

Knowledge Quality Diagnosis of Spatial GraphRepresentations: When Does GeographicProximity Fail in Complex Terrain?

Shuai Yan

yanshuai1@cdjcc.edu.cn

Chengdu Jincheng College

Xiaodong Huang

Chengdu Jincheng College

Chuanyuan Kang

Chengdu Jincheng College

Research Article

Keywords: Knowledge quality diagnosis, Spatial graph representation, Graph neural networks, Elevation gating, Spatial extrapolation, Aboveground biomass estimation

Posted Date: October 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11042501/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Knowledge Quality Diagnosis of Spatial Graph Representations: When Does Geographic Proximity Fail in Complex Terrain?

Shuai Yan ^{1*}, Xiaodong Huang¹ and Chuanyuan Kang¹

^{1*}School of Computer and Software, Chengdu Jincheng College, 1 Xiyuan Avenue, Chengdu, 611731, Sichuan Province, China.

*Corresponding author(s). E-mail(s): yanshuail@cdjcc.edu.cn;

Abstract

Knowledge discovery in geographic information systems depends on the quality of spatial representations, yet the assumption that geographic proximity encodes reliable predictive knowledge is systematically violated in complex terrain. We introduce a knowledge-quality diagnostic that quantifies graph reliability through edge-level metrics (mean absolute label and elevation differences, $\mathbf{MAD}_y$ and $\mathbf{MAD}_h$) and a split-level support metric, the Local Support Index (LSI). Building on this diagnosis, we propose the elevation-gated spatial graph (EGSG), which integrates topographic domain knowledge into graph topology by filtering edges that violate elevation-consistency constraints. On a southeastern Tibetan Plateau aboveground biomass dataset ($N = 6,638$; ten seeds), EGSG improves local interpolation over the standard geographic k -nearest-neighbor graph ($\Delta R^2 = +0.048$, $p < 0.001$, 95% CI $[+0.039, +0.056]$) and outperforms tuned gradient-boosting baselines, but under strict spatial extrapolation (30-km buffer holdout) it degrades more ($\Delta R^2 = -0.054$, $p = 0.027$), showing that spatial knowledge transfer depends on local support availability. The framework transfers to further targets (normalized-difference vegetation index, land surface temperature) with target-dependent gating thresholds. Spatial graph representations should therefore undergo knowledge-quality diagnosis before deployment, with reliability bounded by explicit support metrics.

Keywords: Knowledge quality diagnosis, Spatial graph representation, Graph neural networks, Elevation gating, Spatial extrapolation, Aboveground biomass estimation

1 Introduction

Knowledge discovery in geographic information systems rests on an assumption that is rarely examined: that a spatial representation built from geographic proximity encodes reliable predictive knowledge. Graph neural networks (GNNs) have been widely adopted for geospatial prediction in recent years, including urban flow forecasting [3] and spatially-informed estimation of surface variables [2, 4], with general graph learning surveyed in [1]. In these applications the input graph is almost always built as a k -nearest-neighbor graph from geographic coordinates (Geo kNN), connecting each sample to its k spatially nearest neighbors. This construction embodies the first law of geography [5]—“near things are more related than distant things”—as an implicit assumption that geographic proximity encodes target similarity. Whether this encoding produces reliable knowledge depends on how well horizontal distance predicts ecological similarity—and in complex terrain this relationship breaks down. Elevation gradients cause rapid changes in vegetation type, microclimate, and biomass over short horizontal distances: a valley-bottom forest and an adjacent alpine meadow may be only a few hundred meters apart horizontally yet differ by more than a thousand meters in elevation. The Geo kNN graph then produces systematic *misconnections*—spurious edges connecting ecologically dissimilar nodes—and the graph representation itself encodes misleading relational knowledge. **The knowledge quality of the spatial representation, rather than the power of the learning model, becomes the bottleneck of the prediction system.**

From the graph learning perspective, this failure is a breakdown of the graph homophily assumption under a non-stationary spatial process. Standard GNN message passing aggregates features over connected neighborhoods and implicitly assumes homophily—that connected nodes have similar labels [6]. When Geo kNN uses geographic proximity as a proxy for homophily, complex terrain induces spatial non-stationarity [7]: horizontally close nodes may span different elevation and ecological zones, and their target values may differ sharply. Research on heterophilic graphs shows that architecture-level remedies (contrastive homophily-seeking objectives [8], curvature-aware geometric reweighting [9], oversmoothing-mitigating designs [10], wavelet high-pass filtering [11], and architecture search over heterogeneous message-passing mechanisms [52]) can mitigate—but not remove—the damage caused by misleading edges. This paper takes a complementary knowledge-quality approach: rather than designing heterophily-robust architectures, we diagnose the quality of the knowledge encoded in the graph representation and correct the representation itself at the construction stage.

A subtler issue concerns evaluation. We distinguish two generalization scenarios: (1) **local spatial interpolation**—test samples lie in the same geographic region as the training data, and the graph can supply knowledge through local neighbors; and (2) **strict spatial extrapolation**—test regions are entirely absent from training, and the graph cannot provide local support. Under a random train–test split, spatially adjacent samples may appear in both partitions, so the graph provides test nodes with local training support and the evaluation measures local interpolation. Spatial cross-validation strategies, including block cross-validation [12], buffer holdout [13, 14], and area-of-applicability estimation [15], were developed to mitigate exactly this optimistic

bias, yet they are not standard practice in spatial GNN evaluation. The knowledge encoded in a spatial graph representation can therefore only be validated relative to the support structure of the evaluation protocol.

This paper investigates these issues on an aboveground biomass (AGB) estimation task in the complex terrain of the southeastern Tibetan Plateau (elevations 146–5,383 m), an extreme testbed for diagnosing when geographic proximity encodes reliable knowledge. The contributions are organized along the knowledge discovery pipeline—diagnosis, construction, validation, and boundary analysis:

- **Knowledge quality diagnosis.** We introduce edge-level knowledge metrics (MAD_y , MAD_h) and a split-level support metric (LSI) that quantify the degree to which a spatial graph representation encodes reliable knowledge. On the formal clean sample ($N = 6,638$), 21.41% of the Geo kNN edges span elevation differences larger than 800 m and 12.24% connect nodes whose AGB differs by more than 70 Mg/ha—systematic misconnection caused by terrain complexity, not data errors (Sec. 5.2).
- **Knowledge-driven graph construction.** We propose the elevation-gated spatial graph (EGSG), which prunes terrain-violating edges to respect elevation consistency (Sec. 4.4).
- **Empirical knowledge validation.** Under a strict clean protocol (scaler fitted on training data only, validation-based early stopping, ten seeds), EGSG improves R^2 over the Geo kNN graph by $\Delta R^2 = +0.048$ ($p < 0.001$, 95% CI [+0.039, +0.056]) and outperforms tuned gradient-boosting baselines; the gain is robust across $k \in \{3, 5, 10, 15, 20\}$ and across GCN, GAT, and APPNP backbones, and the framework transfers to further target variables—normalized difference vegetation index (NDVI) and land surface temperature (LST)—with target-dependent gating thresholds (Secs. 5.4–5.5, 5.8).
- **Knowledge transfer boundary and deployment principles.** Under strict spatial extrapolation (buffer holdout), the knowledge-transfer benefit of EGSG reverses ($\Delta R^2 = -0.054$, $p = 0.027$ at 30 km; $\Delta R^2 = -0.083$, $p = 0.007$ at 50 km), and the LSI shows local knowledge support is exactly zero in these settings—identifying when proximity-based spatial knowledge is and is not transferable (Sec. 5.6). From this we derive deployment guidelines for spatial knowledge systems: diagnose graph quality before deployment, bound transfer by support metrics, and calibrate domain constraints to the target variable’s spatial structure.

Together, these contributions characterize when geographic proximity encodes reliable knowledge for prediction, how unreliable encodings can be corrected with domain knowledge (EGSG), how that knowledge is validated empirically, and where it stops being transferable (the support boundary identified by LSI).

2 Related Work

2.1 Spatial Graph Learning for Geographic Regression

GNNs are widely used for geospatial regression; spatially-informed architectures augment the message-passing aggregation with geographic side information [2],

homogeneity-aware transfer models propagate knowledge across comparable regions for flow prediction [3], and graph convolutional decoders interpolate correlated spatial fields directly from observed samples [4]. In all of these the input graph is built from geographic proximity, encoding the assumption that adjacent locations share similar targets. Complementary lines shape the graph and its message flow with domain structure: structure-enhanced temporal graph operators [16] and spatiotemporal indexes over road networks [17] constrain propagation along transport topology, while a further family learns the structure directly from the data, jointly predicting and discovering latent spatiotemporal relationships [48] and managing the massive spatiotemporal grids such graphs live on [47]. Whereas these methods either infer both the graph and the model jointly from data or inherit a fixed proximity graph, EGSG constrains the graph using domain knowledge rather than learning it from data (Sec. 2.3). Whether the graph actually encodes reliable predictive knowledge, however, has received far less attention—especially in complex terrain.

2.2 Graph Homophily, Heterophily, and Edge Quality

GNN message passing relies on graph homophily—the degree to which connected nodes share similar labels [6]. In heterophilic graphs standard architectures may underperform, and most remedies are architectural: curvature-aware geometric reweighting [9], wavelet-driven high-pass filtering that separates low- and high-frequency signals [11], contrastive objectives that explicitly seek the homophilic signal within a heterophilic graph [8], and soft-label propagation that relaxes the neighbor-labelling assumption [24]; recent analyses and surveys attribute much of the residual degradation to over-smoothing and to its interaction with heterophily [10, 43]. The root cause often lies in construction itself: if the graph connects dissimilar nodes, no architecture-level remedy fully compensates. This paper instead reduces heterophilic edges at the construction stage via an explicit knowledge constraint, so these architecture-level remedies are complementary rather than a substitute.

2.3 Graph Structure Learning

A separate line of work learns the graph topology jointly with the network parameters: implicit-depth architectures embed structure learning in a fixed-point solver [40], diffusion-sampled dynamic models infer the graph and its temporal anomalies together [41], multi-head spectral attention designs learn the propagation weights directly [42], and rewiring approaches restructure the edges themselves for weakly supervised graphs [51]. EGSG belongs to this family but takes a different route: rather than learning topology from data, it constrains topology using domain knowledge, making it interpretable (each removal has a physical explanation), efficient (one closed-form pass, no iterative or bi-level optimization), and robust to small samples (no overfitting from a learned topology). Whereas learned-topology methods infer a task- or representation-specific notion of good structure, EGSG externalizes that criterion: it encodes an explicit, physically grounded constraint and evaluates it as a single closed-form edge filter, so the diagnostic is measurable, the constraint testable, and no surrogate for knowledge quality needs to be learned. This positions EGSG within the data-quality

“fit-for-use” tradition [46] applied to the geometric setting—rather than asking what a graph embedding represents, we ask whether the retained edges are fit for the prediction task at hand. EGSG fixes a closed-form constraint yet selects its threshold by a label-free rule (the mutual-information knee, Sec. 5.8) rather than tuning it on a target, so only one scalar, τ , is set without labels and the mechanism can be audited.

2.4 Knowledge Discovery in Spatial Data

Knowledge discovery in databases (KDD) frames data mining as extracting implicit, previously unknown, potentially useful knowledge [18], and recent studies in this journal map how intelligent computing methods reshape that pipeline [49] and embed geographic semantics into the discovery process through ontologies [19]. Geographic knowledge discovery emerged as a dedicated field [50], with foundational work establishing the primitives of spatial data mining—neighborhood relations, graphs, and indices [20]. A substantial body of this journal’s work concerns spatial co-location pattern mining [21], and further spatial-specific primitives flag anomalies in space: ontologically grounded deep models detect abnormal spatial transactions [44], and geographic attention methods localize geotagged observations from text [45]. Such primitives detect anomalous *samples* or *transactions*; EGSG instead diagnoses and repairs the anomalous *edges* of the representation, shifting the object of knowledge-quality scrutiny from points to the connections a GNN propagates.

This literature has concentrated on *what knowledge can be mined from spatial data*. The complementary question—*whether the representation itself encodes reliable knowledge*—has received little attention. Pattern mining and GNN prediction both inherit the neighborhood structure of the underlying representation; if that structure encodes misleading relational knowledge (e.g., edges between ecologically dissimilar sites), any downstream step is contaminated. This paper addresses that gap by treating the spatial graph representation as the object of a knowledge quality diagnosis and proposing a knowledge-driven construction that corrects the representation before learning.

2.5 Spatial Autocorrelation and Leakage-Aware Evaluation

Spatial autocorrelation—the tendency of nearby locations to have similar values—is a fundamental property of geospatial data [22]. It motivates spatial graphs but also creates optimistic bias in evaluation: under random splits adjacent samples appear in both partitions and the graph can transfer information between them. Spatial cross-validation (block [12], buffer holdout [13, 14], leave-one-region-out [14]) and area-of-applicability methods [15] mitigate this bias. Spatial graph learning evaluation should therefore report both local-interpolation (random split) and strict-extrapolation (spatial holdout) performance, with split diagnostics and local support metrics such as the LSI.

2.6 Terrain Constraints in Remote Sensing Regression

Terrain complexity is a well-known challenge in remote sensing biomass estimation [23]: elevation gradients cause sharp ecological transitions that violate smoothness

assumptions, and the signal–biomass relationship is known to vary steeply with terrain [23, 25]. Prior terrain-aware approaches use DEM-derived features as inputs or stratify by elevation band [25]. The difference here is that elevation *corrects the graph topology* rather than serving as another input feature: DEM features let the model learn elevation relations but cannot stop the graph connecting ecologically dissimilar nodes, whereas EGSG removes terrain-violating edges at the source of the unreliable encoding.

3 Study Area and Data

3.1 Study Area

The study area is the southeastern Tibetan Plateau, covering part of the Tibet Autonomous Region of China. The terrain relief is extreme: across the region the DEM spans roughly 146 m to 5,383 m, and the 6,638 clean samples used in this paper span 282–5,279 m (median 3,850 m), transitioning rapidly from subtropical forests on the valley floor to alpine meadows and bare rock at altitude. This pronounced elevation gradient makes it an ideal testbed for diagnosing knowledge quality in spatial graph representations: horizontal proximity does not guarantee ecological similarity.

Fig. 1 shows the geographic location, elevation distribution, and spatial distribution of the sampling points.

3.2 Data Sources

The primary prediction target—aboveground biomass (AGB)—is taken from the ESA Climate Change Initiative (CCI) Biomass product v5.0, with a spatial resolution of 100 m, providing global AGB estimates for 2010 and 2015–2021 [27]. This paper mainly uses the 2021 snapshot.

Label provenance. The AGB labels are not field measurements but a derived product: ESA CCI v5.0 is generated by random-forest upscaling of radar and lidar reference data followed by spatial aggregation [27]. Two signatures of this provenance are visible in the clean sample. First, the label is quantized to integer Mg/ha—only 194 distinct values across the 6,638 clean samples, versus 6,546 distinct values for the NDVI feature. Second, its empirical semivariogram shows short-range-to-long-range semivariance ratios (0.77 at 1–2 km vs. 40–80 km) comparable to those of the smoothed NDVI feature (0.72). All accuracies here should be interpreted against this derived-product origin; relative comparisons (EGSG vs. Geo kNN) remain valid because both predict the same label, while absolute R^2 interpretation is discussed in Sec. 6.

Remote sensing predictor features are derived from multiple satellite products—Moderate Resolution Imaging Spectroradiometer (MODIS) surface products and the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) Global Digital Elevation Model (GDEM)—[28–32], summarized in Table 1. The feature set comprises NDVI composites, leaf area index (LAI) and the fraction of absorbed photosynthetically active radiation (FPAR), LST, evapotranspiration (ET), DEM derivatives, and land cover.

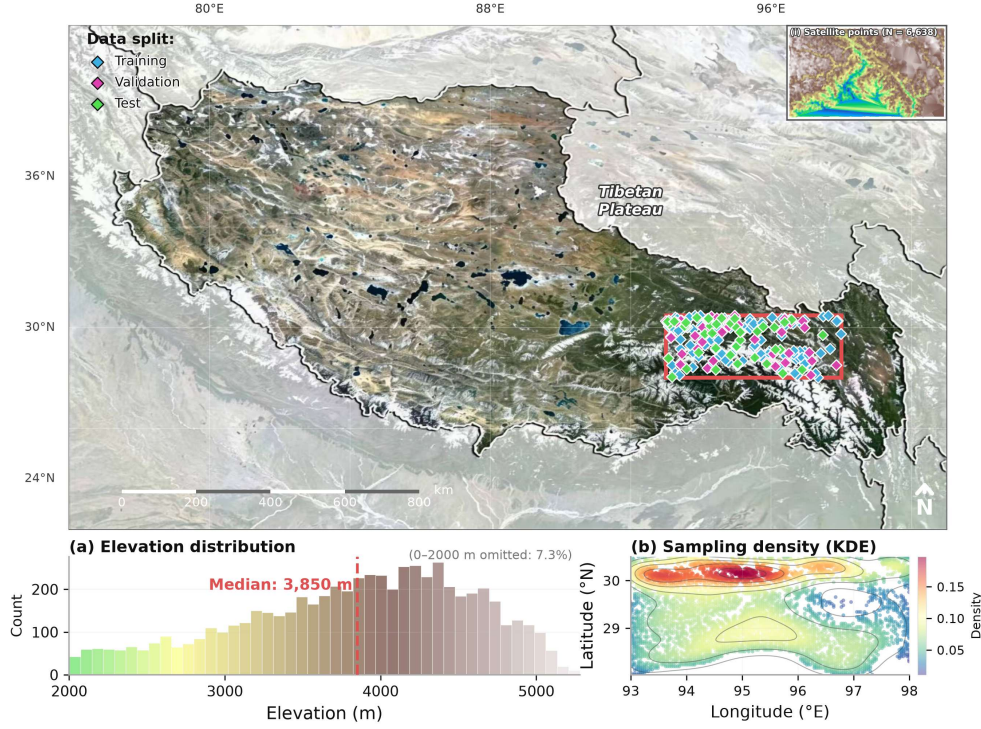


Fig. 1 Overview of the study area. The main panel shows the southeastern Tibetan Plateau on a satellite base map, with the study region outlined in red and the 6,638 clean sampling points colored by a 60/20/20 random train/validation/test partition (top-left legend); a 0–800 km scale bar and a graticule anchor the location. The inset (i) zooms into the study region to show the satellite sampling points ($N = 6,638$). (a) Elevation distribution of the 6,638 clean samples (terrain color scale), spanning 282–5,279 m with the median (3,850 m) marked by the dashed line. (b) Sampling density (KDE) across the study region, showing a dense north and a sparser south. Samples are drawn by stratified sampling of AGB quantiles $\times$ spatial grid (Sec. 3); over the region vegetation and terrain change rapidly over short horizontal distances, so horizontal proximity does not guarantee ecological similarity—the core property that makes this an extreme testbed for studying failure modes of spatial-graph construction

The NDVI layer is a provider-supplied annual composite delivered already clipped to the study region, and it is the one predictor whose source product the archive does not name: the GeoTIFFs carry no product identifier, and no download or processing script for them exists in our pipeline. We therefore report what the raster metadata does state—a 0.0002695° grid (≈ 30 m at the equator, $\approx 26 \times 30$ m at this latitude) on which each of the seven statistics is computed over a 7×7 -pixel window (a support of roughly 200 m)—rather than attributing it to a MODIS product, whose NDVI layers are no finer than 250 m and are therefore inconsistent with this grid. This provenance gap does not affect any comparison in this paper: the NDVI layer enters only as a predictor, and every result reported here compares graph constructions evaluated on the identical feature matrix.

Table 1 Data sources and feature list

Data source	Product	Resolution	Features	Use	Citation
NDVI	Annual clipped composite (as delivered)	~30 m grid; 7×7 -pixel window (~200 m support)	mean, std, median, Q25, Q75, Q90, range (7 dims)	Vegetation features	—
LAI/FPAR	MODIS MCD15A3H.061	500 m	lai_mean, fpar_mean (2 dims)	Vegetation auxiliary	[28]
LST	MODIS MOD11A2.061	1 km	lst_mean (1 dim)	Thermal features	[29]
ET	MODIS MOD16A2GF.061	500 m	et_mean (1 dim)	Moisture features	[30]
DEM	ASTER GDEM v3	~30 m	elevation_mean, slope_mean, aspect_sin, aspect_cos, curvature_mean (5 dims)	Terrain features	[31]
LandCover	MODIS MCD12Q1.061	500 m	landcover type	Edge-quality diagnostics	[32]
AGB label	ESA Biomass CCI v5.0	100 m	agb_label	Prediction target	[27]

In addition, three binary missing-value flags (veg_missing, hydro_missing, landcover_missing) indicate feature availability. The total predictor dimensionality is 16 features (9 vegetation, 5 terrain, 2 thermal/moisture) plus 3 missingness flags.

Note on SAR data: ALOS-2 PALSAR-2 L-band SAR data [33] (HH, HV polarizations) cover a single scene (roughly 95°E , $28\text{--}29^\circ\text{N}$), yielding 649 samples with SAR backscatter; of these, 578 also have complete vegetation predictors and 335 have all 16 predictors plus the AGB label. Because of the limited spatial coverage and the small fully-valid subset, they are used only for exploratory analysis.

3.3 Sample and Feature Construction

The main experiments use the 2021 single-year snapshot. Samples are drawn by stratified sampling (AGB quantiles $\times$ spatial grid) to ensure coverage of the full AGB range and geographic distribution. The formal clean sample retains the $N = 6,638$ samples for which all 16 predictor features and the AGB label are complete (1,362 rows with missing values are discarded), exactly reproducible from the archived data. All

features are standardized with StandardScaler, **fitted on the training set only** to prevent leakage. Early cross-year aggregated samples carry temporal consistency risks between features and year-specific AGB labels, so they are not included; single-year independently constructed sample sets are used instead.

Note on sample versions. The early edge-misconnection diagnosis (Sec. 5.2) was performed on a pre-cleaning subset of $N = 7,728$ samples whose upstream filter is no longer reproducible from the archive. To avoid parallel-sample confusion, **all diagnostic statistics quoted here are recomputed on the formal clean sample** ($N = 6,638$), with the archived 8,000-sample pool as a robustness check: the high-risk edge fractions differ by at most 0.8 percentage points between versions (e.g., 21.41% vs. 21.49% for $|\Delta\text{Elevation}| > 800$ m), so no diagnostic conclusion depends on the sample version. This paper does not compare absolute R^2 across sample versions, only relative trends within each experiment.

4 Methods

4.1 Notation

Let $G = (V, E)$ be a graph, where:

- $V = \{1, 2, \dots, N\}$ is the set of nodes, and N is the total number of nodes ($N = 6,638$ in the formal experiments);
- $E \subseteq V \times V$ is the set of edges, where each edge (i, j) represents a connection between nodes i and j ;
- each node $i \in V$ is associated with the following attributes:
 - (λ_i, φ_i) : geographic coordinates (longitude, latitude), in degrees;
 - h_i : elevation, in meters (m);
 - y_i : target label (AGB in this paper), in Mg/ha.

The number of geographic neighbors is $k = 10$ unless stated otherwise.

4.2 Geographic k -Nearest-Neighbor Spatial Graph (Geo kNN)

The baseline representation in this paper is the geographic k -nearest-neighbor (Geo kNN) graph. For each node i in the node set V , it is connected to its k nearest neighbors in geographic space ranked by great-circle distance, forming the edge set:

$$E_{\text{geo}} = \{(i, j) \mid j \in \mathcal{N}_k(i)\} \quad (1)$$

where $\mathcal{N}_k(i)$ is the set of the k nearest neighbors of node i ranked by great-circle distance d_{ij} (in km). On the formal clean sample ($N = 6,638$, $k = 10$), the directed edge set contains 66,380 edges.

Coordinate and distance material. All Geo kNN computations rank neighbors by great-circle distance. Implementation-wise, distances are computed as 3-D chord lengths on a spherical Earth ($R = 6,371$ km): $\text{chord}(d) = 2R \sin(d/2R)$ is strictly monotonic in the great-circle distance d , so the kNN ranking is provably identical to the Haversine ranking, verified empirically on 2,000 randomly sampled nodes (0/2,000 mismatches in the $k = 10$ neighbor sets). The archived coordinates contain no missing

or duplicated entries, lying within $\text{lon} \in [93.00^\circ, 98.00^\circ]$, $\text{lat} \in [28.02^\circ, 30.50^\circ]$. On the formal clean sample the mean $k = 10$ edge length is 4.99 km (P90 = 7.78 km); the stratified sampling grid imposes a minimum inter-node spacing of 1.19 km.

The implicit assumption of Geo kNN is that geographic proximity implies target similarity—reasonable in flat terrain, but in the southeastern Tibetan Plateau the extreme elevation gradient (146 m to 5,383 m) means two locations very close horizontally can be ecologically very different, for example a valley-bottom forest and an adjacent alpine meadow.

4.3 Edge-Quality Diagnostics: Knowledge Quality in Spatial Graph Representations

To quantify the severity of Geo kNN misconnections we define two edge-level knowledge metrics, the **mean absolute label difference** $\text{MAD}_y(E)$ and the **mean absolute elevation difference** $\text{MAD}_h(E)$:

$$\text{MAD}_y(E) = \frac{1}{|E|} \sum_{(i,j) \in E} |y_i - y_j| \quad (2)$$

$$\text{MAD}_h(E) = \frac{1}{|E|} \sum_{(i,j) \in E} |h_i - h_j| \quad (3)$$

Lower $\text{MAD}_y(E)$ indicates that connected nodes have more similar AGB (more homogeneous edges); lower $\text{MAD}_h(E)$ indicates more similar elevations (better terrain consistency). Based on these definitions, this paper reports: the mean and P90 of $|\Delta\text{AGB}|$ (i.e., MAD_y and its 90th percentile), the mean and P90 of $|\Delta\text{Elevation}|$, the land-cover mismatch ratio (the proportion of edges connecting nodes with different MODIS land-cover types), and the high-risk edge ratio (the proportion with $|h_i - h_j| > 800$ m or $|y_i - y_j| > 70$ Mg/ha).

Threshold provenance. The risk thresholds are set so that a violating edge is, with high probability, a vegetation-belt flip. The 70 Mg/ha threshold lies in the biomass gap between the forest classes (class medians 73–87 Mg/ha) and the grass/shrub classes (19–50.5 Mg/ha); the 800 m threshold approximates the vertical separation between the forest belt (class median elevations 1,514–3,319 m) and the alpine belt (3,910–4,806 m). Consequently, an edge with $|\Delta\text{AGB}| > 70$ Mg/ha or $|\Delta\text{Elevation}| > 800$ m almost certainly connects a forest site to an alpine non-forest site—the dominant misconnection mode (60.6% of the high-AGB-difference edges on the formal clean sample connect different land-cover types, dominantly forest–meadow transitions; 63.9% on the archived 8,000-sample pool).

Connection to GNN performance. The link between edge quality and GNN performance is intuitive: message passing aggregates features over connected neighborhoods, so edges connecting ecologically dissimilar nodes inject noisy information into the aggregation. Reducing MAD_y directly reduces this noise source. The empirical relationship between MAD_y and R^2 (Sec. 5.7) confirms this mechanism, and Sec. 5.8 shows it generalizes across target variables.

4.4 Elevation-Gated Spatial Graph (EGSG)

Design problem. We pose the graph-construction task as a constrained sparsification: among all subgraphs of E_{geo} that retain sufficient connectivity, we seek the one with greatest edge homogeneity,

$$\min_{E \subseteq E_{\text{geo}}} \text{MAD}_h(E) \quad \text{s.t.} \quad \deg_{G'}(i) \geq d_{\min}, \quad \forall i \in V, \quad (4)$$

where $G' = (V, E)$ and $d_{\min}$ is a target minimum degree. This paper introduces a knowledge-constrained construction, the Elevation-Gated Spatial Graph (EGSG), that instantiates Eq. 4 through a single closed-form filter—retain only edges whose endpoint elevation difference does not exceed a threshold τ :

$$E_{\text{EGSG}}(\tau) = \{(i, j) \in E_{\text{geo}} \mid |h_i - h_j| \leq \tau\} \quad (5)$$

where τ is the elevation gate threshold (in m). When $\tau \rightarrow \infty$, E_{EGSG} degenerates to E_{geo} (no gating); when $\tau \rightarrow 0$, only nodes with nearly identical elevations are retained. EGSG does not solve Eq. 4 by search; it implements the purity side directly through Eq. 5 and calibrates the connectivity side through τ , making no claim of optimality but instantiating the homogeneity–connectivity trade-off in a single interpretable step. This contrasts with architecture-level remedies, which act after construction rather than on the representation itself.

Property 1 (monotone purity–connectivity trade-off). The construction is monotone: for $\tau_1 \leq \tau_2$, $E_{\text{EGSG}}(\tau_1) \subseteq E_{\text{EGSG}}(\tau_2) \subseteq E_{\text{geo}}$, so MAD_h is nondecreasing in τ while connectivity improves. On the 6,638 clean samples this is exact (Fig. 4): MAD_h rises from 49.7 m ($\tau = 100$) to 122.6 m ($\tau = 250$) to 404.8 m ($\tau = 1000$) to 512.4 m (no gate), retained edges from 37,142 to 60,698 to 66,380, and isolated nodes from 112 to 6 to 0 (Table 3). τ therefore traces a monotone feasibility curve between edge purity and graph connectivity, and $\tau = 250$ m is the empirically balanced point.

Property 2 (connectivity feasibility). At the calibrated $\tau = 250$ m the drop-only graph is already near-connected: 6 of 6,638 vertices (0.09%) are isolated and the mean degree is 18.3. An explicit connectivity-repair step (re-adding edges so that every node has degree $\geq d_{\min} = 5$) guarantees $\deg_{G'}(i) \geq \min(d_{\min}, |N_{\kappa}(i)|)$ and removes all isolated nodes, but is statistically neutral under both protocols (random $\Delta = -0.0004$, $p = 0.685$; buffer-30 $\Delta = +0.0009$, $p = 0.853$; Table 10): the calibrated τ already sits on the feasibility boundary, so connectivity is a means, not a goal.

Greedy construction and the scope of the approximation. Because Eq. 4 is a combinatorial subgraph-selection problem, it is not solved to optimality in general. EGSG instead constructs a feasible subgraph by a closed-form, stage-wise heuristic—the drop stage (Eq. 5) plus an optional connectivity-repair stage—and we do not claim an approximation ratio. We support the construction as the mechanism the heuristic is meant to implement: a random-pruning control at the same edge count yields $R^2 = 0.364$ ($\Delta = -0.045$ vs EGSG, $p = 10^{-6}$), the add-back dose response degrades monotonically (b30 $p = 0.024$, b50 $p = 0.009$), and the connectivity-repair control is neutral (Sec. 5.9, Table 10)—so at least $\approx 95\%$ of the benefit tracks *which* edges are removed, exactly the quantity Eq. 4 formalizes.

The key difference between EGSG and simply using the DEM as an input feature is that EGSG **directly corrects the graph topology**. Adding elevation features lets the model learn elevation-related relationships but cannot prevent the graph from connecting ecologically dissimilar nodes; EGSG directly removes these problematic edges, preventing contaminated knowledge from propagating.

The threshold τ controls the trade-off between edge-pruning strength and connectivity. An overly small τ removes too many edges, producing an overly sparse graph; an overly large τ retains most edges, including terrain-violating connections. A 10-seed threshold sensitivity analysis on the 6,638 clean-protocol samples identifies $\tau = 250$ m as the empirical trade-off under the current study area, sampling density, and clean sample (see Sec. 5.3); different regions, sampling densities, or target variables require recalibration.

Fig. 2 illustrates the Geo kNN misconnection mechanism—horizontal proximity with ecological dissimilarity—and how EGSG prunes cross-elevation-band heterogeneous edges through elevation gating.

4.5 Local Support Index (LSI)

To quantitatively characterize the local adjacency between training and test data under different split protocols, this paper defines the Local Support Index (LSI). For each test node i , its LSI is defined as:

$$\text{LSI}_i(r, \tau) = \sum_{j \in V_{\text{train}}} \mathbb{I}[d_{ij} \leq r] \cdot \mathbb{I}[|h_i - h_j| \leq \tau] \quad (6)$$

where V_{train} is the set of training nodes, r is the local neighborhood radius (in km), τ is the elevation gate threshold consistent with Eq. (5), and $\mathbb{I}[\cdot]$ is the indicator function. LSI_i counts the number of training neighbors of node i that lie within geographic radius r and whose elevation difference from i does not exceed τ . Averaging over the entire test set gives the split-level LSI:

$$\overline{\text{LSI}}(r, \tau) = \frac{1}{|V_{\text{test}}|} \sum_{i \in V_{\text{test}}} \text{LSI}_i(r, \tau) \quad (7)$$

This paper mainly reports $\overline{\text{LSI}}$ at $r = 30$ km and $\tau = 250$ m.

LSI provides a mechanistic explanation: the gains of EGSG depend on message passing extracting knowledge from training neighbors; when LSI is 0, graph propagation has no available local neighbors and topology correction cannot produce gains. Its interpretation is bounded, however. A sample-level correlation analysis under random split shows only 4 samples have $\text{LSI} = 0$ and about 93% have $\text{LSI} > 10$; LSI has no significant correlation with single-sample prediction error (Spearman $r \approx 0.001$, $p \approx 0.933$), and multi-seed re-examinations find no stable correlation with single-point errors or the sample-level EGSG improvement. LSI is therefore a *split-level* data topology descriptor—of the knowledge support structure at the partition level—not a sample-level error predictor, which limits but does not negate its mechanistic role at the split level. Two caveats complete its scope: under splits that also induce covariate

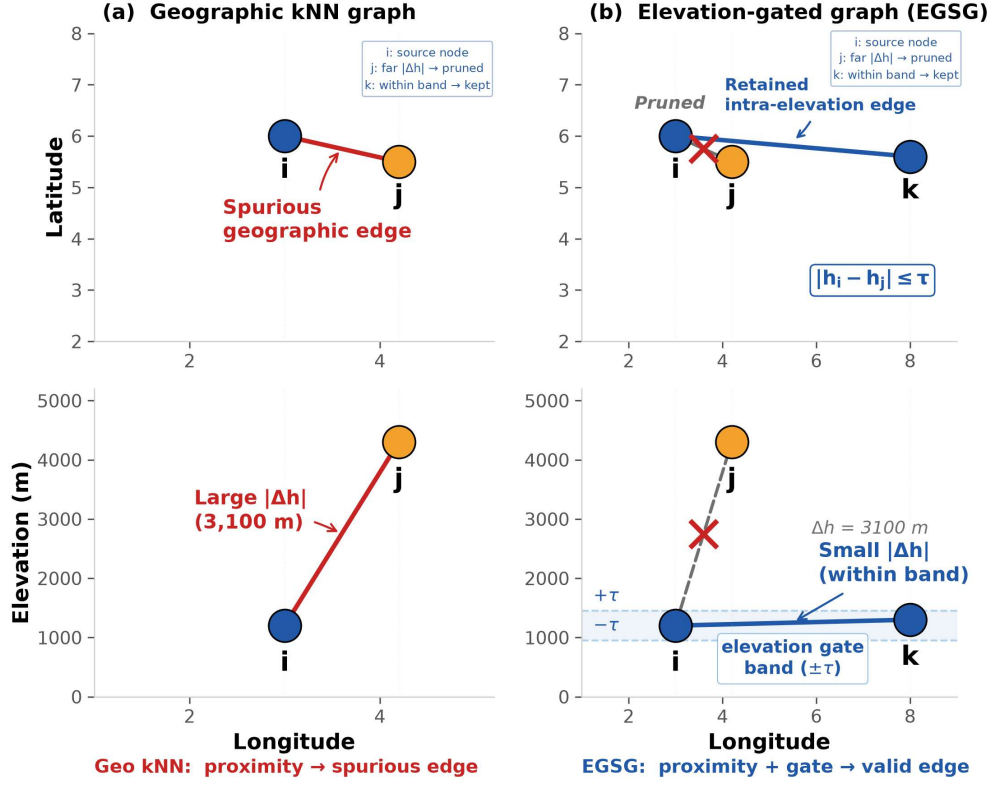


Fig. 2 Schematic of the Geo kNN misconnection mechanism and the EGSG gating correction. The schematic is arranged in two rows: the upper row gives the geographic-space view (latitude vs longitude) and the lower row the corresponding elevation profile (elevation vs longitude); the node legend of each panel lists i as the source, j as a neighbour with large $|\Delta h|$ and k as one within the band. (a) Geo kNN graph (left column): edges are formed from two-dimensional horizontal distance alone, so nodes i and j are connected because of horizontal proximity (solid red line), and the elevation profile of this same edge shows an endpoint elevation difference of 3,100 m—a spurious geographic edge, i.e. a cross-elevation-band misconnection. (b) EGSG (right column): elevation gating $|h_i - h_j| \leq \tau$ ($\tau = 250$ m in this paper) is introduced into graph construction, the misconnected edge $i-j$ is pruned (dashed grey line, marked $\times$), and the intra-elevation edge $i-k$ is retained (solid blue line); the shaded band marks the gate $\pm \tau$ around the elevation of node i , so only node pairs whose elevation difference falls within that band remain connected

shift (the elevation holdouts) the degradation exceeds what support removal alone predicts, so LSI-based explanations must be conditioned on shift (Sec. 5.7); and LSI is a property of the split, not of the model, so it cannot distinguish EGSG from Geo kNN.

4.6 Evaluation Protocols

Models are evaluated under four protocols representing different levels of spatial generalization:

1. **Random split (local spatial interpolation)**: samples are randomly divided into 60% training, 20% validation, and 20% test. Spatially adjacent samples may appear

in different partitions, providing local knowledge support through the graph; this tests **local spatial interpolation**—predicting held-out locations within the same region.

2. **Block split (moderate spatial partitioning)**: the study area is divided into 50 km or 100 km blocks, entire blocks are held out for testing, removing local knowledge support within the test region; this tests **regional generalization**.
3. **Buffer split (strict spatial extrapolation)**: a 30 km or 50 km buffer is placed around each test sample and all training samples within it are removed, eliminating local graph support; this tests **strict spatial extrapolation**.
4. **Elevation holdout (terrain distribution extrapolation)**: samples are partitioned by elevation bands (low, middle, high) and the high-elevation samples are held out, testing generalization to terrain regions unrepresented in the training data.

All protocols follow the following conventions:

- StandardScaler is fitted **on the training set only**, preventing feature leakage;
- early stopping is based on **validation loss** (not test loss), evaluated every 10 epochs, patience = 10; the optimal model is selected at the lowest-validation-loss epoch;
- under the random split (the headline local-interpolation protocol) the full graph is used *transductively*: all N nodes, including test nodes, participate in message passing and only test labels are masked—the standard setting for GNN evaluation under a random split. By contrast, *in strict extrapolation protocols*, test-region nodes are restricted from participating in graph message passing, preventing test information leaking through the graph (inductive inference);
- all split, backbone, terrain-aware, threshold-sensitivity, and buffer-radius comparisons use 10 random seeds (seeds = {42, 123, 2025, 7, 13, 37, 61, 73, 97, 101}). Each table states its seed count in the caption. Mean $\pm$ standard deviation is reported;
- statistical significance is assessed with paired t -tests; for the headline comparisons we report the 95% confidence interval (CI) of the per-seed paired ΔR^2 .

To measure the **performance gain** (R^2 gain) of EGSG relative to Geo kNN under a split s , we use $\Delta R^2(s) = R_{\text{EGSG}}^2(s) - R_{\text{Geo}}^2(s)$; to measure the **performance degradation** of a split s relative to the random split, we use $\mathcal{D}(s) = R^2(\text{random}) - R^2(s)$. The former fills the ΔR^2 columns of the EGSG-vs-Geo-kNN comparisons, and the latter fills the degradation column of Table 8. The protocol review identified leakage problems in the original implementation regarding feature standardization and graph transductivity; the repaired protocols (scaler fitted on training only; and, in extrapolation protocols, test-region nodes excluded from propagation) ensure the reliability of the reported results.

Use of AI tools. During the preparation of this manuscript, large language models (DeepSeek, ChatGPT, Claude) were used to assist with organizing the manuscript structure, improving English expression, and formatting references. All scientific content, data analysis, experimental design, and conclusions are the original work of the authors, who take full responsibility for the accuracy and integrity of the reported results.

5 Experiments and Results

This section is organized along the knowledge discovery pipeline: we first diagnose the knowledge quality of the Geo kNN representation (Sec. 5.2), then determine the empirical gate threshold (Sec. 5.3), validate the knowledge-driven construction under local spatial interpolation (Sec. 5.4) and across backbones (Sec. 5.5), analyze the boundary at which the encoded knowledge stops transferring (Sec. 5.6), explain the mechanism through edge quality and local support (Sec. 5.7), examine generalization across target variables (Sec. 5.8), and report ablations, controls, terrain-aware comparisons, and efficiency (Sec. 5.9).

5.1 Experimental Setup

All GNN models use a GCN backbone [34] with hidden dimension 64, learning rate 0.001, 300 epochs, and the clean protocol of Sec. 4.6, on a single CPU core (Intel Core i9-14900HX). Ablation and control experiments follow the same protocol unless stated otherwise.

5.2 Systematic Misconnection of Geo kNN in Complex Terrain

A Geo kNN graph ($k = 10$) built on the formal clean sample ($N = 6,638$) revealed severe misconnection problems (Table 2). More than one fifth of the edges (21.41%) span elevation differences larger than 800 m, and 12.24% correspond to AGB differences larger than 70 Mg/ha. Among the high-AGB-difference edges, 60.6% connect different land-cover types, and the dominant pattern is forest–grassland transitions (low-elevation forest $\leftrightarrow$ high-elevation grassland/meadow).

Table 2 Knowledge quality diagnosis of the Geo kNN graph ($k = 10$): formal clean sample ($N = 6,638$, primary) and archived pool ($N = 8,000$, robustness check)

Metric	$N = 6,638$ (formal)	$N = 8,000$ (archive)
Directed edges ($k = 10$)	66,380	80,000
Mean edge length	4.99 km	4.94 km
P90 edge length	7.78 km	7.76 km
$MAD_h(E_{\text{geo}})$	512.4 m	512.2 m
P90 $ \Delta\text{Elevation} $	1,085.0 m	1,087.9 m
$MAD_y(E_{\text{geo}})$	33.4 Mg/ha	34.2 Mg/ha
P90 $ \Delta\text{AGB} $	76.0 Mg/ha	78.0 Mg/ha
Edges with $ \Delta\text{Elevation} > 800$ m	14,210 (21.41%)	17,191 (21.49%)
Edges with $ \Delta\text{AGB} > 70$ Mg/ha	8,125 (12.24%)	10,408 (13.01%)
Land-cover mismatch among high-AGB-difference edges	60.6%	63.9%

The distance computation is correct and the coordinate data contain no anomalies (Sec. 4); the root cause is not data quality but a methodological limitation of Geo kNN—in complex terrain, two-dimensional proximity does not equal ecological proximity. Fig. 3 shows the edge elevation- and AGB-difference distributions, the joint scatter of $|\Delta\text{Elev}|$ vs $|\Delta\text{AGB}|$, and the high-risk edge proportions, on the formal clean sample.

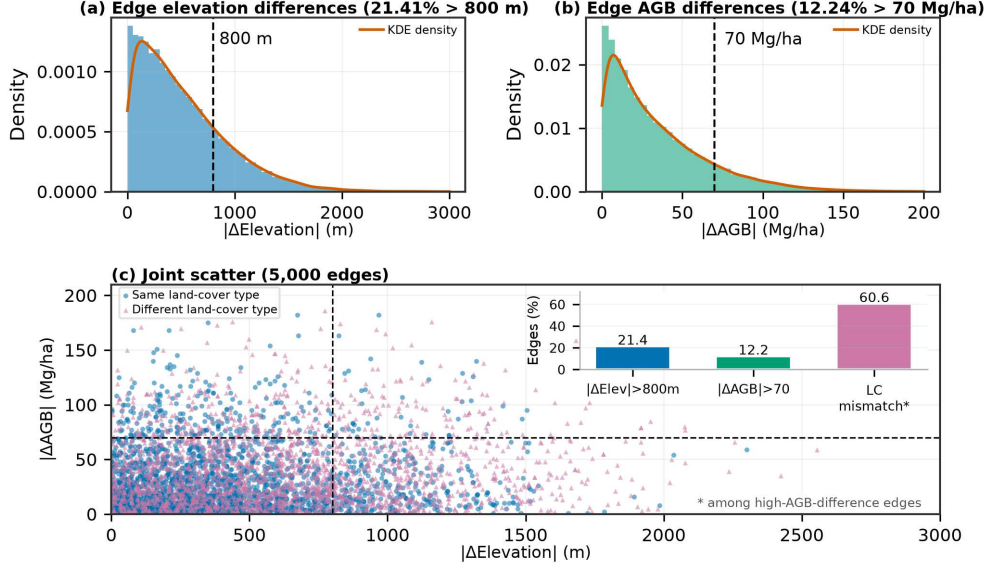


Fig. 3 Geo kNN edge-quality diagnostics ($k=10$) on the formal clean sample ($N=6,638$, 66,380 directed edges). (a) Edge elevation differences $|\Delta\text{Elevation}|$: 21.41% exceed the 800 m high-risk threshold (dashed). (b) Edge AGB differences $|\Delta\text{AGB}|$: 12.24% exceed the 70 Mg/ha high-risk threshold (dashed). (c) Left: $|\Delta\text{Elev}|$ vs $|\Delta\text{AGB}|$ scatter over 5,000 sampled edges, colored by whether the end-points share a MODIS land-cover type; the upper-right quadrant holds the double-high-risk edges (terrain-induced misconnections). Right: the three high-risk edge proportions; the land-cover mismatch rate is 60.6% when restricted to edges with a high AGB difference, the dominant pattern being forest–meadow transitions

5.3 Sensitivity of EGSG to the Elevation Gate

To select the EGSG elevation gate threshold τ , we rely on two independent sources of evidence.

Label-free primary justification. The mutual-information knee rule of Sec. 5.8 selects τ from the retained-edge mutual-information curve (Fig. 11) without consulting the test set; on the clean-protocol samples it independently recovers $\tau = 250$ m for AGB with full recovery of the achievable gating gain (recovery 1.00). A fixed 250 m therefore coincides with a data-driven rule rather than a single-point optimum tuned on the test set.

Confirmatory 10-seed sweep. A 10-seed threshold sensitivity analysis on the 6,638 clean-protocol samples (Table 3) confirms this inverted-U optimum independently: $\tau = 250$ m attains the highest mean R^2 (0.409), with $\tau = 500$ m statistically indistinguishable (0.408), while an overly small τ (100 m; too sparse, 112 isolated nodes, $R^2 = 0.394$) and an overly large τ (1000 m; too permissive, $R^2 = 0.379$, near the no-gate baseline of 0.362) both underperform. Fig. 4 visualizes the sweep and the associated graph-structure changes.

Robustness. The discussion following Table 3 establishes the 250–500 m plateau as a stable effective range rather than a reliance on precise single-point selection. Throughout, 250 m is not interpreted as theoretically optimal or universally transferable; different regions, sampling densities, or target variables require recalibration.

Table 3 EGS threshold sensitivity (10 seeds, $N = 6,638$, clean protocol). Edge counts and isolated nodes are computed on the 6,638 clean samples

Threshold τ	Edges	Isolated nodes	$\text{MAD}_y(E)$ (Mg/ha)	R^2 (mean $\pm$ std)
100 m	37,142	112	30.0	0.394 ± 0.024
250 m	60,698	6	30.7	0.409 ± 0.023
500 m	65,717	0	31.5	0.408 ± 0.020
1000 m	66,376	0	32.9	0.379 ± 0.024
no gate (Geo kNN)	66,380	0	33.4	0.362 ± 0.019

Paired tests show that the gain of 250 m over Geo kNN (no gate) is $\Delta R^2 = +0.048$ ($p = 6.0 \times 10^{-7}$), confirming that the elevation gating mechanism is effective. The 250–500 m plateau is a stable effective range: $\tau = 500$ m yields $R^2 = 0.408 \pm 0.020$ ($\Delta R^2 = +0.046$, $p = 7.8 \times 10^{-8}$), statistically indistinguishable from the 250 m optimum, whereas both the tighter $\tau = 100$ m and the looser $\tau = 1000$ m settings are significantly worse ($\Delta R^2 = +0.033$, $p = 9.4 \times 10^{-5}$ and $\Delta R^2 = +0.017$, $p = 2.2 \times 10^{-4}$, respectively).

Why is 250 m reasonable? The role of τ is to trade off edge homogeneity against graph connectivity. An overly small τ (100 m) filters cross-elevation edges and improves homogeneity (MAD_y drops to 30.0 Mg/ha) but fragments the graph—112 isolated nodes, only 37,142 edges remain (44% of the Geo kNN edges pruned), so many nodes lose most of their message-passing neighborhood, and R^2 drops to 0.394. An overly large τ (no gate) retains all Geo kNN edges, including many heterogeneous edges spanning different elevation bands—at no gate $\text{MAD}_h = 512.4$ m, 4.2 times the 122.6 m of 250 m—so connectivity is good but edge quality degrades, and R^2 drops to 0.362. The 250 m threshold sits at the trade-off point: only 6 isolated nodes remain (connectivity essentially intact) while heterogeneous edges are effectively filtered (MAD_y to 30.7 Mg/ha, MAD_h from 512.4 m to 122.6 m).

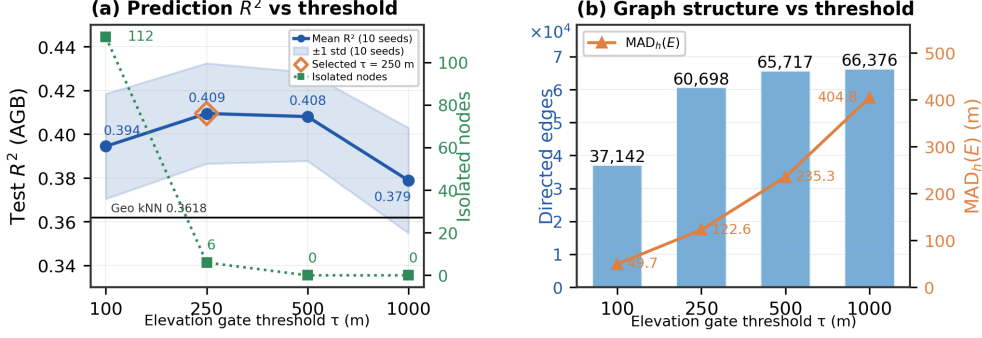


Fig. 4 EGSG threshold sensitivity (10 seeds, $N=6,638$, clean protocol). (a) Left axis: R^2 vs τ curve (solid line, mean $\pm 1\sigma$ shading over 10 seeds), with an inverted-U shape— $\tau=250$ m achieves the highest $R^2=0.409$ (diamond marker), and the no-gate Geo kNN baseline is $R^2=0.362$ (solid line); right axis: number of isolated nodes (dotted line with square markers), reaching 112 at $\tau=100$ m and 0 for $\tau \geq 500$ m. (b) Left axis: directed edge count vs τ (bars), rising from 37,142 at $\tau=100$ m to 66,376 at $\tau=1000$ m (the full Geo kNN edge set has 66,380); right axis: $MAD_h(E)$ vs τ (line with triangle markers), monotonically increasing from 49.7 m at $\tau=100$ m to 404.8 m at $\tau=1000$ m (512.4 m at no gate). This figure characterizes the trade-off mechanism of τ : when τ is too small (100 m), edge filtering is overly strict, the graph fragments (112 isolated nodes, 44% of edges pruned), and R^2 drops to 0.394; when τ is too large (≥ 1000 m), gating becomes ineffective, terrain-heterogeneous edges re-enter the graph, and R^2 falls back toward the no-gate level; $\tau=250$ m achieves a trade-off between edge homogeneity ($MAD_h=122.6$ m) and graph connectivity (only 6 isolated nodes)

5.4 Effectiveness of EGSG for Local Spatial Interpolation

Under the clean random-split protocol (60/20/20 random partition, 10 seeds, $N = 6,638$), EGSG-250m + GCN achieves $R^2 = 0.409 \pm 0.023$, significantly outperforming all baselines—random forest (RF), multilayer perceptron (MLP), and gradient-boosting models (XGBoost, LightGBM; default and tuned) (Table 4). The headline improvement over the Geo kNN graph is $\Delta R^2 = +0.048$ ($p = 6.0 \times 10^{-7}$, paired t -test), 95% CI $[+0.039, +0.056]$ over the per-seed paired differences; all ten are positive (minimum $+0.027$, maximum $+0.065$). Notably, the uncorrected Geo kNN + GCN is not significantly better than random forest ($\Delta R^2 = -0.002$, $p = 0.49$), so the improvement comes entirely from elevation-gated edge pruning, not the GNN architecture.

k sensitivity. To verify that the conclusion does not depend on $k = 10$, we sweep $k \in \{3, 5, 10, 15, 20\}$ under the clean protocol (Table 5). The EGSG advantage over Geo kNN is significant for every k and grows monotonically ($\Delta R^2 = +0.034$ at $k = 3$ to $+0.055$ at $k = 20$), confirming the knowledge-driven construction benefits any reasonable neighborhood size rather than a single tuned choice.

Performance under random split reflects **local spatial interpolation** ability—predicting held-out locations within the same region—rather than strict extrapolation. Fig. 5 ranks the methods by mean test R^2 and, for each baseline, reports the paired ΔR^2 of EGSG minus that baseline over the same ten seeds.

To make the spatial effect of the gate concrete, Fig. 6 maps the per-sample prediction difference $\Delta \hat{y} = \hat{y}_{\text{EGSG}} - \hat{y}_{\text{geo}}$ across the study region for AGB and NDVI, under both the clean random split (local interpolation) and the buffer-30 km split (extrapolation). Under the random split the gate shifts predictions through the region and

Table 4 Clean random-split results (10 seeds, $N = 6,638$). ΔR^2 is computed as $R^2_{\text{EGSG}} - R^2_{\text{method}}$; p values from paired t -tests. Headline comparison: $\Delta R^2 = +0.048$, $p = 6.0 \times 10^{-7}$, 95% confidence interval $[+0.039, +0.056]$. RMSE (root mean squared error) and MAE (mean absolute error) are back-transformed to AGB units (Mg/ha) from the standardized target; Train (s) single training run on one CPU core (i9-14900HX); hyperparameter-search time excluded

Method	R^2 (mean $\pm$ std)	RMSE	MAE	ΔR^2 vs EGSG	p	Train (s)
RF	0.364 ± 0.023	33.27	25.55	+0.045	3.4×10^{-6}	3.79
MLP	0.378 ± 0.021	32.90	25.19	+0.031	1.9×10^{-5}	2.54
XGBoost	0.352 ± 0.029	33.57	25.55	+0.057	6.8×10^{-6}	0.88
XGBoost (tuned)	0.361 ± 0.029	33.35	25.38	+0.049	1.5×10^{-5}	0.88
LightGBM (default)	0.346 ± 0.029	33.72	25.68	+0.063	1.2×10^{-6}	0.39
LightGBM (tuned)	0.362 ± 0.028	33.33	25.38	+0.048	1.9×10^{-5}	0.39
Geo kNN + GCN	0.362 ± 0.019	33.33	25.87	+0.048	6.0×10^{-7}	29.15
EGSG-250m + GCN	0.409 ± 0.023	32.06	24.57	—	—	26.80

Table 5 k sensitivity (clean protocol): EGSG vs Geo kNN at $k \in \{3, 5, 10, 15, 20\}$

k	ΔR^2 (EGSG – Geo)	p
3	+0.034	1.0×10^{-6}
5	+0.038	< 0.001
10	+0.048	< 0.001
15	+0.052	< 0.001
20	+0.055	< 0.001

raises R^2 ; under the buffer split the shown test locations form a contiguous sub-region and the sign of the effect flips, coinciding with the extrapolation boundary of Sec. 5.6. The spatial maps thus provide the granular counterpart of the aggregate headline gain of Table 4.

5.5 Generalization of Graph Construction Across Backbones

To examine whether EGSG is merely an artifact of the GCN architecture, we compare Geo kNN and EGSG under four backbones: GCN [34], GAT [35], GraphSAGE [36], and APPNP [37] (Table 6), all under the clean random-split protocol (10 seeds, $N = 6,638$).

EGSG brings stable improvements with GCN, GAT, and APPNP, indicating its gains come mainly from topology correction at the construction stage rather than a fortuitous property of GCN. With GraphSAGE the improvement is not significant ($\Delta R^2 = -0.001$, $p = 0.642$): different message-passing mechanisms have different sensitivities to edge-quality improvement, so EGSG is an inductive bias at the

(a) Ranked performance on the clean random split

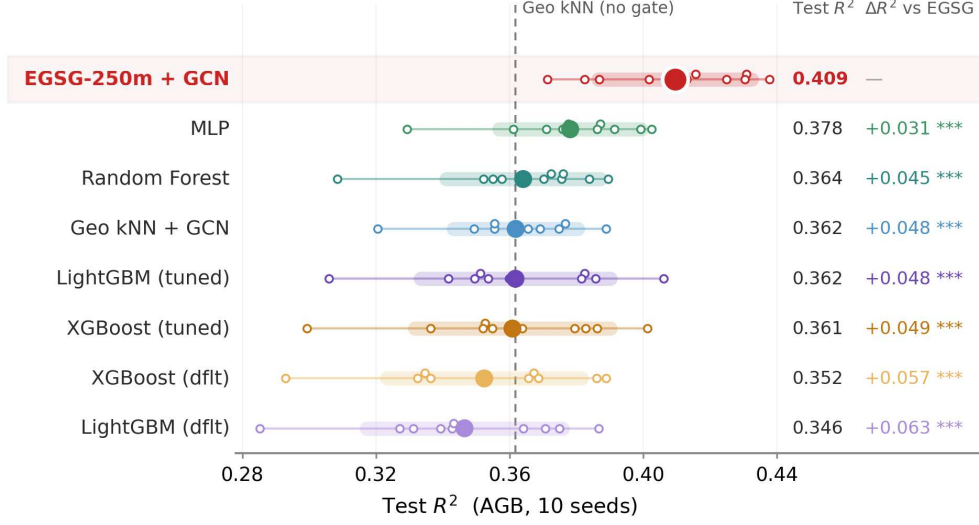


Fig. 5 Performance comparison of the methods under the clean random split (10 seeds, $N=6,638$, 60/20/20 random partition). Methods are ranked by mean test R^2 ; for each method the thin rail spans the seed minimum to the seed maximum, the shaded band is ± 1 standard deviation, the small open circles are the ten individual seeds and the large filled circle is their mean. The two right-hand columns give the mean test R^2 and the paired ΔR^2 of EGSG minus that method, with significance stars (***) $p < 0.001$, ** $p < 0.01$, * $p < 0.05$). EGSG-250m + GCN reaches 0.409 ± 0.023 (highlighted, RMSE=32.06 Mg/ha), while the remaining methods range from LightGBM default (0.346 ± 0.029) through XGBoost default (0.352 ± 0.029), XGBoost tuned (0.361 ± 0.029), LightGBM tuned (0.362 ± 0.028), Geo kNN + GCN (0.362 ± 0.019) and RF (0.364 ± 0.023) to MLP (0.378 ± 0.021); all seven paired differences are significantly positive. The dashed vertical line marks the mean of Geo kNN (no gate). Statistical tests: EGSG vs Geo kNN gives $\Delta R^2 = +0.048$ ($p = 6.0 \times 10^{-7}$, 95% confidence interval $[+0.039, +0.056]$), with all 10 seeds agreeing in direction; Geo kNN is not significantly different from RF ($\Delta R^2 = -0.002$, $p = 0.49$), indicating that the improvement comes entirely from elevation-gated edge pruning, rather than from the GNN architecture itself

Table 6 EGSG vs Geo kNN across backbones (10 seeds, $N = 6,638$).

Backbone	Geo R^2 (mean $\pm$ std)	EGSG R^2 (mean $\pm$ std)	ΔR^2	p (t -test)
GCN	0.362 ± 0.019	0.409 ± 0.023	+0.048	6.0×10^{-7}
GAT	0.368 ± 0.015	0.427 ± 0.022	+0.058	2.3×10^{-6}
APNP	0.374 ± 0.016	0.403 ± 0.023	+0.029	8.1×10^{-6}
GraphSAGE	0.415 ± 0.025	0.413 ± 0.024	-0.001	0.642

graph construction level whose gains do not automatically transfer to all propagation mechanisms.

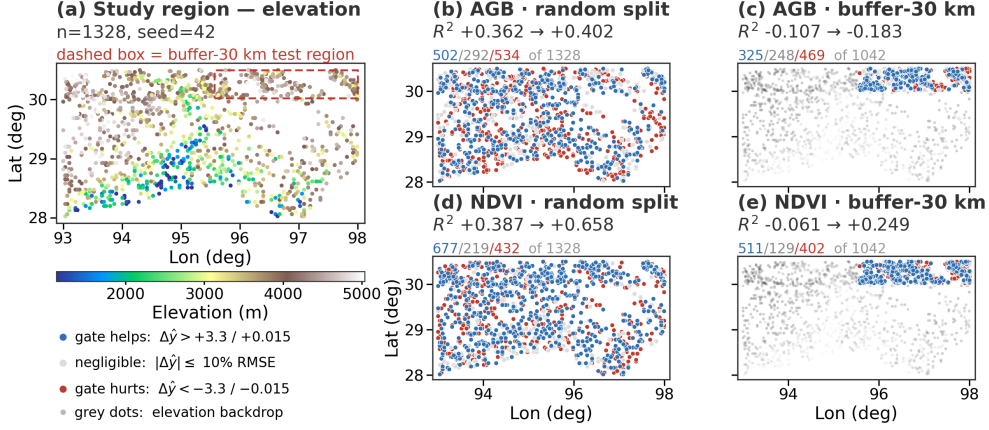


Fig. 6 Spatial distribution of the per-sample prediction difference $\Delta\hat{y} = \hat{y}_{\text{EGSG}} - \hat{y}_{\text{geo}}$ (single run, seed 42; multi-seed values in Tables 4, 7 and 9). (a) Elevation of the study region (m), shared by all panels; the dashed box marks the buffer-30 km test region. (b) AGB under the clean random split and (c) AGB under the buffer-30 km split; (d, e) the same two splits for NDVI. The annotation below each map title gives the single-run R^2 of Geo kNN and of EGSG (geo $\rightarrow$ EGSG), and the line above each map counts the test samples in the three classes. Dot colour encodes the sign of $\Delta\hat{y}$ in each target’s native units, in three classes separated by a fixed threshold of 10% of the Geo kNN random-split RMSE (± 3.3 Mg/ha for AGB, ± 0.015 for NDVI)—blue = gate helps, grey = negligible, red = gate hurts—so that only displacements large enough to matter are counted as a change; small grey dots additionally show the elevation backdrop of panel (a). The gate shifts predictions across the region and improves R^2 under the random split; under the buffer split the test locations form a contiguous sub-region and for AGB the gate fails to help, coinciding with the extrapolation boundary of Sec. 5.6, whereas for NDVI it still protects, consistent with the 10-seed $\Delta R^2 = +0.239$

5.6 Failure of EGSG Under Strict Spatial Extrapolation

To examine the applicability boundary of EGSG, we evaluate EGSG and Geo kNN under block split, buffer split, and elevation holdout (Table 7). All eight splits use 10 seeds under the clean protocol with the leakage-free training variant of Sec. 4.6: edges connecting test-region and non-test-region nodes are deleted, so the test region relies only on its internal edges.

EGSG significantly outperforms Geo kNN under random and both block splits ($\Delta R^2 = +0.047$ to $+0.066$, $p < 10^{-4}$): the topology correction brings gains when local training support is preserved or only moderately reduced. **Under buffer split the sign flips: EGSG is consistently worse than Geo kNN**— $\Delta R^2 = -0.054$ ($p = 0.027$) at buffer 30 km and -0.083 ($p = 0.007$) at buffer 50 km, with all 10 seeds agreeing in direction. This is the **primary strong evidence** for the strict-extrapolation failure boundary. Under Elevation mid holdout EGSG still gains ($+0.082$, $p = 1.1 \times 10^{-5}$). The low/high elevation holdouts are hard failures for both methods: under Elevation low the uncorrected geo graph collapses to -7.05 ± 3.42 and EGSG to -0.57 ± 0.48 (the large positive ΔR^2 reflects the collapse, not usable accuracy); under Elevation high EGSG shows a negative trend ($\Delta R^2 = -0.068$) but is non-significant ($p = 0.594$), serving only as **auxiliary evidence** of terrain-extrapolation risk. Notably, the non-topological baselines (RF, MLP) degrade least

Table 7 Spatial generalization results (clean protocol, 10 seeds, leakage-free variant). All rows use 10 seeds; block-row N_{test} is averaged over seeds. $\Delta R^2(s)$ is defined in Sec. 4.6; p values from paired t -tests. Negative R^2 indicates performance below predicting the mean

Split s	N_{test}	Geo kNN R^2	EGSG R^2	$\Delta R^2(s)$	p (t -test)
Random	1,328	0.362 ± 0.019	0.409 ± 0.023	+0.048	6.0×10^{-7}
Block 50 km	1,043	0.257 ± 0.068	0.305 ± 0.056	+0.047	1.6×10^{-5}
Block 100 km	1,797	0.226 ± 0.123	0.292 ± 0.112	+0.066	4.0×10^{-5}
Buffer 30 km	1,042	-0.137 ± 0.075	-0.191 ± 0.034	-0.054	0.027
Buffer 50 km	1,042	-0.140 ± 0.062	-0.223 ± 0.060	-0.083	0.007
Elevation low	2,191	-7.053 ± 3.419	-0.573 ± 0.477	+6.480 [†]	1.3×10^{-4}
Elevation mid	2,256	0.041 ± 0.028	0.123 ± 0.014	+0.082	1.1×10^{-5}
Elevation high	2,191	-0.449 ± 0.213	-0.517 ± 0.388	-0.068	0.594

Note: [†]In the elevation-low split both methods fail ($R^2 < 0$); the large positive ΔR^2 reflects the Geo kNN collapse (-7.05) rather than usable EGSG accuracy.

under the holdout splits (Fig. 7(a))—the topology that helps in distribution hurts most once support and distribution shift.

Buffer phase diagram. A 6-point buffer scan (random, 10, 20, 30, 50, 80 km) shows the EGSG advantage decays monotonically as the buffer radius grows: $\Delta R^2 = +0.048 \rightarrow -0.009$ (10 km, n.s.) $\rightarrow -0.017$ (20 km, n.s.) $\rightarrow -0.054$ (30 km, $p = 0.027$) $\rightarrow -0.083$ (50 km, $p = 0.007$) $\rightarrow -0.117$ (80 km, $p < 0.0001$). The Geo kNN baseline is insensitive to the buffer radius (a plateau near $R^2 \approx -0.13$), while EGSG degrades monotonically; the EGSG advantage drops to near zero at 10 km ($\Delta R^2 = -0.009$, n.s.) and turns negative beyond, even though the mean LSI is still 2.31 at that radius; the mean LSI reaches zero only at 30 km and beyond, so the sign flip precedes complete support removal.

Comparison with classical spatial models. The buffer failure is not specific to GNNs. Under the same buffer-30 protocol, geographically weighted regression (GWR) [7] drops from $R^2 = 0.338$ (random) to -0.191 —a degradation almost identical to EGSG’s (-0.191)—and ordinary kriging [38] collapses from 0.325 to -1.169 . Strict-extrapolation failure is a shared property of local spatial methods, graph-based or classical; what EGSG changes is only the *slope* of the degradation relative to the uncorrected Geo kNN graph.

The EGSG gain therefore depends on preserving some local knowledge support between training and test; when the buffer split severs this support, the gains consistently fail. Support removal is, however, compounded by covariate shift in the elevation holdouts (Sec. 5.7). Fig. 7 shows the R^2 comparison between EGSG and Geo kNN under the different spatial splits.

Completeness is not a substitute for edge quality. Any construction that restores connectivity—whether by relaxing the gate or by backfilling gated-out edges—re-introduces edge-quality violations. The add-back dose-response control (Table 10) supplies the direct test: re-adding 0/25/50/100% of the gated-out elevation-violating edges produces monotonic degradation under both buffer protocols (b30 $p = 0.024$,

b50 $p = 0.009$), because each re-inserted edge injects elevation-inconsistent noise into a graph whose local training support has already been removed (mean LSI = 0 at 30 km, Table 8). At the other extreme, eliminating every cross-elevation edge ($\tau = 100$ m, Table 3) minimizes MAD_y yet isolates 112 nodes and drops R^2 to 0.394. Connectivity is therefore a means of enabling message passing, not a goal in itself. Among the alternative constructions we examined (Table 10), the connectivity-repairing (back-fill) variant is neutral because EGSG is already connected (mean degree 18.3, 0.09% isolated); the alternate domain gates (land-cover, DEM residual, tree-canopy height) do not match elevation gating; and an attention-based learned structure (LooseGAT) does not reproduce its reported gain once repair is isolated and trades interpretability for additional parameters.

Target dependence of the boundary. The buffer failure above is specific to AGB. Repeating the buffer-30 km protocol on the two second validation targets (10 seeds, same leakage-free variant; Sec. 5.8) shows the EGSG-geo difference under strict extrapolation is itself target-dependent. On LST—a smoothly varying field—neither graph model collapses (Geo $+0.463 \pm 0.035$ vs. EGSG $+0.445 \pm 0.022$, $\Delta R^2 = -0.017$, $p = 0.082$, n.s.), while the non-topological baselines do (RF $+0.122 \pm 0.023$, MLP $+0.032 \pm 0.029$). On NDVI—a locally varying vegetation index—the uncorrected geo graph collapses to -0.050 ± 0.035 under the buffer split, whereas EGSG retains $+0.188 \pm 0.026$ ($\Delta R^2 = +0.239$, $p < 10^{-6}$). Extrapolation degradation is therefore a property of the target field structure (AGB: EGSG worse; LST: no difference; NDVI: EGSG better), not of the graph method.

5.7 Mechanistic Explanation: Edge Quality and Local Knowledge Support

The binary pattern—EGSG works under local interpolation but fails under strict extrapolation—is not coincidental. This section reveals its mechanism through two analyses.

Quantitative relationship between edge quality and performance (Fig. 8). Across the five AGB gate settings (no gate, $\tau \in \{100, 250, 500, 1000\}$ m), R^2 is negatively correlated with both $\text{MAD}_y(E)$ (Pearson $r = -0.78$, $p = 0.12$) and $\text{MAD}_h(E)$ (Pearson $r = -0.82$, $p = 0.09$): the more homogeneous the edges, the better the performance, with a strong effect size though the five-point statistics do not reach conventional significance. Cleaner edges are not always better: an overly strict threshold (100 m) reduces MAD_y (30.0 Mg/ha) but fragments the graph (112 isolated nodes, 44% of edges pruned), dropping R^2 to 0.394. The 250 m threshold sits at the trade-off point between edge quality and connectivity. The 12-point diagnostic-benefit correspondence of Sec. 5.8 corroborates this with adequate power (Spearman $\rho = -0.741$, $p = 0.006$). The mechanism is therefore **moderate filtering of terrain misconceptions while preserving sufficient connectivity**, not maximizing edge purity.

Local Support Index (LSI) analysis. The LSI (Eqs. (6)–(7)) characterizes the number of local training neighbors available to test samples under each split (Table 8).

Under random split, $\bar{\text{LSI}}$ is 28.1—each test sample has, on average, about 28 training neighbors within a 30 km radius—from which EGSG obtains knowledge through

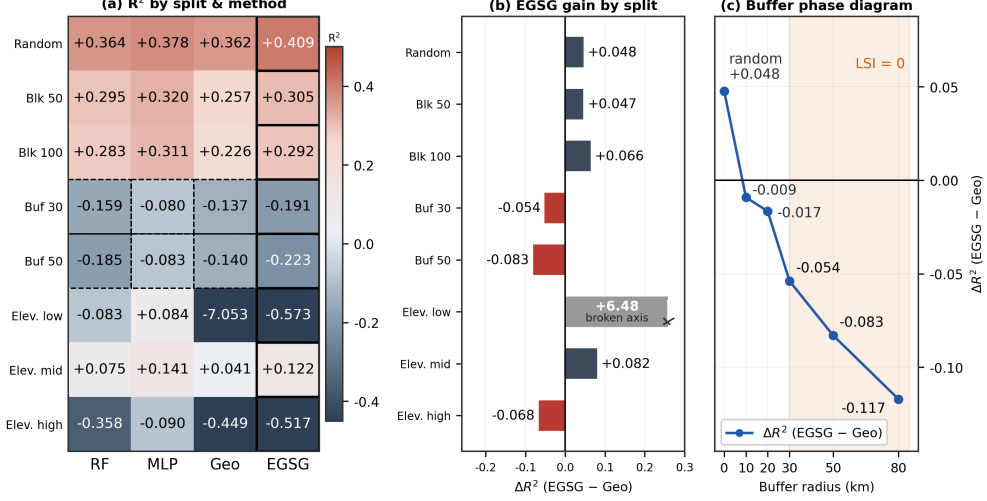


Fig. 7 Spatial generalization under strict protocols (clean protocol, 10 seeds, leakage-free variant). (a) R^2 heatmap (8 splits $\times$ 4 methods; cell entries are the mean over the 10 seeds to three decimals, the standard deviations being omitted from the panel for legibility and reported in Table 7; the EGSG column is bold-bordered and the buffer rows are dashed). The colour scale spans -0.45 to $+0.50$, so the strongly negative Elevation-low Geo cell sits at the floor of the scale and its value (-7.053) is printed in the cell. (b) Paired ΔR^2 (EGSG - Geo kNN) per split; the paired t -test significance is stated here rather than drawn as symbols in the panel: positive under random and both block splits ($+0.047$ to $+0.066$, p from 6.0×10^{-7} to 4.0×10^{-5}), negative under both buffers (-0.054 , $p = 0.027$; -0.083 , $p = 0.007$), positive under Elevation mid ($+0.082$, $p = 1.1 \times 10^{-5}$) and nonsignificant under Elevation high (-0.068 , $p = 0.594$). The Elevation-low bar ($+6.48$, grey) is drawn with a broken axis, both methods failing there (EGSG -0.57 vs Geo -7.05). (c) Buffer phase diagram: the radius-0 entry is the random split, and the paired gain decays monotonically with the buffer radius ($+0.048$ at random to -0.117 at 80 km); the shaded region marks the radii at which the mean LSI has already reached zero (30 km and beyond)

message passing to achieve effective local interpolation. Under buffer split, $\overline{\text{LSI}}$ is strictly 0: the training samples around each test sample are forcibly removed, graph propagation has no available local neighbors, and the topology correction cannot produce gains ($\mathcal{D} = 0.60$ – 0.63). The eight-split degradation is best described by *two* factors rather than LSI alone: $\mathcal{D}(s)$ correlates negatively with LSI (Pearson $r = -0.72$, $p = 0.044$; Spearman $\rho = -0.78$, $p = 0.023$) and, more strongly, with the magnitude of the train–test elevation shift (Pearson $r = +0.976$, $p = 3.4 \times 10^{-5}$; Spearman $\rho = +0.976$, $p = 3.3 \times 10^{-5}$). The buffer splits isolate the support-removal factor but also carry a moderate shift ($\Delta_{\text{elev}} = -683$ m and -678 m); the elevation holdouts combine near-zero support (LSI 2.4–7.9) with extreme shifts ($+1,672$ m for Elevation low, AGB shift -39.6 Mg/ha; $-1,395$ m for Elevation high, AGB shift $+36.1$ Mg/ha), giving the largest degradations ($\mathcal{D} = 0.93$ – 0.98). Under Elevation low the few available neighbors (LSI = 2.37) lie in a shifted distribution: the knowledge they provide is not merely useless—for the uncorrected geo graph it is actively harmful (collapse to -7.05 vs EGSG’s -0.57). Split diagnostics support this: the training–test minimum distance under buffer split is 30 km (set by the protocol), and the buffer test quadrant

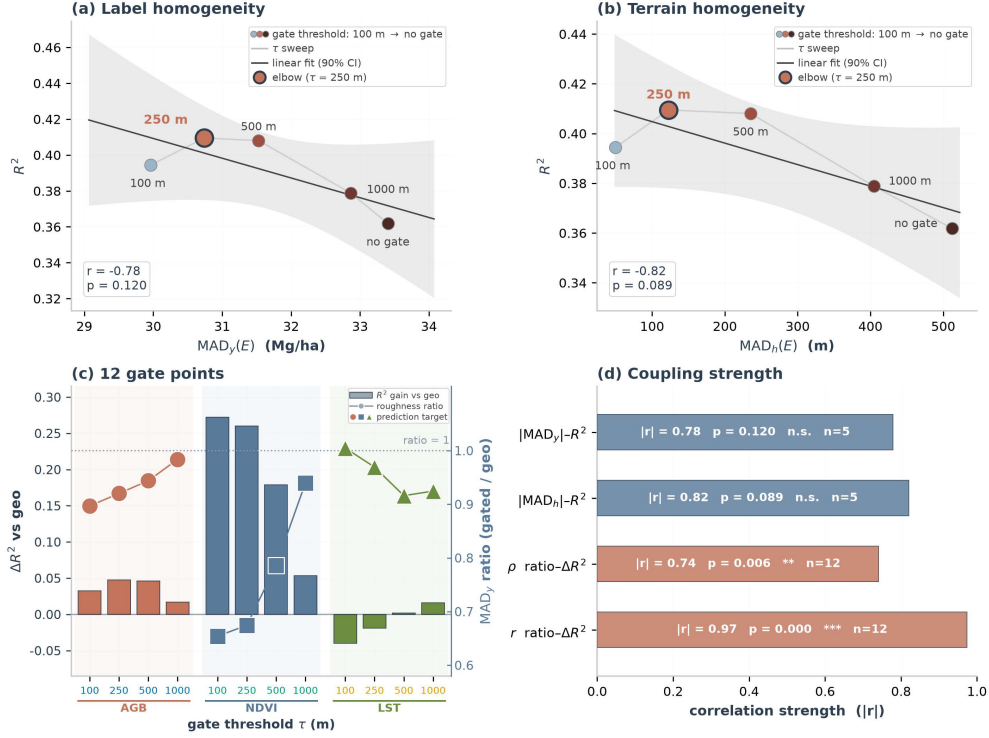


Fig. 8 Quantitative relationship between edge-quality metrics and R^2 (10 seeds). (a) $MAD_y(E)$ and (b) $MAD_h(E)$ against R^2 for the five AGB gate settings (no gate and $\tau = 100/250/500/1000$ m); the grey line follows the sweep in order of τ , the dark line is a linear fit with its 90% confidence band, and the enlarged point marks the elbow ($\tau = 250$ m). Both relationships are negative (Pearson $r = -0.78$, $p = 0.12$ and $r = -0.82$, $p = 0.09$), with a large effect size that the five-point statistics do not carry to conventional significance. (c) The twelve cross-target gate points (AGB, NDVI, LST $\times$ four thresholds): bars give ΔR^2 over Geo kNN (left axis) and the line gives the roughness ratio (MAD_y of the gated graph over that of the geo graph; right axis), with the dotted line at ratio= 1 marking no purification. (d) Coupling strength: the roughness ratio is strongly negatively correlated with the gating gain, both by rank ($\rho = -0.74$, $p = 0.006$) and linearly ($r = -0.97$, $p < 0.001$), whereas neither of the two five-point AGB relationships reaches significance

lies on average 683 m above the training pool. Fig. 9 shows LSI, Δelev , and EGSG degradation.

This analysis refines the root cause of the EGSG boundary: its gain mechanism—obtaining knowledge from local training neighbors through message passing—is systematically severed under strict spatial extrapolation, and the degradation is compounded when the retained neighbors come from a covariate-shifted distribution, as in the elevation holdouts. As noted in Sec. 4.5, LSI is a split-level descriptor of knowledge support, not a sample-level error predictor; it explains degradation fully only when covariate shift is controlled.

Table 8 Mean LSI under different spatial splits ($r = 30$ km, $\tau = 250$ m), EGSG performance degradation $\mathcal{D}(s)$ (Sec. 4.6), and the mean elevation shift Δelev (m; mean elevation of the training pool, i.e. all samples outside the test set, minus that of the test set, so a negative value means the test samples lie higher). All values computed within the 10-seed protocol experiments on the 6,638 clean samples; LSI and $\mathcal{D}(s)$ under buffer and elevation splits are seed-independent (fixed test sets), so their Δelev is exact, whereas the random and block rows report mean $\pm$ std over the 10 seeds—their shift is small on average but varies strongly from seed to seed

Split s	$\overline{\text{LSI}}$ (30 km)	$\mathcal{D}(s)$ (EGSG)	Δelev (m)
Random	28.07 ± 0.50	0.000 (baseline)	$+5 \pm 30$
Block 50 km	18.02 ± 4.03	0.105	$+7 \pm 216$
Block 100 km	6.79 ± 1.47	0.118	$+41 \pm 622$
Elevation low	2.37	0.982	$+1,672$
Elevation mid	7.94	0.287	-273
Elevation high	5.81	0.926	$-1,395$
Buffer 30 km	0.00	0.600	-683
Buffer 50 km	0.00	0.633	-678

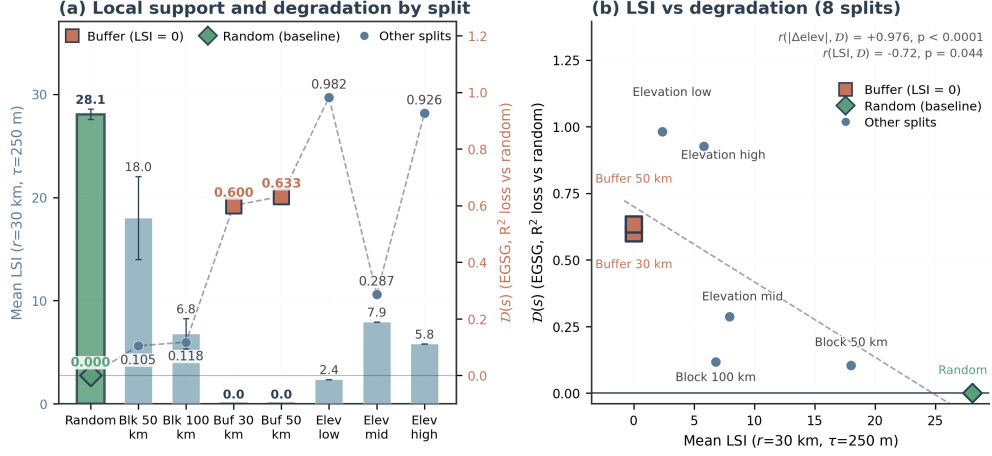


Fig. 9 Relationship between the Local Support Index (LSI), covariate shift, and EGSG performance degradation ($r=30$ km, $\tau=250$ m; 10-seed protocol). (a) Left axis (bars): mean LSI under each spatial split—28.1 under Random, 18.0 under Block 50 km, 2.4 under Elevation low, and strictly 0 under Buffer 30/50 km (the protocol removes all training samples within the buffer); right axis (scatter): the corresponding EGSG degradation $\mathcal{D}(s)$ —0 under Random, 0.10–0.12 under the block splits, 0.60–0.63 under the buffer splits, and 0.93–0.98 under Elevation low/high. (b) LSI vs $\mathcal{D}$ scatter across the 8 splits (Pearson $r = -0.72, p = 0.044$): degradation is dominated by covariate shift ($r(|\Delta\text{elev}|, \mathcal{D}) = +0.976, p = 3.4 \times 10^{-5}$; Spearman $\rho = +0.976, p = 3.3 \times 10^{-5}$) and secondarily by support removal—the buffer splits combine zero support with a moderate shift ($-683/-678$ m), while Elevation low/high combine near-zero support with extreme shifts ($+1,672/-1,395$ m). LSI remains a valid split-level descriptor of knowledge support, but it explains degradation fully only when covariate shift is controlled

5.8 Generalization Across Target Variables

The diagnosis and the gating construction are formulated for AGB, but the underlying claim is about the knowledge quality of spatial representations in general. To test generalization, we repeat the τ -sweep protocol (10 seeds) on two further targets from the same domain: mean growing-season NDVI (a vegetation index) and mean daytime land surface temperature (LST) (Table 9).

Table 9 Gate threshold sweep across three target variables (10 seeds, clean protocol). ΔR^2 is relative to the Geo kNN graph of the same target; p from paired t -tests

Target	τ	R^2 (mean $\pm$ std)	ΔR^2 vs Geo	p
AGB	Geo (no gate)	0.362 ± 0.019	—	—
AGB	100 m	0.394 ± 0.024	+0.033	0.0001
AGB	250 m	0.409 ± 0.023	+0.048	< 0.0001
AGB	500 m	0.408 ± 0.020	+0.046	< 0.0001
AGB	1000 m	0.379 ± 0.024	+0.017	0.0002
NDVI	Geo (no gate)	0.378 ± 0.019	—	—
NDVI	100 m	0.650 ± 0.015	+0.273	< 0.0001
NDVI	250 m	0.638 ± 0.012	+0.260	< 0.0001
NDVI	500 m	0.557 ± 0.016	+0.179	< 0.0001
NDVI	1000 m	0.431 ± 0.018	+0.053	< 0.0001
LST	Geo (no gate)	0.809 ± 0.012	—	—
LST	100 m	0.770 ± 0.013	−0.040	< 0.0001
LST	250 m	0.791 ± 0.013	−0.019	0.0001
LST	500 m	0.812 ± 0.012	+0.002	0.148
LST	1000 m	0.825 ± 0.012	+0.016	< 0.0001

Three findings emerge. First, the gating benefit is far from AGB-specific: on NDVI the gain reaches +0.27 at $\tau = 100$ m (a geo baseline of only 0.378, i.e., 72% relative improvement), and the NDVI geo collapse is stable across learning rates (3 seeds at $\text{LR} \in \{0.0005, 0.001, 0.002\}$ give $\Delta R^2 = +0.247/+0.258/+0.264$), excluding a hyperparameter artifact. Second, the *optimal gate is target-dependent*: NDVI—a locally varying vegetation index—prefers a tighter gate (100–250 m) than AGB (250–500 m), while LST—a smoothly varying temperature field—is *hurt* by the AGB-calibrated 250 m gate ($\Delta R^2 = -0.019$, $p = 0.0001$) and only benefits from a wide 1000 m gate ($\Delta R^2 = +0.016$, $p < 0.0001$). This is decisive against exporting a single gate threshold across variables. Third, the diagnostic–benefit correspondence generalizes: across all 12 gate points (3 targets $\times$ 4 thresholds), the edge-purity ratio (MAD_y of the gated graph relative to the geo graph) is strongly negatively correlated with the gain (Spearman $\rho = -0.741$, $p = 0.006$)—the more the gate purifies edges, the larger the gain, confirming edge quality as the operative mechanism, not a property of AGB. Consistent with this, the extrapolation boundary is likewise target-specific (Sec. 5.6): LST shows no AGB-style collapse for either graph model ($\Delta R^2 = -0.017$, n.s.), while

on NDVI the gate still protects ($\Delta R^2 = +0.239$, $p < 10^{-6}$). Fig. 10 visualizes the target-dependent τ curves and the per-point gains.

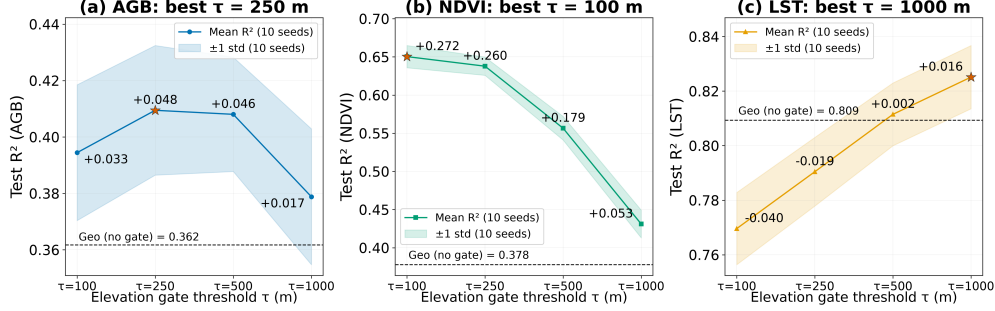


Fig. 10 Gate threshold sensitivity across three target variables (10 seeds, clean protocol; same data as Table 9). Each panel shows R^2 vs τ (mean $\pm 1\sigma$ shading over 10 seeds), the no-gate Geo kNN baseline of the same target (dashed line), the best τ (star), and the per-point ΔR^2 over the geo baseline. AGB peaks at $\tau=250$ m (0.409), NDVI at $\tau=100$ m (0.650, a +0.272 gain over its geo baseline), and LST at $\tau=1000$ m (0.825); the AGB-calibrated 250 m gate actually *hurts* LST (-0.019 , $p = 0.0001$). The optimal gate is therefore target-dependent and must be calibrated per target variable

Threshold-selection protocols. Because the optimal gate depends on the target field, we evaluate three τ -selection protocols across all three targets (10 seeds): a fixed 250 m default, a label-free quantile gate (QG; τ set from elevation-distribution quantiles to keep a fixed edge fraction, $\tau = 482$ m for AGB and NDVI and 486 m for LST), and a mutual-information knee rule (MI-knee; τ at the knee of the retained-edge MI curve, $\tau = 250$ m for AGB, 200 m for NDVI, 500 m for LST). Relative to an oracle choosing the per-target best τ , the mean recovery of the achievable gain is 0.251 (Fixed250), 0.596 (QG), and 0.704 (MI-knee); on LST the fixed-250 default achieves a recovery of -1.201 (worse than not gating). MI-knee matches or exceeds the fixed default on every target: it is identical on AGB, where the MI knee coincides with the 250 m default, and improves on NDVI ($+0.007$, $p = 0.0008$) and LST ($+0.021$, $p = 1.7 \times 10^{-6}$), and QG—which needs no labels—significantly rescues LST ($+0.022$, $p < 0.0001$) while leaving AGB unchanged. Consistent with this, the mutual information of retained edges exceeds that of gated-out edges by large effect sizes on all three targets (Cohen’s $d = 2.89$ – 8.66), and the MI-curve knee recovers the empirical optimal gate ordering exactly (Spearman $\rho = +1.000$, $p = 0.000$). Practical guidance: with labels, calibrate τ by the MI knee; without labels, use the label-free quantile gate; a universal fixed gate is not defensible. Figs. 11 and 12 visualize the mutual-information knee curves and the protocol comparison, respectively.

5.9 Ablations, Controls, Terrain-Aware Comparisons, and Efficiency

Ablations and controls. Table 10 summarizes the ablation and control experiments that defend the causal reading of the gating benefit.

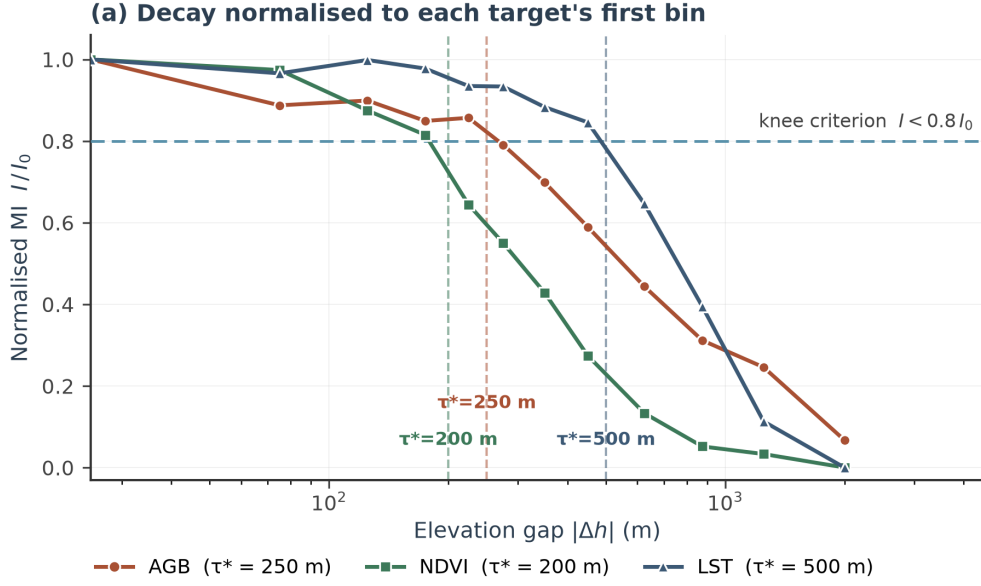


Fig. 11 Edge-level mutual information of the retained edges as a function of the elevation gate gap τ (Gaussian MI proxy, normalized to each target's first bin; 10 seeds, clean protocol). The knee of each curve—where the normalized MI first falls below the $0.8 I_0$ criterion (horizontal dashed line)—selects τ for that target: 250 m for AGB (rust), 200 m for NDVI (green) and 500 m for LST (blue), recovering the empirical optimal gate ordering exactly (Spearman $\rho = +1.000$). The mutual information of retained edges exceeds that of gated-out edges with large effect size on every target (Cohen's $d = 2.89\text{--}8.66$), so the knee acts as a label-free surrogate for the threshold

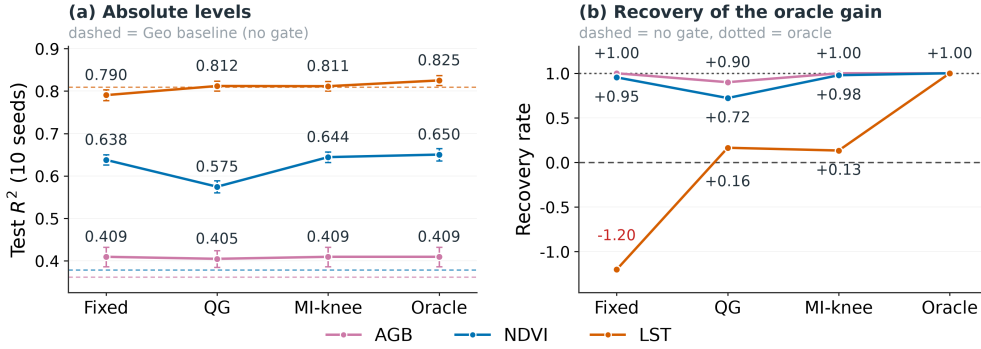


Fig. 12 Evaluation of three τ -selection protocols across the three target variables (10 seeds): fixed 250 m default (Fixed), label-free quantile gate (QG), mutual-information knee rule (MI-knee), and the per-target oracle (best τ from the sweep). (a) Mean R^2 achieved by each protocol, one line per target (error bars: ± 1 std; dashed lines: the no-gate Geo kNN baseline of each target). (b) Recovery of the oracle-achievable gating gain (dotted line at 1.0 marks the oracle; negative values labeled in red). Mean recoveries over the three targets: Fixed250 0.251, QG 0.596, MI-knee 0.704; on LST the fixed 250 m default achieves a recovery of -1.201 (worse than not gating), whereas MI-knee selects 500 m and the label-free QG rescues LST

Table 10 Ablations and controls (clean protocol unless stated). ΔR^2 is relative to EGSG-250m; negative values mean the variant loses against EGSG-250m. The connectivity-repair (backfill), learned-construction, random-pruning and alternative-gate rows form the family of alternative constructions we examine; each targets completeness or learned structure, and is accordingly neutral or penalized precisely where local support is absent

Condition	ΔR^2 vs EGSG-250m	p
Elevation-feature ablation: EGSG_noelev vs Geo_noelev (EGSG_noelev 0.400 ± 0.025 ; Geo_noelev 0.340 ± 0.019 ; topology recovery 163.6% vs -11.0% of the elevation-feature marginal contribution 0.034)	+0.060	$< 10^{-6}$
Random edge pruning (same directed edge count 60,698 and the same per-node degree distribution, random selection; mean over 3 instantiations): 0.364	-0.045	10^{-6}
Add-back dose response: 0/25/50/100% of gated-out edges re-added	monotonic degradation	b30 0.024 / b50 0.009
Backfill (connectivity repair to $d_{\min} = 5$): 0.409 ± 0.022 (random), -0.181 ± 0.029 (b30)	-0.0004 (random) / +0.0009 (b30)	0.685 / 0.853
Land-cover gating (soft, replace elevation gate)	+0.000	0.997
Land-cover gating (hard)	-0.010	0.018
DEM-residual gating (replace elevation gate)	-0.003	0.595
Tree-canopy-height gating (replace elevation gate) [39]	-0.006	0.311
SAR texture features added to EGSG input ($N = 335$)	+0.054	0.235
Label-aware outlier removal ($N = 332$ samples, Sec. 3)	+0.069 (R^2 : $0.414 \rightarrow 0.483$)	—

Three controls deserve emphasis. (i) *Elevation-feature ablation* (mirror experiment): with the elevation feature removed from both graphs’ inputs, the gated topology still beats the geo topology by +0.060 ($p < 10^{-6}$). The MLP baseline quantifies the elevation feature’s marginal contribution as +0.034; relative to this, the gated topology without the feature recovers 163.6% while the geo topology recovers -11.0% —the gate does not merely help the model exploit the elevation feature, it *encodes* the topographic knowledge in the topology. (ii) *Random-pruning control*: a geo graph pruned to the identical directed edge count (60,698) and the identical per-node degree distribution by random deletion yields $R^2 = 0.364$ (mean over 3 random-deletion instantiations), indistinguishable from the unpruned geo graph and -0.045 below EGSG ($p = 10^{-6}$): at least $\approx 95\%$ of the benefit comes from *which* edges are removed, not sparsity. (iii) *Add-back dose response*: re-adding 0/25/50/100% of the gated-out elevation-violating edges produces a monotonic degradation under both buffer protocols (b30 $p = 0.024$, b50 $p = 0.009$)—the violating edges carry the extrapolation harm, and the gate’s effect

is dose-dependent rather than binary. (iv) *Connectivity repair*: re-adding edges so that every node has degree $\geq d_{\min} = 5$ —the complement of the add-back control—changes nothing ($\Delta = -0.0004$, $p = 0.685$, random; $\Delta = +0.0009$, $p = 0.853$, buffer-30): EGSG is already well connected (mean degree 18.3; 6 of 6,638 vertices, 0.09%, isolated), so the benefit derives from *which* edges are removed rather than how many survive—completeness is a means, not a goal. Alternative gate attributes (land-cover type, DEM residual, tree canopy height) all fail to reproduce the benefit (none yields a positive improvement: the soft land-cover gate is null, $\Delta = +0.000$, $p = 0.997$; the remaining gates have $\Delta \leq 0$ with $p \geq 0.018$, land-cover hard gating being significantly *worse*), supporting elevation consistency as the operative domain constraint. SAR texture features do not help ($p = 0.235$, $N = 335$). The label-aware outlier removal of Sec. 3 is conservative: it removes only the 332 samples whose label lies far outside their feature-space neighborhood, and it *increases* the EGSG–geo gap from +0.048 to +0.069 ($R^2 = 0.414 \rightarrow 0.483$), so the headline effect does not ride on outliers.

Terrain-aware method comparison. We compare EGSG with three terrain-aware deep-learning baselines that encode elevation continuously rather than as a discrete gate: NNConv with elevation-consistent continuous kernels, soft gating (edge weights decaying with $|h_i - h_j|$), GATv2 with elevation-augmented features, and a 3D kNN built in (lon, lat, elevation) space (Table 11).

Table 11 Terrain-aware comparison (10 seeds; clean protocol). In-distribution: random split; extrapolation: buffer 30 km. R^2 values are mean $\pm$ std; ΔR^2 relative to EGSG-250m is computed in the text

Method	Random R^2 (mean $\pm$ std)	Buffer 30 km R^2
Geo kNN + GCN	0.362 $\pm$ 0.019	−0.137
EGSG-250m + GCN	0.409 $\pm$ 0.023	−0.191
NNConv (continuous elevation kernel)	0.417 $\pm$ 0.022	−0.123
Soft gating (exponential edge weights)	0.406 $\pm$ 0.024	−0.215
GATv2 (elevation-augmented features)	0.404 $\pm$ 0.026	−0.093
3D kNN (lon, lat, h)	0.368 $\pm$ 0.018	−0.170

In distribution, the strongest continuous encoder (NNConv) is statistically equivalent to the hard gate ($\Delta = +0.008$, $p = 0.060$), while requiring $\approx 17\times$ the parameters (EGSG-GCN: 5,313). The 3D kNN is significantly the weakest terrain-aware option ($\Delta = -0.042$ vs EGSG, $p < 0.001$): mixing elevation into the distance geometry is not a substitute for gating. (The 3D kNN is built in (x, y, z) space with the elevation axis rescaled to $s = 2.0 \text{ km} \times$ its standard score, then re-selects k neighbors by three-dimensional distance; because it applies no elevation-consistency gate, it removes no terrain-violating edges and stays near the no-gate baseline rather than reaching

gated purity.) Under buffer-30 extrapolation the source of the penalty is the *support-modifying* form of the constraint rather than “using terrain at all”: the variants that keep every geometric edge and fold elevation into features or kernels are the most graceful (GATv2 -0.093 , $p < 0.0001$ vs EGSG’s -0.191 ; NNConv -0.123 , $p = 0.032$), whereas both forms that attenuate support degrade sharply—the discrete gate to -0.191 and its soft analogue (edge weights decaying with $|h_i - h_j|$) even further to -0.215 ($p = 0.069$, n.s.). The 3D kNN, which retains all edges but mixes elevation into the distance metric, lies in between (-0.170 , $p = 0.153$, n.s.). Strict extrapolation remains hard for all terrain-aware variants, but encoders that do not remove support degrade more gently than the gate.

Efficiency. The gating step adds essentially no computational burden: graph construction (gating included) takes 0.095 s, and pruning makes training slightly *cheaper* than the uncorrected geo-GCN (26.8 s vs. 29.15 s on a single CPU core, Intel Core i9-14900HX) because it removes edges. A single GNN run is slower than a fast ML baseline (random forest: 3.79 s); inference over the full test set takes 0.037 s; and the GCN has only 5,313 parameters. The method is directly deployable in operational mapping pipelines.

6 Discussion

6.1 Principal Findings

This paper diagnoses the knowledge quality of the Geo kNN spatial graph representation in complex terrain and proposes a knowledge-driven repair. Four findings form the spine of the paper. (1) The default representation is measurably unreliable: 21.41% of the Geo kNN edges span elevation differences above 800 m and 12.24% span AGB differences above 70 Mg/ha, with 60.6% of the high-AGB-difference edges connecting different land-cover types (Sec. 5.2). (2) The knowledge-driven repair works under local interpolation: EGSG improves R^2 by $\Delta R^2 = +0.048$ ($p < 10^{-6}$, 95% CI $[+0.039, +0.056]$), outperforming tuned gradient-boosting baselines, and the gain survives across GCN/GAT/APPNP backbones and $k \in \{3, \dots, 20\}$ (Secs. 5.4–5.5). (3) The benefit has a hard boundary on AGB: under buffer holdout EGSG degrades *more* than the uncorrected graph ($\Delta R^2 = -0.054$, $p = 0.027$ at 30 km), and the boundary aligns with the radius at which LSI reaches zero; the direction of this degradation is target-field dependent—LST shows no collapse, NDVI still benefits (Sec. 5.6). (4) The mechanism is edge quality, not architecture: random pruning with identical sparsity does not reproduce the benefit, the add-back dose response is monotonic, and the diagnostic–benefit correspondence holds across three target variables (Spearman $\rho = -0.741$, $p = 0.006$; Sec. 5.7).

6.2 Relation to Prior Work

The knowledge-quality perspective connects this paper to the spatial data mining literature. Fayyad et al. [18] framed knowledge discovery as deriving actionable patterns from data; the spatial-data-mining lineage established that proximity, autocorrelation,

and scale change what “reliable knowledge” means, spanning the foundational primitives of spatial mining [20], co-location pattern mining with and without thresholds [19, 21], and broad surveys of geographic knowledge discovery [50]. This paper contributes to that lineage a concrete diagnostic: for spatial *graph* representations—the substrate of the current graph-neural-network generation—edge-level knowledge metrics can be measured, reported, and repaired before deployment. In the graph learning literature, our diagnosis instantiates the heterophily critique [8, 10, 11, 24] in a geographic setting with an identifiable physical cause (terrain), and the EGSG gate is a graph-construction-level analog of geometric pre-processing [9]. In ecological modeling, the buffer failure generalizes the spatial-cross-validation warnings [12–15] from classical to graph models and gives them a support-based mechanistic index (LSI). In remote sensing, the results echo the movement toward deep models for Earth observation [26] while cautioning that graph-based AGB mapping inherits the extrapolation fragility of its representation.

Simplicity as a methodological advantage. EGSG is transparent: the gating rule is interpretable, and every edge removal is explainable by a single physical quantity (elevation difference). In knowledge systems, where domain experts need to audit and trust automated decisions, this is a feature rather than a limitation. The simplicity also enables the diagnostic analyses of Secs. 5.2 and 5.7, which isolate one monotone mechanism (edge quality) and would be intractable for a learned graph structure whose edge set has no physical interpretant.

6.3 Implications for Knowledge Systems

Our results speak to how spatial knowledge representations should be engineered, validated, and deployed in geographic information systems. The practical reading is decision support: for an interpolation-like task (gap filling, within-region densification, wall-to-wall mapping inside a sampled domain), a diagnosed, knowledge-driven graph is the better representation; for an extrapolation-like task (unsampled regions, terrain out of coverage, cross-region transfer), the uncorrected Geo kNN graph is not more accurate, but the gated graph’s sharper degradation means the knowledge-support structure—not the model—should drive deployment.

6.4 Implications for Remote Sensing AGB Estimation

For operational AGB mapping, three recommendations follow. First, use the diagnosed and gated graph for interpolation tasks: at $\tau = 250$ m it reaches $R^2 = 0.409$ with 5,313 parameters and 26.8 s training, and it is robust to label noise (label-aware outlier removal raises the benefit to +0.069, $R^2 = 0.483$). Second, calibrate the gate to the target’s spatial structure (MI knee with labels, quantile gate without), never importing a threshold from another variable—the LST counterexample (−0.019 at 250 m vs +0.016 at 1000 m) shows the cost. Third, do not claim extrapolation beyond local support: under buffer holdout all local AGB methods, classical and graph-based (GWR −0.191, kriging −1.169, EGSG −0.191), fail; reporting interpolation numbers as extrapolation numbers would mislead end users.

6.5 Limitations and Future Work

Several limitations bound the conclusions. The study uses one region (southeastern Tibetan Plateau), one year (2021), and a fixed sampling density (minimum spacing 1.19 km); the empirical threshold $\tau = 250$ m is region- and density-specific, and its transferability needs multi-region validation. The GraphSAGE null result ($\Delta R^2 = -0.001$, $p = 0.642$, Sec. 5.5) reveals a boundary condition: topology-correction benefits depend on the message-passing design. GraphSAGE’s neighbor sampling introduces stochasticity that dilutes the deterministic edge-quality improvement, and its learnable aggregator can implicitly reweight edges, absorbing some of the heterophilic edge damage EGSG removes explicitly. The practical implication is that EGSG should be paired with deterministic, full-neighborhood message passing (GCN, GAT, APPNP) rather than stochastic sampling-based propagation; a systematic account of which propagation mechanisms absorb topology corrections is open. In the terrain-aware comparison (Table 11), continuous encoders such as NNConv and GATv2 fold elevation into features or kernels without removing edges, and their smoother extrapolation profile (e.g. GATv2 -0.093 vs EGSG -0.191 under buffer-30) suggests that a hybrid design combining the discrete gate’s in-distribution purity with a continuous residual encoding may transfer more gracefully; this direction, together with theoretical analysis of why discrete support removal is costly under extrapolation, remains open.

A further caveat concerns the AGB label itself. The labels are not field measurements but an ESA CCI derived product, produced by random-forest upscaling of radar and lidar reference data followed by spatial aggregation [27]; its provenance signatures (integer quantization, spatial smoothing) are quantified in Sec. 3. Because the production pipeline includes spatial aggregation, edges between geographically close, elevation-consistent nodes may carry artificially similar labels, and part of the EGSG gain could reflect alignment with the label-smoothing pattern rather than ecological signal alone. The cross-target validation (NDVI, LST), the random-pruning control, and the buffer-based spatial isolation (Secs. 5.6–5.9) mitigate but do not fully eliminate this concern; absolute AGB accuracies should therefore be interpreted against this derived-product origin.

7 Conclusion

This paper shows that spatial graph representations can be unreliable in complex terrain. The default geographic k -nearest-neighbor graph connects nodes whose predictions can differ greatly: 21.41% of its edges span elevation differences above 800 m and 12.24% span AGB differences above 70 Mg/ha. The elevation-gated spatial graph (EGSG) prunes these terrain-violating edges before learning, improving local interpolation by $\Delta R^2 = +0.048$ ($p < 10^{-6}$, 95% CI $[+0.039, +0.056]$) over the uncorrected graph while remaining at 5,313 parameters and 26.8 s of training time.

Methodologically, the paper contributes a reusable diagnostic: edge-level metrics (MAD_y , MAD_h) quantify how reliably a spatial graph encodes knowledge, and the split-level Local Support Index (LSI) bounds where that knowledge transfers. Together they decompose performance into representation quality and support availability, explaining both the interpolation gain and the strict-extrapolation failure on AGB;

the same boundary analysis shows this failure is target-field dependent (no collapse on LST, gating still protects on NDVI). The diagnostic–benefit correspondence generalizes across target variables, and the threshold-selection protocols (MI knee with labels, quantile gate without) make it operational.

The practical implications are concrete: spatial graph representations should be diagnosed before deployment, transfer reliability bounded by support metrics, and knowledge-driven topology construction preferred over proximity heuristics when domain constraints are available. We release the diagnostic–repair–bound cycle as a template for knowledge quality assessment in spatial graph learning.

Supplementary information. No supplementary material accompanies this manuscript.

Declarations

Funding

This work was supported by the Science and Technology Projects of Xizang Autonomous Region, China (grant number XZ202502ZY0073, project title “An Intelligent Question-Answering System for the Ecological Environment of the Qinghai-Tibet Plateau Based on Large Language Model-Powered Knowledge Graph Agents”).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics approval and consent to participate

Not applicable (the study uses publicly available remote sensing products and contains no human or animal participants; consent to participate is therefore not applicable).

Consent for Publication

Not applicable.

Data Availability

All data used in this study are publicly available remote sensing products: ESA CCI Biomass v5 (<https://doi.org/10.5285/02e1b18071ad45a19b4d3e8adafa2817>) [27], MODIS products (MCD15A3H, MOD11A2, MOD16A2GF, MCD12Q1; DOIs in the reference list) [28–30, 32], and ASTER GDEM v3 (<https://doi.org/10.5067/ASTER/ASTGTM.003>) [31]. The derived clean-sample table used in this study is available at <https://github.com/GEN1225/Knowledge-Quality-Diagnosis-of-Spatial-Graph-Representation>.

Materials Availability

Not applicable (no physical materials were used in this study).

Code Availability

The analysis code is available from the corresponding author on reasonable request.

Author Contributions

Conceptualization: Shuai Yan; Methodology: Shuai Yan; Software: Shuai Yan; Formal analysis: Shuai Yan; Investigation: Xiaodong Huang; Resources: Shuai Yan; Data curation: Xiaodong Huang, Chuanyuan Kang; Validation: Xiaodong Huang; Writing — original draft: Shuai Yan, Xiaodong Huang; Writing — review and editing: Shuai Yan, Xiaodong Huang, Chuanyuan Kang; Visualization: Shuai Yan, Xiaodong Huang, Chuanyuan Kang; Supervision: Shuai Yan; Project administration: Shuai Yan; Funding acquisition: Shuai Yan. All authors read and approved the final manuscript.

References

- [1] Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS (2021) A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst* 32(1):4–24. <https://doi.org/10.1109/TNNLS.2020.2978386>
- [2] Ma X, Zhang Z, Song Y (2025) Geographically informed graph neural networks. *Spatial Stat* 69:100920. <https://doi.org/10.1016/j.spasta.2025.100920>
- [3] Liu Y, Shen G, Fu Y, Feng Z, Zhao Z, Kong X (2025) Spatial homogeneity-aware transfer learning for urban flow prediction. *Knowl Inf Syst* 67(5):4349–4371. <https://doi.org/10.1007/s10115-025-02363-z>
- [4] Abémgnigni Njifon M, Schuhmacher D (2024) Graph convolutional networks for spatial interpolation of correlated data. *Spatial Stat* 60:100822. <https://doi.org/10.1016/j.spasta.2024.100822>
- [5] Tobler WR (1970) A computer movie simulating urban growth in the Detroit region. *Econ Geogr* 46(sup1):234–240. <https://doi.org/10.2307/143141>
- [6] McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. *Annu Rev Sociol* 27:415–444. <https://doi.org/10.1146/annurev.soc.27.1.415>
- [7] Fotheringham AS, Brunson C, Charlton M (2002) Geographically weighted regression: the analysis of spatially varying relationships. Wiley, Chichester
- [8] Liu C, Yu C, Gui N, Yu Z, Deng S (2023) SimGCL: graph contrastive learning by finding homophily in heterophily. *Knowl Inf Syst* 66(3):2089–2114. <https://doi.org/10.1007/s10115-023-02022-1>
- [9] Fu X, Li J, Wu J, Qin J, Sun Q, Ji C, Wang S, Peng H, Yu PS (2023) Adaptive curvature exploration geometric graph neural network. *Knowl Inf Syst* 65(5):2281–2304. <https://doi.org/10.1007/s10115-022-01811-4>

- [10] Jin Y, Zhu X (2025) Oversmoothing alleviation in graph neural networks: a survey and unified view. *Knowl Inf Syst* 67(12):11259–11285. <https://doi.org/10.1007/s10115-025-02548-6>
- [11] Huang B, Zheng W, Qian F, Zhao S, Chen J, Zhang Y (2023) GWNN-HF: beyond assortativity in graph wavelet neural network. *Knowl Inf Syst* 65(11):5005–5024. <https://doi.org/10.1007/s10115-023-01900-y>
- [12] Valavi R, Elith J, Lahoz-Monfort JJ, Guillerá-Arroita G (2019) 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>
- [13] Roberts DR, Bahn V, Ciuti S, Boyce MS, Elith J, Guillerá-Arroita G, et al (2017) Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* 40(8):913–929. <https://doi.org/10.1111/ecog.02881>
- [14] Ploton P, Mortier F, Réjou-Méchain M, Barbier N, Picard N, Rossi V, et al (2020) 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>
- [15] Meyer H, Pebesma E (2021) Predicting into unknown space? Estimating the area of applicability of spatial prediction models. *Methods Ecol Evol* 12(9):1620–1633. <https://doi.org/10.1111/2041-210X.13650>
- [16] Fu J, Guo X, Hou J, Yu W, Shi H, Zhao Y (2024) SEGODE: a structure-enhanced graph neural ordinary differential equation network model for temporal link prediction. *Knowl Inf Syst* 67(2):1713–1740. <https://doi.org/10.1007/s10115-024-02261-w>
- [17] Min X, Wu M, Sheng Y (2025) Efficient spatiotemporal index for geostreaming data on road networks based on graph hierarchical partition. *Knowl Inf Syst* 68(1). <https://doi.org/10.1007/s10115-025-02641-w>
- [18] Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. *AI Mag* 17(3):37–54. <https://doi.org/10.1609/aimag.v17i3.1230>
- [19] Wang L, Chang L, Bao X, Zhu C, Gu T (2024) Knowledge-based discovery of multi-level co-location patterns using ontology. *Knowl Inf Syst* 66(10):6463–6491. <https://doi.org/10.1007/s10115-024-02155-x>
- [20] Ester M, Frommelt A, Kriegel HP, Sander J (2000) Spatial data mining: database primitives, algorithms and efficient DBMS support. *Data Min Knowl Discov* 4(2–3):193–216. <https://doi.org/10.1023/A:1009843930701>

- [21] Qian F, He Q, Chiew K, He J (2012) Spatial co-location pattern discovery without thresholds. *Knowl Inf Syst* 33(2):419–445. <https://doi.org/10.1007/s10115-012-0506-9>
- [22] Cliff AD, Ord JK (1973) Spatial autocorrelation. Pion, London
- [23] Lu D (2006) The potential and challenge of remote sensing-based biomass estimation. *Int J Remote Sens* 27(7):1297–1328. <https://doi.org/10.1080/01431160500486732>
- [24] Yang H, Fang J, Wu J, Li D, Wang Y, Zheng Z (2025) Soft label enhanced graph neural network under heterophily. *Knowl-Based Syst* 309:112861. <https://doi.org/10.1016/j.knosys.2024.112861>
- [25] Lu D, Chen Q, Wang G, Liu L, Li G, Moran E (2016) A survey of remote sensing-based aboveground biomass estimation methods in forest ecosystems. *Int J Digit Earth* 9(1):63–105. <https://doi.org/10.1080/17538947.2014.990526>
- [26] Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, et al (2019) Deep learning and process understanding for data-driven Earth system science. *Nature* 566(7743):195–204. <https://doi.org/10.1038/s41586-019-0912-1>
- [27] Santoro M, Cartus O (2024) ESA Biomass Climate Change Initiative (Biomass_cci): global datasets of forest above-ground biomass for the years 2010, 2015, 2016, 2017, 2018, 2019, 2020 and 2021, v5. NERC EDS Centre for Environmental Data Analysis. <https://doi.org/10.5285/02e1b18071ad45a19b4d3e8adafa2817>
- [28] Myneni R, Knyazikhin Y, Park T (2021) MCD15A3H MODIS/Terra+Aqua Leaf Area Index/FPAR 4-Day L4 Global 500 m SIN Grid V061. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MCD15A3H.061>
- [29] Wan Z, Hook S, Hulley G (2021) MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1 km SIN Grid V061. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MOD11A2.061>
- [30] Running S, Mu Q, Zhao M (2021) MOD16A2GF MODIS/Terra Net Evapotranspiration Gap-Filled 8-Day L4 Global 500 m SIN Grid V061. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MOD16A2GF.061>
- [31] NASA/METI/AIST/Japan Spacesystems and US/Japan ASTER Science Team (2019) ASTER Global Digital Elevation Model V003. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/ASTER/ASTGTM.003>
- [32] Friedl M, Sulla-Menashe D (2022) MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500 m SIN Grid V061. NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MCD12Q1.061>

- [33] Shimada M, Itoh T, Motooka T, Watanabe M, Shiraishi T, Thapa R, et al (2014) New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sens Environ* 155:13–31. <https://doi.org/10.1016/j.rse.2014.04.014>
- [34] Kipf TN, Welling M (2017) Semi-supervised classification with graph convolutional networks. In: *International Conference on Learning Representations (ICLR)*
- [35] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y (2018) Graph attention networks. In: *International Conference on Learning Representations (ICLR)*
- [36] Hamilton WL, Ying R, Leskovec J (2017) Inductive representation learning on large graphs. In: *Advances in Neural Information Processing Systems 30 (NeurIPS)*, pp 1024–1034
- [37] Klicpera J, Bojchevski A, Günnemann S (2019) Predict then propagate: graph neural networks meet personalized PageRank. In: *International Conference on Learning Representations (ICLR)*
- [38] Cressie N (1993) *Statistics for spatial data*, revised edition. Wiley, New York
- [39] Potapov P, Li X, Hernandez-Serna A, Tyukavina A, Hansen MC, Komareddy A, et al (2021) Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens Environ* 253:112165. <https://doi.org/10.1016/j.rse.2020.112165>
- [40] Zhang L, Zhang Q, Chen Z, Sun Y, Lu C-T, Zhao L (2025) Implicit graph neural network for deep graph transformation. *Knowl Inf Syst* 67(9):7871–7900. <https://doi.org/10.1007/s10115-025-02468-5>
- [41] He J, Wang M, Zhang Z, Chen L, Dong Y (2025) Graph representation learning method using diffusion sampling for dynamic graph anomaly detection. *Knowl Inf Syst* 67(12):12637–12660. <https://doi.org/10.1007/s10115-025-02606-z>
- [42] Lv Y, Xiong H, Liu J (2025) MSAD-GNN: multi-head spectral attention-driven deep propagation graph neural network. *Knowl Inf Syst* 67(8):6803–6824. <https://doi.org/10.1007/s10115-025-02406-5>
- [43] Wang R, Jin D, Yu Z (2026) Unifying heterophily and oversmoothing in graph neural networks via feature spectral diversity. *Inf Sci* 746:123437. <https://doi.org/10.1016/j.ins.2026.123437>
- [44] Sambandam P (2025) Detection of abnormal transaction using spatial ontology-based deep learning approach. *Knowl Inf Syst* 67(12):11875–11902. <https://doi.org/10.1007/s10115-025-02579-z>

- [45] Kumar A, Dubey M, Sharma S (2026) GAttE: geographic attention model for extraction of users' current locations from social media texts. *Knowl Inf Syst* 68(1). <https://doi.org/10.1007/s10115-026-02755-9>
- [46] Wang RY, Strong DM (1996) Beyond accuracy: what data quality means to data consumers. *J Manag Inf Syst* 12(4):5–33. <https://doi.org/10.1080/07421222.1996.11518099>
- [47] Vitale VN, Martino SD, Peron A, Russo M, Battista E (2023) How to manage massive spatiotemporal dataset from stationary and non-stationary sensors in commercial DBMS? *Knowl Inf Syst* 66(3):2063–2088. <https://doi.org/10.1007/s10115-023-02009-y>
- [48] Xie Y, Wu J, Zhou Y (2024) GTHP: a novel graph transformer Hawkes process for spatiotemporal event prediction. *Knowl Inf Syst* 66(7):4043–4062. <https://doi.org/10.1007/s10115-024-02080-z>
- [49] Ahamad R, Mishra KN (2024) Enhancing knowledge discovery and management through intelligent computing methods: a decisive investigation. *Knowl Inf Syst* 66(7):3719–3771. <https://doi.org/10.1007/s10115-024-02099-2>
- [50] Sowmiya Narayanan KJ, Manimaran A (2023) Recent developments in geographic information systems across different application domains: a review. *Knowl Inf Syst* 66(3):1523–1547. <https://doi.org/10.1007/s10115-023-01969-5>
- [51] Messias GH, Iasulaitis S, Valejo ADB (2025) Graph neural networks for positive and unlabeled learning: a rewiring approach. *Knowl Inf Syst* 67(11):10727–10749. <https://doi.org/10.1007/s10115-025-02547-7>
- [52] Wang Y, Chen J, Li Q, He C, Gao J (2024) Graph neural architecture search with heterogeneous message-passing mechanisms. *Knowl Inf Syst* 66(7):4283–4308. <https://doi.org/10.1007/s10115-024-02090-x>