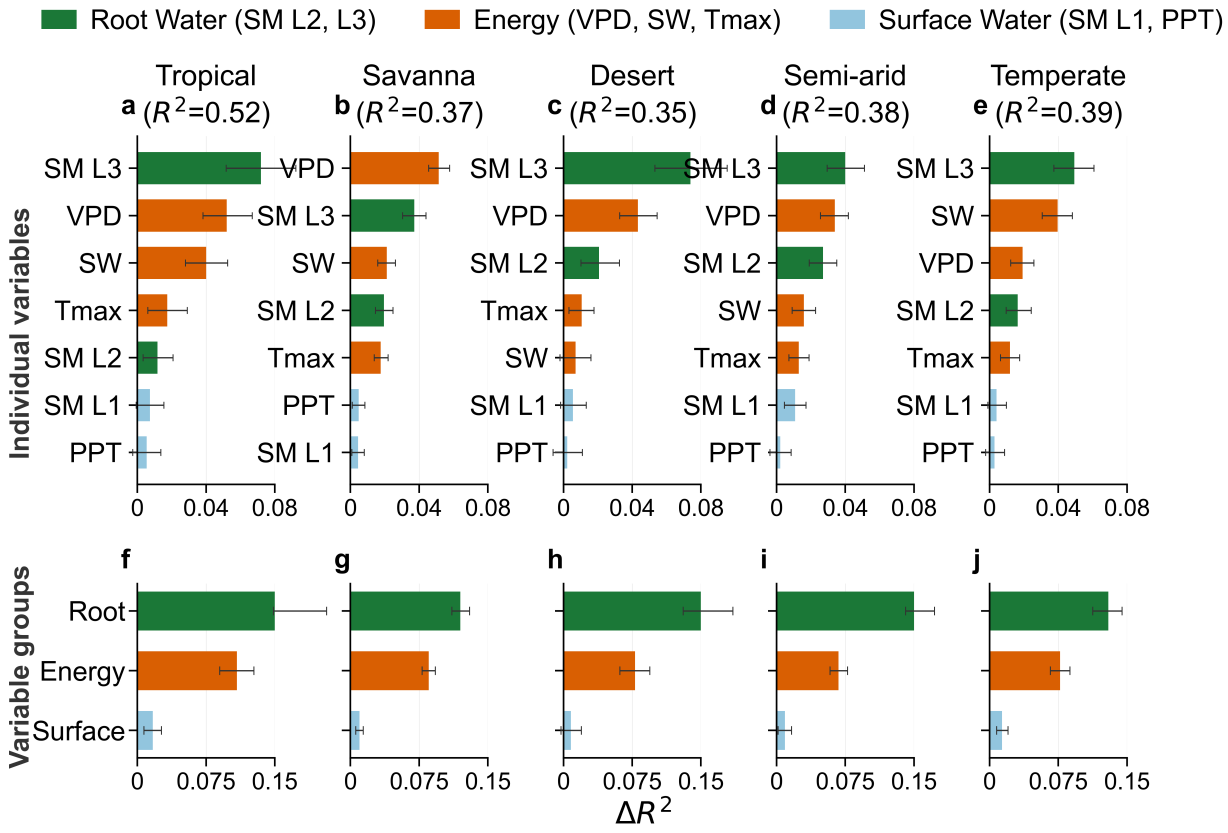

Figure S10: **SIF attribution using hard energy filtering.** Same as Fig. 4 but restricting observations to energy-sufficient conditions (maximum temperature > 15°C, VPD > 0.5 kPa, and shortwave radiation > 200 W m^{-2}). (a-e) Individual-variable drop-column importance (ΔR^2) with bootstrap 95% CIs across Köppen climate zones. (f-j) Group-level ΔR^2 (Root Water vs Energy vs Surface Water).