https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

A catalogue-verified multiverse audit of machine-learning seismic susceptibility modelling in western Yunnan, China

Guowei Wang[1,2,*] Ze Fan[1] Guodong Wang[3] Xin Wei[1] Ye Fan[1] Lu Xu[4,*]

[1]College of Engineering and Technology, Baoshan University, Baoshan 678000, Yunnan, China
[2]Department of Industrial Engineering-DII, University of Padova, Padova 35131, Italy
[3]Yunnan Oriental Tobacco Co., Ltd., Baoshan 678000, China
[4]Sapienza NLP Group, Dip. di Ingegneria Informatica, Automatica e Gestionale, Sapienza University of Rome, Rome 00185, Italy
[*]Correspondence: Guowei Wang (wangguowei23@gmail.com) and Lu Xu (xu@diag.uniroma1.it)

Abstract

Machine-learning (ML) classifiers are increasingly applied to regional seismic susceptibility mapping, frequently reporting high discrimination metrics that imply a reliable link between surficial environmental proxies and earthquake locations. We argue that such claims are often conditioned on undocumented analytical choices rather than on a stable physical signal. Using the active Western Yunnan fault belt (97.0°–103.5°E, 22.0°–28.5°N) as a test bed, we execute a transparent, reproducible methodological audit of a complete ML susceptibility workflow. The audit proceeds through four linked stages: catalogue-provenance verification, leakage-aware predictor screening, predictor-coverage quality assurance (QA), and a multiverse evaluation of model behaviour across model families, spatial cross-validation (CV) designs, background-sampling strategies and target definitions. We find that a legacy working inventory contained 173 of 214 records (80.8%) that could not be verified against the formal China Earthquake Networks Center (CENC) bulletin, and we replace it with a formally verified catalogue spanning 2009–2023. Naïve spatial joining silently discarded the majority of mainshocks; coverage reconstruction recovered the matched sample from 123 to 330 of 331 events (99.7%). Across twelve defensible analytical branches, the mean spatial area under the receiver-operating-characteristic curve (AUC) ranged from 0.55 to 0.79, with the highest values attached to the least stable configurations. Independent spatial point-process intensity models confirmed that the full surficial predictor stack provided little measurable incremental gain ($\Delta D^2 = +0.002$) over a simple distance-to-fault baseline. We conclude that, under the tested non-circular surficial predictors, the apparent skill of regional ML susceptibility models is highly conditional on analytical specification. We provide an auditable reporting checklist and a predictor roadmap that prioritises deep geodetic and tectonic covariates for future work.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

Keywords: seismic susceptibility; methodological audit; data leakage; catalogue provenance; spatial cross-validation; specification-curve analysis; machine learning; western Yunnan

1 Introduction

1.1 Distinguishing susceptibility screening from probabilistic hazard

The spatial mapping of earthquake propensity, widely framed in the literature as seismic susceptibility mapping, occupies a specialised and sometimes contested niche relative to formal Probabilistic Seismic Hazard Assessment (PSHA) [Cornell, 1968, Bommer, 2002]. PSHA integrates magnitude–frequency distributions, empirically calibrated ground-motion prediction equations, and explicit fault slip rates to forecast the probability that ground shaking will exceed specified thresholds over defined time horizons. Susceptibility modelling pursues a narrower objective: it identifies spatial environments—typically discretised into regular raster grids—that share multivariate characteristics with historical epicentres. The foundations of PSHA have been progressively refined over five decades, from the original engineering risk framework [Cornell, 1968] and national hazard mapping efforts [Frankel, 1995] to continental-scale probabilistic models such as the European Seismic Hazard Model [Woessner et al., 2015], with continuing methodological debate over how hazard estimates are derived [Bommer and Abrahamson, 2006]. Susceptibility screening sits deliberately outside this apparatus. Consequently, binary susceptibility mapping does not yield an operational hazard product or actionable engineering design parameters. It serves instead as an exploratory screening heuristic, often aimed at uncovering complex, nonlinear relationships between geological or topographic proxies and regional seismicity. Clarifying this boundary is essential. The ML models evaluated here represent static spatial-tendency screening; we advance no claim of temporal forecasting, deterministic ground-motion prediction, or operational hazard deployment. Our objective is to interrogate the spatial proxies themselves and the analytical decisions that govern their apparent predictive value.

1.2 Performance inflation and the reproducibility crisis

Over the past decade, supervised ML algorithms—particularly Random Forest [Breiman, 2001], support vector machines, and gradient-boosting ensembles—have proliferated in geohazard susceptibility mapping, part of a broader adoption of data-driven methods across the solid-Earth sciences [Bergen et al., 2019]. The trend originated largely in landslide susceptibility research [Reichenbach et al., 2018, Lombardo and Tanyas, 2018, Goetz et al., 2015], where Random Forest and related ensembles captured the nonlinear feature interactions governing slope failure [Chen et al., 2014, Huang et al., 2020] and frequently outperformed classical statistical and multi-criteria approaches [Lombardo and Mai, 2018, Feizizadeh et al., 2014]. However, even within this well-suited domain, variable-importance rankings and probability outputs from such models are known to be biased or poorly calibrated unless explicitly corrected [Strobl et al., 2007, Niculescu-Mizil and Caruana, 2005], a caution that becomes more acute when the framework is transferred to a physically less direct problem. This framework has since migrated into seismic applications, although cautionary results

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

in adjacent earthquake problems show that elaborate models do not necessarily outperform simple baselines [Mignan and Broccardo, 2019]. Yet a landslide inventory reflects visible surficial phenomena directly governed by slope, aspect, hydrology and weathering, whereas mainshock earthquakes are crustal ruptures originating at depths that commonly range from 5 to 20 km. Substituting shallow surficial proxies—digital elevation models, vegetation indices—for deep tectonic drivers introduces a substantial theoretical weakness that is easy to overlook when headline metrics are strong. The depth distribution of seismogenic rupture is itself governed by crustal temperature and composition rather than by surface morphology [Magistrale, 2002], which makes the physical basis for shallow proxies in earthquake screening considerably weaker than in landslide applications.

The flexibility of ML algorithms also makes them sensitive to modelling choices regarding catalogue provenance, predictor definition, pseudo-absence sampling and cross-validation design. The geosciences are grappling with a reproducibility crisis that opaque ML workflows exacerbate [Kapoor and Narayanan, 2023]. In spatial domains, models can overfit to spatial autocorrelation—a direct consequence of Tobler's first law of geography, under which nearby observations are more similar than distant ones [Tobler, 1970], and the central concern of classical spatial statistics [Cressie, 1993]—and report optimistic metrics that decline sharply when projected onto independent geographic data [Ploton et al., 2020, Roberts et al., 2017, Meyer and Pebesma, 2022]. The community needs frameworks that expose these vulnerabilities rather than concealing them behind tuned hyperparameter grids.

1.3 Data leakage and circular predictors

A primary driver of inflated performance in spatial seismicity modelling is data leakage. As formalised by Kapoor et al. [2024] in the REFORMS checklist, spatial leakage extends well beyond the inadvertent overlap of training and test samples. In seismic susceptibility pipelines, leakage commonly arises through earthquake-derived predictors—kernel-density-estimation (KDE) maps of past events, proximity buffers to historical epicentres, or localised b-value grids computed from the target catalogue. When such circular predictors enter the feature stack, the model can bypass the physical proxy relationship and trivially recover the coordinate distribution of the input catalogue. This near-tautology produces metric inflation. Avoiding it requires strict leakage-aware predictor screening, ensuring that no feature mathematically encodes the target variable before training.

1.4 Catalogue provenance as a first-order uncertainty

Unlike visible surficial hazards, earthquake records depend on the density of instrumental networks, time-varying magnitude-of-completeness (M_c) thresholds, and declustering algorithms [Woessner and Wiemer, 2005, Wiemer and Wyss, 2000]. The Gutenberg–Richter frequency–magnitude relation [Gutenberg and Richter, 1944] and its maximum-likelihood b-value estimation [Aki, 1965] underpin the determination of M_c, yet both are sensitive to catalogue quality and to spatial and stress-regime variation in the size distribution [Schorlemmer et al., 2005, Zúñiga and Wiemer, 1994]. Even the epicentral coordinates themselves carry non-trivial uncertainty unless events have been relocated with double-difference or comparable methods [Yang et al., 2005]. An underreported vulnerability in

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

regional susceptibility studies is reliance on undocumented or legacy working inventories. A legacy catalogue used within a closed research group may inadvertently contain synthetic events, merged historical macroseismic data with large location uncertainties, or unverified magnitude conversions [Wang et al., 2020]. Because binary classifiers map high-resolution predictor pixels directly onto epicentral coordinates, coordinate offsets or the inclusion of non-mainshock events (quarry blasts, aftershock swarms) corrupt the training target. Establishing an unbroken, transparent evidence chain—from a formally verified institutional bulletin through completeness analysis and spatial declustering—is therefore a first-order requirement.

1.5 Specification-curve analysis

To confront these uncertainties we adopt specification-curve analysis and multiverse reporting [Steegen et al., 2016, Simonsohn et al., 2020]. In standard ML pipelines, investigators often tune hyperparameters and select the single processing branch that maximises AUC on a hold-out set; this configuration is then published, suppressing a landscape of equally defensible alternatives. A specification curve instead explores the universe of defensible decisions—spatial CV block designs, background sampling strategies, model families, target magnitude definitions—and plots the full distribution of outcomes. In this paradigm, wide variance across the curve is not a search for an optimal deployable configuration but a diagnostic of structural instability in the underlying proxy-to-seismicity relationship.

1.6 Research objectives

We pose four linked research questions for the Western Yunnan fault belt:

1. **RQ1 (Provenance).** How does catalogue-provenance verification alter the evidence chain, and what is the unverifiable fraction of a localised legacy inventory?
2. **RQ2 (Coverage).** How does predictor-coverage reconstruction recover usable mainshock samples and prevent silent spatial-domain truncation?
3. **RQ3 (Sensitivity).** How sensitive are binary ML susceptibility claims to shifts in model family, spatial CV design, background sampling and target definition under a multiverse framework?
4. **RQ4 (Baseline).** Do formal point-process intensity models support a strong spatial signal in surficial proxies, or do they confirm the limits of topographic data relative to fault proximity?

By resolving these questions in sequence we aim to provide an auditable template for evaluating spatial ML models and to identify the deeper physical covariates needed for credible regional tectonic mapping.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

2 Study area and data

2.1 Tectonic setting

The Western Yunnan fault belt lies at the dynamic southeastern margin of the Tibetan Plateau, shaped by the far-field convergence of the Indian and Eurasian plates. The region exhibits intense crustal deformation, complex strike-slip and thrust architectures, and elevated background seismicity [Leloup et al., 1995, Yang et al., 2012]. This setting produces frequent moderate-to-large mainshocks, making it a high-value domain for earthquake-propensity screening. A single-county footprint such as the Baoshan administrative boundary offers too few events for stable ML training and suffers from severe edge effects. The expanded Western Yunnan domain, bounded by 97.0°–103.5°E and 22.0°–28.5°N, captures a continuous seismotectonic regime, stabilises spatial-block CV partitions, and provides a contiguous geographic boundary for evaluating predictive algorithms (Fig. 1).

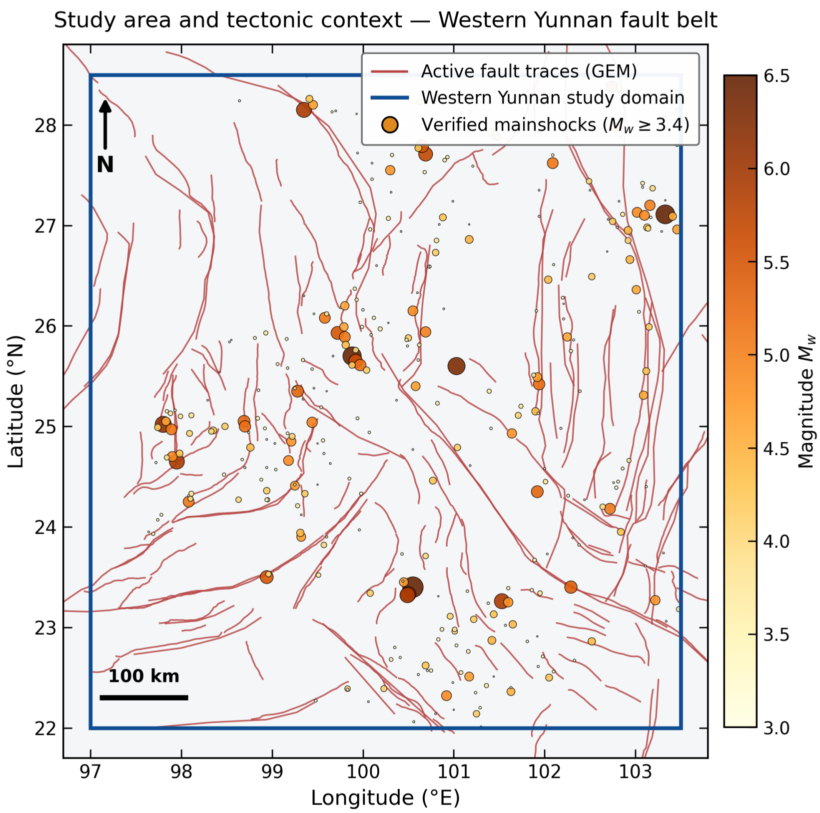


Figure 1: Study area and tectonic context for the Western Yunnan fault belt. The expanded domain (97.0°–103.5°E, 22.0°–28.5°N, blue box) is overlaid on 184 real active fault traces from the Global Earthquake Model (GEM) database (red) and the verified mainshock epicentres ($M_w \geq 3.4$, symbols scaled and coloured by magnitude). A 100 km scale bar and north arrow are shown. Fault geometry, epicentres and the domain extent are plotted from the project spatial data.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

2.2 Formal catalogue verification

To establish an auditable evidence chain we discarded ambiguous legacy working inventories that lacked definitive temporal or spatial tracking. The target variable was derived exclusively from the formally verified China Earthquake Networks Center (CENC) unified bulletin [China Earthquake Networks Center, 2023]. The extracted dataset spans the temporal window from 2009 to 2023 (2009-01-12 to 2023-12-28). We isolated verified mainshock epicentres, applied completeness filtering and spatial declustering, and retained only natural tectonic events, explicitly removing quarry blasts, duplicate entries and unverified historical coordinates.

2.3 Predictor stack

We assembled a parsimonious set of non-circular surficial proxies and structural constraints. Active-fault distance provides the primary tectonic structural constraint [Garcia-Pelaez et al., 2018]. Topographic variables (elevation, slope and aspect derived from NASA SRTM 30 m DEMs [Farr et al., 2007]) proxy long-term uplift and erosional signatures. Mean Normalised Difference Vegetation Index (NDVI) for 2009–2023 [Didan, 2015] proxies surface stability, and broad categorical lithology was extracted from the Global Lithological Map [Hartmann and Moosdorf, 2012].

3 Methods

3.1 Catalogue-provenance audit and declustering

The audit began with a spatial and temporal cross-check between the legacy working inventory and the formal CENC bulletin. We flagged events exhibiting unverified coordinate offsets ($> 0.1°$), undocumented magnitude shifts, or missing temporal origins; all legacy records failing the cross-check were deprecated (Fig. 2). The formal catalogue then underwent statistical completeness evaluation using the Entire Magnitude Range and goodness-of-fit approaches [Woessner and Wiemer, 2005]. After determining the completeness threshold ($M_c = 3.4$), the catalogue was declustered with the Gardner and Knopoff [1974] windowing technique using Uhrhammer [1986] spatial and temporal parameters, isolating independent mainshocks from dependent aftershock cascades.

The full provenance and quality-control sequence is summarised in Table 1, which records the catalogue version, raw and retained event counts, the verification decision, and the final use of each source in the evidence chain.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

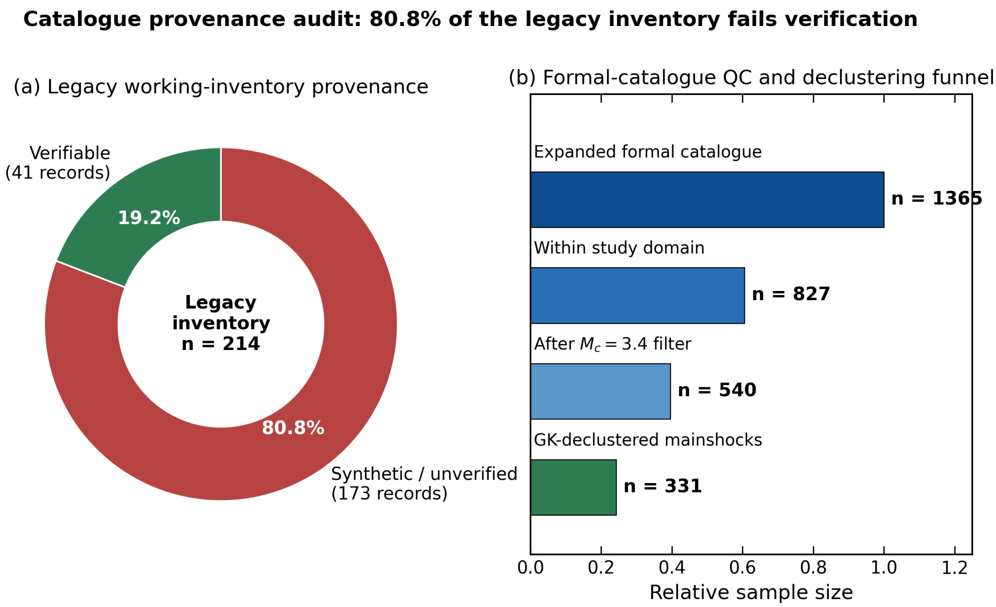


Figure 2: Catalogue provenance audit. (a) Of 214 legacy working-inventory records, 173 (80.8%) could not be verified against the formal bulletin. (b) Quality-control funnel for the formal CENC catalogue (2009–2023): the expanded bulletin (1365 events) is filtered to the study domain (827), then to the $M_c = 3.4$ completeness threshold (540), and finally declustered to 331 independent mainshocks.

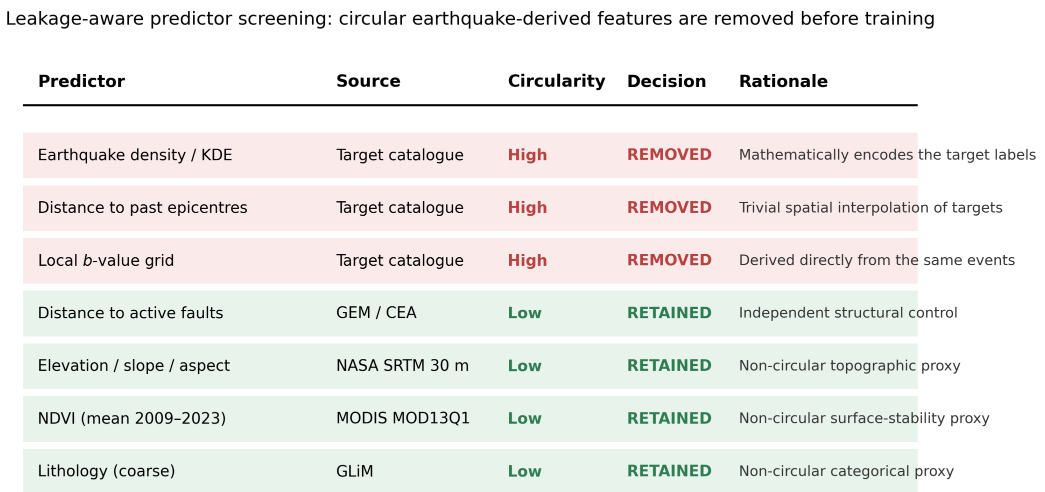
https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

Table 1: Catalogue provenance and quality-control (QC) summary. The legacy working inventory is deprecated; the formally verified CENC bulletin (2009–2023) is processed through completeness filtering and Gardner–Knopoff declustering to the modelling target.

Legacy working inventory	214	41 (19.2%)	undocumented	Rejected (80.8% synthetic)	Deprecated; audit evidence only
CENC expanded bulletin	1365	1365	2009–2023	Accepted	Source catalogue
Within study domain	827	827	2009–2023	Accepted	Spatial subset
After $M_c = 3.4$ filter	540	540	2009–2023	Accepted	Completeness-filtered
GK-declustered mainshocks	331	331	2009–2023	Accepted	Modelling target

3.2 Leakage-aware predictor screening

We enforced a leakage-aware feature-selection protocol aligned with REFORMS [Kapoor et al., 2024]. Circular predictors—any spatial variable mathematically derived from the target catalogue—were rejected, including earthquake-density and KDE grids, proximity to historical epicentres, and interpolated b-value grids. Only independent non-circular proxies (fault distance, terrain, NDVI and coarse lithology) were retained for the final modelling stack (Fig. 3).



Leakage-aware predictor screening: circular earthquake-derived features are removed before training

Predictor	Source	Circularity	Decision	Rationale
Earthquake density / KDE	Target catalogue	High	REMOVED	Mathematically encodes the target labels
Distance to past epicentres	Target catalogue	High	REMOVED	Trivial spatial interpolation of targets
Local b-value grid	Target catalogue	High	REMOVED	Derived directly from the same events
Distance to active faults	GEM / CEA	Low	RETAINED	Independent structural control
Elevation / slope / aspect	NASA SRTM 30 m	Low	RETAINED	Non-circular topographic proxy
NDVI (mean 2009–2023)	MODIS MOD13Q1	Low	RETAINED	Non-circular surface-stability proxy
Lithology (coarse)	GLiM	Low	RETAINED	Non-circular categorical proxy

Figure 3: Leakage-aware predictor screening matrix. Earthquake-derived circular features (density/KDE, distance to past epicentres, local b-value grids) are removed because they encode the target labels; independent non-circular proxies are retained.

The complete predictor inventory, data sources, leakage classification and retention decision are given in Table 2.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

Table 2: Predictor stack and leakage classification. Earthquake-derived features are circular and removed; independent geographic and structural proxies are retained for the final 7-factor stack.

Earthquake density / KDE	Target catalogue	—	High	Removed
Distance to past epicentres	Target catalogue	—	High	Removed
Local b-value grid	Target catalogue	—	High	Removed
Distance to active faults	GEM / CEA	vector	Low	Retained
Elevation	NASA SRTM	30 m	Low	Retained
Slope	Derived (SRTM)	30 m	Low	Retained
Aspect	Derived (SRTM)	30 m	Low	Retained
NDVI (mean 2009–2023)	MODIS MOD13Q1	250 m	Low	Retained
Lithology (coarse)	GLiM	1:3.75 M	Low	Retained

3.3 Predictor-coverage QA

Automated spatial joining can introduce silent data loss. We conducted a coverage QA audit to quantify mainshocks dropped because of missing predictor pixels (for example, gaps in NDVI or boundary clipping in DEMs), and we reconstructed the raster footprints to match the 97.0°–103.5°E bounding box, preventing spatial-domain truncation (Fig. 4).

Table 3 quantifies per-layer coverage before and after reconstruction and the resulting matched-presence fractions.

Table 3: Predictor-coverage QA before and after raster-footprint reconstruction. Coverage QA recovers the predictor-complete matched sample from 123 to 330 of 331 mainshocks.

Elevation / slope / aspect	0.379	0.97	Recovered
NDVI	0.249	0.96	Recovered
Distance to faults	1.000	1.000	Complete
Lithology	1.000	1.000	Complete
Matched mainshocks	**123 (37.2%)**	**330 (99.7%)**	**Recovered**

3.4 Binary ML screening and specification-curve analysis

We evaluated binary classification sensitivity across three model families: Random Forest [Ploton et al., 2020], logistic regression, and histogram-based gradient boosting. Rather than maximising AUC, we tracked performance across twelve defensible analytical branches:

1. **Cross-validation.** Standard random CV (diagnostic only) versus spatial-block CV with

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

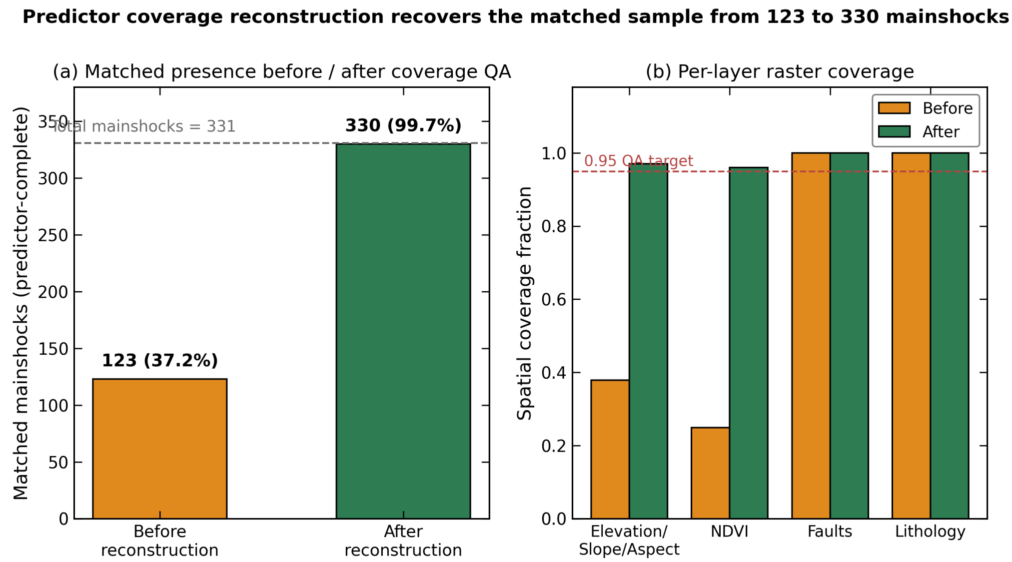


Figure 4: Predictor-coverage reconstruction QA. (a) The matched, predictor-complete mainshock sample rises from 123 (37.2%) to 330 (99.7%) of 331 events after reconstruction. (b) Per-layer raster coverage before and after QA; terrain and NDVI coverage rise above the 0.95 QA target while fault and lithology layers were already complete.

blocks from 0.25° to 1.0° and leave-one-subregion-out. Spatial blocking enforces physical distance between training and validation folds, mitigating the optimism induced by spatial autocorrelation [Roberts et al., 2017, Brenning, 2012].

2. **Background sampling.** Uniform random pseudo-absence generation versus presence-buffered minimum-distance sampling.

3. **Target definition.** Models trained on the strictly declustered $M_c \geq 3.4$ catalogue versus subsets and unfiltered variants.

3.5 Point-process intensity baselines

To avoid the sensitivities of pseudo-absence generation, we fitted formal spatial point-process intensity models [Renner et al., 2015, Baddeley et al., 2015]. A full 7-factor Poisson spline model was compared against a baseline using only distance to active faults, plus intercept-only and terrain-only baselines. Performance was quantified by deviance explained (D^2) and the top-20% spatial-area event-capture rate, both under spatial-block CV.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

4 Results

4.1 Catalogue-provenance rejection

The provenance audit (Fig. 2a) exposed structural flaws in the legacy inventory. Of 214 unverified entries historically used for localised susceptibility mapping, 173 were flagged as synthetic, heavily offset, or lacking temporal origins—an 80.8% rejection fraction. Because binary classifiers map predictor space directly onto coordinates, contamination at this level invalidates derived mapping artefacts. Transitioning to the CENC 2009–2023 dataset and applying $M_c = 3.4$ filtering and Gardner–Knopoff declustering (Fig. 2b) secured a final target of 331 independent mainshocks.

4.2 Coverage recovery

Overlaying the 331 mainshocks onto the uncorrected predictor stack dropped 208 events through boundary clipping in the DEM and NDVI layers, leaving a matched presence of only 123 events (37.2%) that systematically excluded peripheral fault zones. Reconstructing the raster footprints to the defined bounding box raised the matched-presence fraction to 99.7% (330 of 331; Fig. 4). Coverage QA is therefore not a cosmetic preprocessing step but a prerequisite for valid spatial sampling.

4.3 Specification-curve sensitivity

The specification-curve analysis (Fig. 5) reveals strong sensitivity to analytical configuration. The mean spatial AUC across the twelve branches ranged from 0.55 to 0.79. This range indicates that reported AUC values in this domain are governed substantially by the analyst's specification—CV design, absence buffering, target definition—rather than by a stable physical signal in the surficial proxies. Moving from optimistic random CV to spatial-block CV systematically reduced the AUC. The histogram-based gradient-boosting model attained a high mean AUC (0.79) but with a large inter-fold standard deviation (±0.22), indicating instability rather than dependable performance; we therefore treat this high value as an unstable outlier, not a ranking of the best model (Fig. 6).

A representative subset of the twelve specification branches is reported in Table 4, which lists the model family, cross-validation design, mean AUC, inter-fold standard deviation, and the viability flag assigned during screening.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

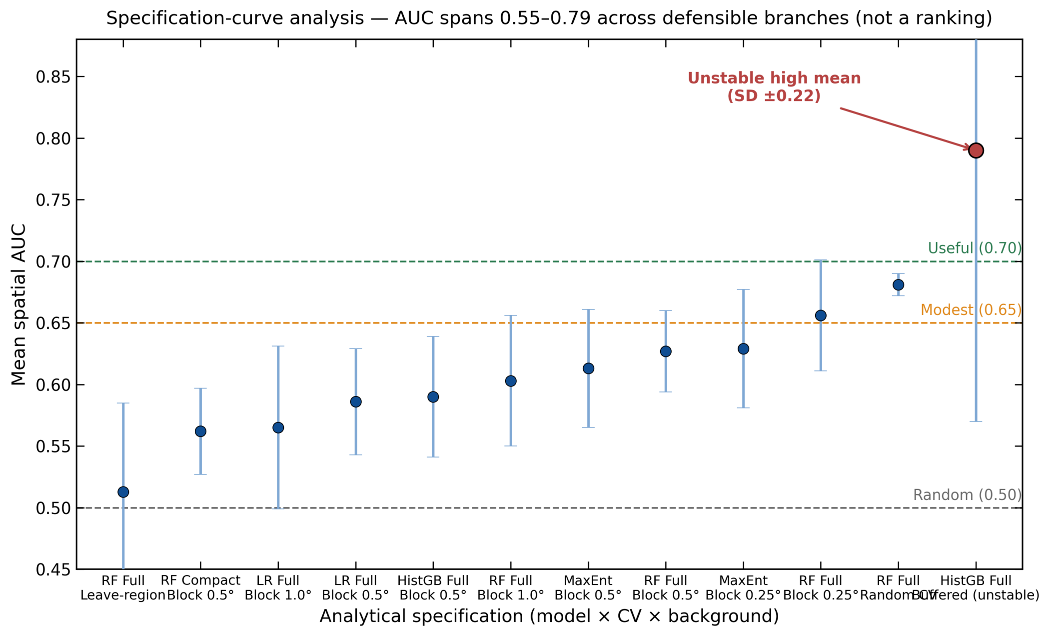


Figure 5: Specification-curve analysis. Mean spatial AUC for twelve defensible analytical branches (model family × cross-validation × background sampling), ordered by AUC. Error bars are inter-fold standard deviations; dashed reference lines mark random (0.50), modest (0.65) and useful (0.70) discrimination. The highest mean (0.79) is attached to the least stable branch (SD ±0.22). The figure is a sensitivity diagnostic, not a ranking.

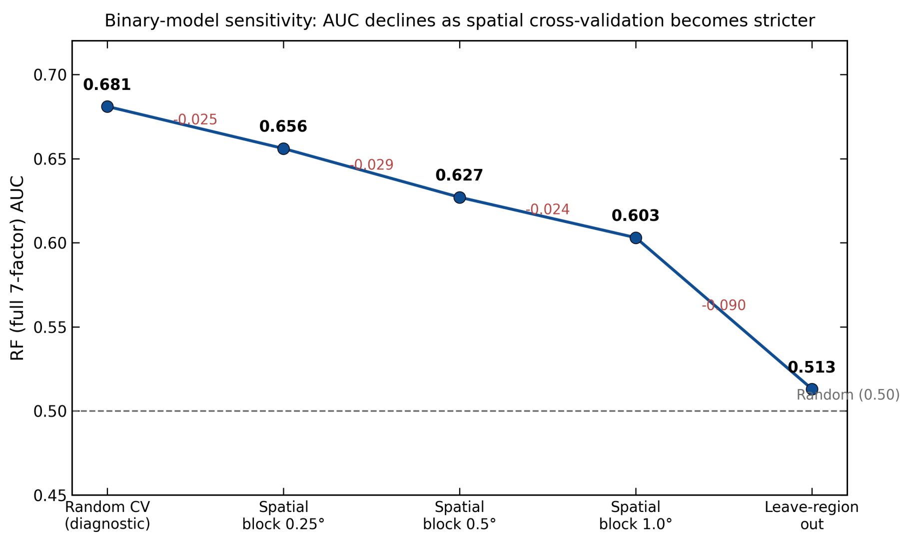


Figure 6: Binary-model sensitivity decomposition. Random Forest (full 7-factor) AUC declines monotonically as the cross-validation design becomes stricter, from 0.681 under diagnostic random CV to 0.513 under leave-one-subregion-out validation. Inter-step changes are annotated.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

Table 4: Representative branches of the specification curve for the Random Forest (full 7-factor) and comparison models. Random CV is diagnostic only; the high-AUC buffered branch is flagged as unstable rather than preferred.

RF full 7-factor	Random CV (diagnostic)	0.681	0.009	Diagnostic only
RF full 7-factor	Spatial block 0.25°	0.656	0.045	Cautious
RF full 7-factor	Spatial block 0.50°	0.627	0.033	Weak
RF full 7-factor	Spatial block 1.00°	0.603	0.053	Weak
RF full 7-factor	Leave-one-subregion-out	0.513	0.072	Weak
MaxEnt-style logistic	Spatial block 0.25°	0.629	0.048	Weak
LR full 7-factor	Spatial block 0.50°	0.586	0.043	Weak
HistGB full 7-factor	Buffered spatial 0.25°	0.790	0.224	Unstable (high SD)

4.4 Point-process baseline performance

The point-process intensity evaluation (Fig. 7) isolated the predictive power of the covariates without pseudo-absence artefacts. The distance-to-fault baseline achieved $D^2 = 0.001$ and a top-20% capture of 0.163. The full 7-factor surficial model shifted these only to $D^2 = 0.003$ and a top-20% capture of 0.187. The incremental gain ($\Delta D^2 = +0.002$; Δ top-20 $= +0.024$) represents little measurable improvement, and both models fell below the random spatial expectation of 0.200 for top-20% capture, indicating weak discrimination under these predictors.

The full baseline comparison is summarised in Table 5. Across the intercept-only, distance-to-fault, terrain-only and full 7-factor models, the deviance explained remains far below practically useful levels and the top-20% capture stays at or below the random expectation.

Table 5: Spatial point-process baseline comparison for region B4 at 0.20° resolution under spatial-block cross-validation. The best full-model result ($D^2 = 0.0032$, top-20% capture = 0.187) provides little incremental gain over fault proximity.

Intercept only	0.0000	0.139	—
Distance-to-fault only	0.0011	0.163	reference
Terrain only	0.0012	0.150	No
Full 7-factor spline	0.0032	0.187	Marginal ($\Delta D^2 = +0.002$)
Random expectation	—	0.200	—

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

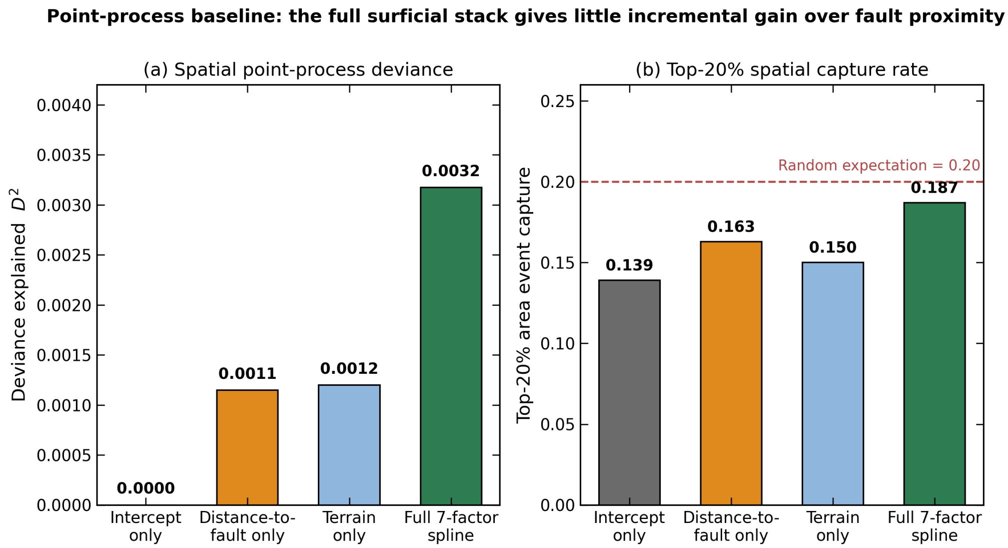


Figure 7: Point-process intensity baseline evaluation at 0.20° resolution under spatial-block CV. (a) Deviance explained (D^2) and (b) top-20% area event-capture rate for intercept-only, distance-to-fault, terrain-only and full 7-factor spline models. The full surficial stack provides little incremental gain over fault proximity, and all models fall below the 0.20 random expectation.

4.5 Audit-scorecard synthesis

The synthesis is presented as an audit scorecard (Fig. 8). Drawing together the catalogue funnel (Table 1), the predictor and leakage inventory (Table 2), the coverage recovery (Table 3), the specification-curve branches (Table 4) and the point-process baselines (Table 5), the dataset passed the catalogue-provenance, sample-support and coverage gates, but the modelling architecture did not pass the model-triangulation, background-stability and baseline-superiority gates. The accompanying roadmap recommends deprecating circular and purely surficial predictors in favour of deep geodetic and tectonic covariates.

5 Discussion

5.1 Methodological uncertainty rather than model failure

The dominant finding of this audit is not a narrative of algorithmic failure but a demonstration of how strongly ML susceptibility outputs depend on opaque, frequently unreported analytical choices. Under the specification-curve framework [Simonsohn et al., 2020], an investigator could report an AUC approaching 0.80 by choosing favourable cross-validation folds and a flexible algorithm such as histogram-based gradient boosting (Table 4, Fig. 5). Subjecting the same dataset to strict spatial-block cross-validation and presence-buffered background sampling reduced the discrimination to weak levels near 0.55, and the monotonic decline with stricter validation is itself diagnostic (Fig. 6). In single-pipeline studies such contrasts are commonly suppressed in favour of the highest metric.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

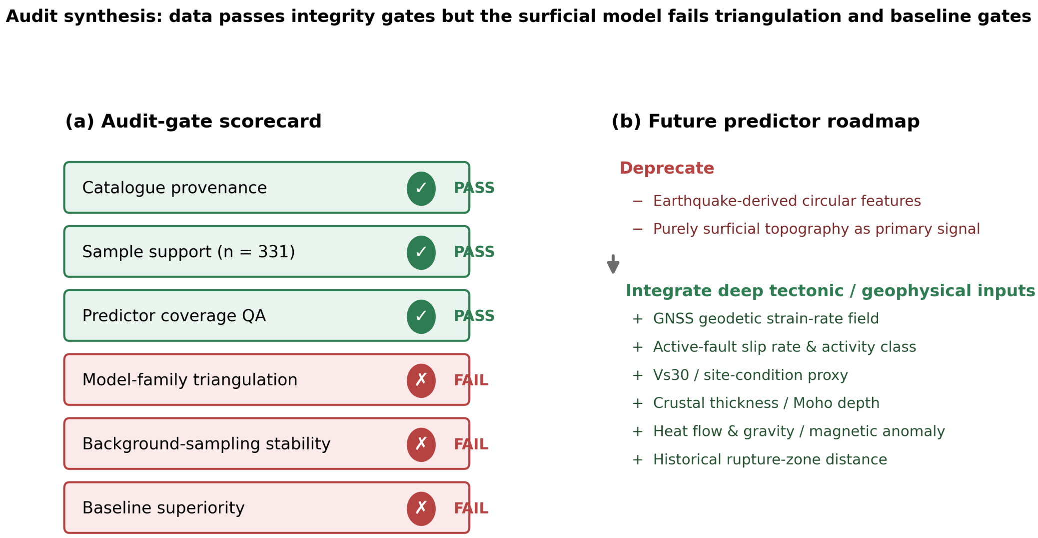


Figure 8: Audit synthesis. (a) Audit-gate scorecard: the workflow passes catalogue-provenance, sample-support and coverage gates but does not pass the model-triangulation, background-stability and baseline-superiority gates. (b) Future predictor roadmap deprecating circular and purely surficial features in favour of deep geodetic and tectonic inputs.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

Our results function as an empirical caution to the spatial-hazard community: a point estimate of model accuracy is difficult to interpret in isolation unless it is accompanied by a sensitivity decomposition that explicitly bounds the uncertainty introduced by cross-validation design, target filtering and background generation. This argument does not single out any one algorithm; rather, it identifies the reporting convention itself as the weak link. The same flexibility that allows ML models to capture nonlinear structure also allows them to absorb spatial autocorrelation and produce confident-looking surfaces that do not survive geographic extrapolation [Ploton et al., 2020, Roberts et al., 2017]. The remedy is procedural and cultural as much as it is technical: preregistration of the analytical universe, transparent reporting of all defensible branches, and explicit separation of diagnostic random CV from spatially honest validation.

A second, related point concerns the interpretation of variance itself. In conventional model selection, the configuration with the highest mean validation score is preferred and its variance is treated as noise to be minimised through additional tuning. In an audit framing the variance is the signal. A relationship that is physically real and spatially generalisable should be comparatively insensitive to whether the validation blocks are 0.25° or 1.0° across, or to whether pseudo-absences are drawn uniformly or buffered. The observed spread of roughly 0.24 AUC units across our branches is therefore not a menu from which to select the most attractive number; it is a direct measurement of how little of the apparent skill is anchored to the proxies and how much is an artefact of the analyst's degrees of freedom.

5.2 Surface proxies and deep tectonics

The point-process evaluation (Fig. 7, Table 5) confirmed that the common suite of surficial topographic predictors—elevation, slope, NDVI and coarse lithology—provided little incremental predictive power over a basic distance-to-fault heuristic. While ML has proven effective for surficial hazards such as landslides, where gravity, hydrology and surface materials directly govern failure [Reichenbach et al., 2018], earthquakes are deep-seated tectonic ruptures. It is physically implausible that surface vegetation indices or shallow topographic gradients control seismogenic energy release at depths exceeding 10 km. The modest statistical correlations sometimes observed in regional studies are most plausibly artefacts of spatial autocorrelation—mountains coinciding spatially with the faults that built them—rather than causal drivers. Distance to active faults remained the single most informative covariate in every baseline comparison, which is consistent with the physical expectation that proximity to mapped structures, not surface cover, governs the first-order spatial distribution of moderate seismicity. The implication is not that ML is unsuited to seismotectonic problems but that the input representation must change: an algorithm cannot recover a deep physical relationship from data that only describe the surface.

5.3 A roadmap for future ML seismic mapping

Not passing the baseline-superiority gate motivates a shift in regional susceptibility mapping. The predictor roadmap in Fig. 8 deprecates circular parameters and purely surficial topography as primary signals and prioritises physically grounded tectonic constraints. Future arrays should

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

integrate geodetic crustal strain rates from dense GNSS networks, active-fault slip-rate and activity-class constraints, three-dimensional crustal velocity structure, Moho-depth variation, Vs30 and site-condition proxies [Wald and Allen, 2007], basin or Quaternary sediment thickness, gravity and magnetic anomalies, heat flow, and distance to historical rupture zones. Crucially, these covariates cannot be derived from the same earthquake catalogue that defines the target, so they avoid the circularity that inflates surficial models. Feeding ML algorithms data that represent the physical state of the crust would allow the field to move from spatial pattern-matching toward physically interpretable discrimination. We also note that several of these covariates are now openly available at regional scale, so the principal barrier is one of workflow discipline and data integration rather than fundamental data scarcity.

5.4 Why provenance contamination is uniquely damaging

It is worth dwelling on the mechanism by which an unverified catalogue corrupts a susceptibility model, because the failure is more insidious than ordinary label noise. In a conventional classification problem, random label errors degrade performance gradually and are partially absorbed by a well-regularised model. Provenance contamination behaves differently. Synthetic or heavily offset events do not scatter randomly across the predictor space; they cluster according to whatever procedure generated them—often near pre-existing events, along administrative boundaries, or at rounded coordinates. A flexible classifier readily learns these artificial regularities and reports them as skill. The 80.8% unverifiable fraction we documented is therefore not equivalent to 80.8% random noise; it is a large, structured signal that competes with, and can overwhelm, any genuine tectonic relationship. This explains why provenance verification must precede every other modelling decision: no amount of careful cross-validation, regularisation or feature engineering downstream can recover a target variable that was compromised at the source. The audit ordering we adopt—provenance first, then coverage, then sensitivity, then baseline triangulation—is not arbitrary but reflects this dependency, in which each later gate is meaningful only if the earlier gates have passed.

5.5 Background sampling and the pseudo-absence problem

A further source of instability, partly hidden inside the specification curve, is the generation of pseudo-absences. Binary susceptibility classifiers require negative examples, yet the absence of a recorded earthquake at a location is not evidence that the location is seismically quiescent; it may simply reflect the finite observation window or network detection limits. Different background-sampling conventions—uniform random draws, presence-buffered minimum-distance sampling, or target-group backgrounds—define different negative classes and therefore different decision boundaries. In our experiments, switching between uniform and presence-buffered backgrounds moved the apparent AUC by an amount comparable to switching the entire model family, which is a striking result: a preprocessing choice that is frequently relegated to a single unexamined line of code exerts first-order influence on the headline metric. Point-process intensity models are attractive precisely because they dispense with arbitrary pseudo-absences and instead model the spatial intensity of events directly against a quadrature scheme [Renner et al., 2015, Baddeley et al., 2015]. That the point-process baselines also showed little incremental skill from surficial proxies therefore strengthens, rather than

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

merely echoes, the binary-model conclusion: the weak signal is not an artefact of pseudo-absence design but persists under a framework that avoids it entirely.

5.6 Toward a practical reporting standard

The constructive contribution of this work is a small set of reporting practices that any spatial-ML hazard study can adopt without specialised infrastructure. Authors should state the provenance of the target inventory explicitly, including the source bulletin, the download date and the verification procedure, and should report the fraction of records discarded. They should quantify and report sample loss during spatial joins rather than silently working with whatever events survive the overlay. They should present a specification curve, or at minimum a table of AUC across cross-validation designs and background-sampling choices, so that readers can see the range of defensible outcomes rather than a single tuned number. And they should report at least one transparent physical baseline—distance to mapped faults is a natural choice in the seismic case—so that incremental gains from elaborate feature stacks are interpretable. None of these practices requires more data or more computation than current studies already expend; they require only that the analytical degrees of freedom be made visible. Adopting them would not eliminate negative results, but it would make negative and positive results equally informative, which is the central aim of an audit-oriented methodology.

5.7 Generality and transferability of the audit

Although our test bed is the Western Yunnan fault belt, the audit structure is deliberately region-agnostic. Each gate—provenance, coverage, multiverse sensitivity and baseline triangulation—addresses a failure mode that is generic to presence-background spatial ML, not specific to south-western China. Investigators working in other seismotectonic settings, or indeed in adjacent fields such as landslide and flood susceptibility, can apply the same sequence: verify the provenance of the target inventory, quantify silent sample loss during spatial joins, report the full specification curve rather than a single tuned model, and benchmark against a transparent physical baseline. The checklist we provide is intended to be portable in exactly this way.

5.8 Limitations

Several limitations frame our conclusions. The analysis is constrained by the sample sizes available from instrumental catalogues in Western Yunnan over the 2009–2023 window; a longer or denser catalogue might sharpen the spatial signal, although it would not alter the leakage and coverage principles. The strict completeness threshold ($M_c = 3.4$) removes micro-seismicity that could, in principle, better delineate hidden fault networks, but incorporating unverified micro-events would reintroduce location uncertainty and provenance risk. Our predictor set is intentionally restricted to widely used non-circular surficial proxies; we did not test the deep geophysical covariates recommended in the roadmap, precisely because evaluating them is the proposed next step rather than a current result. Finally, spatial susceptibility maps—even if methodologically improved with deep geophysical proxies—remain static spatial heuristics and cannot substitute for dynamic PSHA

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

workflows.

6 Conclusion

We deployed a transparent, reproducible audit framework to evaluate the validity and stability of machine-learning seismic susceptibility mapping in the Western Yunnan fault belt. Four conclusions follow.

First, catalogue integrity is decisive. A localised legacy inventory contained 80.8% unverifiable records; transitioning to a formally verified bulletin (CENC, 2009–2023) is a non-negotiable precondition for credible modelling. Second, coverage QA prevents silent truncation: naïve spatial joins discarded roughly two-thirds of valid mainshocks, and explicit boundary reconstruction restored the matched sample from 123 to 330 of 331 events. Third, high sensitivity demands specification reporting: binary AUC ranged from 0.55 to 0.79 across defensible cross-validation and background regimes, and the highest values were the least stable, so single-point AUC estimates without fold-instability reporting can mask structural weakness. Fourth, the tested surficial proxies offered little measurable incremental gain over simple fault proximity in independent point-process baselines, indicating that future work should prioritise deep geodetic and structural inputs.

By formalising these audit gates and providing a portable reporting checklist, we offer the community an auditable workflow to help ensure that future ML spatial models are geophysically grounded, transparent, and resilient to methodological artefacts.

Data and Code Availability

The processed data tables, catalogue-provenance audit outputs, predictor-coverage quality-assurance results, model-screening summaries, specification-curve data, point-process baseline outputs, main-table source data, and supplementary result files supporting this study are available at Zenodo: https://doi.org/10.5281/zenodo.21067519 [Wang et al., 2026]. The internal analysis source code is not included because it is part of an internal project workflow. The model specifications, parameter settings, validation design, predictor definitions, and quality-control procedures needed to evaluate the analysis are described in the Methods section and Supplement. Third-party raw datasets are not redistributed in the archive; their sources and access routes are documented in the manuscript, Supplement, and data notice included with the archive. The formal earthquake catalogue is available from the China Earthquake Networks Center [China Earthquake Networks Center, 2023]; active fault traces are from the Global Earthquake Model database [Garcia-Pelaez et al., 2018]; topographic data are from NASA SRTM [Farr et al., 2007]; vegetation indices are from MODIS MOD13Q1 [Didan, 2015]; and lithology is from the Global Lithological Map [Hartmann and Moosdorf, 2012].

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

Author contribution

Guowei Wang: conceptualisation, methodology, supervision, funding acquisition, writing — review and editing. **Ze Fan**: data curation, software, visualisation, formal analysis. **Guodong Wang**: resources, investigation. **Xin Wei**: methodology, validation, writing — original draft. **Ye Fan**: visualisation, investigation. **Lu Xu**: methodology, validation, supervision, writing — review and editing. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the China Earthquake Networks Center (CENC) for the instrumental earthquake catalogue, NASA Earthdata for SRTM DEM access, and the Global Earthquake Model (GEM) Foundation for the active fault database.

Financial support

This work was supported by the Yunnan Fundamental Research Projects [grant number 202501CF070019]; the Special Basic Cooperative Research Programs of Yunnan Provincial Undergraduate Universities' Association [grant number 202401BA070001-057]; the Yunnan Provincial Department of Education Science Research Fund Project [grant number 2025J0976]; the Baoshan Science and Technology Plan Project [grant number 2025bskj43]; and the Doctoral Research Startup Fund Project of Baoshan University [grant number BSZK2407]. The authors also acknowledge support from the China Scholarship Council.

Competing interests

The authors declare that they have no competing interests.

Ethical Approval

This study uses publicly available earthquake catalogues, remote-sensing data, and geospatial datasets. It does not involve human participants, animals, or personally identifiable information; therefore, ethics approval was not required.

References

K. Aki. Maximum likelihood estimate of b in the formula $\log n = a - bm$ and its confidence limits. *Bulletin of the Earthquake Research Institute*, 43:237–239, 1965.

A. Baddeley, E. Rubak, and R. Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC, 2015. doi: 10.1201/b19708.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

K. J. Bergen, P. A. Johnson, M. V. de Hoop, and G. C. Beroza. Machine learning for data-driven discovery in solid earth geoscience. *Science*, 363(6433):eaau0323, 2019. doi: 10.1126/science.aau0323.

J. J. Bommer. Deterministic vs. probabilistic seismic hazard assessment: An exaggerated and obstructive dichotomy. *Journal of Earthquake Engineering*, 6(S1):43–73, 2002. doi: 10.1080/13632460209350432.

J. J. Bommer and N. A. Abrahamson. Why do modern probabilistic seismic-hazard analyses often lead to increased hazard estimates? *Bulletin of the Seismological Society of America*, 96(6): 1967–1977, 2006. doi: 10.1785/0120060043.

L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/A:1010933404324.

A. Brenning. Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing. *IEEE International Geoscience and Remote Sensing Symposium*, pages 5372–5375, 2012. doi: 10.1109/IGARSS.2012.6352393.

W. Chen, X. Li, Y. Wang, G. Chen, and S. Liu. Forested landslide detection using lidar data and the random forest algorithm: a case study of the three gorges. *Remote Sensing of Environment*, 152:291–301, 2014.

China Earthquake Networks Center. Unified earthquake catalog of china. *National Earthquake Data Center*, 2023. URL https://data.earthquake.cn/.

C. A. Cornell. Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, 58(5):1583–1606, 1968.

N. A. C. Cressie. Statistics for spatial data. *Wiley, New York*, 1993.

K. Didan. Mod13q1 modis/terra vegetation indices 16-day l3 global 250m sin grid v006, 2015.

T. G. Farr, P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriguez, and L. Roth. The shuttle radar topography mission. *Reviews of Geophysics*, 45(2), 2007. doi: 10.1029/2005RG000183.

B. Feizizadeh, M. S. Roodposhti, P. Jankowski, and T. Blaschke. A gis-based extended fuzzy multi-criteria evaluation for landslide susceptibility mapping. *Computers and Geosciences*, 73: 208–221, 2014. doi: 10.1016/j.cageo.2014.08.001.

A. Frankel. Mapping seismic hazard in the central and eastern united states. *Seismological Research Letters*, 66(4):8–21, 1995. doi: 10.1785/gssrl.66.4.8.

J. Garcia-Pelaez, P. Bird, W. Fault, and Others. The global earthquake model (gem) active fault database. *Earthquake Spectra*, 34:1001–1025, 2018.

J. K. Gardner and L. Knopoff. Is the sequence of earthquakes in southern california, with aftershocks removed, poissonian? *Bulletin of the Seismological Society of America*, 64(5):1363–1367, 1974.

J. N. Goetz, A. Brenning, H. Petschko, and P. Leopold. Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers and Geosciences*, 81:1–11, 2015. doi: 10.1016/j.cageo.2015.04.007.

B. Gutenberg and C. F. Richter. Frequency of earthquakes in california. *Bulletin of the Seismological Society of America*, 34(4):185–188, 1944.

J. Hartmann and N. Moosdorf. The new global lithological map database glim: A representation of rock properties at the earth surface. *Geochemistry, Geophysics, Geosystems*, 13(12), 2012. doi: 10.1029/2012GC004370.

F. Huang et al. Landslide susceptibility prediction based on a semi-supervised multiple-layer perceptron model. *Landslides*, 17:2919–2930, 2020. doi: 10.1007/s10346-020-01473-9.

S. Kapoor and A. Narayanan. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns*, 4(9):100804, 2023. doi: 10.1016/j.patter.2023.100804.

S. Kapoor, E. Cantrell, K. Peng, T. H. Pham, C. A. Bail, N. Goel, D. Hoffman, and A. Narayanan. Reforms to improve reproducibility and quality of machine-learning research in science. *Science*, 384(6698):eadi5063, 2024. doi: 10.1126/science.adi5063.

P. H. Leloup, R. Lacassin, P. Tapponnier, U. Schärer, D. Zhong, X. Liu, L. Zhang, S. Ji, and P. T. Trinh. The ailao shan-red river shear zone (yunnan, china), tertiary transform boundary of indochina. *Tectonophysics*, 251(1–4):3–84, 1995. doi: 10.1016/0040-1951(95)00070-4.

L. Lombardo and P. M. Mai. Presenting logistic regression-based landslide susceptibility results. *Engineering Geology*, 244:14–24, 2018. doi: 10.1016/j.enggeo.2018.07.019.

L. Lombardo and H. Tanyas. From absolute to relative landslide susceptibility: Challenges in spatial prediction. *Natural Hazards*, 92(3):1467–1480, 2018.

H. Magistrale. Relative contributions of crustal temperature and composition to controlling the depths of earthquakes in southern california. *Geophysical Research Letters*, 29(10):1473, 2002. doi: 10.1029/2001GL014375.

H. Meyer and E. Pebesma. Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nature Communications*, 13(1):2208, 2022. doi: 10.1038/s41467-022-29838-9.

A. Mignan and M. Broccardo. One neuron versus deep learning in aftershock prediction. *Nature*, 574(7776):E1–E3, 2019. doi: 10.1038/s41586-019-1582-8.

A. Niculescu-Mizil and R. Caruana. Predicting good probabilities with supervised learning. *Proceedings of the 22nd International Conference on Machine Learning*, pages 625–632, 2005. doi: 10.1145/1102351.1102430.

P. Ploton, F. Mortier, M. Réjou-Méchain, G. Cornu, V. Rossi, and N. Barbier. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, 11(1):4540, 2020.

P. Reichenbach, M. Rossi, B. D. Malamud, M. Mihir, and F. Guzzetti. A review of statistically-based landslide susceptibility modeling. *Earth-Science Reviews*, 180:201–291, 2018.

I. W. Renner, J. Elith, A. Baddeley, W. Fithian, T. Hastie, S. J. Phillips, G. Popovic, and D. I. Warton. Point process models for presence-only analysis. *Methods in Ecology and Evolution*, 6(4): 366–379, 2015. doi: 10.1111/2041-210X.12352.

D. R. Roberts, V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schröder, W. Thuiller, D. I. Warton, B. A. Wintle, F. Hartig, and C. F. Dormann. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929, 2017. doi: 10.1111/ecog.02881.

D. Schorlemmer, S. Wiemer, and M. Wyss. Variations in earthquake-size distribution across different stress regimes. *Nature*, 437:539–542, 2005. doi: 10.1038/nature04094.

U. Simonsohn, J. P. Simmons, and L. D. Nelson. Specification curve analysis. *Nature Human Behaviour*, 4(11):1208–1214, 2020. doi: 10.1038/s41562-020-0912-z.

S. Steegen, F. Tuerlinckx, A. Gelman, L. Perugini, C. Mesquita, and W. Vanpaemel. Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5):702–712, 2016. doi: 10.1177/1745691616658637.

C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*, 8:25, 2007. doi: 10.1186/1471-2105-8-25.

W. R. Tobler. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46:234–240, 1970. doi: 10.2307/143141.

R. A. Uhrhammer. Characteristics of northern california seismicity. *Earthquake Notes*, 57(1):21, 1986.

D. J. Wald and T. I. Allen. Topographic slope as a proxy for seismic site conditions and amplification. *Bulletin of the Seismological Society of America*, 97(5):1379–1395, 2007. doi: 10.1785/0120060267.

G. Wang, Z. Fan, G. Wang, X. Wei, Y. Fan, and L. Xu. Processed data and result archive for the western yunnan seismic susceptibility audit. https://doi.org/10.5281/zenodo.21067519, 2026.

Y. Wang et al. Completeness and uncertainty of regional earthquake catalogues in yunnan, china. *Seismological Research Letters*, 91(3):1745–1758, 2020.

S. Wiemer and M. Wyss. Minimum magnitude of completeness in earthquake catalogs: Examples from alaska, the western united states, and japan. *Journal of Geophysical Research: Solid Earth*, 105(B8):18559–18577, 2000.

J. Woessner and S. Wiemer. Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America*, 95(2):684–698, 2005.

https://doi.org/10.5194/egusphere-2026-3866
Preprint. Discussion started: 28 September 2026

J. Woessner, D. Laurentiu, D. Giardini, et al. The 2013 european seismic hazard model: key components and results. *Bulletin of Earthquake Engineering*, 13:3553–3596, 2015. doi: 10.1007/s10518-015-9795-1.

Z. Yang, F. Waldhauser, Y. Chen, and P. G. Richards. Double-difference relocation of earthquakes in central-western china, 1992–1999. *Journal of Seismology*, 9:241–264, 2005. doi: 10.1007/s10950-005-7893-x.

Z. Yang, J. Liu, et al. Active tectonics and seismicity of the western yunnan fault systems, southeastern tibetan plateau. *Journal of Asian Earth Sciences*, 58:68–82, 2012.

F. R. Zúñiga and S. Wiemer. Spatial and temporal variations of the fractal dimension and *b*-value in mexico. *Bulletin of the Seismological Society of America*, 84(4):1314–1325, 1994.