Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Shared raster support reveals pseudo-replication in settlement- centered remote-sensing monitoring

Peng Yue[1]

yp6253344@163.com

Kangwon National University

Jia Hou[2]

Zhengzhou Business University

Xiao Zhang[3]

Christian-Albrechts-Universität zu Kiel

Article

Keywords: remote sensing, landscape fragmentation, blocked cross-validation, pseudo-replication, land-use change, rural settlements

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11241271/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Remote-sensing studies often treat buffers around nearby settlements as separate observations even when they reuse many of the same raster cells. This shared spatial support can create pseudo-replication that is distinct from ordinary geographic autocorrelation. We evaluated this problem for five 2000–2022 landscape-change metrics around 210 historically documented rural settlements in China's Yellow River Delta. Among 21,945 pairs of 3-km buffers, 2,036 overlapped, and outcome-independent 4-km thinning reduced the cohort to 43–50 sites. We combined overlap auditing with 15-km blocked spatial cross-validation, coordinate-only baselines, mathematical-coupling tests, residual Moran's I, and low-overlap thinning. Fragmentation was the only metric with positive out-of-fold R^2 across both predictor branches and both tree algorithms (best $R^2 = 0.388$; Spearman $\rho = 0.761$), while a quadratic coordinate model alone achieved $R^2 = 0.340$. Fragmentation remained positive across all six thinning orders, but every full-cohort model retained significant residual spatial autocorrelation. Built-up change, cropland loss, spectral stress, and the composite stress metric were weak, negative, or coupling-sensitive. Shared raster support therefore changes how settlement-centered monitoring should be interpreted: nominal place counts can overstate spatial replication, and predictive screening requires evidence beyond direct descriptive change.

Introduction

Remote sensing and geographic information systems are widely used to document land-use and landscape change over large areas, increasingly in combination with machine learning [1, 2]. The resulting maps and landscape metrics are useful for monitoring, but their interpretation depends on scale, spatial resolution, metric redundancy, and the way observation units are defined. These issues have been discussed for more than two decades in the landscape-metrics literature [3–6]. A metric can therefore be meaningful as a description of observed change without necessarily supporting prediction beyond the locations from which it was calculated.

A separate literature has shown that model performance can be overstated when spatially structured observations are split at random rather than separated geographically [7–10]. Place-centered monitoring adds another complication. Buffers around nearby villages or other fixed locations may reuse the same raster cells, so nominal sample size is larger than the number of effectively distinct spatial supports. Baseline variables can also enter change metrics mathematically, and both predictors and responses may follow coast–inland or urban–rural gradients. In such settings, a model may predict nearby places well while adding little information beyond location itself.

The Yellow River Delta provides a useful setting in which to examine these issues because land-use change, landscape fragmentation, and salinity-related spatial gradients have already been studied extensively. Previous work has mapped long-term land-use change with machine learning [2], examined the scale dependence and drivers of coastal fragmentation [11], and mapped salinity or saline-land dynamics from field and remote-sensing data [12–15]. Those studies establish the regional

environmental context. The question addressed here is narrower: when change is summarized around clustered settlement locations, which conclusions survive controls for shared raster support, broad geography, mathematical coupling, and spatially separated validation?

Accordingly, we treat shared spatial support as a separate reliability problem rather than as a synonym for residual spatial autocorrelation. When neighboring settlement buffers reuse the same raster cells, nominally distinct places can contain partially duplicated environmental information before any model is fitted. Spatial blocking can reduce optimistic interpolation across nearby locations, but it does not by itself quantify how much observational support is shared. We therefore combine an overlap audit and outcome-independent low-overlap thinning with blocked cross-validation, coordinate-only baselines, mathematical-coupling checks, and residual spatial diagnostics. These checks are used to distinguish direct description at observed settlements, diagnostic use when construction or dependence is problematic, and conditional comparative screening when spatial transfer remains positive.

We apply this framework to five 2000–2022 landscape-change metrics around 210 historically documented rural settlements in the YRD. We first quantify pairwise overlap among the 3-km settlement buffers and then evaluate the five metrics using locked spatial folds and environmental-versus-coordinate baselines. For the metric that retains blocked skill, we test whether the signal persists after six outcome-independent 4-km thinnings that sharply reduce shared raster support, together with regional consistency, coupling, block-size, MGWR, permutation-importance, and residual-autocorrelation checks. The study therefore asks not whether spatial cross-validation can improve model assessment in general, but how duplicated raster support changes the evidential status of settlement-centered monitoring results.

Materials and methods

Study area

The study area is the Yellow River Delta of northern Shandong Province, China, along the Bohai Sea and Laizhou Bay. The broader delta is low-relief, coastally influenced, and affected by shallow saline groundwater and strong fluvial–marine interactions [12, 16–18]. The reconstructed cohort spans 118.194–118.892° E and 37.494–37.984° N. WorldClim 1970–2000 annual precipitation was used as the baseline climate covariate [19].

Repeated river-channel shifts, coastal accretion, reclamation, oil-field development, aquaculture, urbanization, and ecological restoration have produced a heterogeneous cropland–built land–wetland–water mosaic. The analysis focuses on the historical settlement corridor rather than the full modern administrative extent of Dongying.

The reconstructed settlement cohort and study extent are shown in Fig. 1.

Historical settlement reconstruction and cohort definition

Settlement reconstruction preceded all landscape analysis. The staged source universe contained 1,416 documentary/source-era records: 422 for Zhanhua, 490 for Lijin, 214 for Dongying District, 178 for Hekou, and 112 in a Kenli subregional source pool. Principal documentary anchors were the Zhanhua Gazetteer of Geographical Names [20], Lijin County Gazetteer [21], Dongying District Gazetteer [22], and Hekou District Gazetteer [23]. The Kenli pool is retained as a subregional source frame rather than relabeled as a full-county gazetteer universe. “Historically documented” denotes independent documentary provenance, not a uniform founding-date threshold.

Eligibility was based on pre-specified criteria: traceable documentary provenance, source-era rural-settlement identity, a defensible physical point, and point-in-polygon study-area membership. Chronology was descriptive rather than an eligibility requirement. The 249 coordinate-bearing settlements comprised 31 PRE1855, 135 POST1855, 82 chronology-HOLD, and one AMBIGUOUS-1855 record. Regional totals were Zhanhua 17, Lijin 136, Dongying District 3, Hekou 8, and Kenli 85, reflecting source recovery and spatial eligibility rather than proportional sampling. The earlier 714-settlement benchmark remained external to cohort membership.

Regional accounting from staged documentary/source-era records to the final coordinate-bearing cohorts is summarized in Table 1.

Table 1
Regional accounting from staged documentary/source-era records to the final coordinate-bearing cohorts.

Region	Staged source records	MAIN	Sensitivity	Final coordinate-bearing	MAIN / staged source records (%)
Zhanhua (ZH)	422	2	15	17	0.47
Lijin (LJ)	490	121	15	136	24.69
Dongying District (DY)	214	0	3	3	0.00
Hekou (HK)	178	7	1	8	3.93
Kenli source lane (KL)	112	80	5	85	71.43
Total	**1416**	**210**	**39**	**249**	**14.83**

Remote-sensing and geospatial data used to construct indicators

Primary land-change outcomes were derived from remotely sensed land-cover and spectral products, with ancillary environmental covariates obtained from open or documented geospatial sources. GLC_FCS30D supplied 30-m land cover [24], and Landsat Collection 2 Level-2 surface reflectance

supplied the 2000/2022 spectral composites [25]. JRC Global Surface Water provided water/coastal context [26]. WorldClim 2, WorldPop, and Copernicus DEM provided climate, population, and terrain baselines [19, 27, 28], while the China coastal aquaculture dataset supplied 1998 aquaculture context [29]. The analysis therefore combines multi-temporal remote-sensing products with baseline environmental context rather than treating any single sensor or index as a complete explanation of landscape change.

Detailed geospatial data sources and their designated temporal roles are listed in Supplementary Table S9 (Online Resource 1).

Only baseline, lagged, or static information was permitted in the primary predictor matrices.

Spatial support and stable-land mask

Each settlement point was buffered at 1, 3, and 5 km in projected space; 3 km was the pre-specified primary scale and 1/5 km were sensitivities. Geometry QA verified point containment, buffer nesting, and raster coverage. Endpoint land-cover outcomes used a stable-land mask retaining pixels classified as land at both endpoints while excluding open water and permanent snow/ice. Newly accreted land was mapped separately and never entered the stable-land denominator, preventing shoreline accretion from mechanically altering inland change denominators.

Shared raster support was quantified for all 21,945 pairs of MAIN 3-km buffers using center distance, intersection area, and overlap relative to mean buffer area. Low-overlap sensitivity used a deterministic greedy 4-km thinning: candidates were retained only if at least 4.0 km from every previously retained center. Six outcome-independent orderings were tested (ID, east/west, north/south, and fixed pseudorandom seed 20260822), retaining 43–50 sites. For equal 3-km circles, 4-km separation limits theoretical overlap to 21.91%. Additional 3- and 6-km checks are supplementary; inference emphasizes consistency across all six 4-km orderings.

Landscape-change monitoring metrics

Five remotely sensed landscape-change metrics were retained regardless of predictive success. Y1 measures 2000–2022 impervious/built-up share change on stable land; Y2 measures proportional loss of baseline cropland; Y3 summarizes fragmentation change from configuration metrics reduced before modeling; Y4 summarizes vegetation/spectral-stress change; and HSLSI is a descriptive PCA synthesis. These metrics were treated as complementary monitoring outputs rather than as co-equal predictive indicators. Their recommended use was determined after the reliability checks described below.

For Y4, Landsat 5 TM (2000) and Landsat 8 OLI (2022) were restricted to WRS 121/034, 1 September–31 October, cloud cover ≤ 10%, masked with QA_PIXEL, scaled according to Collection 2 Level-2 specifications, and summarized by buffer. Because its two components later proved nearly rank-identical

and SRSI contains NDVI, Y4 is treated as an exploratory indicator whose construction is explicitly tested for coupling.

HSLSI was the PC1 synthesis of Y1–Y4 (75.10% of joint variance at 3 km). Because it inherits Y4 coupling and component dependence, it is evaluated as a composite indicator but is not granted a higher evidential status simply because it integrates multiple components.

The five candidate metrics and their recommended monitoring roles are summarized in Table 2.

Table 2
Candidate landscape-change metrics and recommended monitoring roles after reliability assessment.

Indicator	Meaning	Frozen definition	Analysis n	Recommended monitoring role
Y1	Impervious/built change	2022 share − 2000 share on stable land	210	Direct descriptive reporting
Y2	Cropland loss rate	2000 cropland converted to non-cropland / 2000 cropland	210	Descriptive / diagnostic only
Y3	Fragmentation pressure	PC1 of ΔED and − ΔLPI	210	Conditional comparative screening
Y4	Vegetation/spectral-salinity PC1	PC1 of NDVI degradation and ΔSRSI	208	Exploratory diagnostic only
HSLSI	Composite landscape stress	PC1 of Y1–Y4	208	Descriptive composite only

Reliability checks and monitoring interpretation

The analysis used five established checks. Temporal independence was protected by excluding response-period and outcome-derived predictors. Support dependence was quantified from neighborhood overlap and, for the metric that retained blocked skill, from outcome-independent low-overlap thinnings. Out-of-area performance was evaluated with blocked spatial cross-validation. Environmental information was compared with coordinate-only and coordinate-plus-environment baselines. Mathematical-coupling sensitivities, regional consistency, block-size sensitivity, thinning robustness, and residual spatial dependence were then used to define the limits of interpretation.

The component analyses, predictor matrices, spatial folds, and sensitivity rules were fixed before interpretation. None of these checks was used to select endpoints, predictors, or model settings after seeing the results. We distinguish direct descriptive reporting at observed locations, diagnostic use when metric construction or inference is sensitive, and conditional comparative screening when spatial transfer is positive but important limits remain.

The five reliability checks and the limitation addressed by each are summarized in Table 3.

Table 3
Reliability checks used in the analysis and the limitation addressed by each check.

Check	Monitoring-quality threat	Operational test	Interpretive purpose
Temporal independence	Temporal leakage	Use only static, lagged, or baseline predictors; lock folds before change-metric linkage.	Protect chronology of predictive monitoring claims.
Support dependence	Shared raster support / pseudo-replication	Quantify overlap; use outcome-independent low-overlap thinning for the metric that retains transfer skill.	Reveal effective spatial replication and sensitivity to shared support.
Blocked spatial transfer	Spatial interpolation	Evaluate pooled and fold-level out-of-area skill under grouped spatial cross-validation.	Separate direct description from metrics suitable for comparative screening.
Beyond geography	Broad spatial gradients	Compare coordinate-only, coordinate + companion covariates, and coordinate + focal predictor models.	Test whether environmental context adds held-out information beyond location.
Robustness boundary	Coupling and unresolved spatial dependence	Audit mathematical coupling, regional consistency, block scale, thinning, MGWR/importance consistency, and residual Moran's I.	Bound extrapolation and management interpretation when dependence remains.

Baseline predictor matrix and temporal leakage control

The predictor set was fixed before outcome attachment and restricted to static, lagged, or baseline states. Both primary branches included baseline cropland share, distance to the 2000 coastline, 1970–2000 annual precipitation, 2000 built-up share, and 1998 aquaculture share. Path A additionally included SRSI2000, whereas Path B substituted elevation. These two physical-context variables were pre-specified alternatives and were not selected post hoc according to outcome performance.

Potentially attractive but temporally invalid predictors were excluded. No 2000–2022 change variable, 2022 land-cover state, current road network, post-2000 restoration layer, or outcome-derived quantity was allowed to re-enter X. A current protected-area boundary was also held out because historical boundary comparability had not been established. This restriction separated baseline susceptibility or context from contemporaneous response.

Spatial cross-validation and tree models

Geographic predictive skill was evaluated with grouped spatial cross-validation rather than random folds because outcomes and predictors were spatially structured [7–10]. Coordinates were projected to UTM 50N (EPSG:32650). A 15-km grid, defined before outcome attachment, contained 17 occupied blocks assigned deterministically to five outer folds (48, 41, 41, 40, 40 observations) with four grouped inner folds. Pre-defined 10- and 20-km geometries served as sensitivities; the latter produced three outer folds. The 15-km distance is a design choice, not an estimated decorrelation optimum.

XGBoost [30] and random forest [31] were fitted in raw predictor units. Out-of-fold RMSE, MAE, R^2, and Spearman correlation were retained for every branch/algorithm combination, with fold-level metrics used to expose heterogeneity. Held-out permutation importance refitted the archived fold-specific hyperparameters and permuted each feature 30 times within the outer test fold; importance was summarized as the held-out RMSE increase and number of folds with positive importance.

Three sensitivities separated environmental information from broad geography. First, descriptive 10/15/20-km reruns held fixed the candidate selected by mean archived inner-CV RMSE across the original 15-km outer folds. Second, a quadratic-coordinate baseline standardized UTM x/y within each outer training fold and predicted the held-out fold. Third, matched 15-km nested comparisons separately fitted coordinates only, coordinates plus the five non-SRSI Path A covariates, and coordinates plus all six Path A covariates; parallel linear models added the same terms stepwise. These tests ask whether the environmental set, and SRSI2000 specifically, adds held-out information beyond location.

Regional consistency and predictor-only spatial diagnostics

Because 201/210 MAIN sites were in Lijin or the Kenli source lane, Path A consistency was checked separately within those regions using raw SRSI2000–Y3 Spearman correlations and correlations after residualizing both variables against within-region quadratic coordinate surfaces. SRSI2000 spatial structure was also summarized against easting, northing, coastline distance, and by blocked coordinate-only prediction. An outcome-free predictor-similarity diagnostic z-standardized the six Path A predictors and summarized mean pair similarity by distance band; it describes baseline spatial structure but does not retrospectively justify the 15-km fold size.

Mathematical-coupling audit

Several outcomes contain a baseline quantity that is mathematically related to a candidate baseline predictor. This was treated as mathematical coupling rather than future-information leakage. Before model interpretation, mandatory sensitivity models removed 2000 built-up share from Y1, baseline cropland from Y2, SRSI2000 from Y4 Path A, and corresponding baseline quantities from HSLSI analyses. These tests determined whether an apparently predictive association remained after removing the shared baseline term.

MGWR and spatial stability

MGWR examined multiscale spatial associations [32, 33], not out-of-sample prediction. Coordinates were UTM x/y and X/y were standardized within each fit. The archived mgwr 2.2.1 analysis used multiscale bandwidth selection (minimum 30 neighbors) and a full fit plus five leave-one-fold-out refits for each branch/outcome. These refits are perturbation sensitivities, not independent replications; “6/6 sign consistency” therefore means agreement of one full fit and five overlapping omissions. Minimum-limited bandwidths were not interpreted as process radii.

Spatial autocorrelation and software

Global structure and out-of-fold residual dependence were assessed with Moran's I [34] using an 8-nearest-neighbor graph and 199 permutations. Analyses were run in Python 3 with pandas, NumPy, SciPy, GeoPandas, rasterio/GDAL, scikit-learn, XGBoost 3.1.3, and mgwr 2.2.1; stochastic algorithms used seed 20260822. Versioned scripts and outputs for overlap, thinning, regional checks, predictor similarity, incremental comparisons, block geometry, coordinate trends, and permutation importance are supplied as Supplementary Materials.

OpenAI ChatGPT (GPT-5.6 Sol; accessed 13 September 2026) was used as an assistive tool for drafting and language refinement, manuscript organization, documentation, code troubleshooting, and quality-control scripting. All scientific questions, analytical decisions, literature selection and interpretation, code, parameter choices, numerical results, figures, and final wording were independently checked and approved by the authors. The tool was not used to generate, alter, or fabricate observational data, statistical results, figures, or citations. The authors take full responsibility for the content of the manuscript.

The overall analysis workflow is summarized in Fig. 2.

Results

Temporal controls and shared spatial support

Temporal independence was satisfied by design: the primary predictor matrices contained only static, lagged, or 2000 baseline information, and spatial folds were locked before the change metrics were linked. The five metrics were retained regardless of later performance. MAIN contained 210 settlements plus 39 sensitivity sites. Lijin contributed 121 MAIN sites and the Kenli source lane 80 (201/210; 95.7%); Hekou contributed seven, Zhanhua two, and Dongying District none. This regional imbalance defines the population boundary. Descriptive distributions are reported in Supplementary Table S10.

The support audit identified strong dependence among monitoring units. Of 21,945 MAIN-settlement pairs, 2,036 had intersecting 3-km buffers; 1,054 exceeded 25% overlap, 505 exceeded 50%, and 123 exceeded 75%. Median nearest-neighbor distance was 1.000 km and median maximum per-site overlap was 0.789, so n = 210 is not equivalent to 210 independent raster supports. The six outcome-independent 4-km thinnings retained 43–50 sites. For fragmentation, which retained blocked skill, Path A

R^2 remained positive for both algorithms in every ordering (XGBoost 0.086–0.347; random forest 0.232–0.422), while residual Moran's I fell markedly but was not always eliminated (Figure S7).

Descriptive distributions for all five candidate indicators are provided in Supplementary Table S10.

Spatially blocked evaluation of five monitoring metrics

All five landscape-change metrics were spatially autocorrelated before modeling (Moran's I = 0.626–0.840; all permutation p = 0.001), supporting blocked rather than random validation. Only fragmentation (Y3) met the strict transfer screen of positive pooled out-of-fold R^2 in both physical-context branches and both algorithms. Path A XGBoost achieved R^2 = 0.388 and Spearman ρ = 0.761 (random forest R^2 = 0.339); Path B random forest achieved R^2 = 0.252 and ρ = 0.697 (XGBoost R^2 = 0.241). The other metrics reached at most R^2 = 0.074. The outcome-free Path A predictor-similarity diagnostic decayed with distance, from 0.659 at 0–5 km to 0.006 at 15–20 km, confirming that spatial separation materially changes the monitoring problem.

Fragmentation retained information beyond broad geography

Positive pooled transfer did not imply uniformly strong performance. Y3 fold-specific R^2 varied sharply (Path A XGBoost − 0.837 to 0.513; Path B random forest − 1.766 to 0.434), so the pooled result represents average out-of-area screening skill rather than reliable prediction in every held-out region.

Broad geography remained a strong competing explanation. The quadratic-coordinate baseline achieved Y3 R^2 = 0.340. In linear models, adding the five non-SRSI Path A covariates increased R^2 to 0.458 and adding SRSI2000 increased it to 0.563 (SRSI-specific ΔR^2 = +0.106). In matched tree comparisons, SRSI2000 increased R^2 from 0.333 to 0.533 for XGBoost and from 0.311 to 0.451 for random forest. Model class still mattered: coordinate-only random forest reached R^2 = 0.493, slightly above its coordinate-plus-all-environment counterpart. Environmental context therefore added held-out information in some comparisons, but location alone already captured substantial structure.

The compact 15-km spatial-CV transfer screen is summarized in Table 4.

Table 4
Compact 15-km nested spatial-CV transfer screen for the candidate indicators. Within each indicator/branch, the displayed algorithm has the lower pooled OOF RMSE; all 20 pre-specified XGBoost/random-forest runs are reported in Supplementary Table S1 and Fig. 3.

Indicator	Branch	Algorithm	n	OOF RMSE	OOF R^2	Spearman ρ	Residual Moran's I
Y1	Path A	XGB	210	3.829	-0.114	0.439	0.820
Y1	Path B	RF	210	4.045	-0.243	0.234	0.845
Y2	Path A	RF	210	5.117	0.074	0.435	0.657
Y2	Path B	RF	210	5.265	0.020	0.479	0.658
Y3	Path A	XGB	210	0.780	0.388	0.761	0.570
Y3	Path B	RF	210	0.863	0.252	0.697	0.514
Y4	Path A	XGB	208	1.000	-0.004	0.393	0.676
Y4	Path B	XGB	208	0.975	0.044	0.469	0.657
HSLSI	Path A	XGB	208	0.977	0.042	0.402	0.741
HSLSI	Path B	RF	208	1.112	-0.243	0.303	0.772

The compact table summarizes the spatial-transfer screen; no tree algorithm is treated as an inferential winner.

Robustness checks supported conditional monitoring use

Regional, coupling, scale, thinning, and residual-dependence checks were then used to bound monitoring interpretation. SRSI2000 showed the most consistent Path A evidence while remaining spatially structured. Its MGWR association with Y3 was negative (mean standardized β = −0.719; local range −1.735 to − 0.290; bandwidth 48 neighbors), and the sign agreed in the full fit and five leave-one-fold-out refits. Held-out permutation importance ranked SRSI2000 first in both tree algorithms, with positive importance in all five folds. SRSI2000 correlated with easting (ρ = 0.665) and coastline distance (ρ = −0.672), and coordinates alone predicted it with blocked R^2 = 0.411. Yet the Y3 association remained negative within Lijin and Kenli and after within-region coordinate residualization. Together with the six low-overlap thinnings, these checks support conditional use without establishing a salinity mechanism.

Path B elevation had a positive MGWR association (mean β = 0.721; 6/6 sign consistency; bandwidth 46) but weaker predictive and thinning support. Coastline distance was negative in both branches at broad bandwidths. Because all 20 primary tree runs nevertheless retained significant out-of-fold residual clustering (Moran's I = 0.514–0.845; all permutation p = 0.005), fragmentation is appropriate only for conditional comparative screening within a similar spatial domain. Detailed MGWR, held-out importance, and sensitivity evidence are provided in Supplementary Table S11 and associated sheets.

Detailed Y3 MGWR and sensitivity evidence is reported in Supplementary Table S11 and the supplementary workbook.

Full-fit MGWR coefficient patterns for both physical-context branches are shown in Fig. 4a and Fig. 4b.

Monitoring-use categories separated descriptive, diagnostic, and screening roles

The monitoring-use classification separates different practical roles. Y1 supports direct descriptive reporting because blocked R^2 was negative. Y2 is diagnostic/coupling-sensitive because its weak signal did not survive removal of denominator-related baseline cropland. Y4 is exploratory because it depended on a mathematically related SRSI component and its two components were nearly rank-identical (ρ = 0.999813). HSLSI remains a descriptive composite because it inherits those dependencies. Y3 alone supports conditional comparative screening; no metric supports unrestricted extrapolation because residual spatial dependence remains significant.

Residual dependence limits extrapolation

All 20 primary tree runs retained significant out-of-fold residual spatial clustering (Moran's I = 0.514–0.845; all permutation p = 0.005). Feature importance, MGWR coefficients, and SRSI2000 are therefore interpreted as predictive or spatial-association evidence within the reconstructed cohort, not causal effects and not a basis for unrestricted delta-wide prediction.

Discussion

Shared raster support changes monitoring reliability

Landscape-metric research has long emphasized that scale, resolution, redundancy, and metric construction constrain ecological interpretation [3–6]. Our results identify an additional sampling issue for place-centered remote sensing: the named observation unit and the raster support are not necessarily the same unit. Among 210 settlements, the 3-km buffers frequently reused the same pixels, and six outcome-independent 4-km thinnings retained only 43–50 sites. Thus, nominal place counts can substantially exceed the number of weakly overlapping spatial supports. This matters because a metric may be valid for describing change at each named place while still providing less independent evidence for comparative screening. In our case, four metrics remained descriptive or diagnostic, whereas fragmentation alone retained positive blocked skill across both predictor branches and both algorithms.

Spatial validation is necessary but not sufficient

Blocked spatial cross-validation is already established as a safeguard against optimistic model assessment in structured data [7–10]. Shared-support auditing addresses a different question: whether nominal observations partially duplicate the same raster evidence before training-test separation. Blocking addresses geographic proximity between training and test locations, whereas overlap auditing

and thinning address reused support among place-centered units. In this study, the combination of these checks materially changed interpretation. Fragmentation retained positive skill after reducing overlap and in spatially separated folds, but location alone explained substantial structure and residual Moran's I remained significant. The evidence therefore supports conditional screening rather than unrestricted prediction.

Because the predictor matrices, spatial folds, and thinning rules were fixed before interpretation, the main conclusion does not depend on choosing a favorable overlap threshold or validation split after observing outcomes. Nevertheless, the thinned samples are sensitivity analyses, not independent replications, and the design does not substitute for external validation. The reported R^2 values, block size, and screening interpretation should therefore be regarded as properties of the reconstructed YRD cohort until the same shared-support and spatial-transfer tests are repeated in an independent monitoring region.

Fragmentation and the limits of spatial transfer

Fragmentation is the strongest candidate for comparative screening, but the result is narrower than a general fragmentation model for the YRD. Kong et al. [11] showed that fragmentation patterns and their apparent drivers in the delta depend on analytical scale. Here, positive blocked skill, incremental information beyond coordinates, within-region consistency, and persistence across six low-overlap thinnings indicate that the settlement-centered signal is not explained solely by short-range overlap or a simple coordinate trend. Significant residual Moran's I nevertheless remained in every full-cohort model. The evidence therefore supports cautious comparison within similar spatial settings, not universal prediction or causal attribution.

Physical interpretation must also remain limited. Field and remote-sensing studies consistently report strong salinity gradients in the YRD [12–15, 18, 35]. Our SRSI2000 layer was not calibrated against local electrical conductivity, and the index contains NDVI. Its association with fragmentation may therefore reflect a mixture of vegetation condition, background reflectance, moisture, land cover, and salinity-related variation. We use it as a baseline spectral context variable, not as a measurement of soil salinity.

Shared support and effective replication

The overlap analysis makes the effective-sample problem concrete. Although the cohort contained 210 settlements, outcome-independent 4-km thinning retained only 43–50 sites. For metrics derived from continuous raster surfaces, the observation unit is therefore defined partly by spatial support rather than by the number of place names. This matters whenever monitoring relies on buffers, moving windows, catchments, or neighboring administrative units. Fragmentation remained positive in all six thinning orders, which supports persistence of the result, but the thinned samples are sensitivity analyses rather than independent replications.

Geographic structure and environmental interpretation

The coordinate benchmark shows why broad geography cannot be treated only as a nuisance term. A quadratic coordinate model explained $R^2 = 0.340$ for fragmentation, and SRSI2000 itself was spatially predictable. Some of the apparent environmental signal was therefore already available from location. This is consistent with the wider spatial-validation literature, in which geographic structure can inflate apparent prediction when training and test data are not sufficiently separated [7, 8, 10]. Environmental variables used for prioritization should consequently add held-out information beyond a spatial-only benchmark.

Coastline distance was negatively associated with fragmentation in both branches at broad MGWR bandwidths. That pattern is compatible with previous descriptions of coast–inland differences in wetland configuration, salinity, and land use [2, 11, 17], but the sampling frame prevents a delta-wide interpretation. Lijin and the Kenli source lane contribute 201 of the 210 MAIN sites, whereas Dongying District contributes none. The regional checks reduce the chance that the Path A association is created only by mixing the two dominant regions; they do not correct the underlying imbalance.

What the weak metrics contribute

The four weaker metrics clarify what information is lost when descriptive measures are treated as predictors. Built-up change can be reported directly even though the baseline models do not support spatial ranking. The cropland-loss result depends partly on a denominator-linked baseline term. The two components of the spectral metric are almost rank-identical, and that dependence is carried into the composite HSLSI. These are not failed measurements; they are warnings against using the same metric interchangeably for description, diagnosis, and prediction.

Implications for monitoring design

Remote-sensing monitoring studies increasingly combine land-cover products and machine learning to support environmental management [1, 2]. Our results suggest that the intended management use should be specified before a model is interpreted. Directly observed metrics can support reporting at monitored settlements. Metrics affected by coupling or construction problems are more useful as diagnostics. Fragmentation can be used as a secondary screening layer within a similar spatial domain, provided that residual dependence and uncertainty are retained in the decision process. Supplementary Table S12 states the corresponding restrictions.

From monitoring evidence to operational use

For implementation, the distinction is between observing change and transferring a model. The first requires consistent remote-sensing products and transparent metric definitions. The second additionally requires evidence that a metric retains skill when locations are spatially separated and that performance is not explained largely by geography or shared support. Expansion to new districts or deltas would therefore require a new assessment of support size, block geometry, spatial baselines, and residual dependence rather than carrying the present model forward unchanged.

Application in the Yellow River Delta

In the present YRD cohort, fragmentation is the only metric that justifies conditional comparison among spatially separated settlements. Built-up change and cropland loss remain useful as observed land-use-change measures, while the spectral and composite metrics need further calibration before they are used predictively. This interpretation is deliberately narrower than a single delta-wide ranking: it separates quantities that can be reported from those that can support screening and from those that still need methodological or field validation.

Limitations and external validation needs

The main limitation is external validity. The cohort is regionally imbalanced, the 3-km supports overlap extensively, chronology is incomplete, and SRSI2000 is an uncalibrated spectral proxy that contains NDVI. The 15-km blocks were fixed as a design choice rather than estimated from an empirical decorrelation range; the 20-km sensitivity contains only three outer folds; and residual spatial autocorrelation persists. Most importantly, the analysis has not been repeated in an independent region. The numerical performance reported here should therefore not be transferred directly to another monitoring network. A new application would need to reassess support size and block scale, confirm data comparability, and repeat the geographic-baseline and residual-dependence analyses.

Conclusions

The five metrics did not support the same kind of monitoring inference. Built-up change, cropland loss, spectral stress, and the composite stress measure are most defensible as descriptive or diagnostic outputs in this cohort. Fragmentation alone retained positive blocked skill in both predictor branches and both tree algorithms, added information beyond broad geography in matched comparisons, and remained positive after overlap-reducing thinning. Residual spatial dependence, however, rules out unrestricted extrapolation. The result is therefore a conditional screening signal rather than a delta-wide prediction model. For settlement-centered remote-sensing studies, the main implication is methodological but conventional: the spatial support of the observations, the timing of predictors, a geographic baseline, and residual autocorrelation should be examined before model outputs are used to rank or prioritize places.

Online Resource 1 (ESM_1_Supplementary_Materials.zip): The peer-review supplementary package contains complete nested spatial-CV metrics, outer-fold results, coupling sensitivities, MGWR diagnostics, overlap analyses, all six 4-km thinning selections, and regional SRSI2000–Y3 checks. It also includes predictor-similarity diagnostics, matched coordinate/environment models, held-out permutation importance, block-geometry sensitivities, analysis scripts, Supplementary Tables S1–S12, Supplementary Figures S1–S7, and the monitoring-use matrix. A private repository link to the derived settlement-level cohort, fixed predictor matrices, model outputs, and provenance/quality-control tables will be supplied through the submission system. The repository will be made public with a permanent DOI upon acceptance.

Declarations

Funding:

This research received no external funding.

Ethical responsibilities: All authors have read, understood, and have complied as applicable with the statement on “Ethical responsibilities of Authors” as found in the Instructions for Authors.

Ethics approval: Not applicable. The study used historical documentary sources and geospatial/environmental data and did not involve human participants, animals, interventions, or identifiable personal data.

CRediT authorship contribution statement: Conceptualization, P.Y.; Methodology, P.Y. and J.H.; Software, X.Z.; Validation, P.Y. and X.Z.; Formal analysis, P.Y.; Investigation, P.Y.; Resources, J.H.; Writing – original draft, P.Y., J.H. and X.Z.; Writing – review & editing, P.Y., J.H. and X.Z.; Visualization, P.Y. and X.Z.

Data availability: Public source datasets are available from their respective providers, including GLC_FCS30D, USGS Landsat Collection 2, JRC Global Surface Water, WorldClim 2, WorldPop, Copernicus DEM, and the China coastal aquaculture dataset. Derived settlement-level data, fixed predictor matrices, spatial-CV folds, model outputs, and provenance/quality-control materials are available to editors and reviewers through a private Zenodo link provided in the submission system. Upon acceptance, the repository will be made public and its permanent DOI will be added to the article. Copyrighted gazetteer scans or page images are not redistributed.

Author Contribution

Conceptualization, P.Y.; Methodology, P.Y. and J.H.; Software, X.Z.; Validation, P.Y. and X.Z.; Formal analysis, P.Y.; Investigation, P.Y.; Resources, J.H.; Writing – original draft, P.Y., J.H. and X.Z.; Writing – review & editing, P.Y., J.H. and X.Z.; Visualization, P.Y. and X.Z. All authors read and approved the final manuscript.

Acknowledgments:

None.

Data Availability

Public source datasets are available from their respective providers, including GLC_FCS30D, USGS Landsat Collection 2, JRC Global Surface Water, WorldClim 2, WorldPop, Copernicus DEM, and the China coastal aquaculture dataset. Derived settlement-level data, fixed predictor matrices, spatial-CV folds,

model outputs, and provenance/quality-control materials are available to editors and reviewers through a private Zenodo link provided in the submission system. Upon acceptance, the repository will be made public and its permanent DOI will be added to the article. Copyrighted gazetteer scans or page images are not redistributed.

Code availability:

Analysis scripts supporting the reported sensitivity, spatial-robustness, and incremental-value checks are included in Online Resource 1 and in the peer-review repository. The repository will be made public with a permanent DOI upon acceptance.

Competing interests: The authors declare no competing financial or non-financial interests related to this work.

References

1. Lizcano Toledo, M. V. et al. Machine learning for land use change analysis in environmental protection areas. *Environ. Monit. Assess.* **198**, 481. https://doi.org/10.1007/s10661-026-15280-7 (2026).
2. Zhu, Y. et al. Monitoring land use changes in the Yellow River Delta using multi-temporal remote sensing data and machine learning from 2000 to 2020. *Remote Sens.* **16** (11), 1946. https://doi.org/10.3390/rs16111946 (2024).
3. Lausch, A. & Herzog, F. Applicability of landscape metrics for the monitoring of landscape change: issues of scale, resolution and interpretability. *Ecol. Indic.* 2(1–2), 3–15 (2002). https://doi.org/10.1016/S1470-160X(02)00053-5
4. Li, H. & Wu, J. Use and misuse of landscape indices. *Landsc. Ecol.* **19** (4), 389–399. https://doi.org/10.1023/B:LAND.0000030441.15628.d6 (2004).
5. Cushman, S. A., McGarigal, K. & Neel, M. C. Parsimony in landscape metrics: strength, universality, and consistency. *Ecol. Indic.* **8** (5), 691–703. https://doi.org/10.1016/j.ecolind.2007.12.002 (2008).
6. Wu, J. Effects of changing scale on landscape pattern analysis: scaling relations. *Landsc. Ecol.* **19** (2), 125–138. https://doi.org/10.1023/B:LAND.0000021711.40074.ae (2004).
7. Roberts, D. R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929. https://doi.org/10.1111/ecog.02881 (2017).
8. Meyer, H., Reudenbach, C., Hengl, T., Katurji, M. & Nauss, T. Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. *Environ. Model. Softw.* **101**, 1–9. https://doi.org/10.1016/j.envsoft.2017.12.001 (2018).
9. Valavi, R., Elith, J., Lahoz-Monfort, J. J. & Guillera-Arroita, G. blockCV: an R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods Ecol. Evol.* **10** (2), 225–232. https://doi.org/10.1111/2041-210X.13107 (2019).

10. Ploton, P. et al. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nat. Commun.* **11**, 4540. https://doi.org/10.1038/s41467-020-18321-y (2020).
11. Kong, X. et al. Assessment of coastal landscape fragmentation and its driving factors based on optimal scale: a case study of the Yellow River Delta, China. *Ecol. Indic.* **166**, 112537. https://doi.org/10.1016/j.ecolind.2024.112537 (2024).
12. Guo, B. et al. Salinization information extraction model based on VI–SI feature space combinations in the Yellow River Delta based on Landsat 8 OLI image. *Geomat. Nat. Hazards Risk*. **10** (1), 1863–1878. https://doi.org/10.1080/19475705.2019.1650125 (2019).
13. Guo, B. et al. Mapping soil salinity using a combination of vegetation index time series and single-temporal remote sensing images in the Yellow River Delta, China. *Catena* **231**, 107313. https://doi.org/10.1016/j.catena.2023.107313 (2023).
14. Li, Y., Chang, C., Wang, Z. & Zhao, G. Upscaling remote sensing inversion and dynamic monitoring of soil salinization in the Yellow River Delta, China. *Ecol. Indic.* **148**, 110087. https://doi.org/10.1016/j.ecolind.2023.110087 (2023).
15. Qiu, H. et al. Evolutionary patterns of soil salinization and saline-alkali land types in the Yellow River Delta: a spatiotemporal analysis from 2002 to 2022. *Soil. Tillage Res.* **262**, 107222. https://doi.org/10.1016/j.still.2026.107222 (2026).
16. Liu, G. et al. GIS-mapping spatial distribution of soil salinity for eco-restoring the Yellow River Delta in combination with electromagnetic induction. *Ecol. Eng.* **94**, 306–314. https://doi.org/10.1016/j.ecoleng.2016.05.037 (2016).
17. Qiu, M. et al. Spatio-temporal changes and hydrological forces of wetland landscape pattern in the Yellow River Delta during 1986–2022. *Landsc. Ecol.* **39**, 51. https://doi.org/10.1007/s10980-024-01850-y (2024).
18. Song, C., Ren, H. & Huang, C. Estimating soil salinity in the Yellow River Delta, Eastern China—an integrated approach using spectral and terrain indices with the generalized additive model. *Pedosphere* **26** (5), 626–635. https://doi.org/10.1016/S1002-0160(15)60071-6 (2016).
19. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol*. **37**, 4302–4315. https://doi.org/10.1002/joc.5086 (2017).
20. Zhanhua County Place Names Committee Office. □□□□□□□□□□ *[Gazetteer of Geographical Names of Zhanhua County, Shandong Province]* (Ocean, 1988). (In Chinese).
21. Shandong Province Lijin County Local History Compilation Committee. □□□□ [Lijin County Gazetteer] (Oriental, (1990). (In Chinese).
22. Dongying District Local History Compilation Committee. *Dongying City.* □□□□ *[Dongying District Gazetteer]* (Zhonghua Book Company, 2000). (In Chinese).
23. Hekou District Local History Compilation Committee. *Dongying City.* □□□□□□□ *[Hekou District Gazetteer]* (Zhonghua Book Company, 2002). (In Chinese).
24. Zhang, X. et al. GLC_FCS30D: the first global 30 m land-cover dynamics monitoring product with a fine classification system for the period from 1985 to 2022 generated using dense-time-series

Landsat imagery and the continuous change-detection method. *Earth Syst. Sci. Data*. **16**, 1353–1381. https://doi.org/10.5194/essd-16-1353-2024 (2024).

25. Wulder, M. A. et al. Current status of Landsat program, science, and applications. *Remote Sens. Environ.* **225**, 127–147. https://doi.org/10.1016/j.rse.2019.02.015 (2019).
26. Pekel, J. F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540**, 418–422. https://doi.org/10.1038/nature20584 (2016).
27. Copernicus Data Space Ecosystem. Copernicus DEM GLO-30 [Data set]. Copernicus Data Space Ecosystem. https://doi.org/10.5270/ESA-c5d3d65 (2021).
28. Lloyd, C. T. et al. Global spatio-temporally harmonised datasets for producing high-resolution gridded population distribution datasets. *Big Earth Data*. **3** (2), 108–139. https://doi.org/10.1080/20964471.2019.1625151 (2019).
29. Duan, Y. et al. Tracking changes in aquaculture ponds on the China coast using 30 years of Landsat images. *Int. J. Appl. Earth Obs Geoinf.* **102**, 102383. https://doi.org/10.1016/j.jag.2021.102383 (2021).
30. Chen, T. & Guestrin, C. XGBoost: a scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 785–794 (2016). https://doi.org/10.1145/2939672.2939785
31. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32. https://doi.org/10.1023/A:1010933404324 (2001).
32. Fotheringham, A. S., Yang, W. & Kang, W. Multiscale geographically weighted regression (MGWR). *Ann. Am. Assoc. Geogr.* **107** (6), 1247–1265. https://doi.org/10.1080/24694452.2017.1352480 (2017).
33. Oshan, T. M., Li, Z., Kang, W., Wolf, L. J. & Fotheringham, A. S. mgwr: a Python implementation of multiscale geographically weighted regression for investigating process spatial heterogeneity and scale. *ISPRS Int. J. Geo-Inf*. **8** (6), 269. https://doi.org/10.3390/ijgi8060269 (2019).
34. Moran, P. A. P. Notes on continuous stochastic phenomena. *Biometrika* **37** (1–2), 17–23. https://doi.org/10.1093/biomet/37.1-2.17 (1950).
35. Weng, Y. L., Gong, P. & Zhu, Z. L. A spectral index for estimating soil salinity in the Yellow River Delta region of China using EO-1 Hyperion data. *Pedosphere* **20** (3), 378–388. https://doi.org/10.1016/S1002-0160(10)60027-6 (2010).

Figures

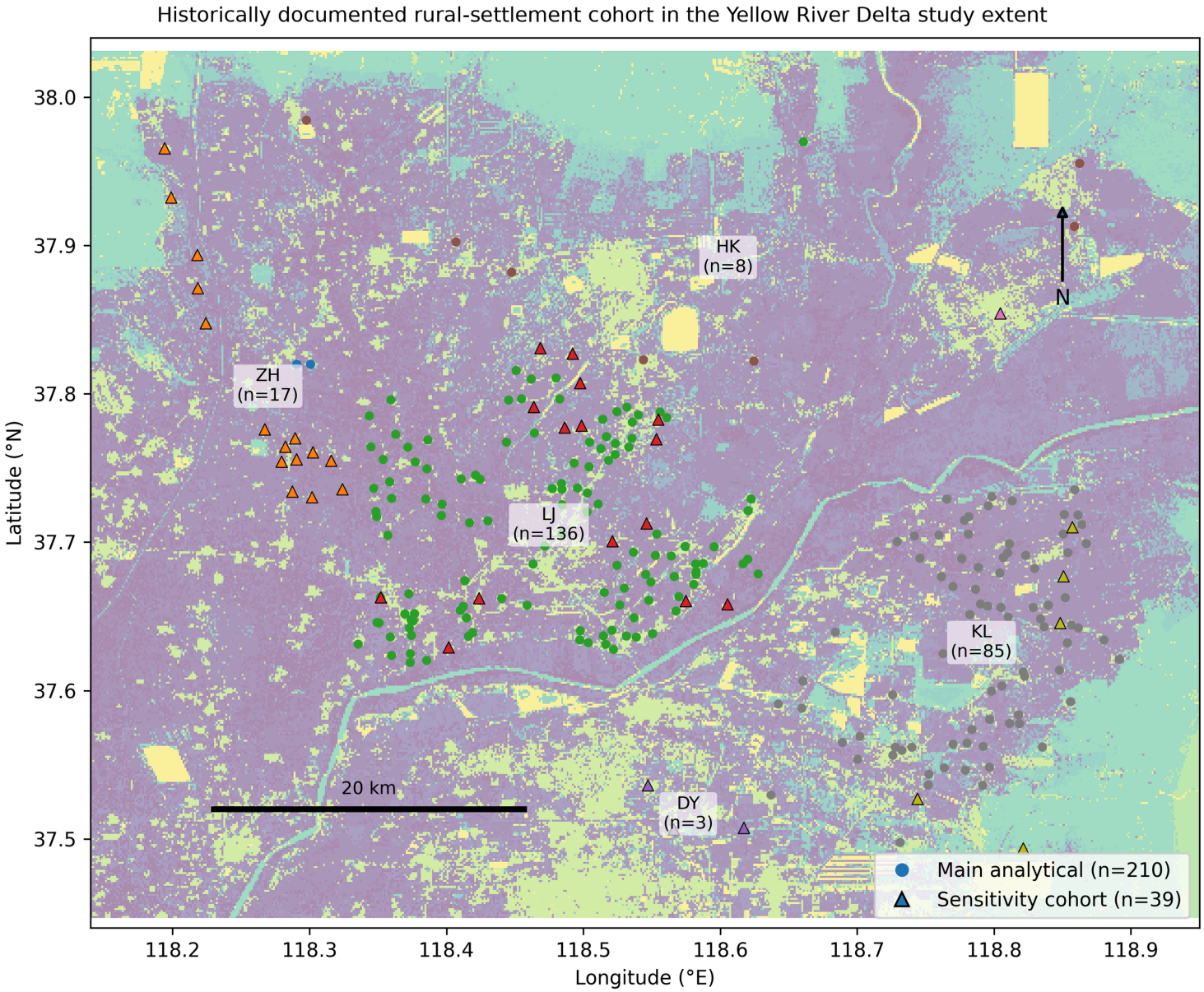


Figure 1

Fig. 1 Historically documented rural-settlement cohort in the Yellow River Delta study extent. Circles show the 210 MAIN analytical settlements; triangles show the 39 sensitivity-cohort settlements. Regional labels give MAIN + sensitivity totals: ZH, Zhanhua (n = 17); LJ, Lijin (n = 136); DY, Dongying District (n = 3); HK, Hekou (n = 8); KL, Kenli subregional source lane (n = 85). The 2000 GLC_FCS30D background is shown for spatial context only

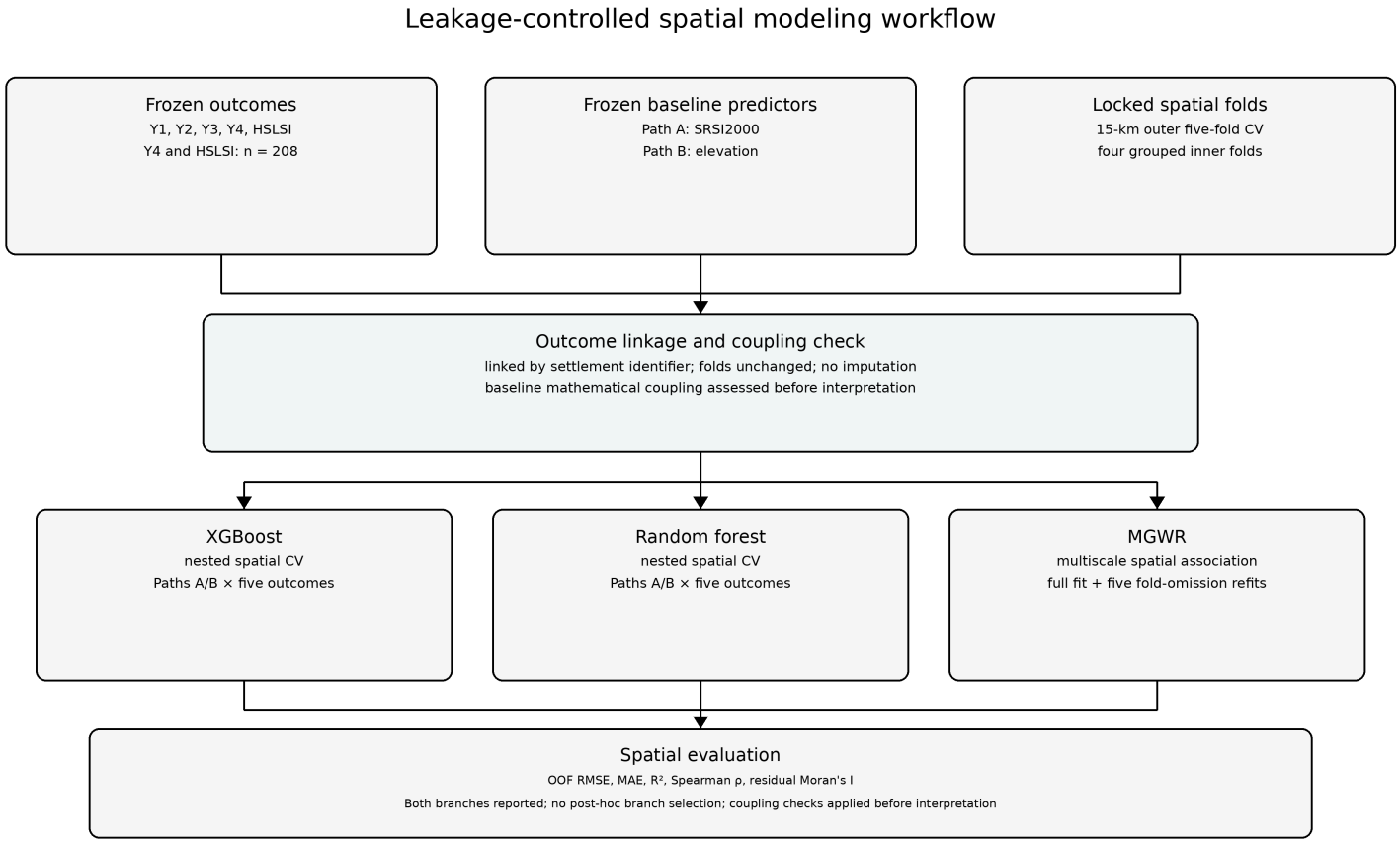


Figure 2

Fig. 2 Sequence of analyses used to evaluate the spatial reliability of settlement-centered monitoring metrics. Predictor matrices and spatial folds were fixed before the change metrics were linked

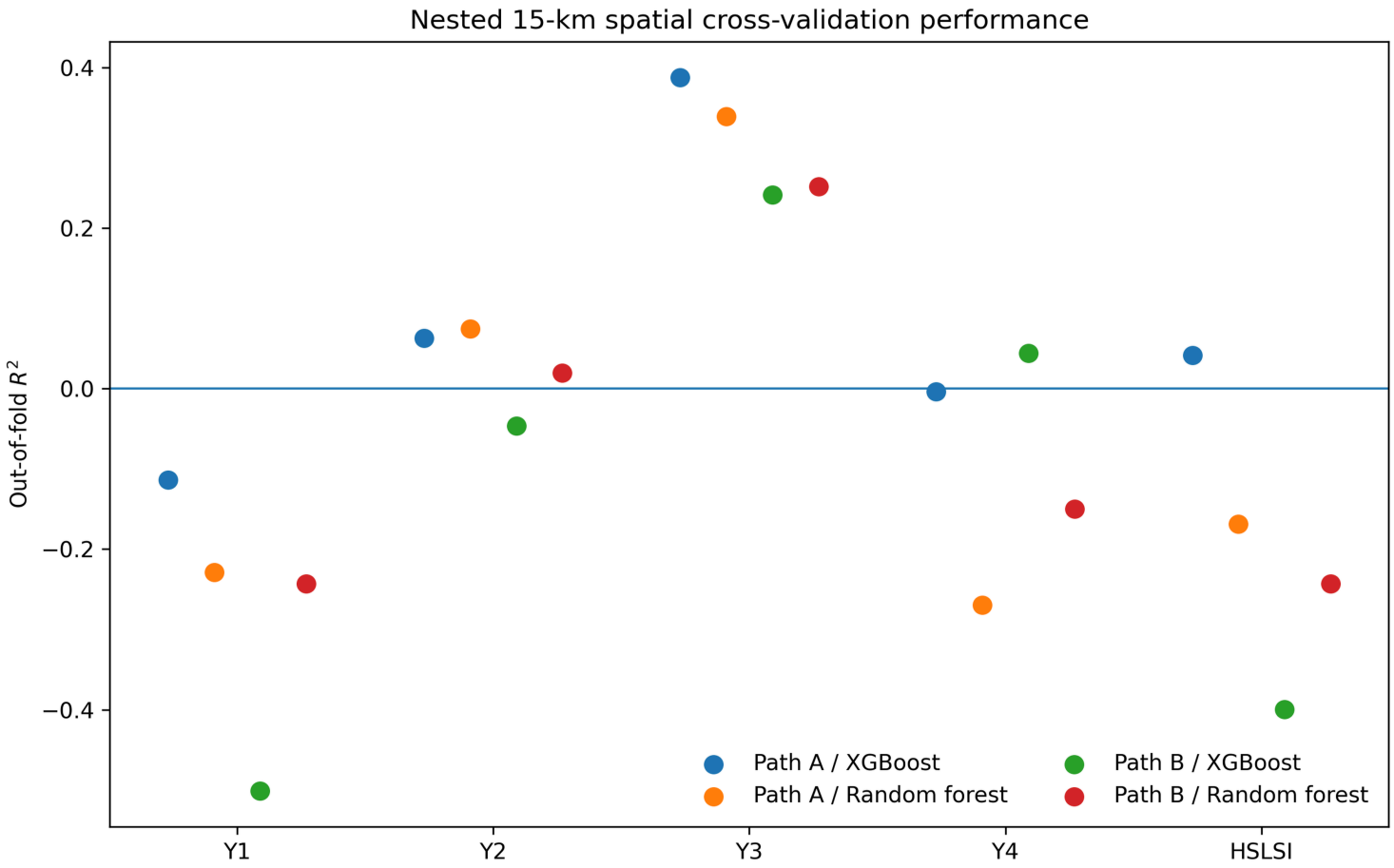


Figure 3

Fig. 3 Nested 15-km spatial cross-validation performance for both pre-specified physical-context branches

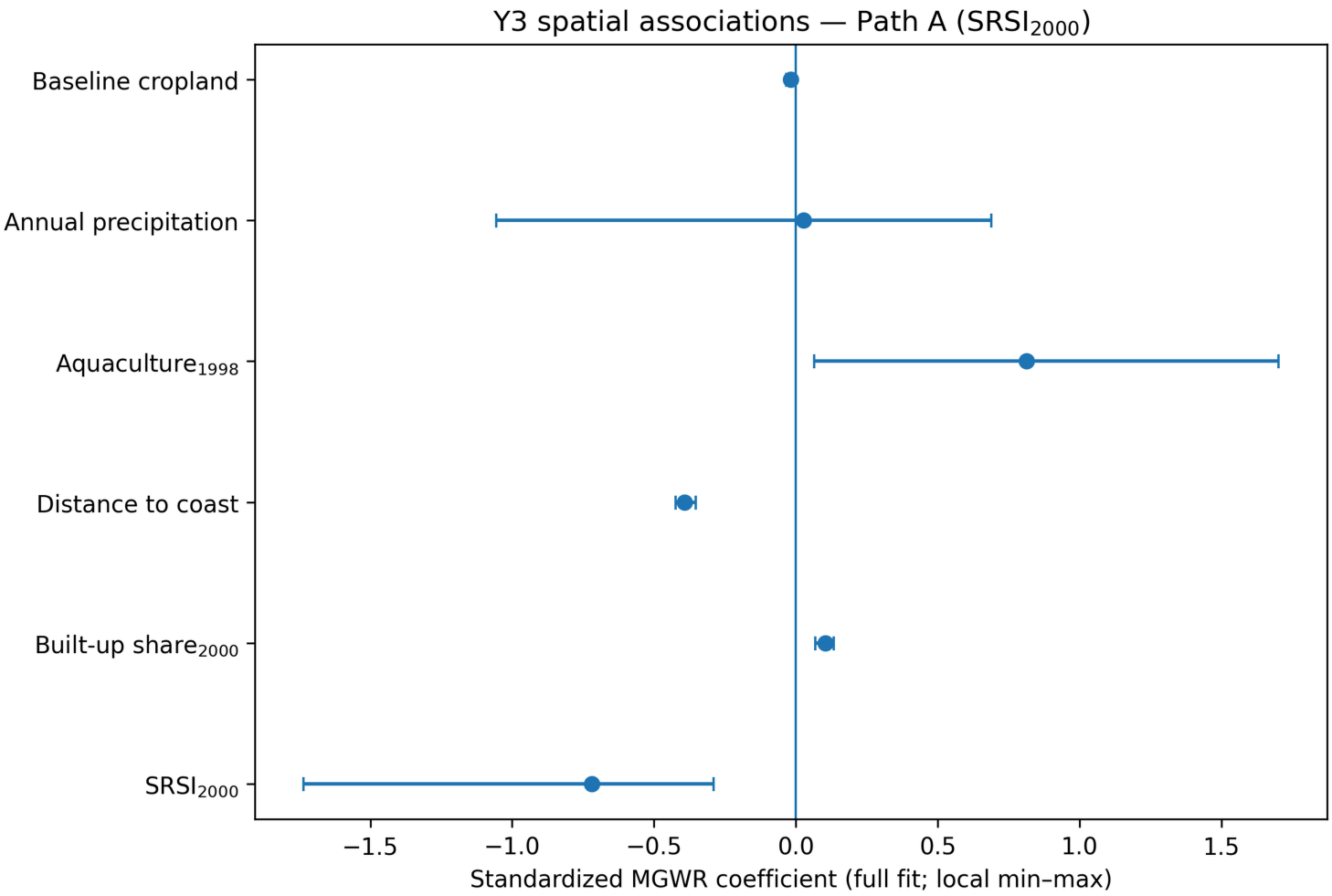


Figure 4

Fig. 4a Full-fit standardized MGWR coefficient mean and local min–max range for Y3 under Path A (SRSI2000)

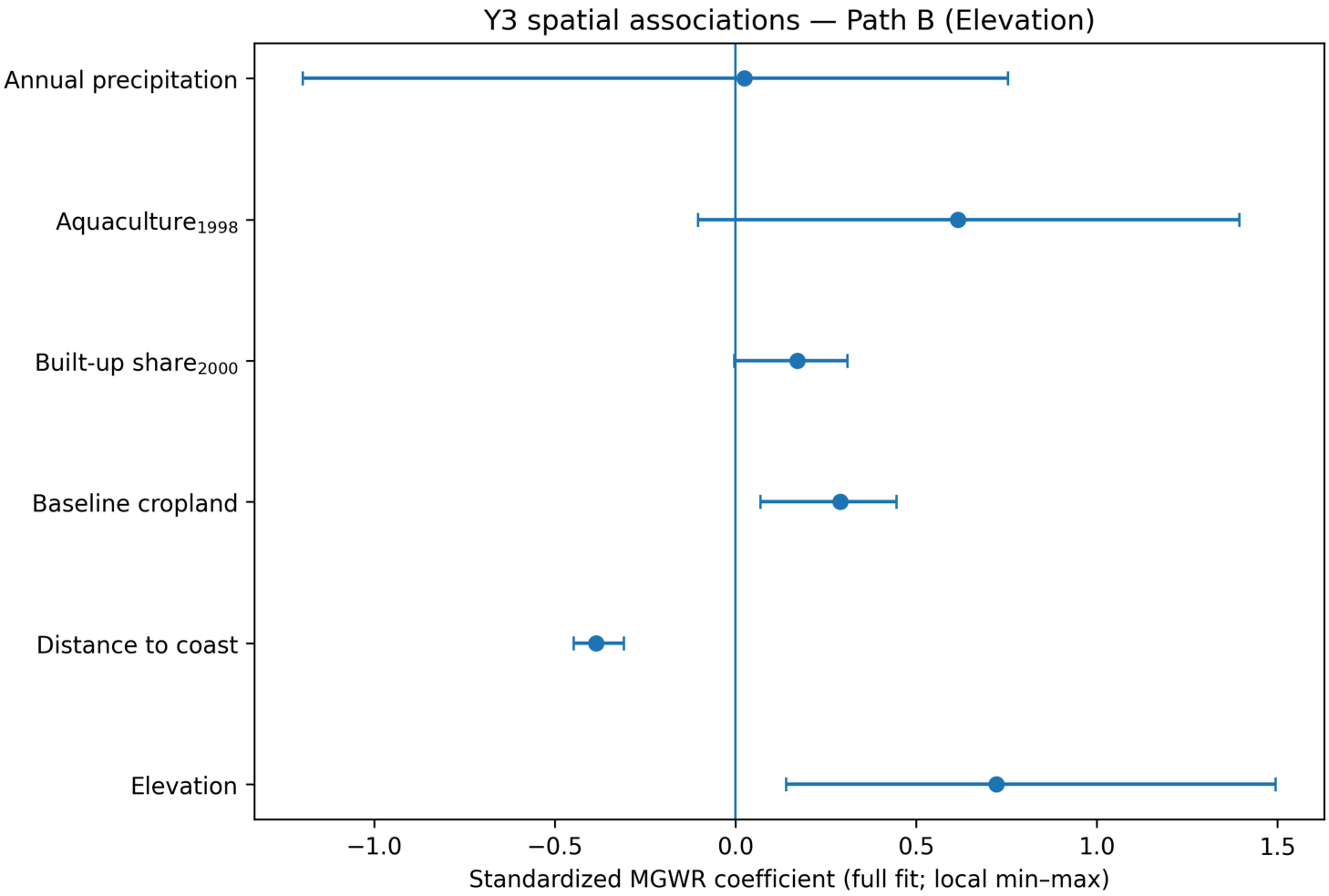


Figure 5

Fig. 4b Full-fit standardized MGWR coefficient mean and local min–max range for Y3 under Path B (elevation)

Supplementary Files

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

- ESM1SupplementaryMaterialsSRFINAL.zip