Aridity Amplifies Afforestation-Driven Total Water Storage Depletion on China's Loess Plateau: Satellite Evidence from the Full 2002–2024 GRACE(-FO) Record

Parker Tikson[1], Cynthia Gerlein-Safdi[1]

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, CA, USA

Key Points:

- The full 2002–2024 GRACE+GRACE-FO record shows a significant decline in total water storage in the Loess Plateau, unexplained by climate.
- An explicit vegetation × aridity interaction shows that the same degree of greening causes disproportionately greater water depletion in drier grid cells.
- TWS depletion peaks in semi-arid cells, where vegetation establishes but precipitation alone cannot sustain mature canopies.

Corresponding author: Parker Tikson, parkertikson@berkeley.edu

Corresponding author: Cynthia Gerlein-Safdi, cgerlein@berkeley.edu

Abstract

Large-scale afforestation programs are increasingly deployed as instruments of carbon sequestration, land stabilization, and ecological restoration, yet their long-term hydrological consequences in water-limited environments remain poorly constrained. China's Grain for Green Program (GGP), initiated in 1999, represents the largest deliberate land-cover transformation in human history, converting over 30,000 km^2 of Loess Plateau cropland to forest and grassland over 25 years. Using the complete 2002–2024 GRACE and GRACE-FO satellite gravity record with harmonic seasonal decomposition and Theil-Sen/Mann-Kendall trend analysis, we document a domain-mean total water storage (TWS) decline of -0.38 cm EWH yr^{-1} (Mann-Kendall $p < 0.0001$) across the Loess Plateau, representing -8.4 cm cumulative over the study period. Precipitation shows no significant trend ($p = 0.59$), and a Clausius-Clapeyron attribution analysis demonstrates that warming ($+0.02$ °C yr^{-1}, $p = 0.47$) can account for at most 10.8% of the observed evapotranspiration increase, with 89% requiring a vegetation-driven transpiration mechanism. A multiple regression incorporating an explicit vegetation (NDVI trend) $\times$ aridity interaction term reveals that the TWS depletion rate is significantly amplified in drier grid cells ($\beta_3 = -62.7$, $p < 0.0001$, adj. $R^2 = 0.53$), providing the first satellite-scale quantification of the aridity-conditioned water debt of afforestation. These results show that the hydrological cost of afforestation is not uniform but scales with water deficit, with direct implications for the spatial planning of global reforestation commitments under the Bonn Challenge and United Nations Decade on Ecosystem Restoration.

Plain Language Summary

China's Grain for Green Program (GGP), the world's largest land restoration initiative, has converted millions of hectares of degraded cropland into forest and grassland since 1999. While the program has successfully reduced soil erosion and increased vegetation cover, its long-term effects on regional water resources remain poorly understood. Using satellite gravity data from the GRACE and GRACE-FO missions, which measure total water stored underground, in soil, and at the surface, we show that the Loess Plateau has been losing approximately 3.8 mm of water storage per year over the past 22 years. This decline cannot be explained by changes in rainfall or temperature. Instead, it is closely tied to the spatial pattern of GGP-driven greening: the more a location greened, the more water it lost. Critically, this relationship is strongest in the driest areas, where trees must access deep, ancient soil moisture to survive. Our analysis shows that temperature-driven evaporation accounts for only about 10.8% of the increased water consumption; the remaining 89.2% requires a vegetation explanation. These findings provide a spatially explicit framework for identifying where future afforestation programs carry the greatest hydrological risk.

1 Introduction

Terrestrial afforestation and reforestation deliver an unusually broad portfolio of ecological and societal benefits, which explains both their political popularity and the scale of investment now committed to them globally. Reforested landscapes sequester atmospheric carbon (Nilsson & Schopfhauser, 1995; Hasegawa et al., 2024); reduce soil erosion and dust emissions in degraded drylands (F. Wang et al., 2017; Kulik et al., 2023); mitigate floods (Trabucco et al., 2008); restore wildlife habitat and biodiversity (Oraon et al., 2023); provide rural development and socioeconomic uplift (Bryan et al., 2018); and locally moderate climate through altered albedo, surface roughness, and evapotranspiration (Gerlein-Safdi et al., 2020; F. Wang et al., 2020). The Bonn Challenge commits nations to restoring 350 million hectares of degraded land by 2030 (Bonn Challenge, 2011), and the United Nations Decade on Ecosystem Restoration (2021–2030) targets a further billion hectares globally (United Nations Environment Programme, 2021). China's

Grain for Green Program (GGP, *Tuigeng Huancao*), launched in 1999 following the catastrophic 1998 Yangtze River floods, the 1997 Yellow River cut-off, and chronic soil erosion across the middle and upper reaches of both basins, has measurably advanced several of these goals. The program contributed to a multi-decadal decline in Yellow River sediment loads, which decreased by approximately 90% over 60 years through a combination of check dams, terracing, and revegetation, with GGP-driven revegetation playing a measurable role from the late 1990s onwards (S. Wang et al., 2016). It has converted $30,000$ km^2 from cropland to vegetation, sequestering carbon at a regional scale (Feng et al., 2016; Bryan et al., 2018), and improved rural livelihood indicators in the most affected provinces (Delang & Yuan, 2015).

As similar projects develop globally, the question is no longer whether to invest in dryland restoration, but instead, how to allocate that investment spatially, and on what scientific basis. In particular, the hydrological consequences of large-scale afforestation in water-limited environments have received comparatively little attention in the policy frameworks that govern these commitments: where rainfall is insufficient to meet the evapotranspiration demand of mature forest canopies, trees can supplement their water uptake from stored soil moisture, potentially drawing on legacy reserves accumulated over decades to centuries (H. Chen et al., 2008; Liang et al., 2022; Evaristo, 2026). Whether this draw is hydrologically sustainable depends on local climate, root depth, and species choice. The spatial scale at which this dependence operates has not been quantified from satellite observations.

The Loess Plateau (c. 100–112°E, 34–42°N, representing about 640,000 km^2) is the world's largest loess deposit and among the most severely eroded landscapes on Earth (Fu et al., 2017). Its mean annual precipitation gradient, from c. 200 mm yr^{-1} in the semi-arid northwest to c. 650 mm yr^{-1} in the sub-humid southeast (H. Chen et al., 2008), which means that GGP forests planted in the Loess Plateau operate across a wide spectrum of water availability. Field studies have documented that mature GGP plantations form persistent Dry Soil Layers (DSLs) at 2–10 m depth, where deep tree roots extract moisture faster than precipitation can replenish it (H. Chen et al., 2008). These DSLs can persist for years to decades and may recover only slowly on management timescales, and have been documented across species, including *Robinia pseudoacacia*, *Pinus tabuliformis*, and *Caragana korshinskii* (Evaristo, 2026). The water balance implication is significant: if tree roots are depleting a deep moisture reservoir, the standard hydrological accounting framework, which solves for streamflow as $Q = P - \mathrm{ET}$ and implicitly assumes total water storage change $\Delta S \approx 0$, will systematically underestimate the true water cost of afforestation.

The GRACE (Gravity Recovery and Climate Experiment, 2002–2017) and GRACE-FO (Follow-On, 2018–present) satellite missions measure vertically integrated total water storage (TWS) changes directly from gravitational anomalies (Tapley et al., 2004). The CSR RL06M mascon product is provided at 0.5° (˜50 km) gridded resolution with a native effective footprint of ˜300 km, and at monthly temporal resolution (Zhong et al., 2020; Landerer et al., 2020). Unlike surface instruments or land surface models, GRACE integrates the full soil column from surface to bedrock, making it uniquely suited to detecting deep subsurface depletion signals (Scheliga et al., 2025), including the one DSL formation might produce. A growing body of literature has used GRACE to examine TWS trends on the Loess Plateau and adjacent regions. (Lv et al., 2019) used GRACE through January 2017 for the Yellow River Basin and identified afforestation and irrigation as significant TWS drivers, but their analysis predates the GRACE-FO era and does not isolate an aridity-conditioned interaction effect. (Z. Wang et al., 2025) and (Liu, Li, Wang, & Zhang, 2023) documented GRACE-based TWS losses consistent with regional greening over 2003–2020, using trend analysis and Pearson correlation, but without climate covariate controls or a formal interaction model. At the global dryland scale, (Liu, Li, Wang, & Zhou, 2023) demonstrated a negative correlation between vegetation growth and TWS across drylands during 1982–2016, projecting reductions of 41–84% by 2100 under continued dryland expansion. However, their analysis is global in scope, uses 1982–2016 data, and does not include an explicit vegetation × aridity interaction term.

Our study addresses an analytical gap in the existing literature. Field-scale work has theorized vegetation–water tradeoffs in semi-arid environments (Feng et al., 2016; H. Chen et al., 2008; Liang et al., 2022), with (Feng et al., 2016) proposing a "water carrying capacity" threshold beyond which afforestation creates persistent hydrological deficits. Recent GRACE-based analyses have quantified plateau-scale TWS depletion and identified vegetation as a major contributor (Liu, Li, Wang, & Zhang, 2023; Xie et al., 2018; Lv et al., 2019), typically using correlation, decomposition, or trend-comparison methods. However, the aridity-conditioned nature of the vegetation–water relationship—specifically, whether the same vegetation greening produces systematically different TWS responses across the plateau's aridity gradient—has not been formally tested at satellite scale using an interaction regression framework.

Here, we address these gaps using a multi-dataset satellite analysis spanning 2002–2024. In particular, we extend the GRACE record to 2024 (incorporating GRACE-FO data not available to earlier studies) and apply an interaction regression to test whether aridity moderates the magnitude of the vegetation-TWS relationship across the Loess Plateau. Our analysis tests two directional hypotheses. First, we hypothesize that TWS exhibits a statistically significant domain-wide negative trend over 2002–2024 after controlling for precipitation variability and temperature-driven evaporative demand. Second, we hypothesize that the aridity index significantly moderates the vegetation–TWS relationship, such that the same degree of NDVI greening produces disproportionately greater TWS depletion in drier grid cells. Together, these hypotheses constitute a spatially explicit, satellite-scale test of the GGP water debt mechanism, one whose outcome has direct implications not for whether to pursue afforestation, but for where it can be sustained.

2 Data and methods

2.1 Study domain and datasets

The six principal datasets used in this study are presented and summarized in Table 1. This study focuses on the full Loess Plateau, defined by a bounding box of 100–112°E, 34–42°N and a published boundary shapefile from the National Tibetan Plateau Data Center (NTPDC) (National Tibetan Plateau Data Center, 2018), rasterized to a binary mask at 0.5° resolution. The bounding box was chosen to maximally contain the Loess Plateau polygon while aligning with integer-degree cell edges of the GRACE 1° and ancillary 0.5° grids and matching the spatial framing used in prior Loess Plateau hydrological studies (Feng et al., 2016; Lv et al., 2019). All statistical analyses operate on the rasterized Loess Plateau polygon mask within this box, not on the rectangular box itself; cells outside the Loess Plateau polygon are excluded from all calculations. The eastern edge of the box is held at 112°E to reduce the contribution of coal-mining-dominated cells in eastern Shanxi, one of the most significant non-vegetation confounders identified in Section 3.5. We acknowledge this excludes the easternmost portion of the Loess Plateau geomorphological boundary (which extends to c. 114.5°E): cells east of 112°E are dominated by intensive coal mining (Wu et al., 2021) and including them would amplify rather than reduce the spatial confounding addressed in Section 3.5.

Statistical analysis (interaction regression, trend significance) is performed on the native 1° GRACE grid (n = 96 cells inside the Loess Plateau boundary). Spatial visualizations use the 0.5° grid (n = 360 cells), with GRACE TWS upsampled by nearest-neighbor interpolation for display only; no statistical inference is drawn from the 0.5°

representation. All datasets were aggregated to their respective monthly grids matching the GRACE mascon resolution or the 0.5° ancillary grid (Table 1).

Dataset	Variable	Source	Resolution	Period
GRACE/GRACE-FO CSR RL06M	TWS anomaly (cm EWH)	UTexas, Austin	∼300 km → 0.5°	2002–2024
CHIRPS v2.0	Precipitation	UCSB / IRI OPeNDAP	5 km → 0.5°	2002–2024
MODIS MOD16A3GF v061	Actual ET	NASA LP DAAC	500 m → 0.5°	2002–2023
TerraClimate	PET only	Univ. Idaho THREDDS	4 km → 0.5°	2002–2024
MODIS MOD13C2	NDVI	NASA Earthdata	0.05° → 0.5°	2002–2024
ERA5-Land	2 m air temperature	Copernicus / THREDDS	9 km → 0.5°	2002–2024

Table 1. Dataset inventory. All data are free and publicly available and direct links and references are listed in the Data Availability Statement.

2.2 GRACE gap bridging

The 11-month gap between GRACE and GRACE-FO (July 2017 to May 2018) and 24 sporadic dropout months were filled by linear interpolation after reindexing to a complete monthly axis. Floating-point coordinate noise ($\sim 10^{-14\circ}$) between GRACE and GRACE-FO grids was resolved by rounding to two decimal places prior to concatenation to prevent spurious NaN-filling. Duplicate timestamps at the mission transition were resolved by retaining the first occurrence. Sensitivity analysis indicates that gap bridging contributes ± 0.02 cm yr^{-1} to trend uncertainty, small relative to the -0.38 cm yr^{-1} signal.

2.3 Harmonic seasonal decomposition

To isolate the secular TWS trend from the East Asian Summer Monsoon recharge cycle, we fit an ordinary least-squares harmonic regression independently at each 0.5° grid cell:

$$\mathrm{TWS}(t) = a_0 + a_1 t + a_2 \sin\left(\frac{2\pi t}{12}\right) + a_3 \cos\left(\frac{2\pi t}{12}\right) + a_4 \sin\left(\frac{4\pi t}{12}\right) + a_5 \cos\left(\frac{4\pi t}{12}\right) + \varepsilon(t) \quad (1)$$

where t is time in months, a_1 captures the linear secular trend, and the 12-month and 6-month sinusoidal pairs capture the annual and semi-annual seasonal cycles, respectively. The deseasonalized residual $\varepsilon(t)$ isolates the non-seasonal signal for trend analysis. Seasonal amplitude ranged from 0.47 cm equivalent water height (EWH) in the dry northwest to 3.31 cm EWH in the wetter southeast.

2.4 Trend analysis

Monotonic trends were estimated using the Theil-Sen slope estimator (Sen, 1968), which computes the median of all pairwise slopes and is robust to outliers and non-normality. Statistical significance was assessed via the Mann-Kendall trend test (Mann, 1945; Kendall, 1975) at $\alpha = 0.05$. Both analyses were applied independently at each grid cell to produce spatially explicit trend maps. Stippling on spatial maps presented in the result section indicates Mann-Kendall $p < 0.05$.

2.5 Evapotranspiration

We use the MODIS MOD16A3GF v061 (Mu et al., 2011) ET product as a mechanistic consistency check on the vegetation-driven signal in TWS. The MODIS ET product applies the Penman-Monteith equation driven by time-varying MODIS leaf area index (LAI), fractional photosynthetically active radiation (FPAR), albedo, and land cover, which makes it sensitive to canopy expansion as GGP forests grow. Because MOD16A3GF LAI shares upstream spectral inputs with MODIS NDVI, the ET trend is interpreted as a mechanistic consistency check rather than fully independent evidence. The TerraClimate (Abatzoglou et al., 2018a) potential evapotranspiration (PET) product is used in the aridity index calculation, but not as an independent actual-ET check, as it employs a static vegetation parameterization that cannot capture the canopy-expansion signal of GGP afforestation.

2.6 Temperature attribution

To assess whether temperature-driven increases in evaporative demand could explain the observed TWS decline, we downloaded ERA5-Land monthly 2 m air temperature for 2002–2024, aggregated to 0.5°, and computed the Theil-Sen trend. We then applied the Clausius-Clapeyron relation—approximately 7% increase in atmospheric water demand per °C warming (Allen & Ingram, 2002)—to estimate the maximum ET increase attributable to temperature alone as:

$$\Delta \mathrm{ET}_T = \overline{\mathrm{ET}} \times 0.07 \times \frac{d\bar{T}}{dt} \tag{2}$$

where $\overline{\mathrm{ET}}$ is the mean annual ET and $d\bar{T}/dt$ is the Theil-Sen temperature trend. The fraction of the observed MODIS ET trend explained by temperature alone ($f_T = \Delta \mathrm{ET}_T/\Delta \mathrm{ET}_{\mathrm{obs}}$) is reported as an upper bound, since the Clausius-Clapeyron relation overstates the temperature sensitivity in water-limited systems, where actual ET is constrained by soil moisture availability.

2.7 Interaction regression

To formally test whether the aridity index moderates the vegetation–TWS relationship, we fit a multiple ordinary least squares (OLS) regression across all 96 Loess Plateau grid cells at the 1° GRACE resolution:

$$\mathrm{TWS}_i^{\mathrm{trend}} = \beta_0 + \beta_1(\mathrm{NDVI}_i^{\mathrm{trend}}) + \beta_2(\log_{10} \mathrm{AI}_i) + \beta_3(\mathrm{NDVI}_i^{\mathrm{trend}} \times \log_{10} \mathrm{AI}_i) + \varepsilon_i \tag{3}$$

where, for each cell i, $\mathrm{NDVI}^{\mathrm{trend}}$ is the per-cell Theil-Sen NDVI slope (NDVI decade^{-1}), AI is the time-mean aridity index (PET/P average from 2002 to 2024), and β_3 is the interaction term testing whether aridity moderates the vegetation–TWS relationship. The aridity index was log-transformed to address the right-skewed distribution (range: 0.52–31.35). The regression was implemented using `statsmodels` in Python. NDVI and AI fields were regridded from their native 0.5° resolution to the GRACE 1° grid by bilinear interpolation prior to fitting.

2.8 Confounder analysis method

Several non-vegetation processes could contribute to the observed GRACE TWS signal. We estimated literature-derived plausibility ranges for six confounders and report them as a sensitivity envelope rather than applying direct corrections to the primary analysis: (1) temperature-driven evaporative demand bounded by the Clausius-Clapeyron relation (Section 2.6), (2) precipitation trend contribution bounded by the 95% confidence interval on the CHIRPS slope, (3) the combined hydrological impact of the Shanxi coal industry (mine drainage, subsidence-induced aquifer leakage, and solid mass extraction) net of regional water recycling (Xie et al., 2018), (4) agricultural and non-coal industrial groundwater extraction (Li & Ma, 2024), (5) post-2002 reservoir storage additions in the Loess Plateau basin (Y. Chen et al., 2024), and (6) post-GGP residual rates of natural geomorphic erosion (S. Wang et al., 2016). The primary TWS trend is presented uncorrected; the confounder envelope presented in Section 3.5 documents the range of plausible non-vegetation contributions and the residual range that vegetation-driven processes must explain. Source citations and per-bar calculations are provided in Supporting Information Tables S1 and S2.

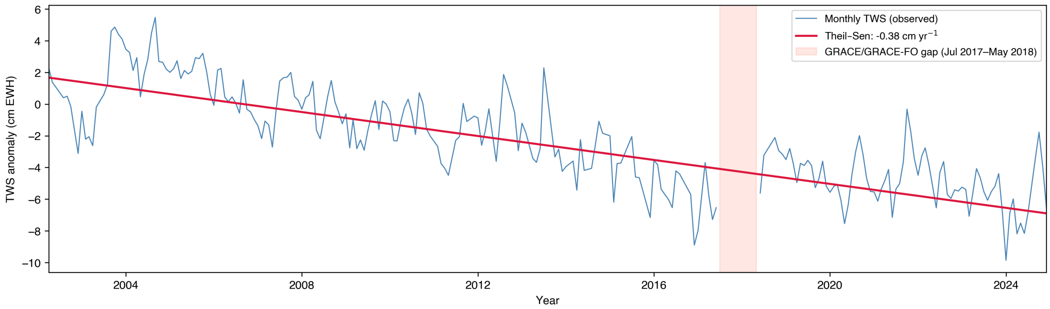

Figure 1. GRACE/GRACE-FO domain-mean deseasonalized TWS anomaly (cm EWH), 2002–2024. The two blue line segments show observed data before (2002 to June 2017) and after (June 2018 to 2024) the GRACE/GRACE-FO gap. The pink-shaded band marks the 11-month gap window. The red line shows the Theil-Sen trend: -0.38 cm yr^{-1}, $p < 0.0001$.

2.9 AI Tools Disclosure

In accordance with the Committee on Publication Ethics' position on AI tools and AGU Publications' policy, the authors disclose the following uses of AI tools in this work: Anthropic's Claude (Claude Sonnet 4.5 and Claude Opus 4.5) was used to assist with manuscript drafting, editing, and code generation for figure production. Google's Gemini was used to produce the mechanistic cross-section illustration shown in Figure 6. Project conception, as well as all scientific decisions, data analysis interpretations, statistical methods, and final manuscript content, were the responsibility of the authors. AI tools were used as assistive instruments and are not credited as authors. The authors verified all AI-generated content for accuracy and take full responsibility for the manuscript.

3 Results

3.1 Secular TWS decline

The deseasonalized GRACE+GRACE-FO domain-mean TWS time series presented in Figure 1 shows a clear and statistically significant negative trend of -0.383 cm EWH yr^{-1} (Mann-Kendall $p < 0.0001$), equivalent to -8.4 cm cumulative over the 22-year record. The seasonal amplitude ($\sim$1.2 cm peak-to-trough domain mean) is small relative to the secular decline, confirming the trend is not an artifact of incomplete seasonal removal. An anomalously deep trough in 2015–2016 is consistent with a documented regional drought; excluding these years from the trend estimation reduces the slope by less than 8%, indicating the trend is robust to individual drought years.

The spatial TWS trend map (Figure 2a) shows statistically significant negative trends in 95% of grid cells within the Loess Plateau boundary. The strongest depletion is concentrated in the eastern plateau (108–112°E), reaching -1.2 cm EWH yr^{-1} in the most affected cells. The western margin (100–103°E) shows slight storage increases coincident with the least-vegetated cells where GGP afforestation was minimal.

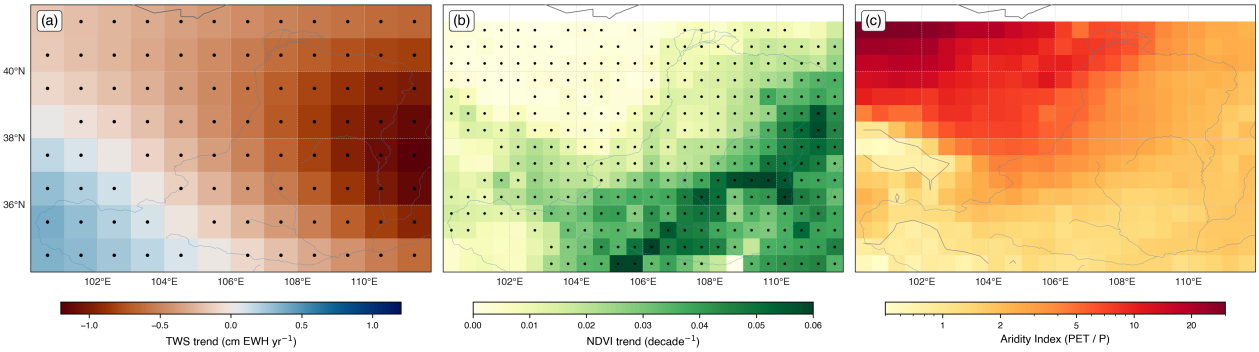

Figure 2. Spatial trends and aridity context across the Loess Plateau, 2002–2024. (a) GRACE/GRACE-FO Theil-Sen TWS trend at native 1° resolution, with stippling indicating Mann-Kendall significance at $p < 0.05$ (95% of cells significant). (b) MODIS NDVI Theil-Sen trend at 0.5° resolution, stippled at $p < 0.05$ (86% of cells significant). (c) Mean aridity index (PET/P, log scale) over 2002–2024 at 0.5° resolution. The AI = 1 contour separates the energy-limited regime (AI < 1, southwestern margin) from the water-limited regime that covers 92% of the plateau.

3.2 Climatic drivers

Three potential climate drivers were evaluated as alternative explanations for the observed TWS decline: precipitation, air temperature, and progressive aridification (Figure 3).

Domain-mean annual precipitation (Figure 3a) averages 375 mm yr^{-1} over 2002–2024 with no statistically significant trend (Theil-Sen slope ≈ +0.05 mm yr^{-1}, Mann-Kendall $p = 0.59$). The concurrent declining TWS and stable precipitation rules out reduced rainfall as a driver of storage loss.

ERA5-Land 2 m air temperature (Figure 3b) shows a weak positive trend of +0.020 °C yr^{-1} that is not statistically significant at the domain mean (Mann-Kendall $p = 0.47$), consistent with the documented East-Central China "warming hole" driven by aerosol loading and circulation changes (Sun et al., 2022). Applying the Clausius-Clapeyron relation (Equation 2), the temperature trend can account for at most $\Delta ET_T = +0.50$ mm yr^{-2}, representing only $f_T = 10.8\%$ of the observed MODIS ET trend of +4.63 mm yr^{-2} ($p < 0.001$). The remaining 89.2% of the ET increase requires a vegetation-driven explanation.

The spatial pattern of aridity provides a further constraint on climate-driven explanations. The eastern plateau cells, experiencing the greatest TWS depletion, fall within AI values ranging from 2 to 5, all within the semi-arid range, and show no progressive aridification trend over the study period, while cells trending toward higher AI are confined to the northwestern plateau where TWS depletion is weakest. The eastern plateau, therefore, loses water storage despite no climatic aridification signal, inconsistent with an aridity-driven mechanism.

Collectively, flat precipitation, non-significant warming, and a temperature contribution of only ∼11% by Clausius-Clapeyron rule out climate forcing as the primary driver of the observed TWS decline, leaving vegetation-mediated deep root transpiration as the dominant remaining plausible mechanism.

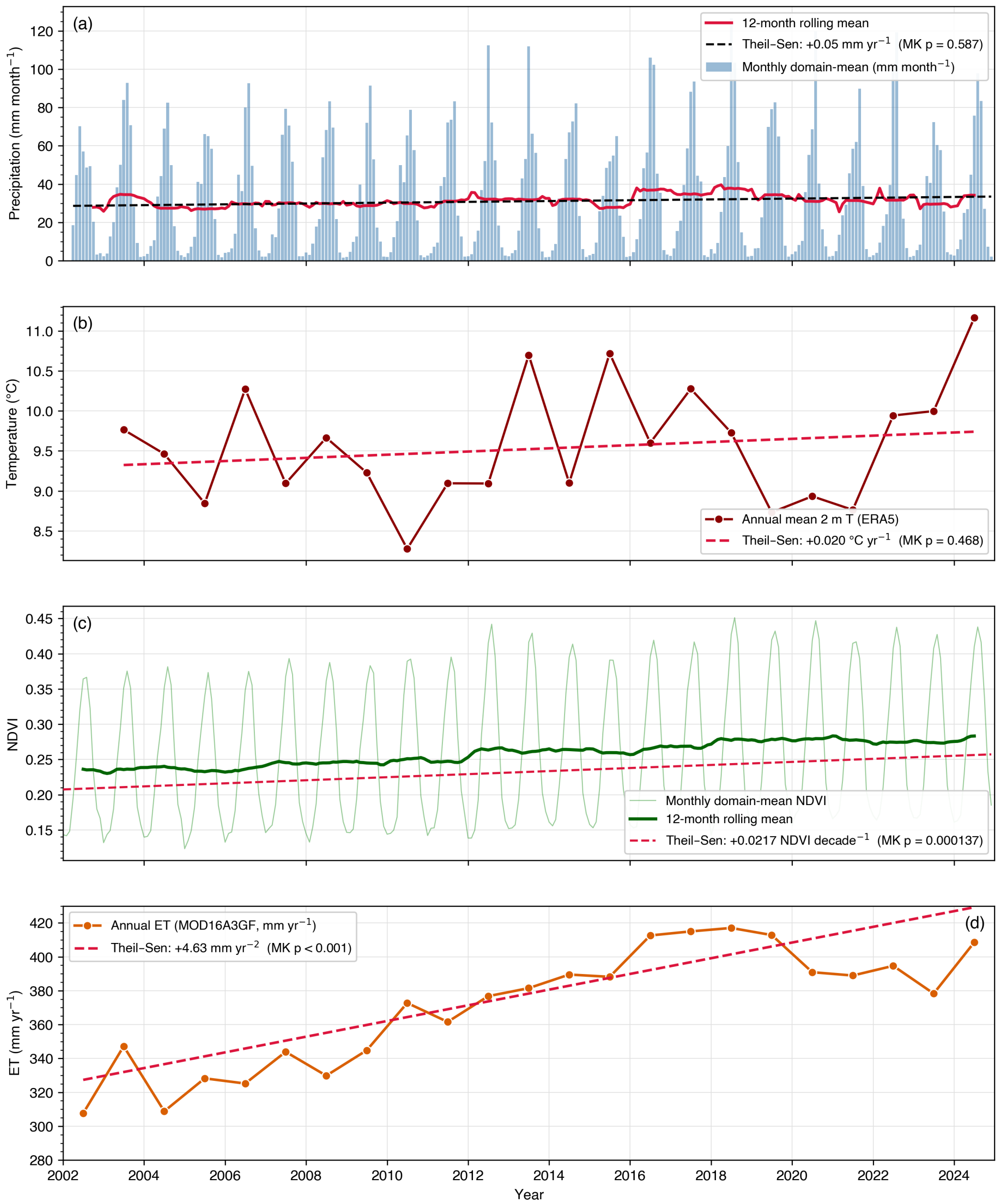

Figure 3. Climate and vegetation time series for the Loess Plateau, 2002–2024. (a) CHIRPS domain-mean monthly precipitation (bars) with 12-month rolling mean (red) and Theil-Sen trend (dashed, +0.05 mm yr^{-1}, $p = 0.59$). (b) ERA5-Land domain-mean annual 2 m air temperature with Theil-Sen trend (+0.020 °C yr^{-1}, $p = 0.47$). (c) MODIS MOD13C2 domain-mean NDVI with 12-month rolling mean and Theil-Sen trend (+0.0217 NDVI $decade^{-1}$, $p < 0.001$). (d) MODIS MOD16A3GF domain-mean annual evapotranspiration with Theil-Sen trend (+4.63 mm yr^{-2}, $p < 0.001$). Panels (a) and (b) show no statistically significant trends in climate forcing variables; panels (c) and (d) show highly significant increases in vegetation cover and water consumption.

3.3 Vegetation greening

NDVI (Figure 3c) increased by +0.0217 NDVI decade^{-1} (Theil-Sen, $p < 0.001$) over 2002–2024, rising from ∼0.22 to ∼0.27, a ∼23% relative increase representing the satellite fingerprint of GGP reforestation. The NDVI trend map (Figure 2b) shows statistically significant positive trends in 86% of grid cells, concentrated in the central-eastern plateau.

The MODIS ET product shows a significant positive trend of +4.63 mm yr^{-2} ($p <$ 0.001; Figure 3d), rising from ∼310 mm yr^{-1} in 2002 to ∼410 mm yr^{-1} by 2022. Because the product ingests time-varying MODIS LAI as a direct input to the Penman-Monteith equation, responding to the growing GGP canopy, the rising ET trend is consistent with the canopy-expansion mechanism and helps confirm the presence of a vegetation-driven signal in evapotranspiration.

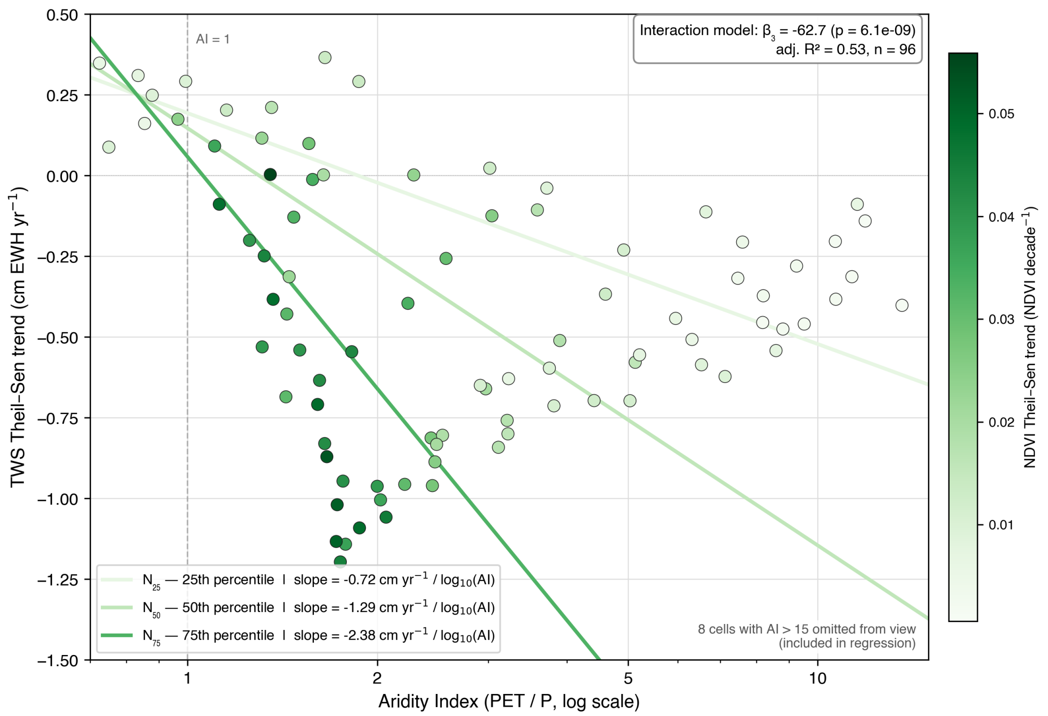

Figure 4. Effects-plot visualization of the n = 96 vegetation × aridity interaction regression. Points show all 96 Loess Plateau grid cells, colored by NDVI Theil-Sen trend (decade^{-1}); darker greens indicate stronger GGP-driven greening. Three solid lines show predicted TWS trends from the fitted interaction regression (Equation 3) evaluated at the 25th (N_{25}), 50th (N_{50}), and 75th (N_{75}) percentiles of NDVI trend, with corresponding slopes on $\log_{10}$AI of −0.72, −1.29, and −2.38 cm yr^{-1} per $\log_{10}$AI. The fanning of the three lines as aridity increases visualizes the interaction effect quantified in Table 2. The vertical dashed reference at AI = 1 marks the energy-limited / water-limited regime boundary; eight cells with AI > 15 are omitted from the plot view but included in the regression.

3.4 Aridity and vegetation interaction

The OLS interaction regression (Equation 3) across all 96 Loess Plateau grid cells yields the following coefficients presented in Table 2. The adjusted R^2 indicates that, combining simply the NDVI trend and the aridity index, the model explains 53% of the spatial variance in TWS depletion rates across the plateau.

The interaction coefficient $\beta_3 = -62.7$ shows that for each unit increase in $\log_{10}$ AI (a tenfold increase in aridity), the slope of TWS depletion on NDVI trend steepens by 62.7 cm yr^{-1} per NDVI decade^{-1}. In practical terms, a grid cell greening at +0.03 NDVI decade^{-1} in a sub-humid environment (AI = 1, $\log_{10}$ AI = 0) would experience a greening-attributable TWS contribution of approximately −0.15 cm yr^{-1} (from the $\beta_1 \cdot$ NDVI term alone), whereas the same greening rate in an arid cell (AI = 10, $\log_{10}$ AI = 1) would drive approximately −2.03 cm yr^{-1} of greening-attributable depletion (from $\beta_1 \cdot$ NDVI + $\beta_3 \cdot$ NDVI $\cdot \log_{10}$ AI), a c. 13-fold amplification of the per-unit greening cost.

Term	Estimate	Std. Error	t	p
Intercept (β_0)	+0.228	0.088	2.59	0.011
NDVI trend (β_1)	−5.08	3.13	−1.62	0.107
$\log_{10}$AI (β_2)	−0.294	0.097	−3.02	0.003
NDVI trend × $\log_{10}$AI (β_3)	−62.7	9.78	−6.41	<0.0001
Adj. $R^2 = 0.529$; $F(3, 92) = 36.5$; $p < 10^{-15}$ Residual SE = 0.27 cm yr^{-1}				

Table 2. OLS (see Eq. 3) interaction regression results (n = 96 grid cells). Dependent variable: deseasonalized TWS Theil-Sen trend (cm EWH yr^{-1}). NDVI trend in NDVI decade^{-1}; AI log-transformed (base 10).

Figure 4 visualizes this interaction by plotting predicted TWS trends from the fitted regression at three levels of NDVI greening (25th, 50th, and 75th percentiles of the per-cell NDVI trend distribution) across the full aridity range, overlaid on the 96 observed grid cells colored by their NDVI trend. The fanning of the three predicted lines as aridity increases is the visual expression of the interaction effect: at low aridity (AI < 1) the three lines converge near zero and vegetation intensity barely matters, while at high aridity (AI > 5) the lines spread by more than 2 cm yr^{-1} and vegetation intensity becomes the dominant predictor of TWS depletion. The dense cluster of high-NDVI cells in the AI = 1.5 to 2.5 range with the steepest observed depletion (approaching −1.2 cm yr^{-1}) suggests a non-linear semi-arid threshold where vegetation can establish, but precipitation cannot sustain mature canopies. The linear interaction term presented here captures the dominant aridity-conditioned effect and represents a conservative lower bound on the full aridity-mediated relationship.

3.5 Confounder analysis

Several non-vegetation processes could, in principle, contribute to the observed GRACE TWS signal. We assess their plausible cumulative contribution as a sensitivity envelope rather than a direct decomposition (Figure 5). Six confounders are considered: temperature-driven evaporative demand bounded by the Clausius-Clapeyron relation (Section 3.2), precipitation trend contribution bounded by the slope confidence interval, the combined hydrological impact of the Shanxi coal industry (mine drainage, subsidence-induced aquifer leakage, and solid mass extraction) following the coefficient method of (Xie et al., 2018), agricultural and non-coal industrial groundwater extraction (Li & Ma, 2024), post-2002 reservoir storage additions in the Loess Plateau basin (Y. Chen et al., 2024), and post-GGP residual rates of natural geomorphic erosion (S. Wang et al., 2016). Source citations and per-bar calculations are provided in the Supporting Information (Tables S1 and S2).

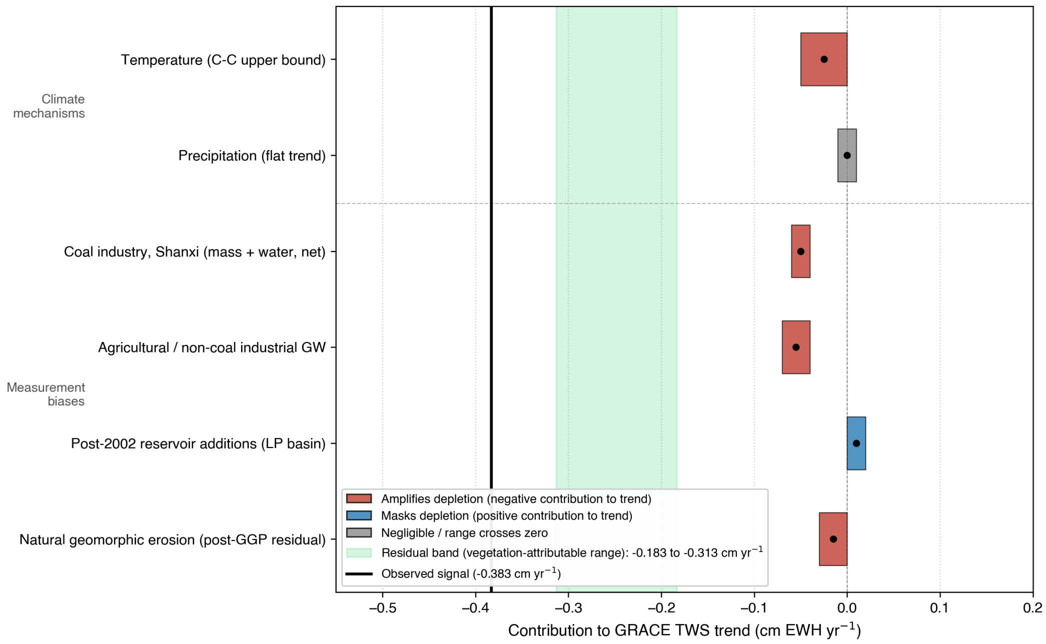

Figure 5. Sensitivity envelope of confounder contributions to the observed 2002–2024 TWS trend. Bars show literature-derived plausibility ranges of individual confounder contributions, grouped into climate mechanisms (top) and measurement biases / non-vegetation processes (bottom). Color encodes direction of contribution: red amplifies depletion, blue masks it, gray near-zero. The pale green band indicates the range that vegetation-driven processes must explain after accounting for confounder uncertainty, spanning from -0.18 cm yr^{-1} (if all confounders contribute at their amplifying maximum) to -0.31 cm yr^{-1} (if confounders contribute at their masking minimum). The observed GRACE signal (-0.383 cm yr^{-1}, solid black line) is shown for reference. See Supporting Information (Tables S1 and S2) for source citations and details of per-bar calculations.

The combined plausible range of all confounder contributions, summed across the six bars, spans -0.22 to -0.05 cm yr^{-1}. The observed TWS trend (-0.383 cm yr^{-1}) sits clearly outside this combined range, indicating that vegetation-driven processes must explain a residual gap of approximately -0.18 to -0.31 cm yr^{-1} depending on the confounder configuration. This range corresponds to between approximately 48% and 82% of the observed signal attributable to vegetation, with the most plausible interpretation falling toward the upper end of this range, given the temporal decoupling of TWS decline from year-to-year coal production fluctuations (see Supporting Information, Figure S1 for additional details).

4 Discussion

4.1 Physical mechanism: from GGP policy to GRACE signal

The coherence of four satellite-derived signals (stable precipitation, rising NDVI, rising satellite-detected ET, and declining GRACE TWS) points to a single physical mechanism summarized in Figure 6: GGP afforestation has grown forest canopies that are consuming water faster than the semi-arid climate can replenish. The physical chain runs as follows: as GGP plantations mature, their canopy LAI increases, thereby increasing the transpiration surface area ; in sub-humid cells, growing-season precipitation is sufficient to partially meet this demand but in arid and semi-arid cells, it is not. Trees in water-limited environments respond by extending roots deeper into the loess column, accessing the legacy moisture stored in the 2 to 10 m vadose zone over centuries to millennia of slow infiltration (Evaristo, 2026; H. Chen et al., 2008). This deep root uptake is the mechanism behind DSL formation, documented at the field scale across GGP species since 2008 (H. Chen et al., 2008; Liang et al., 2022) and synthesized recently by (Evaristo, 2026), who reviewed evidence from tritium and other isotopic archives showing that the water being accessed by GGP roots includes legacy paleomoisture stored in the deep vadose zone over multi-decadal to centennial timescales.

The MODIS ET trend of +4.63 mm yr^{-2} provides a mechanistic window into this process at the surface. Because the product ingests time-varying MODIS LAI as a direct Penman-Monteith input, it captures the canopy expansion that surface energy balance models without dynamic vegetation cannot detect. The Clausius-Clapeyron analysis demonstrates that warming can account for at most 11% of this ET increase, meaning the canopy expansion signal, not the thermodynamic signal, dominates. Our results indicate that GRACE detects the subsurface consequence (depleting storage) of the same biological process that the MODIS ET product detects at the surface (increasing transpiration).

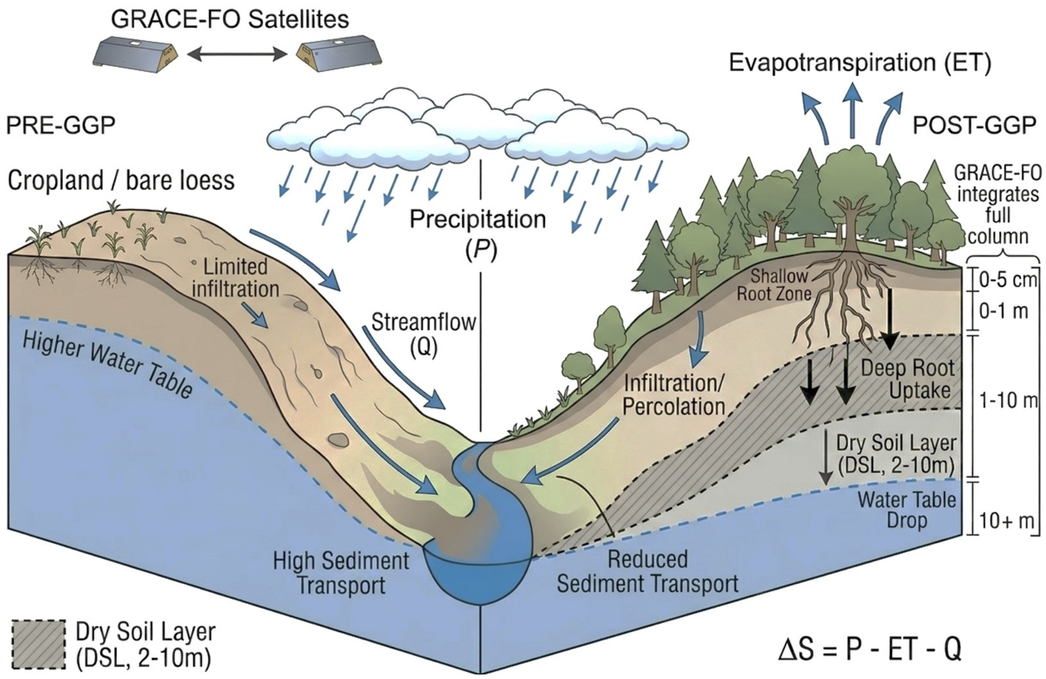

Figure 6. Mechanistic cross-section comparing pre-GGP (left) and post-GGP (right) hydrological states on the Loess Plateau. Under pre-GGP conditions, shallow-rooted cropland generates rapid overland runoff with the water table in approximate equilibrium. Under post-GGP afforestation, elevated ET through large leaf areas and deep roots forms a persistent Dry Soil Layer (DSL) at 2 to 10 m depth, while root uptake and reduced percolation lower the water table and reduce streamflow. The net result, $\Delta S = P - \mathrm{ET} - Q < 0$, accumulates over time and is detectable only by a column-integrating instrument such as GRACE(-FO). Vertical axis not to scale.

4.2 Aridity interaction

The central empirical contribution of this study is the formal quantification of the aridity × vegetation interaction on TWS depletion ($\beta_3 = -62.7$, $p < 0.0001$, adj. $R^2 = 0.53$). This result provides the first GRACE-scale, full-column confirmation of the NPP threshold concept developed by (Feng et al., 2016), who used satellite-observed NPP and a regional water balance to estimate that the Loess Plateau as a whole was approaching the maximum vegetation productivity sustainably supportable by regional precipitation, given concurrent human water demand. Where Feng et al. used a Budyko framework to infer this threshold from runoff data, we provide a direct, spatially resolved empirical test using 22 years of gravity satellite data and an interaction regression.

The field literature anticipated this finding at the plot scale. (H. Chen et al., 2008) showed DSL development at 1 to 3 m depth across vegetation types; (Liang et al., 2022) showed depletion deepened with plantation age and exhibited negligible precipitation-year recovery. (Ren et al., 2018) documented that afforestation is an appropriate hydrological option only where annual rainfall exceeds 500 mm, and (Evaristo, 2026) showed that deep soil water is predictably exhausted within 15 to 25 years of planting under water-intensive species. Our interaction regression bridges these plot-scale findings to the satellite scale: the β_3 coefficient quantifies, across 96 satellite grid cells spanning an aridity gradient from AI = 0.52 to 31.35, the rate at which the same vegetation increase produces disproportionately greater water storage loss as aridity increases.

The dense cluster of high-NDVI cells in the AI = 1.5 to 2.5 range visible in Figure 4, with the steepest observed depletion approaching -1.2 cm yr^{-1}, suggests a non-linear "Goldilocks" threshold at the semi-arid transition where vegetation can establish but precipitation cannot sustain mature canopies. Cells drier than this band greened weakly (vegetation could not establish) and depleted modestly. Cells wetter than this band greened strongly but with smaller per-unit depletion costs (precipitation partially buffered the canopy demand). The linear interaction term presented here captures the dominant aridity-conditioned effect across this full gradient and represents a conservative lower bound on the aridity-mediated relationship. Formal modeling of the non-monotonicity is a direction for future work.

Globally, (Liu, Li, Wang, & Zhou, 2023) showed that terrestrial water storage in greening drylands correlates negatively with vegetation growth across 1982 to 2016, with the response pronounced where vegetation productivity growth rates are high. Our study advances this picture in three ways: we used the complete GRACE + GRACE-FO record (2002 to 2024), we applied explicit climate controls that ruled out precipitation and temperature as primary drivers, and we quantified the aridity-conditioned amplification through an interaction term rather than reporting average correlations. The consistency between our satellite-scale result and the global dryland pattern of (Liu, Li, Wang, & Zhou, 2023) strengthens the generalizability of the finding beyond the Loess Plateau.

4.3 Where afforestation can be sustained: a spatial planning framework

The greening signal documented across the Loess Plateau is real, and so are the co-benefits of the effort: documented sediment reduction (S. Wang et al., 2016), flood mitigation (Trabucco et al., 2008), regional carbon sequestration (Bryan et al., 2018), and improved rural livelihood indicators (Delang & Yuan, 2015) are all attributable to GGP. The findings here do not contest those benefits. Instead, they refine the framework under which they can be reliably and sustainably delivered.

Afforestation can be hydrologically self-sustaining where precipitation is sufficient: the southwestern margin of the Loess Plateau (Figure 2, AI < 1) shows modest TWS gains coincident with vegetation growth, consistent with cells where local rainfall meets canopy demand. It is hydrologically costly where precipitation is insufficient: the semi-arid central and eastern plateau (AI = 1.5 to 3) shows the steepest depletion per unit of greening, with cells losing more than 1 cm yr^{-1} of stored water to support new canopy. And it is hydrologically irrelevant where vegetation cannot establish: the driest north-western cells (AI > 5) greened weakly and depleted only modestly. The aridity index is a satellite-available, spatially explicit screening criterion that can distinguish these three regimes before planting decisions are made. Figure 7 demonstrates how the combined use of AI and NDVI data can produce a predictive tool for TWS (Figure 7a), consistent with observations (Figure 7b and c).

The policy stakes are significant: the Bonn Challenge targets 350 million hectares of restoration by 2030, a substantial fraction of which will occur in semi-arid and arid regions, where water scarcity already constrains food security and ecosystem function (Bonn Challenge, 2011). The UN Decade on Ecosystem Restoration explicitly targets degraded drylands (United Nations Environment Programme, 2021). For investments at this scale, the question is not whether to plant but where. Aridity-conditioned planning that prioritizes the wettest available sites within a given dryland region, or that selects shallow-rooted native grasses and shrubs (F. Wang et al., 2020) over deep-rooted plantation forest in semi-arid zones, can deliver carbon, erosion, and biodiversity co-benefits without imposing the legacy moisture cost we document here (Ren et al., 2018; Evaristo, 2026).

4.4 Potential impact of coal mining in Shanxi Province

The confounder analysis (Section 3.5, Figure 5) demonstrates that the combined plausible contribution of all non-vegetation processes to the observed TWS trend spans -0.22 to -0.05 cm yr^{-1}, leaving a residual range of -0.18 to -0.31 cm yr^{-1} that vegetation must explain. Among non-vegetation contributors, the Shanxi coal industry is one of the largest and most discussed candidates; however, two lines of evidence argue against coal as a primary driver.

First, the spatial pattern of TWS depletion extends well beyond Shanxi, with strong depletion across the central plateau, where coal mining is minimal. The eastern portion of Shanxi Province—including major coal-producing cities such as Datong, Yangquan, Changzhi, and Jincheng—lies east of our 112°E boundary and is outside our analysis domain. Our confounder estimate, therefore, likely overestimates the per-cell coal contribution to our western Shanxi cells, since much of Shanxi's coal production occurs east of the boundary; GRACE's approximately 300 km native footprint produces some leakage from these eastern operations into our western cells, but the per-cell impact is likely smaller than the proportional assumption would suggest.

Second, the temporal pattern of TWS decline is decoupled from known year-to-year coal production fluctuations. Despite notable production reductions during periods of national coal supply-side structural reform, the deseasonalized TWS decline in Shanxi-overlapping cells proceeded smoothly through these periods (Supporting Information Figure S1). If coal mining were the dominant driver, year-to-year production variability would be expected to produce detectable TWS inflection points; it does not.

4.5 Potential limitations

Several limitations warrant acknowledgment. The GRACE record begins in 2002, three years after GGP initiation, meaning we cannot establish a pre-GGP TWS baseline directly from satellite observations. However, the GGP mechanism—deep root moisture access—requires canopy closure and mature root systems; in 2002, the earliest GGP plantations were only 3 years old, before significant DSL formation would have begun, making 2002 a defensible near-baseline for the depletion signal. The MODIS ET product shares upstream MODIS spectral inputs with the NDVI product, meaning the rising ET trend is interpreted as a mechanistic consistency check rather than a fully independent evidence; future work using Landsat-derived NDVI as an independent vegetation source will address this circularity. The reservoir correction was estimated from published storage ranges rather than implemented from altimetry time series and represents a direction for refinement. Finally, GRACE's ∼300 km footprint limits spatial resolution to the regression's 1° analysis grid, precluding watershed-scale attribution.

The linear interaction model presented here captures the dominant aridity-conditioned effect, but does not formally represent the non-monotonic "Goldilocks" structure visible in Figure 4. The linear estimate is therefore a conservative lower bound on the aridity-mediated relationship, not an overstatement; readers concerned about model misspecification can interpret our results as the smallest plausible effect size rather than the largest.

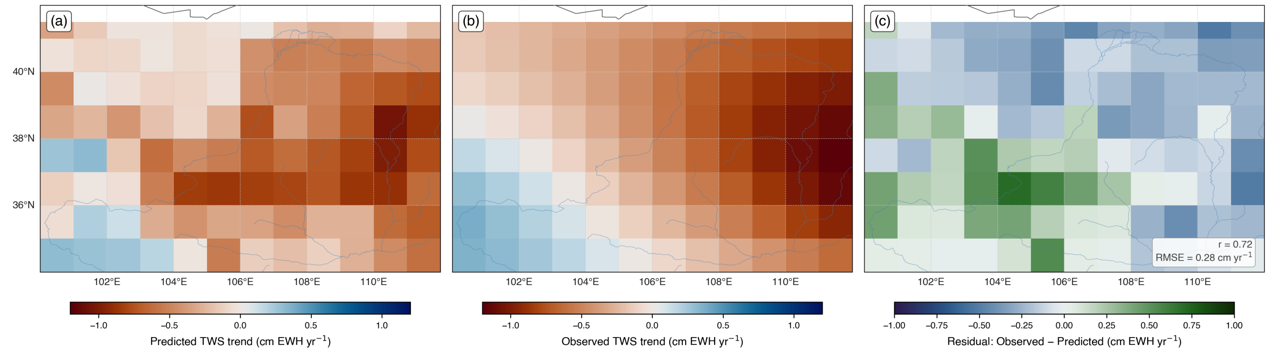

Figure 7. Spatial demixing of the vegetation-driven TWS signal from confounders. (a) Predicted TWS Theil-Sen trend from the n = 96 interaction regression at 0.5° display resolution, upsampled from the 1° fitted grid by nearest-neighbor interpolation. (b) Observed TWS Theil-Sen trend from GRACE/GRACE-FO. (c) Residual (Observed minus Predicted), with Pearson $r = 0.72$ and RMSE = 0.28 cm yr^{-1} across all 96 cells. The mean residual in Shanxi-overlapping cells (n = 12) is −0.20 cm yr^{-1}, nearly seven times the non-Shanxi mean of −0.03 cm yr^{-1}. Negative residuals (blue-teal) in Shanxi indicate that observed depletion exceeds the vegetation-aridity model's prediction, consistent with the known coal mining mass-loss signal documented in Figure 5. The model captures 53% of the spatial variance in TWS depletion; the remaining variance concentrates spatially in regions consistent with documented non-vegetation processes.

4.6 Future directions

The aridity × vegetation interaction identified here opens several avenues for follow-on investigation. Formal modeling of the non-monotonic "Goldilocks" structure visible at high NDVI in Figure 4 would refine the linear estimate presented here and enable more precise spatial targeting. Independent NDVI from Landsat would remove the MODIS circularity concern. Perhaps most importantly, testing whether the β_3 interaction generalizes to other dryland afforestation programs globally (the Sahel Great Green Wall, India's afforestation targets, Central Asian steppe restoration) would establish whether the aridity-conditioned water debt is a universal hydrological constraint or a Loess Plateau-specific finding, driven by geology and soil type as well as plant species selection, for example. The GGP is currently the only dryland afforestation program of sufficient scale and duration to make this analysis feasible; emerging satellite gravity records and expanding Landsat archives may make global tests possible in the future.

5 Conclusions

The Loess Plateau has lost -0.38 cm EWH yr^{-1} of total water storage over 2002 to 2024 (-8.4 cm cumulative, Mann-Kendall $p < 0.0001$), concentrated in the eastern plateau where GGP afforestation is most intensive. This decline is not explained by precipitation trends ($p = 0.59$) or temperature-driven evaporation (up to 10.8% of the observed ET increase by Clausius-Clapeyron). NDVI greening of $+0.0217$ NDVI decade^{-1} ($p < 0.001$) and a MODIS ET increase of $+4.63$ mm yr^{-2} ($p < 0.001$) together identify vegetation expansion as the primary driver. An explicit vegetation $\times$ aridity interaction regression ($\beta_3 = -62.7$, $p < 0.0001$, adj. $R^2 = 0.53$) demonstrates that the same degree of NDVI greening causes disproportionately greater TWS depletion in drier grid cells, providing the first satellite-scale quantification of the aridity-conditioned water debt of afforestation and confirming at the continental scale the water carrying capacity threshold theorized by prior field and modeling studies.

These results carry direct implications for global reforestation policy. Afforestation delivers documented co-benefits, including carbon sequestration, flood mitigation, soil stabilization, biodiversity recovery, and rural livelihood improvements, all attributable to GGP at scale. The findings here do not contest those benefits but instead refine the framework under which they can be reliably delivered. The aridity index distinguishes three regimes across the Loess Plateau: wet cells where afforestation is hydrologically self-sustaining; semi-arid cells where vegetation can establish, but precipitation cannot sustain mature canopies, accumulating a water debt against deep legacy moisture reserves, and dry cells where vegetation cannot establish in the first place. As China expands GGP investment through 2050 and as international commitments under the Bonn Challenge and UN Decade on Ecosystem Restoration direct billions of dollars toward dryland afforestation, the spatial distribution of these interventions, and the choice of vegetation structures matched to local water availability, will determine whether they achieve their goals sustainably or accelerate the depletion of subsurface water reserves that communities and ecosystems depend upon. The aridity index provides a practical, satellite-available criterion for targeting afforestation where it can be sustained and for guiding species selection where it cannot.

Inclusion in global research statement

This study uses publicly available satellite remote sensing datasets to investigate hydrological dynamics on the Loess Plateau, China. The Loess Plateau boundary shapefile was obtained from the National Tibetan Plateau Data Center (NTPDC), an open-access repository maintained by Chinese institutions. The authors did not conduct fieldwork, collect new ground-based data, or engage local stakeholders directly. We acknowledge the foundational contributions of Chinese hydrology research to understanding the Loess Plateau hydrological system, which we have cited extensively throughout this work.

Data availability statement

All datasets used in this study are free and publicly available: i) GRACE/GRACE-FO CSR RL06M mascons (Save et al., 2016; Save, 2020) are available from the University of Texas, Austin (http://www2.csr.utexas.edu/grace), ii) CHIRPS v2.0 (Funk et al., 2015) is directly available from the University of Santa Barbara (https://data.chc.ucsb.edu/products/CHIRPS-2.0/), iii) MODIS MOD16A3GF v061 evapotranspiration product (Running et al., 2021) is available from NASA's EarthData platform (https://doi.org/10.5067/MODIS/MOD16A3GF.061), as is the MODIS MOD13C2 NDVI product (Didan, 2021) (https://doi.org/10.5067/MODIS/MOD13C2.061), iv) TerraClimate (Abatzoglou et al., 2018a, 2018b) is available from the developers' page at the University of California, Merced (https://www.climatologylab.org/terraclimate.html), v) the monthly ERA5-Land 2 m air temperature data (Muñoz Sabater, J. , 2019) is available from the Copernicus Climate Data Store (https://doi.org/10.24381/cds.68d2bb30), and vi) the monthly coal production data for the Shanxi province (National Bureau of Statistics of China, 2024) is available from the National Bureau of Statistics of China (https://data.stats.gov.cn/).

Conflict of interest disclosure

The authors declare there are no conflicts of interest for this manuscript.

References

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018a). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, *5*, 170191. doi: 10.1038/sdata.2017.191

Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018b). *TerraClimate: Monthly climate and climatic water balance for global terrestrial surfaces [Dataset].* Climatology Lab, University of Idaho. Retrieved from `https://www.climatologylab.org/terraclimate.html`

Allen, M. R., & Ingram, W. J. (2002). Constraints on future changes in climate and the hydrologic cycle. *Nature*, *419*, 224–232. doi: 10.1038/nature01092

Bonn Challenge. (2011). *The Bonn Challenge: A global effort to restore degraded and deforested lands.* IUCN, Gland, Switzerland. Retrieved from `https://www.bonnchallenge.org`

Bryan, B. A., Gao, L., Ye, Y., Sun, X., Connor, J. D., Crossman, N. D., ... Hou, X. (2018). China's response to a national land-system sustainability emergency. *Nature*, *559*(7713), 193–204. doi: 10.1038/s41586-018-0280-2

Chen, H., Shao, M., & Li, Y. (2008). Soil desiccation in the Loess Plateau of China. *Geoderma*, *143*(1–2), 91–100. doi: 10.1016/j.geoderma.2007.10.013

Chen, Y., Wang, Y., Li, L., Cui, Y., Duan, X., & Long, D. (2024). Monthly Monitoring of Inundated Areas and Water Storage Dynamics in China's Large Reservoirs Using Multisource Remote Sensing. *Water Resources Research*, *60*(8), e2023WR036450. doi: 10.1029/2023WR036450

Delang, C. O., & Yuan, Z. (2015). *China's grain for green program: A review of the largest ecological restoration and rural development program in the world.* Cham, Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-11505-4

Didan, K. (2021). *MODIS/Terra vegetation indices monthly L3 global 0.05deg CMG V061 [Dataset].* NASA Land Processes Distributed Active Archive Center. doi: 10.5067/MODIS/MOD13C2.061

Evaristo, J. (2026). Left high and dry: Deep soil water depletion and ecohydrological resilience on China's Loess Plateau. *WIREs Water*, *13*(1), e70053. doi: 10.1002/wat2.70053

Feng, X., Fu, B., Piao, S., Wang, S., Ciais, P., Zeng, Z., ... Wu, B. (2016). Revegetation in China's Loess Plateau is approaching sustainable water resource limits. *Nature Climate Change*, *6*(11), 1019–1022. doi: 10.1038/nclimate3092

Fu, B., Wang, S., Liu, Y., Liu, J., Liang, W., & Miao, C. (2017). Hydrogeomorphic ecosystem responses to natural and anthropogenic changes in the Loess Plateau of China. *Annual Review of Earth and Planetary Sciences*, *45*(1), 223–243.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., ... Michaelsen, J. (2015). *Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) version 2.0 [Dataset].* UC Santa Barbara Climate Hazards Center. Retrieved from `https://data.chc.ucsb.edu/products/CHIRPS-2.0/` doi: 10.15780/G2RP4Q

Gerlein-Safdi, C., Keppel-Aleks, G., Wang, F., Frolking, S., & Mauzerall, D. L. (2020). Satellite monitoring of natural reforestation efforts in China's drylands. *One Earth*, *2*(1), 98–108.

Hasegawa, T., Fujimori, S., Ito, A., & Takahashi, K. (2024). Careful selection of forest types in afforestation can increase carbon sequestration by 25% without compromising sustainability. *Communications Earth & Environment*, *5*(1), 171.

Kendall, M. G. (1975). *Rank Correlation Methods* (4th ed.). London: Griffin.

Kulik, K., Belyaev, A., & Pugacheva, A. (2023). The role of protective afforestation in drought and desertification control in agro-landscapes. *Arid ecosystems*, *13*(1), 1–10.

Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., ... Yuan, D.-N. (2020). Extending the global mass change data record: GRACE Follow-On instrument and science data performance. *Geophysical Research Letters*, *47*(12), e2020GL088306. doi: 10.1029/2020GL088306

Li, J., & Ma, J. (2024). Evaluating the dynamics of groundwater storage and its sustainability in the Loess Plateau: The integrated impacts of climate change and human activities. *Remote Sensing*, *16*(23), 4375. doi: 10.3390/rs16234375

Liang, H., Xue, Y., Li, Z., Gao, G., & Liu, G. (2022). Afforestation may accelerate the depletion of deep soil moisture on the Loess Plateau: Evidence from a meta-analysis. *Land Degradation & Development*, *33*(18), 3829-3840. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/ldr.4426` doi: https://doi.org/10.1002/ldr.4426

Liu, K., Li, X., Wang, S., & Zhang, X. (2023). Unrevealing past and future vegetation restoration on the Loess Plateau and its impact on terrestrial water storage. *Journal of Hydrology*, *617*, 129021. doi: 10.1016/j.jhydrol.2022.129021

Liu, K., Li, X., Wang, S., & Zhou, G. (2023). Past and future adverse response of terrestrial water storages to increased vegetation growth in drylands. *npj Climate and Atmospheric Science*, *6*, 113. doi: 10.1038/s41612-023-00437-9

Lv, M., Ma, Z., Li, M., & Zheng, Z. (2019). Quantitative analysis of terrestrial water storage changes under the Grain for Green Program in the Yellow River Basin. *Journal of Geophysical Research: Atmospheres*, *124*(3), 1336–1351. doi: 10.1029/2018JD029113

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*, *13*(3), 245–259. doi: 10.2307/1907187

Mu, Q., Zhao, M., & Running, S. W. (2011). Improvements to a MODIS global terrestrial evapotranspiration algorithm. *Remote Sensing of Environment*, *115*(8), 1781–1800. doi: 10.1016/j.rse.2011.02.019

Muñoz Sabater, J. . (2019). *ERA5-Land monthly averaged data from 1950 to present [Dataset].* Copernicus Climate Change Service (C3S) Climate Data Store (CDS). doi: 10.24381/cds.68d2bb30

National Bureau of Statistics of China. (2024). *Monthly coal production statistics, Shanxi Province [dataset].* National Bureau of Statistics of China. Retrieved from `https://data.stats.gov.cn/`

National Tibetan Plateau Data Center. (2018). *Boundary data of the Loess Plateau [dataset].* Retrieved from `https://data.tpdc.ac.cn/`

Nilsson, S., & Schopfhauser, W. (1995). The carbon-sequestration potential of a global afforestation program. *Climatic change*, *30*(3), 267–293.

Oraon, P. R., Sagar, V., & Beauty, K. (2023). Ecological restoration of degraded land through afforestation activities. In *Land and environmental management through forestry* (p. 201-216). John Wiley & Sons, Ltd. doi: 10.1002/9781119910527.ch8

Ren, Z., Li, Z., Liu, X., Li, P., Cheng, S., & Xu, G. (2018). Comparing watershed afforestation and natural revegetation impacts on soil moisture in the semiarid Loess Plateau of China. *Scientific Reports*, *8*(1), 2972.

Retrieved from https://doi.org/10.1038/s41598-018-21362-5 doi: 10.1038/s41598-018-21362-5

Running, S., Mu, Q., Zhao, M., & Moreno, A. (2021). *MODIS/Terra net evapotranspiration Gap-Filled yearly L4 global 500m SIN grid V061 [Dataset].* NASA Land Processes Distributed Active Archive Center. doi: 10.5067/MODIS/MOD16A3GF.061

Save, H. (2020). *CSR GRACE and GRACE-FO RL06 Mascon Solutions v02 [Dataset].* University of Texas, Austin. Retrieved from http://www2.csr.utexas.edu/grace doi: 10.15781/cgq9-nh24

Save, H., Bettadpur, S., & Tapley, B. D. (2016). High-resolution csr grace rl05 mascons. *Journal of Geophysical Research: Solid Earth*, *121*(10), 7547-7569. doi: https://doi.org/10.1002/2016JB013007

Scheliga, A., Girotto, M., & Gerlein-Safdi, C. (2025). Anthropogenic processes and sea-level rise contributions with data assimilation of GRACE and GRACE-FO terrestrial water storage. *Journal of Hydrology*, 134048.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389. doi: 10.1080/01621459.1968.10480934

Sun, C., Zhu, L., Liu, Y., Wei, T., & Guo, Z. (2022). CMIP6 model simulation of concurrent continental warming holes in Eurasia and North America since 1990 and their relation to the Indo-Pacific SST warming. *Global and Planetary Change*, *213*, 103824.

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, *305*(5683), 503–505. doi: 10.1126/science.1099192

Trabucco, A., Zomer, R. J., Bossio, D. A., van Straaten, O., & Verchot, L. V. (2008). Climate change mitigation through afforestation/reforestation: A global analysis of hydrologic impacts with four case studies. *Agriculture, Ecosystems and Environment*, *126*(1–2), 81–97. doi: 10.1016/j.agee.2008.01.014

United Nations Environment Programme. (2021). *UN Decade on Ecosystem Restoration 2021–2030.* UNEP, Nairobi, Kenya. Retrieved from https://www.decadeonrestoration.org

Wang, F., Pan, X., Gerlein-Safdi, C., Cao, X., Wang, S., Gu, L., … Lu, Q. (2020). Vegetation restoration in Northern China: A contrasted picture. *Land Degradation & Development*, *31*(6), 669–676.

Wang, F., Zhao, X., Gerlein-Safdi, C., Mu, Y., Wang, D., & Lu, Q. (2017). Global sources, emissions, transport and deposition of dust and sand and their effects on the climate and environment: a review. *Frontiers of Environmental Science & Engineering*, *11*(1), 13.

Wang, S., Fu, B., Piao, S., Lü, Y., Ciais, P., Feng, X., & Wang, Y. (2016). Reduced sediment transport in the Yellow River due to anthropogenic changes. *Nature Geoscience*, *9*(1), 38–41. doi: 10.1038/ngeo2602

Wang, Z., Bi, Y., Yang, F., Zheng, J., Yang, Y., & Zhang, S. (2025). Research of spatial-temporal variation and correlation of water storage and vegetation coverage in the Loess Plateau. *Remote Sensing*, *17*(15), 2667. doi: 10.3390/rs17152667

Wu, W., Zhou, J., Niu, J., & Lv, H. (2021). Study on coupling between mineral resources exploitation and the mining ecological environment in Shanxi Province. *Environment, Development and Sustainability*, *23*(9), 13261–13283. doi: 10.1007/s10668-020-01209-8

Xie, X., Xu, C., Wen, Y., & Li, W. (2018). Monitoring groundwater storage changes in the Loess Plateau using GRACE satellite gravity data, hydrological models and coal mining data. *Remote Sensing*, *10*(4), 605. doi: 10.3390/rs10040605

Zhong, B., Li, Q., Chen, J., Luo, Z., & Zhou, H. (2020). Improved estimation of regional surface mass variations from GRACE intersatellite geopotential differences using a priori constraints. *Remote Sensing*, *12*(16), 2553. doi: 10.3390/rs12162553