

Source-Audited Agricultural Input-Carbon and Irrigation-Water Productivity across China: Hydroclimatic Associations and Hypothetical Carbon-Tax Burdens

Aoqi Xu

Fujian Agriculture and Forestry University

Shan Miao

Shantou University

Zhiming Guo

zhiming_guo2@outlook.com

Wenzhou University of Technology

Article

Keywords: Agricultural resource productivity, Hydroclimatic variability, Ecological informatics, Data provenance, Reproducible data integration, Hypothetical carbon-price scenario

Posted Date: September 17th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10880213/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Source-Audited Agricultural Input-Carbon and Irrigation-Water Productivity across China: Hydroclimatic Associations and Hypothetical Carbon-Tax Burdens

Aoqi Xu¹, Shan Miao², Zhiming Guo^{3*}

¹School of Economics and Management, Fujian Agriculture and
Forestry University, Fuzhou, 350002, China.

²School of Continuing Education, Shantou University, Shantou, 515061,
China.

^{3*}School of Economics and Management, Wenzhou University of
Technology, Wenzhou, 325027, China.

*Corresponding author(s). E-mail(s): zhiming-guo2@outlook.com;

Abstract

Integrating agricultural statistics with gridded water-use and hydroclimatic products requires harmonized spatial support, units, calendars, missingness rules, and source versions. We constructed a province-year panel for 31 mainland Chinese units during 2010–2022. Irrigation-water productivity (IWP) and hydroclimate were available for 403 observations; agricultural input-carbon productivity (AICP) was available for 310 observations during 2010–2019 because agricultural diesel data were unavailable thereafter. In 2020, reconstructed irrigation withdrawals equaled 95.1% of official agricultural water use and showed strong provincial concordance (Pearson $r = \mathbf{0.998}$; mean absolute percentage error = $\mathbf{9.03\%}$). Two-way fixed-effects estimates for climatic water availability and SPEI12 were small, with all four 95% confidence intervals spanning zero and adjusted $q = \mathbf{0.962}$. Quadratic and dry-context interaction blocks were jointly nonsignificant for both outcomes. We found no evidence that the hydroclimatic coefficients differed before and after the 2014 irrigation-product level shift (joint $p = \mathbf{0.766}$). Hypothetical carbon prices produced median full-liability burdens of 0.526%, 0.634%, and 0.698% of real agricultural output. These are accounting scenarios rather than policy-effect estimates. The results show that

contemporaneous hydroclimatic variables add little explanatory power for within-province variation in these productivity ratios, while the validated panel supports reproducible monitoring and finer-scale analysis.

Keywords: Agricultural resource productivity, Hydroclimatic variability, Ecological informatics, Data provenance, Reproducible data integration, Hypothetical carbon-price scenario

1 Introduction

Agricultural production is constrained by the joint need to limit input-related carbon emissions and use irrigation water productively. These dimensions are related but not interchangeable: a province may generate more real agricultural output per unit of accounted input carbon without improving output per unit of irrigation water, or vice versa. Evaluating both outcomes therefore gives a more disciplined account of agricultural resource-use performance than a single composite index [1–3].

The information problem is substantial. Agricultural statistics are reported for administrative units and often combine current-price values with physical activities. Water-use products are gridded volumetric fields, whereas precipitation and potential evapotranspiration are gridded rates that require calendar-aware conversion. Drought indices add a third temporal structure and their own fill-value conventions. Directly joining these sources without documented spatial allocation, unit harmonization, coverage checks, and versioned provenance can produce apparently complete but irreproducible indicators.

China is an informative setting because agricultural production and hydroclimatic conditions vary sharply among provinces. The precipitation-to-potential-evapotranspiration ratio captures differences in climatic water availability, while the Standardized Precipitation Evapotranspiration Index describes anomalous dryness relative to local conditions. The two quantities are complementary: aridity reflects the moisture balance, whereas drought records departures from the local climate. Keeping them separate preserves their physical interpretation.

Ecological informatics provides a framework for integrating heterogeneous environmental observations with transparent data management, analysis, and information processing [4]. The FAIR principles extend this concern to the findability, accessibility, interoperability, and reusability of data and associated digital objects [5]. Environmental research infrastructures and ecological-monitoring workflows have operationalized these principles through harmonized data models, machine-readable metadata, source-aware quality control, and reproducible processing [6–9]. These practices are particularly relevant when official administrative statistics are combined with gridded environmental products that differ in spatial support, units, calendars, and coverage.

Provincial studies in China have examined agricultural water–land–carbon relationships, water–land–energy–emissions systems, agricultural eco-efficiency, and irrigation-related carbon-reduction potential [10–13]. The present study addresses

a complementary methodological need: constructing separate output-to-carbon and output-to-water productivity ratios from traceable source data and linking them to continuous hydroclimatic measures without combining their distinct denominators into a composite index.

The study has four objectives. First, it constructs a province-year analytical panel from official agricultural records, reconstructed irrigation withdrawals, ERA5 hydroclimate, and ERA5-derived drought information. Second, it evaluates source coverage, denominator validity, unit consistency, missingness, and product discontinuities. Third, it estimates within-province associations between the two productivity ratios and climatic water availability and SPEI12, including low-order curvature, continuous dry-context interactions, and prespecified sensitivity analyses. Fourth, it translates the observed input-carbon inventory into a static accounting scenario using externally observed carbon-price benchmarks.

The resulting contribution is a reproducible, versioned source-to-output workflow for two explicitly defined resource-productivity outcomes. Stable spatial keys, calendar-aware conversion, polygon-overlap allocation, source hashes, unit-consistent carbon accounting, missingness-preserving joins, and script-generated inference connect the original inputs to the reported results. The estimands are conditional associations and accounting scenarios; causal responses and observed policy effects lie outside their scope.

2 Study Area, Data Architecture, and Evidence Status

The study area comprises 31 provincial-level units in mainland China observed annually from 2010 through 2022: Beijing, Tianjin, Hebei, Shanxi, Inner Mongolia, Liaoning, Jilin, Heilongjiang, Shanghai, Jiangsu, Zhejiang, Anhui, Fujian, Jiangxi, Shandong, Henan, Hubei, Hunan, Guangdong, Guangxi, Hainan, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, Ningxia, and Xinjiang. Hong Kong, Macao, and Taiwan are excluded because they do not belong to the common mainland statistical framework. The validated hydrological and agricultural files each realize the complete 403-key province-year design. Their one-to-one merge supports 403 IWP records and 310 AICP records at observed source coverage.

China’s provinces differ substantially in precipitation, evaporative demand, drought exposure, and irrigation reliance. This heterogeneity motivates a province-level hydroclimatic assessment centered on the precipitation-to-potential-evapotranspiration ratio and SPEI12. Unsupported groundwater, renewable-water-stress, and irrigation-energy proxies are outside the focused design.

2.1 Source-Audited Data Architecture

The data pipeline comprises four controlled layers (Supplementary Table S1). Original source files are preserved unchanged and documented by relative path, byte size, acquisition timestamp, source identifier, and SHA-256 hash. Source-specific transformations generate separately validated province-year files, which are merged only after their spatial, temporal, unit, coverage, and missingness checks have passed. All manuscript

statistics, tables, and figures are generated from the resulting analytical panel. This layered design provides traceability from each reported result to its source records and processing rules.

The initialized data contracts are:

- `province_hydrology_2010_2022.csv`;
- `province_agriculture_2010_2022.csv`;
- `hydro_ag_resource_productivity_panel_2010_2022.csv`; and
- the excluded `province_cet_2010_2022.csv` exposure contract.

Three machine-readable metadata files document the analytical architecture. The source manifest records the identifiers, acquisition details, and hashes of 13 source objects. The province crosswalk maps the boundary labels to canonical provincial identifiers, retains 31 mainland units, and records all label harmonizations and exclusions. The variable dictionary defines the keys, inputs, derived variables, units, transformations, missing-data rules, producing scripts, and analytical status of each field. Candidate carbon-market variables that did not meet the province-year coverage requirements are documented separately and are not included in the analytical panel.

Figure 1 summarizes the implemented source-to-output sequence and the quality-control checks applied before analysis.

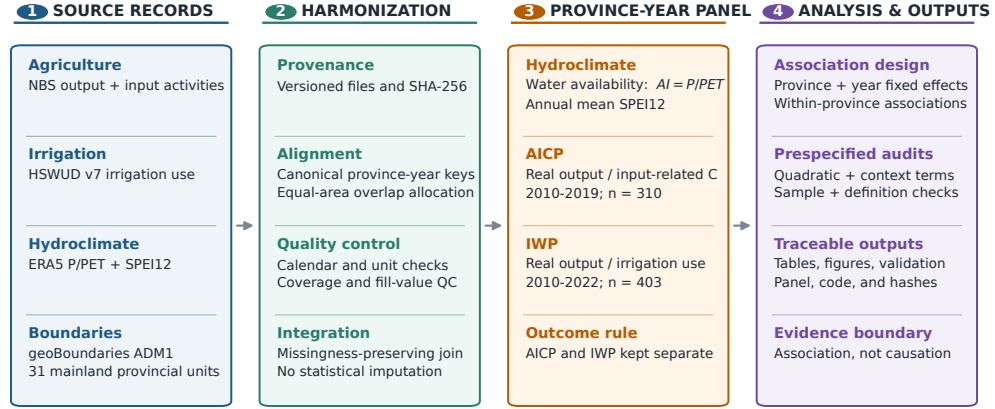


Figure 1 Source-audited province-year integration workflow. Immutable source objects are harmonized to fixed mainland provincial units, common units, and annual calendars. Block-specific products then undergo explicit key, coverage, unit, missingness, and provenance checks before the frozen analytical panel generates descriptive and hydroclimatic-association outputs.

2.2 Verification Status of Data Blocks

Supplementary Table S2 distinguishes file-level inspection from analytical readiness. *File-audited* means that the local object, hash, variables or contents, temporal coverage, and geographic coverage have been inspected. *Constructed and validated*

additionally requires the declared spatial, temporal, unit, missingness, and diagnostic checks to pass. *Audited and excluded* means that a proposed data role failed a prespecified feasibility requirement and is excluded from the analytical panel.

2.3 Lineage, Missingness, and Sample-Formation Rules

The hydrological and agriculture files each realize all 403 intended province–year combinations and have no duplicate keys. Every constructed hydrological value is present. In the agriculture file, real output and IWP are complete, whereas the six-source carbon denominator and AICP are observed for 310 rows. The merge script requires identical `province_id–year` support, matching province labels, unique keys, the intended 31-by-13 design, and agreement between the independently carried irrigation-water fields. Any failed check stops the build. The analytical panel preserves outcome-specific completeness and record-status fields.

Real agricultural output, input-related carbon emissions, reconstructed irrigation-water use, AICP, and IWP are not filled by interpolation, extrapolation, cross-sectional means, forward or backward carry, or statistical or machine-learning imputation. A missing primary value remains missing. Accordingly, any AICP analysis based on the six-source boundary cannot extend beyond 2019 unless the missing official activity records are recovered and separately audited. Alternative missing-data analyses must be identified as sensitivities and cannot redefine the analytical variable.

Every analytical row retains source-record identifiers and missingness flags. Every derived field is traceable to the source hash, original unit, conversion, spatial or temporal aggregation rule, transformation, and producing-script version. A change in source definition, boundary version, conversion rule, or input hash creates a new data version and invalidates downstream outputs until they are regenerated.

2.4 Hydrological Reconstruction and Validation

The hydrological block was generated by a single archived MATLAB build from the Figshare HSWUD version 7 file `HSWUD_irr_v20260711.nc` [14, 15], ERA5 monthly precipitation and potential evaporation [16, 17], the deterministic ERA5–Drought version 1.0 SPEI12 product with a generalized-logistic fit and 1991–2020 reference period [18, 19], and geoBoundaries object `CHN-ADM1-43563684` [20, 21]. HSWUD is a reconstructed gridded product, not a record of reported provincial withdrawal. The build retained 31 mainland units and excluded Hong Kong, Macao, and Taiwan. The geoBoundaries object uses the nonstandard label “Guangzhou Province” for the Guangdong provincial unit; the crosswalk records the original label and harmonizes it to Guangdong before spatial aggregation. Polygon and grid-cell geometries were projected to the EPSG:6933 equal-area coordinate system. Province weights were calculated from polygon–grid intersections; cells intersecting a boundary were allocated by their within-province overlap fraction.

The HSWUD variable `irr` is expressed as 10^8 m^3 per grid-month on a 0.1° grid. Values below -10^{20} were treated as fill values. Valid cell volumes were multiplied by their polygon-overlap fractions, summed within each month, summed across calendar

months, and converted to m³. Missing product areas were not rescaled to a full polygon, so the file preserves the minimum geometric coverage as a diagnostic rather than introducing an implicit imputation.

ERA5 precipitation and potential-evaporation rates were multiplied by the exact number of seconds in each calendar month before annual aggregation. Inspection showed that 90.12% of valid `avg_pevr` values were negative, consistent with the ERA5 upward-flux convention; the sign was therefore reversed before forming positive PET depth. Monthly provincial values were equal-area weighted means over intersecting grid cells, with a minimum 95% spatial-support rule. Annual precipitation and PET are the sums of the 12 monthly depths. The climatic water-availability index is

$$AI_{i,t} = \frac{P_{i,t}}{PET_{i,t}}, \quad (1)$$

and was evaluated only for positive annual PET.

SPEI12 fill values at or below -9000 were excluded. Equal-area provincial means were first calculated for each month and then summarized annually. Annual mean SPEI12 is the primary drought-exposure measure in the association models. Drought-month share, used as the alternative drought definition in the reported sensitivity analysis, is

$$D_{i,t} = \frac{1}{12} \sum_{m=1}^{12} \mathbb{I}(SPEI12_{i,t,m} < -1). \quad (2)$$

Cumulative drought severity, retained as an additional sensitivity measure, is

$$DS_{i,t} = \sum_{m=1}^{12} \max(-SPEI12_{i,t,m} - 1, 0) \quad (3)$$

All SPEI12 summaries require 12 valid province-month values and at least 95% polygon coverage per month.

The build produced 403 complete records. Supplementary Table S3 summarizes the geometric, completeness, conservation, external-comparison, discontinuity, and physical-range checks. Assigned mainland HSWUD totals conserved 99.26–99.33% of the full raw-domain annual totals; the remaining share lies outside the retained polygon allocation or in uncovered product cells. The minimum valid geometric support was 92.0% for HSWUD, 100.0% for ERA5 climate variables, and 99.4% for SPEI12. Because HSWUD supplies cell volumes rather than area-normalized depths, valid volumes were summed without rescaling the uncovered share.

The external comparison was intentionally scope-aware. Reconstructed national HSWUD irrigation withdrawals were 85.3% of reported agricultural water use in 2010 and 95.1% in both 2020 and 2022 [22–24]. For 2020, the comparison was extended to all 31 provinces using Table 7 of the official bulletin. Official agricultural water use includes water beyond the narrower reconstructed irrigation concept, so aggregate ratios and cross-province concordance support plausibility but are not equality tests. Temporal screening identified four province-year irrigation changes above 50%, all in 2014: Tianjin (61.3%), Shanghai (75.3%), Chongqing (79.7%), and Xinjiang (62.1%).

The values were neither deleted nor smoothed. The validation protocol therefore combines 2014 exclusion with a formal level-and-slope break model for reconstructed withdrawals and post-2014 interactions in the IWP association model. All row-level source hashes, coverage measures, missingness fields, and the spatial-rule version are retained in the validated hydrological file.

2.5 Agricultural Output and Primary-Outcome Reconstruction

Eight provincial annual series were retrieved from the National Bureau of Statistics National Data portal: agricultural gross-output value, its matching gross-output index, nutrient-equivalent fertilizer use, pesticide use, agricultural plastic-film use, agricultural diesel use, total crop sown area, and effective irrigated area [25]. The archive retains the request payload, catalog and indicator identifiers, response, source label, unit, annotation, retrieval time, and SHA-256 hash. Each response contains the same 403 intended province–year keys. Values represented as missing by the source remain missing.

The official agricultural gross-output index is reported with the previous year equal to 100. For province i , the observed 2010 nominal output anchors the constant-price series, and later values are chain-linked as

$$AGDP_{i,2010}^{real} = AGDP_{i,2010}^{nom}, \quad AGDP_{i,t}^{real} = AGDP_{i,t-1}^{real} \frac{V_{i,t}}{100}, \quad (4)$$

where $V_{i,t}$ is the province-specific annual gross-output volume index. The corresponding implied deflator is $100 AGDP_{i,t}^{nom} / AGDP_{i,t}^{real}$. This procedure avoids using a national price series for heterogeneous provincial output. The archived NBS annotations record revisions following the Third National Agricultural Census and classification changes; these remain source-definition limitations rather than adjustments made by the authors.

Supplementary Table S4 records the original units, conversions, factor set, and observed coverage. The ploughing component uses the factor–activity pair reported by Li [26]: actual crop-sown area and a coefficient of 312.60 kg C km^{−2}. The coefficient is divided by 100 to obtain 3.126 kg C ha^{−1} before it is applied to the NBS area series. This direct source audit resolves the otherwise 100-fold unit mismatch in the secondary table while retaining repeated cropping in the activity measure. The remaining factors follow the documented six-source inventory reported by Xue et al. [27]. Component masses are converted to tonnes of carbon before summation; neither the factors nor the resulting inventory is labelled as CO₂ equivalent.

The validated outcome file has 403 unique keys. Real output and IWP have no missing records. Agricultural diesel is absent in the NBS response for all 31 provinces in 2020–2022, and pesticide and film are absent for all provinces in 2022. Because a six-source denominator is required, input-related carbon emissions and AICP are available for 310 observed rows in 2010–2019 and remain missing thereafter. No shorter inventory, zero substitution, interpolation, or carry-forward rule is used. The complete panel and raw-archive hashes are recorded in `agriculture_output_hashes.csv`.

3 Integration and Association Design

3.1 Assessment of Province-Level Carbon-Market Exposure

Province-level carbon-market exposure was considered but excluded because the available sources did not support a consistent province-year measure (Supplementary Table S5). National-market prices, traded volumes, and transaction amounts verify market operation in 2021 and 2022 [28, 29], but these national values provide no cross-provincial variation after year fixed effects. The archived sources also did not provide a complete annual province-level measure of covered enterprises or emissions across the national and pilot markets [30, 31]. Assigning national values to every province or replacing missing operational coverage with pilot designation would therefore change the intended estimand. Carbon-market exposure was consequently omitted from the analytical panel. The separate carbon-price calculation described below uses national prices only as hypothetical accounting benchmarks.

3.2 Panel Assembly and Validation

The hydrological and agricultural intermediate files each contain 403 unique province-year keys covering 31 provincial units and 13 years. Row-level indicators distinguish the 310 records with complete AICP inputs from the 93 records in which one or more official carbon-accounting activities are unavailable. Stable province and year identifiers, source-record identifiers, missingness indicators, and processing-rule versions are retained throughout the integration process.

The analytical panel is generated by `scripts/build_hydro_ag_panel.py`. Before merging, the script verifies key uniqueness, identical province-year support, consistent province labels, the intended 31-by-13 design, and agreement between the irrigation-water values carried by the two intermediate files. A failed requirement stops panel construction. The final panel contains one row for every intended province-year, preserves observed source missingness, and records separate AICP and IWP completeness indicators.

Agricultural output is expressed at constant 2010 prices through province-specific chain linking, and all units are harmonized before the productivity ratios are calculated. Descriptive summaries use observed nonmissing values and report their effective sample sizes. All reported estimates, uncertainty intervals, sensitivity analyses, tables, and figures are generated from the validated analytical panel.

3.3 Accounting Boundaries and Primary Productivity Outcomes

The carbon inventory has a fixed activity boundary comprising fertilizer, pesticide, agricultural film, agricultural diesel, ploughing activity, and effective irrigated area. Because it excludes livestock emissions, rice-cultivation methane, soil-carbon change, and other biological processes, it is described as an *input-related* agricultural carbon inventory rather than as total agricultural emissions.

Let $\mathcal{S}_6$ denote the six-source set. Input-related agricultural carbon emissions are

$$ACE_{i,t}^{input} = \sum_{s \in \mathcal{S}_6} Q_{s,i,t} \phi_s, \quad (5)$$

where $Q_{s,i,t}$ is an observed NBS activity quantity and ϕ_s is its compatible carbon-mass factor. The adopted coefficients are 0.89 kg C kg⁻¹ for fertilizer, 4.93 for pesticide, 5.18 for film, 0.59 for diesel, 3.126 kg C ha⁻¹ for crop-sown ploughing activity, and 266.48 kg C ha⁻¹ for effective irrigated area [26, 27]. The ploughing value is the hectare-equivalent of the published 312.60 kg C km⁻² coefficient. The archived factor inventory records every coefficient, unit, citation, conversion, and boundary note. The resulting $ACE_{i,t}^{input}$ is expressed in tonnes of carbon.

Real agricultural output, $AGDP_{i,t}^{real}$, is stored in hundred million constant-2010 CNY. The agricultural input-carbon productivity outcome is therefore

$$AICP_{i,t} = \frac{10^8 AGDP_{i,t}^{real}}{ACE_{i,t}^{input}}, \quad (6)$$

where $ACE_{i,t}^{input}$ is measured in tonnes of carbon. Reconstructed irrigation-water use is denoted by $IWU_{i,t}^{grid}$ and is measured in cubic metres. Irrigation-water productivity is

$$IWP_{i,t} = \frac{10^8 AGDP_{i,t}^{real}}{IWU_{i,t}^{grid}}. \quad (7)$$

Both ratios require positive denominators. AICP is expressed in constant-2010 CNY per tonne of input-related carbon, and IWP is expressed in constant-2010 CNY per cubic metre of reconstructed irrigation water. The factor 10⁸ converts the stored agricultural-output unit from hundred million CNY to CNY before either ratio is calculated. Reciprocal transformations are not treated as separate outcomes.

The accounting boundary supports consistent comparison only within the variables explicitly included. It does not measure farm-level heterogeneity, water quality, return flows, soil-carbon change, livestock emissions, rice methane, or every nitrous-oxide pathway.

3.4 Hypothetical Carbon-Price Accounting Scenario

The carbon-price scenario is an accounting stress test applied to the 310 province-years with a complete six-source inventory; it is not an extension of the rejected province-year carbon-market exposure. Input-related carbon mass is converted stoichiometrically to carbon-dioxide mass as

$$E_{i,t}^{input,CO_2} = ACE_{i,t}^{input} \frac{44}{12}. \quad (8)$$

For carbon-price benchmark τ_s^{2010} expressed in constant-2010 CNY per tonne CO₂, the modeled full-liability amount and its share of real agricultural output are

$$L_{i,t,s} = \tau_s^{2010} E_{i,t}^{input, CO_2}, \quad B_{i,t,s} = 100 \frac{L_{i,t,s}}{10^8 AGDP_{i,t}^{real}}. \quad (9)$$

The low benchmark is the national ETS comprehensive closing price of 79.42 CNY t⁻¹ CO₂ at year-end 2023, which is consistent with the World Bank dashboard’s rounded 80-CNY benchmark [32, 33]. The central benchmark is the analysis-derived 2024 transaction-value-weighted average, 181.14/1.89 = 95.84 CNY t⁻¹ CO₂, and the high benchmark is the 2024 peak closing price of 105.65 CNY t⁻¹ CO₂ [34]. Official national CPI indices convert the three nominal prices to 60.14, 72.43, and 79.84 constant-2010 CNY t⁻¹ CO₂, respectively [35]. The calculation assumes full liability with no exemptions, thresholds, free allocations, pass-through, behavioral response, or market-equilibrium adjustment. It does not estimate an enacted tax, firm incidence, profit, welfare, abatement, or government revenue. Because $B_{i,t,s} = 100\tau_s^{2010}(44/12)/AICP_{i,t}$, the burden is algebraically inverse to AICP and is not used as a separate regression outcome. The scenario range represents carbon-price variation only; the fixed-factor inventory lacks source-compatible parameter distributions for probabilistic uncertainty propagation.

3.5 Hydroclimatic Association Specification

The primary analysis evaluates average within-province associations rather than cross-sectional climatic differences. For outcome $Y_{i,t} \in \{IWP_{i,t}, AICP_{i,t}\}$, the prespecified model is

$$\ln(Y_{i,t}) = \beta_A z(AI_{i,t}) + \beta_S z(SPEI12_{i,t}) + \alpha_i + \lambda_t + \varepsilon_{i,t}, \quad (10)$$

where $AI = P/PET$, α_i denotes province fixed effects, and λ_t denotes year fixed effects. Both hydroclimatic variables are standardized using their pooled means and sample standard deviations over the complete 403-record analytical panel. This common scaling preserves comparable exposure units across the IWP and AICP models; linear rescaling changes the coefficient unit but not the corresponding test statistic. The models are estimated without analytical or population weights, so each observed province–year contributes equally to the estimand. The log transformation reduces the influence of right-skewed productivity values and allows the coefficients to be interpreted as approximate proportional differences per one-standard-deviation increase in the corresponding hydroclimatic measure.

Province fixed effects absorb stable observed and unobserved differences among provinces; year fixed effects absorb shocks common to all provinces. They do not remove time-varying confounding, measurement error, or feedback between agricultural activity and the constructed productivity ratios. Equation 10 therefore estimates conditional associations, not causal effects. Standard errors are clustered by province with the finite-sample correction, and inference uses a t reference distribution with 30 degrees of freedom [36]. The primary family contains four coefficients: AI and $SPEI12$ for each of IWP and AICP. The four primary coefficient tests are adjusted together using the Benjamini–Hochberg procedure [37].

Five association sensitivity checks were fixed before interpreting the estimates. They restrict IWP to the common 2010–2019 outcome support; exclude the four flagged 2014 irrigation records; exclude 2014 for every province; add 30 full-rank province-specific slope deviations; and replace annual mean SPEI12 with drought-month share. The same four-row exclusion is shown for AICP only as a common-sample diagnostic because its denominator does not use the reconstructed HSWUD irrigation series. A separate interrupted-trend model tests a 2014 level shift and post-2014 slope change in log reconstructed irrigation withdrawals. Association stability is tested by interacting both standardized hydroclimatic variables with a post-2014 indicator in the IWP model while retaining province and year fixed effects. Sensitivity estimates are not counted as independent confirmations.

3.6 Observed-Range Shape and Context Analysis

The secondary quadratic analysis adds one degree of curvature per hydroclimatic measure:

$$\ln(Y_{i,t}) = \beta_A z(AI_{i,t}) + \gamma_A z(AI_{i,t})^2 + \beta_S z(SPEI12_{i,t}) + \gamma_S z(SPEI12_{i,t})^2 + \alpha_i + \lambda_t + \varepsilon_{i,t}. \quad (11)$$

The model is estimated on each outcome’s full observed sample, but fitted curves are displayed and interpreted only between the outcome-specific 5th and 95th percentiles of each exposure. The analysis tests $\gamma_A = \gamma_S = 0$ jointly for each outcome and adjusts the four individual quadratic-term tests by Benjamini–Hochberg. This deliberately low-dimensional specification can detect broad curvature; it is not a breakpoint, dose–response, or threshold estimator. Turning points and optimum ranges are not inferred from the plotted curves.

Context dependence is tested without dividing 31 provinces into small, arbitrary groups. Let C_i be the negative of province i ’s mean AI over the common 2010–2019 window, standardized across provinces, so that larger C_i denotes a drier observed climate context. The interaction model is

$$\begin{aligned} \ln(Y_{i,t}) = & \beta_A z(AI_{i,t}) + \beta_S z(SPEI12_{i,t}) \\ & + \theta_A z(AI_{i,t})C_i + \theta_S z(SPEI12_{i,t})C_i + \alpha_i + \lambda_t + \varepsilon_{i,t}. \end{aligned} \quad (12)$$

The main effect of C_i is absorbed by province fixed effects. The two interactions are tested jointly within outcome, and Benjamini–Hochberg adjustment is applied to the four interaction coefficients across outcomes. Both secondary modules retain the primary clustered inference and the same six specification variants: baseline, common 2010–2019 support, flagged-record exclusion, 2014 exclusion, province-specific trends, and the drought-month-share alternative. The modules remain associational and secondary to Equation 10.

4 Quality-Control, Descriptive, and Inferential Protocol

Validation is performed within each source block before integration and again at the merge boundary. Hydrological checks cover raw units, fill values, exact calendar conversion, spatial overlap, annual completeness, polygon coverage, national conservation, the 31-province external comparator, and the 2014 product break. Agricultural checks cover source identity, official units, province-year uniqueness, chain-link recursion, factor–activity compatibility, positive denominators, and source-reported missingness. Carbon-price checks cover the rates observed in 2023 and 2024, CPI-chain arithmetic, C-to-CO₂ conversion, scenario ordering, and the inverse-AICP identity. The merge adds cross-file key equality, province-label agreement, duplicate rejection, irrigation-water equality, completeness counts, and output hashing.

The descriptive protocol uses observed values only. For each analytical variable it records the nonmissing count, mean, sample standard deviation, median, inclusive first and third quartiles, minimum, and maximum. The manuscript reports the median, interquartile range, and full range because several volumetric variables are strongly right-skewed. These summaries characterize the analytical data and should not be read as estimates of policy or hydroclimatic effects.

Table 1 Observed characteristics of the analytical hydro-agricultural panel

Variable	<i>n</i>	Median (Q1–Q3)	Range	Unit
Real agricultural output	403	1299.4 (685.8–2173.0)	41.726–5793.4	hundred million constant-2010 CNY
Input-related carbon	310	3.225e+06 (1.492e+06–4.350e+06)	133785.3–9.959e+06	tonnes C
Agricultural input- carbon productivity	310	41919.7 (35163.8–49841.1)	24260.5–82694.5	constant-2010 CNY per tonne C
Reconstructed irrigation-water use	403	8.882e+09 (4.066e+09–1.539e+10)	2.521e+08–5.487e+10	cubic metres
Irrigation-water productivity	403	14.075 (10.104–21.382)	1.712–51.373	constant-2010 CNY per cubic metre
Annual precipitation	403	932.6 (600.0–1498.0)	211.9–2322.9	mm
Annual potential evapotranspiration	403	1083.4 (918.6–1175.8)	650.4–1392.2	mm
Climatic water-availability ratio	403	1.000 (0.5452–1.442)	0.2330–2.289	ratio
Annual mean SPEI12	403	-0.2149 (-0.6299–0.1312)	-1.779–1.637	index
Drought-month share	403	0.0000 (0.0000–0.3333)	0.0000–1.000	share

Note: Q1 and Q3 are the inclusive sample quartiles. AICP has 310 observed province-years; all other displayed variables have 403.

The analysis scripts verify the frozen panel hash before estimation, require 31 province clusters, preserve the outcome-specific samples, and read the four flagged irrigation records from the generated quality-control register. They produce the linear, shape, context, formal break, and carbon-price scenario manifests; support records;

primary and sensitivity tables; figures; model formulas; dependency versions; and output hashes. Any change to the panel or an analysis script invalidates the corresponding outputs until they are rerun.

5 Integration, Validation, and Association Results

5.1 Analytical Panel and Observed Coverage

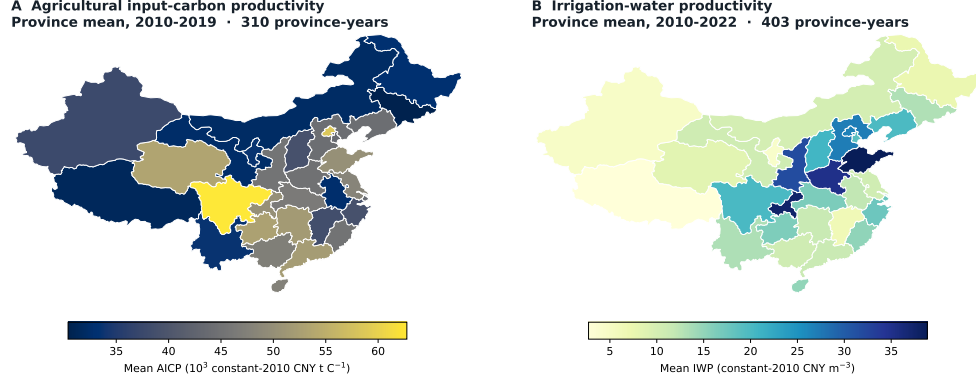
The deterministic merge produced 403 unique province–year records for 31 provincial units over 13 years; its SHA-256 hash is recorded in the analysis-build summary. All 403 records contain real agricultural output, reconstructed irrigation-water use, IWP, precipitation, PET, climatic water availability, SPEI12, and drought summaries. AICP is complete for all 31 provinces in each year from 2010 through 2019, giving 310 observations. Its 93 missing records comprise all provinces in 2020–2022 and arise from the documented source gap in agricultural diesel; pesticide and agricultural film are additionally missing in 2022.

The median province–year IWP is 14.075 constant-2010 CNY m⁻³, with an interquartile range of 10.104–21.382. Across its observed 2010–2019 support, median AICP is 41,919.7 constant-2010 CNY t C⁻¹, with an interquartile range of 35,163.8–49,841.1. These are descriptive properties of heterogeneous provincial observations, not temporal effects or performance rankings. The full distributional summaries and effective counts are generated directly from the analytical panel in Table 1.

Figure 2 maps province-specific means at observed outcome coverage. Because the outcome windows differ, the panels are descriptive spatial summaries rather than matched-period comparisons.

Spatial patterns of agricultural resource productivity across mainland China

Province-level means from the source-audited panel



Periods follow observed outcome coverage; AICP values for 2020-2022 are not imputed.

Figure 2 Spatial distributions of mean agricultural input-carbon productivity (AICP) and irrigation-water productivity (IWP) across 31 provincial-level units in mainland China. Panel A reports province-specific mean AICP for 2010–2019 (310 province–year observations), whereas Panel B reports province-specific mean IWP for 2010–2022 (403 observations). AICP is expressed in constant-2010 CNY per tonne of input-related carbon, and IWP in constant-2010 CNY per cubic metre of reconstructed irrigation-water use. The unequal periods reflect observed outcome availability; AICP values for 2020–2022 were not imputed.

5.2 Validation Outcomes and Scope Decision

The hydrological block passed the unit, boundary, calendar, fill-value, spatial-allocation, conservation, and coverage checks. Reconstructed national irrigation withdrawals equaled 85.3% of reported agricultural water use in 2010 and 95.1% in both 2020 and 2022. In the 2020 province-level comparison, the reconstructed total was 0.951 of the broader official agricultural-water quantity. Across 31 provinces, the mean absolute error was 6.855×10^8 m³, the mean absolute percentage error was 9.03%, and the aggregate-weighted absolute percentage error was 5.88%. Pearson and Spearman correlations were 0.998 and 0.995, respectively (Table 2). Because the official category includes agricultural uses beyond reconstructed crop-irrigation withdrawals, these diagnostics assess aggregate plausibility, spatial concordance, and the magnitude of disagreement rather than observation-level equality.

Four year-to-year irrigation changes exceeding 50% occurred in 2014: Tianjin, Shanghai, Chongqing, and Xinjiang. Their common timing was retained as a product-discontinuity flag. The interrupted-trend analysis estimated a 0.119-log-point level shift in 2014 (95% CI: 0.043–0.195; $p = 0.003$), while the estimated post-2014 slope change was -0.014 per year (95% CI: -0.028 to 0.001 ; $p = 0.072$). In the IWP association model, the two post-2014 hydroclimatic interactions were jointly nonsignificant ($F = 0.269$, $p = 0.766$; Supplementary Table S6). Thus, the irrigation-product discontinuity remains a measurement limitation, but we found no evidence that the estimated hydroclimatic coefficients differed before and after 2014 under this diagnostic.

Table 2 Province-scale validation of reconstructed irrigation withdrawals in 2020

Diagnostic	Result
Reconstructed crop-irrigation withdrawals (10^8 m^3)	3435.4
Official agricultural water use (10^8 m^3)	3612.5
Aggregate reconstructed/official ratio	0.951
Mean absolute error across provinces (10^8 m^3)	6.855
Mean absolute percentage error across provinces	9.03%
Aggregate-weighted absolute percentage error	5.88%
Pearson correlation across provinces	0.998
Spearman rank correlation across provinces	0.995
Median province ratio (Q1–Q3)	0.954 (0.900–0.990)
Provinces exceeding the broader official quantity by > 5%	1 of 31

Notes: The official comparator is agricultural water use in Table 7 of the 2020 China Water Resources Bulletin. It includes uses beyond the reconstructed crop-irrigation withdrawals, so the comparison is a scope-aware plausibility and spatial-concordance check, not an equality test. Correlations are descriptive and unadjusted.

The agriculture block passed source-identity, unit, key, factor, chain-link, positivity, and missingness checks at observed coverage. Direct inspection of the ploughing-factor source established the factor–activity pair of actual crop-sown area and $312.60 \text{ kg C km}^{-2}$, implemented as $3.126 \text{ kg C ha}^{-1}$; the panel and all AICP-dependent outputs were regenerated after that unit correction. No shorter carbon inventory was substituted after 2019, and no missing activity was set to zero. The carbon-market exposure assessment did not support a province–year measure because national operation is common across provinces and complete provincial coverage is unavailable. Carbon-market exposure is therefore absent from the analytical panel, while the separate static carbon-price scenario uses externally sourced national prices only as hypothetical accounting benchmarks.

5.3 Primary Hydroclimatic Associations

Table 3 reports the four prespecified coefficients. For IWP, the coefficient on climatic water availability is 0.001 log points per standard deviation (95% CI: -0.053 to 0.055), equivalent to a 0.1% difference with a 95% interval from -5.1% to 5.7% . The SPEI12 coefficient is -0.008 (-0.032 to 0.016), equivalent to -0.8% (-3.2% to 1.7%). The two coefficients are jointly indistinguishable from zero ($p = 0.785$), and their partial R^2 relative to province and year fixed effects is 0.0022.

For AICP, climatic water availability has a coefficient of -0.005 (-0.023 to 0.014), equivalent to approximately -0.5% (-2.2% to 1.4%). The SPEI12 estimate is 0.001 (-0.012 to 0.013), equivalent to approximately 0.1% (-1.2% to 1.3%). The joint p -value is 0.872 and the exposure-block partial R^2 is 0.0004. Benjamini–Hochberg adjustment over the four primary coefficients gives $q = 0.962$ for each. The confidence intervals and near-zero incremental explanatory shares provide no evidence of a detectable average within-province association over the observed periods; they do not establish that hydroclimate is irrelevant at crop, farm, irrigation-district, seasonal, nonlinear, or lagged scales.

Figure 3 shows the same estimates and confidence intervals without significance symbols or directional categorization. The figure makes the difference in precision

Table 3 Within-province associations of hydroclimate with resource productivity

	$\ln(IWP)$	$\ln(AICP)$
Climatic water availability, AI (1 SD)	0.001 [-0.053, 0.055]	-0.005 [-0.023, 0.014]
Annual mean SPEI12 (1 SD)	-0.008 [-0.032, 0.016]	0.001 [-0.012, 0.013]
Observations	403	310
Province fixed effects	Yes	Yes
Year fixed effects	Yes	Yes
Joint p -value	0.785	0.872
Exposure-block partial R^2	0.0022	0.0004

Notes: Entries are log-point coefficients with 95% confidence intervals. Both exposures are standardized once over the complete 403-row analytical panel. Standard errors are clustered by province with the finite-sample correction and inference uses a t_{30} reference distribution. The four primary coefficient tests are adjusted together using the Benjamini–Hochberg procedure; all four adjusted q -values are 0.962. The estimates are conditional associations, not causal effects.

between the complete IWP panel and the shorter AICP panel explicit while preserving zero as the reference.

Within-province hydroclimatic associations with agricultural resource productivity

Point estimates and 95% province-clustered confidence intervals

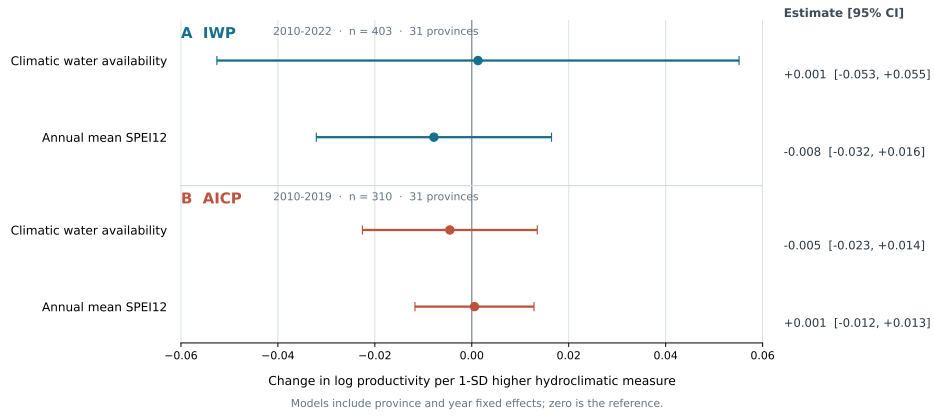


Figure 3 Primary within-province hydroclimatic association estimates. Points represent changes in log productivity per one-standard-deviation increase in climatic water availability or annual mean SPEI12; horizontal lines show 95% confidence intervals based on province-clustered inference. Models include province and year fixed effects. The IWP analysis contains 403 province–year observations from 2010–2022, whereas the AICP analysis contains 310 observations from 2010–2019. Zero is the reference value.

5.4 Quadratic Shape and Context Results

The quadratic analysis is confined to empirically populated ranges. For IWP, the central 90% spans $AI = 0.314$ – 1.966 and annual mean SPEI12 from -1.182 to 0.792 ; for AICP, the corresponding ranges are 0.317 – 1.934 and -1.091 to 0.743 . The full-sample

endpoints are retained in the analytical support file but are not used to characterize curve tails.

Supplementary Table S7 reports the quadratic models. For IWP, the *AI* quadratic coefficient is -0.030 (95% CI: -0.071 to 0.012 ; $q = 0.382$), and the SPEI12 quadratic coefficient is 0.013 (-0.007 to 0.032 ; $q = 0.382$). The two curvature terms are jointly nonsignificant ($p = 0.329$, adjusted $q = 0.659$) and add 0.0164 to the in-sample partial R^2 relative to the linear fixed-effects model. For AICP, the corresponding quadratic coefficients are 0.005 (-0.011 to 0.021 ; $q = 0.548$) and -0.003 (-0.010 to 0.004 ; $q = 0.548$); the joint test gives $p = 0.684$ (adjusted $q = 0.684$) and the added partial R^2 is 0.0026 . Thus, the observed data do not support a low-order curvature qualification of the primary average result.

Supplementary Figure S1 visualizes the same quadratic models over the central 90% of outcome-specific support. The curves are conditional differences from the pooled-mean exposure value, with the other hydroclimatic measure held at its standardized mean. Their pointwise intervals span zero throughout the displayed support. Apparent bends are therefore not interpreted as thresholds or operating ranges.

The continuous context analysis also yields no adjusted evidence of heterogeneity (Supplementary Table S8). For IWP, the *AI*-context and SPEI12-context coefficients are 0.009 (-0.039 to 0.058 ; $q = 0.835$) and 0.018 (-0.008 to 0.043 ; $q = 0.682$), with joint $p = 0.372$ (adjusted $q = 0.634$). For AICP, they are -0.011 (-0.034 to 0.012 ; $q = 0.691$) and -0.001 (-0.012 to 0.010 ; $q = 0.835$), with joint $p = 0.634$ (adjusted $q = 0.634$). These interactions do not show that the average null masks systematically different annual associations along the observed dry-climate gradient.

5.5 Sensitivity Analysis

The primary sensitivity results in Table 4 do not change the interpretation. Restricting IWP to 2010–2019, removing the four flagged irrigation records, removing all of 2014, or adding province-specific linear trends leaves every confidence interval spanning zero. The climatic water-availability coefficient for IWP ranges from -0.030 to 0.008 across these specifications; the corresponding AICP range is -0.008 to -0.002 . Replacing SPEI12 with drought-month share likewise produces intervals spanning zero for both outcomes.

Supplementary Table S9 applies the same specification changes to the secondary modules. Curvature-block joint p -values range from 0.127 to 0.903 and context-block joint p -values from 0.339 to 0.947 . The largest added partial R^2 from curvature is 0.0218 in the common-window IWP model. These checks address the documented sample and product-construction risks but cannot compensate for omitted time-varying determinants.

Table 4 Sensitivity of the hydroclimatic associations

Outcome	Specification	AI (1 SD)	Drought measure (1 SD)	N
IWP	Primary specification	0.001 [-0.053, 0.055]	-0.008 [-0.032, 0.016]	403
	Common 2010–2019 support	-0.030 [-0.100, 0.039]	0.011 [-0.019, 0.041]	310
	Exclude four flagged 2014 rows	0.001 [-0.056, 0.058]	-0.007 [-0.031, 0.016]	399
	Exclude 2014	-0.002 [-0.061, 0.057]	-0.007 [-0.031, 0.017]	372
	Add province-specific linear trends	0.008 [-0.040, 0.056]	0.008 [-0.009, 0.026]	403
	Replace SPEI12 with drought-month share	0.003 [-0.057, 0.064]	0.012 [-0.012, 0.036]	403
AICP	Primary specification	-0.005 [-0.023, 0.014]	0.001 [-0.012, 0.013]	310
	Common 2010–2019 support	-0.005 [-0.023, 0.014]	0.001 [-0.012, 0.013]	310
	Exclude four flagged 2014 rows	-0.008 [-0.029, 0.013]	0.002 [-0.011, 0.014]	306
	Exclude 2014	-0.007 [-0.032, 0.017]	0.001 [-0.012, 0.015]	279
	Add province-specific linear trends	-0.002 [-0.018, 0.015]	0.005 [-0.004, 0.015]	310
	Replace SPEI12 with drought-month share	-0.004 [-0.019, 0.012]	0.000 [-0.016, 0.017]	310

Notes: Entries are log-point coefficients with 95% province-clustered confidence intervals. The drought-measure column reports annual mean SPEI12 except in the named alternative, where higher drought-month share denotes drier conditions. The flagged-row exclusion is primarily relevant to IWP; it is shown for AICP only as a common sample check. Sensitivity estimates are descriptive robustness checks and are not counted as independent confirmations.

5.6 Hypothetical Carbon-Price Accounting Results

The low, central, and high benchmarks correspond to nominal prices of 79.42, 95.84, and 105.65 CNY t⁻¹ CO₂ and to constant-2010 prices of 60.14, 72.43, and 79.84 CNY t⁻¹ CO₂, respectively. Across the 310 complete province-years, the median gross-output burden is 0.526% under the low scenario, 0.634% under the central scenario, and 0.698% under the high scenario (Supplementary Table S10). Under the central scenario, the interquartile range is 0.533–0.755%, the full range is 0.321–1.095%, and the ratio of aggregate modeled liability to aggregate real agricultural output is 0.615%.

Figure 4 summarizes province-level mean burdens under the same three prices. The low–high span reflects only the prespecified price benchmarks; it is neither a sampling interval nor a carbon-inventory uncertainty interval. Because the burden is an inverse rescaling of AICP under a uniform full-liability rate, the figure translates the accounting indicator into a policy-relevant unit without providing an independent empirical result.

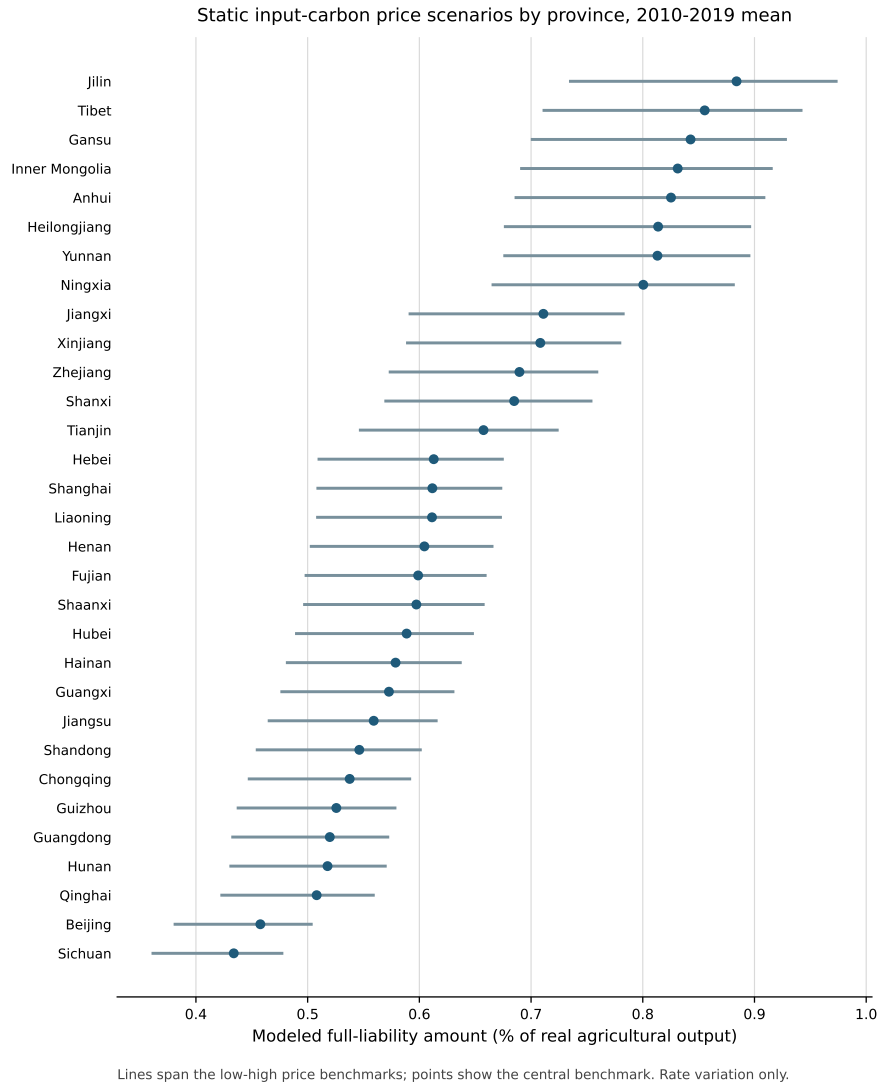


Figure 4 Province-level mean hypothetical carbon-price burden over observed 2010–2019 coverage. Points show the central-price scenario, and horizontal lines span the low and high price scenarios. Burden is modeled full liability as a percentage of real agricultural gross output for the six-source input-related carbon inventory. The scenario span represents price choice only and is not a confidence or carbon-inventory uncertainty interval. The calculation does not represent an enacted agricultural carbon tax, observed payment, behavioral response, or incidence estimate.

6 Limitations

The analysis uses unweighted, aggregated province–year observations and estimates within-province conditional associations. Aggregation obscures variation among farms, crops, seasons, irrigation districts, and water sources. Province and year fixed effects

account for stable provincial characteristics and common annual shocks, but time-varying crop composition, technology, irrigation infrastructure, agricultural policy, and other omitted determinants may still confound the estimates. With 31 province clusters and at most 13 annual observations per province, the analysis also has limited capacity to characterize complex nonlinearities, spatial dependence, or delayed responses.

Outcome coverage and measurement boundaries differ. AICP is observed during 2010–2019, whereas IWP is observed during 2010–2022. The carbon inventory covers six input-related activities and is not a measure of total agricultural greenhouse-gas emissions. Its fixed coefficients inherit uncertainty from the underlying literature, and compatible probability distributions were unavailable for formal propagation of factor uncertainty. IWP uses reconstructed crop-irrigation withdrawals rather than reported provincial withdrawals. The broader official agricultural-water comparator, the 2014 irrigation-product discontinuity, fixed 2019 administrative boundaries, grid resolution, coastal coverage, and polygon-allocation rules all limit the precision of the water denominator.

Climatic water availability and SPEI12 describe hydroclimatic conditions but do not fully represent water access, storage, groundwater dependence, water quality, irrigation infrastructure, or institutional scarcity. The contemporaneous annual models cannot resolve seasonal, crop-specific, lagged, or farm-level responses. The quadratic and context-interaction specifications examine broad departures from linearity over observed support but are not threshold estimators. Finally, the carbon-price calculation is an algebraic accounting transformation of AICP under uniform full liability. It excludes policy thresholds, exemptions, free allocation, pass-through, behavioral adjustment, abatement, market-equilibrium responses, and distributional incidence.

7 Discussion

After adjustment for province and year fixed effects, the estimated associations between the two hydroclimatic measures and both productivity ratios were centered near zero. For IWP, the 95% intervals corresponded to proportional differences of -5.1% to 5.7% for climatic water availability and -3.2% to 1.7% for SPEI12 per standard-deviation increase. The corresponding intervals for AICP were -2.2% to 1.4% and -1.2% to 1.3% . The hydroclimatic variables added less than 0.3% explanatory power beyond the fixed effects. Neither the quadratic terms nor the dry-context interactions provided evidence of systematic departures from the average associations, and the estimates remained similar across the prespecified sensitivity analyses.

These results constrain the magnitude of detectable contemporaneous associations at the annual provincial scale but do not imply that hydroclimate is unimportant for agricultural production. Evidence from physical and lower-scale systems shows that water conditions influence agricultural outcomes [1–3, 27]. Province-year aggregation combines crops, seasons, irrigation systems, and management responses with different sensitivities. Irrigation, storage, crop substitution, and other adaptations may also weaken the relationship between meteorological conditions and annual output-to-input ratios. Seasonal or crop-specific outcomes and a design suited to causal attribution would be required to distinguish among these mechanisms.

The ecological-informatics contribution lies in connecting heterogeneous source records to separately defined carbon- and water-productivity outcomes through a reproducible analytical chain. The workflow preserves source identities, hashes, units, calendars, spatial-allocation rules, missingness, and processing versions and regenerates the reported outputs from the validated panel. These practices operationalize provenance and reproducibility principles developed in environmental research infrastructures and ecological-monitoring workflows [5–9]. They also make the consequences of source revisions, coverage changes, and processing decisions visible during future updates.

For monitoring purposes, AICP and IWP should remain separate because improvement in one ratio does not imply improvement in the other. Provenance records and discontinuity flags can help identify whether changes in reported indicators coincide with changes in source definition or processing. The carbon-price scenarios provide a complementary accounting interpretation of the input-carbon denominator: under the central benchmark, the median province–year burden was 0.634% and the aggregate burden was 0.615% of real agricultural output. Because this burden is algebraically inverse to AICP at a uniform price, its provincial pattern is an accounting translation rather than independent evidence of policy exposure.

8 Conclusion

This study constructed a validated province–year panel linking official agricultural statistics, reconstructed irrigation-water use, ERA5 hydroclimate, and ERA5-derived drought information for 31 mainland Chinese provincial units during 2010–2022. IWP and the hydroclimatic variables are available for all 403 province–years, while the six-source input-carbon denominator and AICP are available for 310 records during 2010–2019. The 2020 external comparison showed close provincial concordance between reconstructed crop-irrigation withdrawals and the broader official agricultural-water quantity, although the two measures are not conceptually identical.

The fixed-effects results provided no evidence of detectable average, low-order quadratic, or dry-context-dependent contemporaneous hydroclimatic associations with either productivity ratio at the annual provincial scale. The confidence intervals nevertheless define the range of effects compatible with the data and leave seasonal, lagged, crop-specific, and farm-level responses unresolved. Hypothetical carbon-price benchmarks produced median full-liability accounting burdens of 0.526%, 0.634%, and 0.698% of real agricultural output. Together, the validated panel, denominator-specific indicators, association estimates, and provenance records provide a reproducible basis for monitoring agricultural resource productivity and for developing finer-scale analyses as additional source data become available.

Declarations

Ethics approval and consent to participate

This study analyzed aggregate province-level secondary data and involved no human participants, identifiable personal data, animals, or biological materials. Ethics approval and consent were therefore not applicable.

Data availability

The derived province-year panel, carbon-price scenario data, processing and analysis code, validation outputs, and file-hash manifests are available from the corresponding author upon reasonable request. Original third-party data remain subject to the access and reuse conditions of their respective providers and may need to be obtained directly from those providers.

Funding Declaration

This work was financially supported by the project of the National Social Science Foundation (23BJL088).

Author Contributions Statement

Aoqi Xu: data collection, data curation, methodology, software, formal analysis, visualization, original draft. Shan Miao: validation, investigation, interpretation, and manuscript review and editing. Zhiming Guo: conceptualization, supervision, project administration, funding acquisition, and manuscript review and editing.

References

- [1] Cheng, H. *et al.* Divergent responses of optimal land and water allocation to different hydrological regimes in the agricultural water-food-carbon nexus system. *Journal of Hydrology* **653**, 132730 (2025).
- [2] Khattak, W. A. *et al.* The role of agricultural land management in modulating water-carbon interplay within dryland ecological systems. *Agriculture, Ecosystems & Environment* **378**, 109315 (2025).
- [3] Jiang, S. *et al.* Spatiotemporal variations of cropland carbon sequestration and water loss across china. *Agricultural Water Management* **287**, 108427 (2023).
- [4] Recknagel, F. Ecological informatics: A discipline in the making. *Ecological Informatics* **6**, 1–3 (2011).
- [5] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. *et al.* The FAIR guiding principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016).

- [6] Huber, R. *et al.* Integrating data and analysis technologies within leading environmental research infrastructures: Challenges and approaches. *Ecological Informatics* **61**, 101245 (2021).
- [7] Wicquart, J. *et al.* A workflow to integrate ecological monitoring data from different sources. *Ecological Informatics* **68**, 101543 (2022).
- [8] Jantzen, C. C. & Vriend, S. J. G. Putting FAIR into practice for ecologists: How to make ecological data more reusable. *Ecological Informatics* **95**, 103712 (2026).
- [9] Jenkins, G. B. *et al.* Reproducibility in ecology and evolution: Minimum standards for data and code. *Ecology and Evolution* **13**, e9961 (2023).
- [10] Zhao, R. *et al.* Impacts of water and land resources exploitation on agricultural carbon emissions: The water–land–energy–carbon nexus. *Land Use Policy* **72**, 480–492 (2018). URL <https://doi.org/10.1016/j.landusepol.2017.12.029>.
- [11] Deng, C. *et al.* Impacts of the integrated pattern of water and land resources use on agricultural greenhouse gas emissions in China during 2006–2017: A water–land–energy–emissions nexus analysis. *Journal of Cleaner Production* **308**, 127221 (2021). URL <https://doi.org/10.1016/j.jclepro.2021.127221>.
- [12] Feng, Y., Zhang, Y., Wu, Z., Ye, Q. & Cao, X. Evaluation of agricultural eco-efficiency and its spatiotemporal differentiation in China, considering green water consumption and carbon emissions based on undesired dynamic SBM–DEA. *Sustainability* **15**, 3361 (2023). URL <https://doi.org/10.3390/su15043361>.
- [13] Xu, Y. Provincial-level carbon-reduction potential for agricultural irrigation in China. *Sustainability* **17**, 5501 (2025). URL <https://doi.org/10.3390/su17125501>.
- [14] Zhang, Y. & Yin, Y. HUSWD datasets (.nc) (2024). URL <https://doi.org/10.6084/m9.figshare.27610524.v7>. Figshare dataset, version 7, published 11 July 2026; CC BY 4.0.
- [15] Zhang, Y., Yin, Y., Yin, M. & Zhang, X. A high-resolution gridded dataset for china’s monthly sectoral water use. *Scientific Data* **12**, 1157 (2025).
- [16] Copernicus Climate Change Service. ERA5 monthly averaged data on single levels from 1940 to present (2019). URL <https://doi.org/10.24381/cds.f17050d7>. Copernicus Climate Data Store dataset; accessed 26 August 2026; CC BY.
- [17] Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049 (2020).
- [18] European Centre for Medium-Range Weather Forecasts. Monthly drought indices from 1940 to present derived from ERA5 reanalysis (2025). URL <https://doi.org/10.24381/9bea5e16>. ERA5–Drought version 1.0; accessed 26 August 2026; CC BY.

- [19] Vicente-Serrano, S. M., Beguería, S. & López-Moreno, J. I. A multiscale drought index sensitive to global warming: the standardized precipitation evapotranspiration index. *Journal of Climate* **23**, 1696–1718 (2010).
- [20] geoBoundaries. China administrative level 1 boundaries (2023). URL <https://www.geoboundaries.org/api/current/gbOpen/CHN/ADM1/>. Boundary ID CHN-ADM1-43563684; representative year 2019; build 12 December 2023; accessed 26 August 2026; public domain.
- [21] Runfola, D. *et al.* geoBoundaries: A global database of political administrative boundaries. *PLOS ONE* **15**, e0231866 (2020).
- [22] National Bureau of Statistics of China. Water supply and water use by region, 2010 (2023). URL https://www.stats.gov.cn/zt_18555/ztsj/hjtjzl/2010/202303/t20230302_1921831.html. Environmental statistics archive; accessed 26 August 2026.
- [23] Ministry of Water Resources of the People’s Republic of China. China water resources bulletin 2020 (2021). URL <https://slqjd.mwr.gov.cn/upload/2022/09-23/08-52-3307561482323230.pdf>. Official annual bulletin; accessed 26 August 2026.
- [24] Ministry of Water Resources of the People’s Republic of China. China water resources bulletin 2022 (2023). URL https://cn.lmcwater.org.cn/knowledge_sharing/water_knowledge/202307/t20230706_39467.shtml. Official bulletin release; accessed 26 August 2026.
- [25] National Bureau of Statistics of China. Provincial annual data: Agriculture, 2010–2022 (2026). URL <https://data.stats.gov.cn/dg/website/page.html#/pc/national/fsYearData>. National Data portal; source indicator identifiers, units, annotations, request payloads, and responses archived; accessed 27 August 2026.
- [26] Li, B. Empirical study of the relationship between economic growth and agricultural carbon emissions. *Ecology and Environmental Sciences* **21**, 220–224 (2012). URL <https://www.jeesci.com/EN/article/downloadArticleFile.do?attachType=PDF&id=2861>. Official journal PDF; Chinese-language article; accessed 29 August 2026.
- [27] Xue, Y. *et al.* Analysis of spatial and temporal characteristics and drivers of agricultural carbon emissions in china. *Chinese Journal of Eco-Agriculture* **32**, 1805–1817 (2024).
- [28] Shanghai Environment and Energy Exchange. Annual national carbon-market trading data, 16 july–31 december 2021 (2021). URL <https://overview.cneeex.com/c/2021-12-31/491811.shtml>. Official annual market record; accessed 27 August 2026.

- [29] Shanghai Environment and Energy Exchange. Annual national carbon-market trading data, 4 january–30 december 2022 (2022). URL <https://overview.cneeeex.com/c/2022-12-30/493617.shtml>. Official annual market record; accessed 27 August 2026.
- [30] Ministry of Ecology and Environment of the People’s Republic of China. Report on the first compliance cycle of the national carbon emissions trading market (2022). URL <https://www.mee.gov.cn/ywgz/xdqhbh/wsqtz/202212/P020221230799532329594.pdf>. Official report; accessed 27 August 2026.
- [31] Wang, K., Li, S., Li, S. & Wang, Z. Reviews of china’s carbon market and prospects of its optimal rolling out plan (2023). *Journal of Beijing Institute of Technology (Social Sciences Edition)* **25**, 36–44 (2023).
- [32] Ministry of Ecology and Environment of the People’s Republic of China. National carbon market development report 2024 (2024). URL <https://www.mee.gov.cn/ywdt/xwfb/202407/W020240722528848347594.pdf>. Official report; accessed 29 August 2026.
- [33] World Bank. China national ETS: Compliance mechanism factsheet (2026). URL https://carbonpricingdashboard.worldbank.org/compliance-factsheet/ETS_CN. Carbon Pricing Dashboard; accessed 29 August 2026.
- [34] Ministry of Ecology and Environment of the People’s Republic of China. National carbon market development report 2025 (2025). URL <https://www.mee.gov.cn/ywgz/xdqhbh/wsqtz/202509/W020250927515316322073.pdf>. Official report; accessed 29 August 2026.
- [35] National Bureau of Statistics of China. National resident consumer price index series, 2010–2024 (2026). URL https://www.stats.gov.cn/zt_18555/ztsj/jzgj/jzsc2023/202404/P020240417553133577139.pdf. Compiled from official annual index tables and 2023–2024 statistical releases; source values and cumulative 2010-price chain archived; accessed 29 August 2026.
- [36] Cameron, A. C. & Miller, D. L. A practitioner’s guide to cluster-robust inference. *Journal of Human Resources* **50**, 317–372 (2015).
- [37] Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformationSR.pdf](#)