

Method-dependent spatiotemporal structure in Pacific Ring of Fire earthquake catalogs: a comparative benchmark and 2026 assessment

Muh Akbar Idris

muhakbaridris@apps.ipb.ac.id

IPB University

Research Article

Keywords: earthquake catalog, spatiotemporal clustering, ST-DBSCAN, ST-SCKM, HDBSCAN, declustering, Pacific Ring of Fire, anomaly analysis

Posted Date: September 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10744628/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Method-dependent spatiotemporal structure in Pacific Ring of Fire earthquake catalogs: a comparative benchmark and 2026 assessment

Muh Akbar Idris^{1,2,*}

¹Department of Statistics and Data Science, IPB University, Bogor 16680, West Java, Indonesia

²Department of Mathematics, Universitas Sulawesi Barat, Majene 91412, West Sulawesi, Indonesia

**Corresponding author: muhakbaridris@apps.ipb.ac.id*

Abstract

Earthquake catalogs mix background activity, swarms, foreshocks, mainshocks, and aftershocks; consequently, apparent clustering depends on both the algorithm and the catalog treatment. We compared six event-clustering paradigms and one seismological reference framework on a common Pacific Ring of Fire catalog, then assessed activity from 1 January through 18 August 2026 at 15:56 UTC against matched historical windows. The refreshed official USGS ComCat query returned 521,939 $M \geq 3$ earthquakes for 2000–2026; quality control retained 521,935 records, and a reproducible 500-km PB2002 boundary mask retained 357,080 events. Maximum curvature with a +0.2 correction supported a working completeness threshold of $M4.6$, although the implemented goodness-of-fit criterion did not reach 90%. K-Means, DBSCAN, HDBSCAN, ST-DBSCAN, spatially constrained spatiotemporal K-Means (ST-SCKM), and a self-organizing map were tuned on 2000–2024, validated on 2025, and evaluated on 2026. Density methods isolated compact retained-event structures but rejected 51.7–54.8% of the common sample as noise. ST-SCKM was highly stable under subsampling (mean adjusted Rand index 0.981), yet adding depth or magnitude reduced separation. At $M \geq 4.6$, partial-2026 activity was 2,690 events versus a historical mean of 2,526.5 (rate ratio 1.065; standardized score 0.230; empirical $p = 0.370$), providing no broad rate anomaly. In contrast, all three spatial masks contained 10 $M \geq 7.2$ events: under the primary mask this was 2.08 times the historical mean, exceeded the historical 97.5th percentile, and had a standardized score of 2.33, although the finite-sample empirical upper-tail probability was 0.074. A raw Philippines elevation disappeared after two alternative declustering procedures, and none of 792 directional regional lag tests survived false-discovery-rate correction. Thus, 2026 shows an elevated great-event tail without a catalog-wide rate anomaly; neither result supports deterministic prediction or cross-Pacific propagation.

Keywords: earthquake catalog; spatiotemporal clustering; ST-DBSCAN; ST-SCKM; HDBSCAN; declustering; Pacific Ring of Fire; anomaly analysis

1 Introduction

Earthquake occurrence is intrinsically organized in space and time, but the word *cluster* has several meanings. Plate geometry and long-term tectonic loading produce persistent spatial concentration along faults and subduction zones. Stress transfer after a large rupture produces short-lived aftershock sequences; magmatic or fluid processes can generate swarms; and changes in network sensitivity alter the number and location of recorded small events. The resulting catalog is neither independent nor identically distributed. A visible concentration can therefore represent a physical sequence, a persistent tectonic province, a detection artifact, or a mixture of all three.

This ambiguity matters because an algorithm does not discover a method-free ground truth. Its output is conditional on magnitude completeness, coordinate representation, spatial and temporal scale, treatment of sparse events, and whether every observation must receive a label. A partitioning algorithm summarizes all events by centroids; a density algorithm identifies connected high-density sets and may classify much of the catalog as noise; a point-process model estimates conditional inten-

sity rather than a partition; and a declustering algorithm estimates which events belong to dependent sequences. Their outputs can be scientifically useful while remaining only partially comparable.

Catalog construction is part of the inference rather than a preliminary clerical step. Detection, phase association, relocation, magnitude estimation, and network expansion can change the apparent density of earthquakes [Kubo et al., 2024, Mousavi and Beroza, 2023, Trugman, 2024]. Gutenberg–Richter inference is meaningful only above a defensible completeness magnitude, but completeness varies in both space and time and may temporarily worsen immediately after large events [GUTENBERG and RICHTER, 1955, Woessner, 2005, Corrado et al., 2024, Wang et al., 2025]. For a global Pacific analysis, a single working threshold is therefore a transparent compromise, not a claim that every regional subcatalog is equally complete.

The Pacific Ring of Fire adds a second scale problem. It is a widely used tectonic description, not a uniquely defined polygon. A global analysis must operationalize the study region without implying that a fixed buffer precisely traces every seismogenic source. It must also

distinguish basin-scale regimes that may span thousands of kilometers and years from local sequences occupying tens to hundreds of kilometers and days to months. A method that is stable for the former can be too coarse for the latter, while a local density method can fragment a global catalog into hundreds of structures.

These issues make comparative evaluation multidimensional. Silhouette, Calinski–Harabasz, and Davies–Bouldin indices assess geometry among assigned points but do not establish tectonic validity. For density methods, apparently favorable values can be achieved after rejecting a large fraction of difficult observations. Cluster count, retained-event fraction, spatial dispersion, temporal dispersion, and stability under resampling must therefore be reported together. Time-aware development and validation are equally necessary: tuning on the 2026 interval and subsequently declaring it unusual would introduce leakage.

The refreshed 2026 catalog provides a particularly instructive case. Through 18 August, the operational mask contains ten $M \geq 7.2$ earthquakes, including multiple events around Indonesia, Japan, the Philippines, the southwest Pacific, and the eastern Pacific margin. This concentration raises a legitimate question, but it does not by itself establish that total seismicity or background rate is anomalous. A high threshold can show an unusual tail while the complete $M \geq 4.6$ count remains typical; conversely, one productive aftershock sequence can inflate a regional low-threshold count without changing independent background activity. The statistical question must therefore be stated at each threshold and catalog variant.

We address four research questions. First, how strongly do conventional, density-based, explicitly spatiotemporal, and neural-representation methods disagree when applied to a common leakage-controlled sample? Second, does spatially constrained spatiotemporal K-Means provide greater stability or cohesion than conventional alternatives? Third, which apparent 2026 anomalies persist under two distinct declustering assumptions? Fourth, are regional lag patterns stronger than circular-shift null expectations after correction for multiple testing?

The study contributes (i) a fully cached and hashed USGS workflow with an explicit PB2002-derived spatial mask; (ii) a common-data benchmark that reports separation, noise rejection, fragmentation, spatial and temporal cohesion, and subsample stability; (iii) a transparent earthquake adaptation of ST-SCKM, including negative results from its spatial penalty and attribute sensitivities; and (iv) a threshold-specific 2026 assessment using matched partial-year windows, empirical and count-model baselines, declustering, false-discovery-rate control, and spatial-mask robustness. Table 1 records the provenance and limitations of the data and literature sources used to support those contributions. The aim is statistical characterization, not earthquake prediction, fault-level attribution, or proof of causal propagation between distant regions.

The manuscript follows that inferential sequence from evidence construction to interpretation. The literature review first distinguishes method families by the statistical object they estimate; the data section then fixes

the catalog, spatial mask, completeness threshold, and controlled variants before any comparison is made. Methods are tuned only on historical development data and assessed through complementary geometric, stability, sequence, and null-model diagnostics. Results then separate algorithmic disagreement, raw-versus-declustered behavior, magnitude-threshold dependence, and regional lag evidence before the Discussion integrates them. This ordering keeps every later claim traceable to an explicit data choice, comparison rule, figure, or table rather than allowing the 2026 observations to determine the analytical design retrospectively.

2 Literature review and methodological background

2.1 Catalog construction and magnitude completeness

Modern earthquake catalogs are statistical products of detection, picking, association, location, and magnitude estimation. Machine-learning systems have increased sensitivity and throughput, particularly for weak or overlapping signals, but they can also introduce sharp changes in catalog density when processing systems or station coverage change [Kubo et al., 2024, Mousavi and Beroza, 2023]. High-precision regional catalogs reveal structures that are absent from standard products [Trugman, 2024], while automated replenishment and homogenization efforts emphasize that completeness and magnitude consistency remain active research problems [Candia and Macedo, 2026, Zhan et al., 2026]. A large event count is consequently not synonymous with a stationary observation process.

The Gutenberg–Richter relationship provides a classical framework for the frequency–magnitude distribution [GUTENBERG and RICHTER, 1955]. Maximum-curvature methods identify the modal rollover, whereas goodness-of-fit approaches evaluate how closely the observed tail follows a fitted distribution [Woessner, 2005]. Comparative work shows that catalog-based estimators can disagree and that their performance depends on sample size and nonstationarity [Wang et al., 2025]. Spatially varying analyses further show why a single global threshold can obscure regional structure [Li et al., 2025]. In addition, short-term incompleteness following large earthquakes can induce magnitude correlations and distort sequence statistics [Corrado et al., 2024]. We therefore report both the chosen threshold and the fact that the stricter goodness-of-fit target was not attained.

2.2 Partitioning and density-based clustering

K-Means remains a useful reference because it is fast, reproducible, and easy to validate under resampling [Hartigan and Wong, 1979]. Its objective favors approximately centroidal groups in the chosen feature space and forces complete assignment. For global seismicity, that behavior is suitable for broad regime summarization but poorly matched to irregular fault geometry or sparse background

events. Results also depend on feature scaling: mixing degrees, dates, depths, and magnitudes without a physical or standardized representation can make a nominally spatiotemporal result arbitrary.

DBSCAN defines clusters through density reachability and explicitly labels low-density observations as noise. HDBSCAN extends this idea across a hierarchy of density levels and can accommodate variable-density structure [Campello et al., 2013]. These properties are attractive for aftershock sequences and swarms, and recent earthquake studies use density approaches for regional zoning and sequence characterization [Sirodj et al., 2025, Essing and Poli, 2024, Bi et al., 2024]. Their principal difficulty in a global catalog is scale: a single neighborhood radius and minimum size must negotiate large variations in network density, tectonic setting, and event productivity. Moreover, internal validation computed only on retained points can reward aggressive rejection, so noise fraction and fragmentation are part of the result rather than nuisance diagnostics.

2.3 Explicit spatiotemporal and neural approaches

ST-DBSCAN separates spatial and temporal neighborhoods instead of embedding kilometers and days in a single unconstrained Euclidean scale [Birant and Kut, 2007]. That separation makes the definition interpretable, but fixed thresholds still encode a particular physical scale. Applications in Chile and Sumatra demonstrate its ability to delineate regional sequences [Nicolis et al., 2024, Putri and Sari, 2025]; integration with neural classifiers illustrates a complementary supervised use after clusters are constructed [Delgado et al., 2025]. These studies are valuable, yet their regional parameterizations should not be assumed to transfer unchanged to the entire Pacific margin.

Self-organizing maps project high-dimensional observations onto a topology-preserving grid [Kohonen, 1990]. They can reveal nonlinear organization and provide prototypes for subsequent clustering, but grid size, training schedule, and metaclustering introduce additional choices. More broadly, neural seismic methods are strongest when the target is operational detection or a labeled prediction task; unsupervised embeddings do not automatically supply a physically unique sequence definition. We include SOM as a neural-representation baseline and assess it

with the same leakage and stability controls used for the other methods.

ST-SCKM augments a weighted space–time centroid objective with a penalty for labels that disagree with spatial nearest neighbors. Spatial regularization can in principle discourage geographically fragmented partitions while preserving a complete assignment. The approach was previously evaluated for wildfire zoning [Idris et al., 2026]; its use here is a cross-domain adaptation, not evidence that wildfire performance generalizes to seismicity. Crucially, the penalty strength must be tuned and is allowed to select zero. That possibility turns the method into a testable question: does the proposed constraint improve the selected earthquake representation, or does ordinary weighted K-Means already fit the benchmark better?

2.4 Declustering, nearest-neighbor structure, and point processes

Declustering estimates which events are dependent members of sequences and which approximate a background process. Window methods such as Gardner–Knopoff use magnitude-dependent space–time rules, while nearest-neighbor methods combine space, time, and magnitude into a proximity metric whose modes can separate clustered and background populations [Zaliapin et al., 2008, Zaliapin and Ben-Zion, 2020]. Neither approach is assumption-free. Declustering choices can change main-shock magnitude distributions and hazard-relevant summaries [Mizrahi et al., 2021, Taroni and Akinci, 2020]; recent work continues to automate or learn these decisions [Bountzis et al., 2026, Shcherbakov and Kothari, 2025]. Agreement across two contrasting procedures is therefore more persuasive than reliance on one declustered catalog.

Epidemic-type aftershock sequence models describe conditional intensity as a background component plus contributions triggered by earlier earthquakes [Ogata, 1988]. Contemporary variants improve scalability or incorporate flexible background structure [Stockman et al., 2024, Molkenhuth et al., 2024]. ETAS is a generative point-process framework rather than a conventional partition and should not be ranked against K-Means by silhouette. In this study it anchors interpretation: strong local space–time dependence is expected in a raw catalog, and a regional count elevation may reflect sequence

Table 1: Data sources and authority.

Source	Role	Period/version	Retrieval	Limitation
USGS ComCat FDSN Event Web Service	Primary earthquake catalog	2000-01-01 through 2026-08-18 15:56:38 UTC; service 2.7.0	Official API cached in six-month CSV chunks	CSV omits tsunami metadata; cached USGS GeoJSON details enrich the major-event inventory. Magnitude types and monitoring capability vary.
USGS Tectonic Plate Boundaries ArcGIS service	Primary boundary geometry	retrieved 2026-08-18; PB2002 derivative	Official paginated GeoJSON query	Operational Ring of Fire corridor and buffer remain analytical choices
Crossref REST API	Literature DOI and metadata verification	retrieved 2026-08-18	DOI-specific and bibliographic queries	Does not independently verify Scopus indexing; abstracts not universal
Official publisher pages	Key-reference validation	retrieved 2026-08-18	Direct publisher URLs	Access and metadata coverage differ by publisher

Table 2: Literature-based taxonomy of clustering approaches.

Family	Representative methods	Earthquake role	Key caution
Conventional partitioning	K-Means; fuzzy C-means; hierarchical clustering	Compact partition baseline	Requires K; favors convex clusters; no native noise
Density-based	DBSCAN; HDBSCAN; OPTICS	Irregular clusters and explicit noise	Density variation and parameter sensitivity
Explicit spatiotemporal	ST-DBSCAN; ST-SNN; space-time density peak	Separate spatial and temporal neighborhoods	Neighborhood definitions are scale-dependent
Spatially constrained	ST-SCKM	Compact partitions with KNN spatial-contiguity penalty	Constraint can smooth real discontinuities; requires K and weights
Neural/unsupervised ML	SOM; modified SOM; embedding plus clustering	Topology-preserving representation	Grid and learning schedule affect output; clusters require post-processing
Seismological/statistical	Nearest-neighbor; ETAS; stochastic declustering	Background-dependent separation and probabilistic sequence description	Not the same estimand as conventional event clustering

Table 3: Selected algorithms, assumptions, and outputs.

Method	Family	Representation	Output	Noise	Parameters
K-Means	Conventional	unit-sphere space plus standardized time	Hard partition	–	K; initialization; spatial/time scaling
DBSCAN	Density-based	scaled unit-sphere space plus time	Clusters plus noise	Native	epsilon; MinPts; feature scaling
HDBSCAN	Density-based	scaled unit-sphere space plus time	Clusters plus noise	Native	min_cluster_size; min_samples
ST-DBSCAN	Explicit spatiotemporal	latitude/longitude/time with Haversine spatial search	Clusters plus noise	Native	spatial epsilon; temporal epsilon; MinPts
ST-SCKM	Spatially constrained	unit-sphere space plus standardized time; attributes in sensitivity models	Hard constrained partition	–	K; spatial/time weights; KNN k; lambda; initialization
SOM	Neural/unsupervised ML	scaled unit-sphere space plus time	Neuron assignment then metaclusters	–	grid; epochs; learning rate; neighborhood radius; metacluster K
Nearest-neighbor	Seismological reference	space-time-magnitude rescaled distance	Parent links and background/cluster labels	Background class	b value; fractal dimension; mixture threshold

productivity rather than a change in background rate.

2.5 Research gap and comparison logic

Recent reviews document rapid growth in clustering and machine learning for seismology [Vijay et al., 2025, Kubo et al., 2024], while regional studies quantify sequence geometry, fluid-related migration, magnitude clustering, and nearest-neighbor networks [Gentili et al., 2024, Talebi et al., 2024, Gossett et al., 2024]. The literature is rich in method-specific applications but thinner in global common-catalog benchmarks that simultaneously reveal completeness uncertainty, noise rejection, fragmentation, temporal leakage, resampling stability, alternative declustering, null models, and current-period sensitivity.

Our comparison therefore avoids a single winner-takes-all score. Partitioning, density, explicit spatiotemporal, and neural-representation methods are evaluated on a common historical sample using a panel of diagnostics. Nearest-neighbor declustering and point-process concepts remain seismological references rather than being forced into the same leaderboard. For 2026, low-threshold rate, high-magnitude tail, regional sequence sensitivity, and lagged association are treated as separate estimands. This structure is essential because evidence for one is not evidence for the others.

Fig. 1 gives a compact visual map of the six method-

Methodological taxonomy and representative paradigms

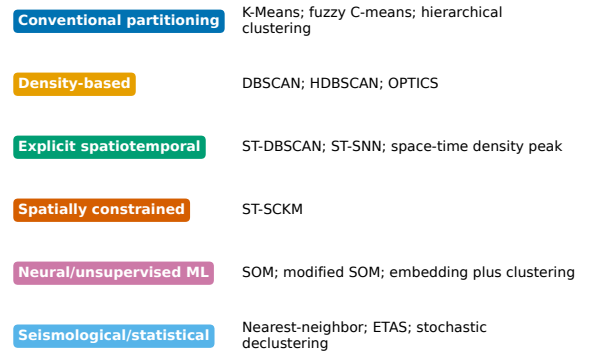


Fig. 1: Literature-based taxonomy of methods considered for earthquake clustering. Statistical point-process and declustering models are treated as reference frameworks rather than forced into a conventional clustering leaderboard.

ological families, and Tables 2 and 3 convert that literature synthesis into the comparison logic used in the analysis. The first table separates the families by estimand, while the second states the representation, output, noise behavior, and principal controls for every selected algorithm. Read together, the visual taxonomy and the two tables explain why point-process and declustering frameworks remain interpretive references instead of being forced into a silhouette-based leaderboard.

3 Data and study area

3.1 USGS catalog and spatial mask

We queried the official USGS FDSN event service in non-overlapping six-month chunks for earthquakes of $M \geq 3$ from 1 January 2000 through the actual retrieval cutoff, 18 August 2026 at 15:56:38 UTC. The refreshed cached download contains 521,939 records and preserves API parameters, retrieval time, chunk row counts, and SHA-256 hashes. Quality control excluded four duplicate event signatures and retained 521,935 unique records. The primary 500-km study mask contains 357,080 events with magnitudes 3.0–9.1 and depths -3.0 –735.8 km.

The Ring of Fire is diffuse rather than a formal polygon. We operationalized it using official USGS PB2002 plate and microplate boundary layers derived from the digital plate model of Bird [2003]. Event-to-boundary geodesic distance used unit-sphere coordinates and a spatial index. The primary mask retained events within 500 km; 300-km and 700-km alternatives test sensitivity. Fig. 2 shows the operational corridor and selected boundary geometry, while Fig. 3 shows how the magnitude-complete observations concentrate along that geometry. Twelve reproducible regional strata are descriptive, not tectonic polygons. The complete sequence of quality-control exclusions is reported in Table 4 rather than being separated from the data description.

The small number of formal exclusions should not be mistaken for uniform catalog quality. Quality control determines whether a record is internally usable, whereas completeness describes whether earthquakes of a given size were consistently observed. An event can therefore pass every row-level check in Table 4 while still belonging to a period or region with different detection capability. We preserve this distinction throughout the analysis: the quality-controlled $M \geq 3$ catalog supports mapping and sensitivity checks, but rate comparisons and the primary clustering benchmark use stricter magnitude thresholds. This ordering connects data provenance to the later inferential design and prevents a clean file from being treated as a homogeneous observation process.

3.2 Completeness and controlled datasets

Maximum curvature placed the modal estimate at M4.35. Adding 0.2 yielded the working threshold M4.6. The 2000–2025 b value above this threshold was 1.153, with 104,076 events. The period-specific estimates in Table 5 show that the same corrected threshold was selected for development, validation, and partial evaluation. The goodness-of-fit calculation did not attain 90% or 95%, so M4.6 is a defensible working threshold rather than a definitive global completeness limit.

The agreement of the corrected maximum-curvature estimate across periods supports using one transparent benchmark threshold, but it does not eliminate regional heterogeneity. For that reason, the analysis does not interpret the M4.6 catalog as perfectly complete everywhere. Instead, it asks whether conclusions survive changes in boundary width, magnitude threshold, and sequence treatment. The frequency–magnitude curve

in Fig. 4 provides the catalog-wide context, while the temporal and depth displays in Figs. 5 and 6 show why a single scalar threshold cannot summarize all observational structure.

Dataset A is the $M \geq 3$ raw masked catalog and is used primarily to describe the available event field. Dataset B applies $M \geq 4.6$ and forms the main rate-analysis catalog. Dataset C contains Gardner–Knopoff-type and nearest-neighbor declustered alternatives, which permit sequence-sensitive findings to be separated from results that persist under contrasting background definitions. Dataset D contains $M \geq 6.5$ events and supports the large-event inventory and high-threshold sensitivity analysis. These variants answer different questions and are never silently pooled; each result therefore carries an explicit magnitude threshold and catalog treatment into the Results and Discussion sections.

4 Methods

4.1 Representation and leakage control

Latitude and longitude were mapped to three-dimensional unit-sphere coordinates. UTC event time was converted to days and standardized. The primary representation was space plus time; magnitude and depth remained attributes except in controlled ST-SCKM sensitivities. Development ended in 2024, calendar year 2025 was reserved for validation, and 2026 was excluded from tuning. A fixed seed and inverse year–region frequency sampling produced a common 20,000-event $M \geq 5$ benchmark.

This representation separates geometric decisions from scientific attributes. Unit-sphere coordinates avoid a discontinuity at the international date line and reduce latitude-dependent distortion, while standardizing time prevents the numerical scale of elapsed days from dominating spatial coordinates. Magnitude and depth were not discarded; they were withheld from the primary distance so that their effect could be evaluated rather than assumed. The year–region sampling weights similarly prevent a few highly productive sequences from defining the entire benchmark. Together, these choices make the comparison a controlled test of method behavior on a shared representation, not a contest among algorithms that received different versions of the catalog.

4.2 Algorithms and tuning

Direct comparison covers K-Means, DBSCAN, HDBSCAN, ST-DBSCAN, ST-SCKM, and SOM. Nearest-neighbor clustering is a seismological reference. We evaluated 176 historical configurations: 4 K-Means, 16 DBSCAN, 16 HDBSCAN, 26 crossed ST-DBSCAN, 6 SOM, and 108 ST-SCKM. Selection required nondegenerate cluster counts and at most 85% noise, then balanced retained-point silhouette against noise and fragmentation. The rule is procedural, not a proposed universal validation index.

The prespecified grids and selection rules are shown in Table 6. ST-SCKM minimizes weighted distance to centroids plus a disagreement penalty for labels differ-

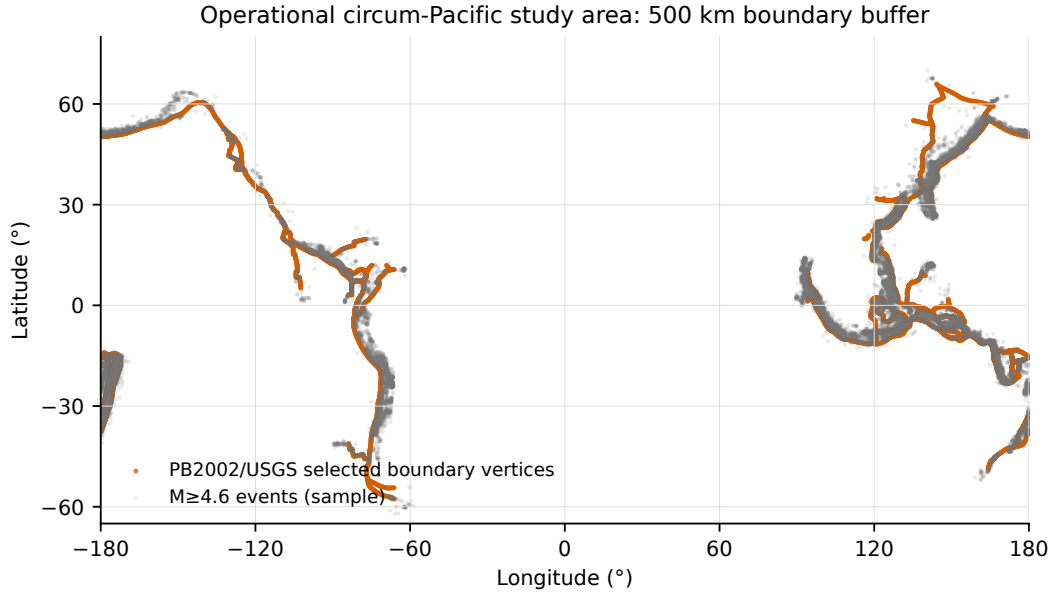


Fig. 2: Operational Pacific Ring of Fire study area based on a 500-km corridor around PB2002 plate-boundary vertices. Alternative 300-km and 700-km masks are used in robustness analysis.

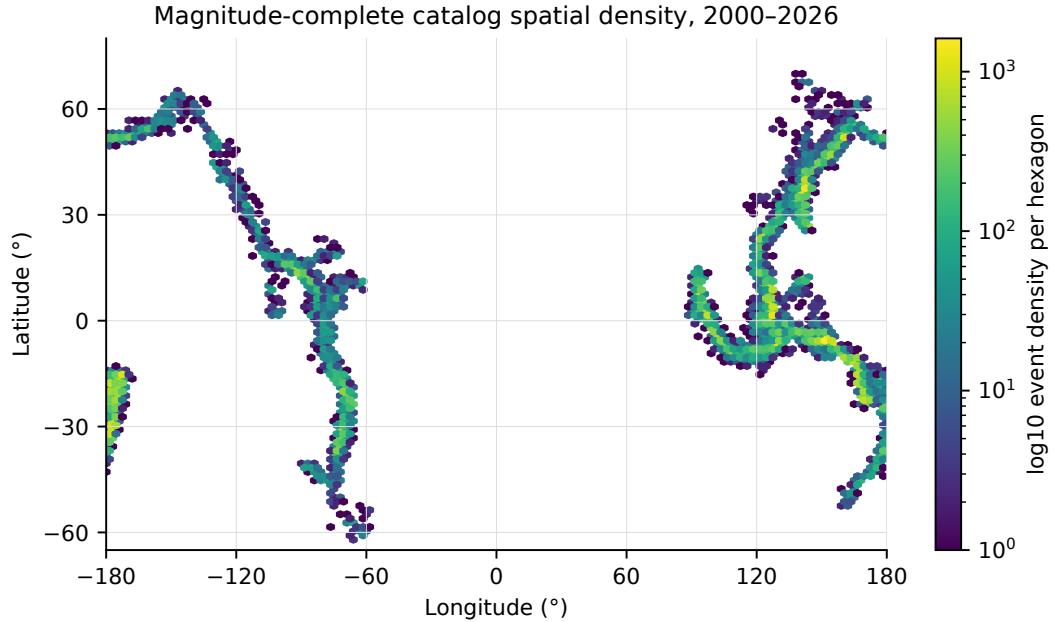


Fig. 3: Spatial distribution of the quality-controlled catalog within the operational study mask.

ing from spatial nearest neighbors. Unit-sphere space and standardized time avoid raw-degree/raw-timestamp Euclidean distance. The selected space–time configuration was $K = 8$, $w_s = 0.7$, $w_t = 0.3$, $k = 10$, and $\lambda = 0$. Specifications were A: space–time; B: A+depth; C: A+magnitude; D: A+depth+magnitude. Table 7 re-

ports the selected development-only configuration for each method. A selected zero penalty is retained as an informative negative result.

Parameter selection was evaluated as a constrained decision rather than a search for the largest silhouette alone. A density model with excellent separation among a

Table 4: Catalog preprocessing and exclusions.

Step	Excluded	Remaining	Reason
invalid_time	0	521939	unparseable preferred origin time
after_cutoff	0	521939	outside prespecified cutoff
missing_magnitude	0	521939	magnitude required for threshold analyses
impossible_coordinates	0	521939	outside geographic coordinate bounds
unrealistic_depth	0	521939	missing or outside -5 to 1000 km
non_earthquake	0	521939	USGS event type is not earthquake
duplicate_event_id	0	521939	retained latest USGS update
duplicate_signature	4	521935	identical time, hypocenter, depth, and magnitude

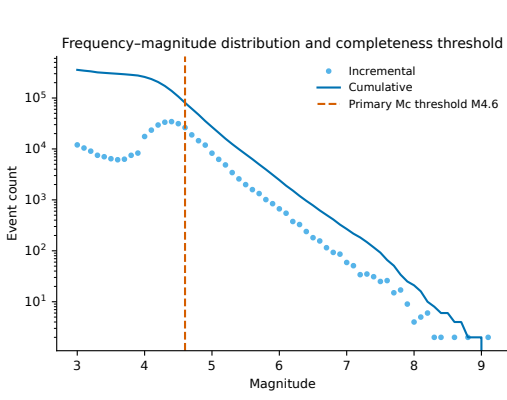


Fig. 4: Frequency–magnitude distribution and the working completeness threshold.

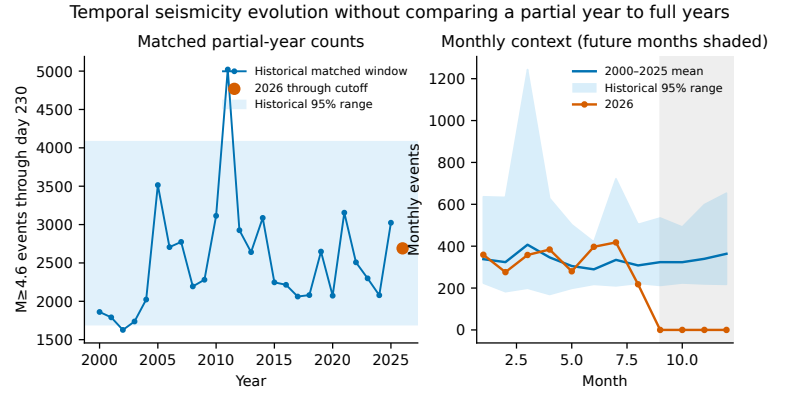


Fig. 5: Matched partial-year counts and monthly context. Historical years and 2026 are truncated to the same day-of-year; unobserved September–December 2026 months are shaded.

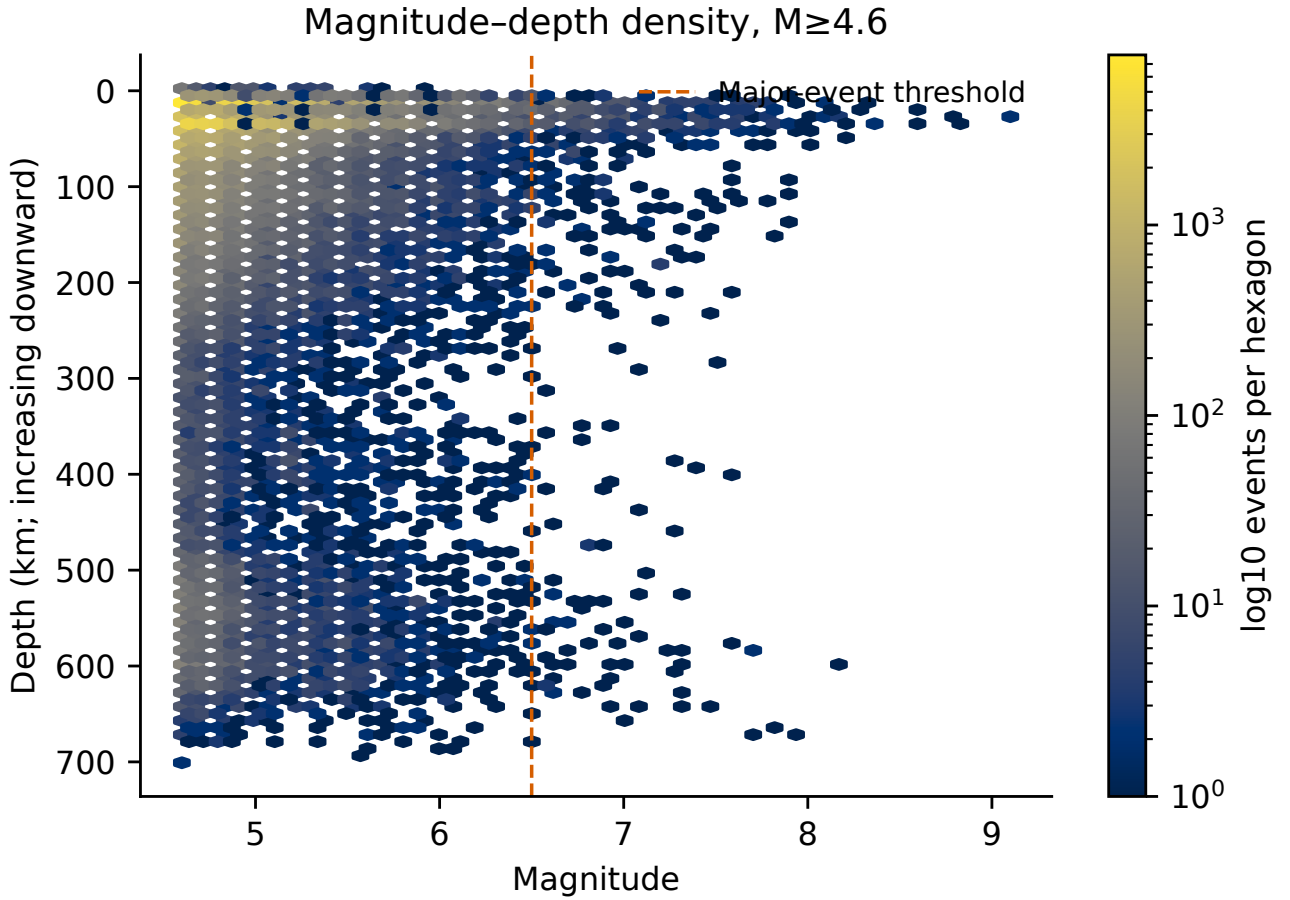


Fig. 6: Magnitude–depth distribution for the magnitude-complete catalog.

small retained subset can be less useful than a weaker but more stable partition if the scientific objective is catalog-wide summarization. Conversely, forcing every event into

a centroidal group can conceal the short-lived structures that a sequence analyst intends to isolate. The selection score therefore serves only as a reproducible filter within

Table 5: Magnitude-completeness estimates by period.

Period	N	MAXC	MAXC+0.2	GFT90	GFT95	b	N above Mc
2000–2025	347358	4.35	4.6	–	–	1.153	104076
2010–2025	217697	4.35	4.6	–	–	1.193	67093
2015–2025	151825	4.35	4.6	–	–	1.198	42756
development_2000–2024	330865	4.35	4.6	–	–	1.152	99256
validation_2025	16493	4.35	4.6	–	–	1.179	4820
evaluation_2026_partial	9722	4.35	4.6	–	–	1.146	2690

Note: The 90% and 95% goodness-of-fit criteria were not attained; blank estimates are retained explicitly.

each method family; final interpretation returns to the full diagnostic panel. This separation between tuning and interpretation is especially important for ST-SCKM, because the selected $\lambda = 0$ shows that the spatial penalty was tested honestly and was not retained merely because it motivated the proposed adaptation.

4.3 Validation and inference

Internal metrics were silhouette, Calinski–Harabasz, and Davies–Bouldin among non-noise events where applicable. Spatial cohesion was mean within-cluster geodesic dispersion; temporal cohesion was mean within-cluster temporal median absolute deviation. Stability used ten seeded 80% subsamples and adjusted Rand agreement.

Knox statistics counted pairs close in space and time against temporal permutations. Regional daily $M \geq 5$ series were tested at 1, 3, 7, 14, 30, and 60 days against circular-shift surrogates. Benjamini–Hochberg adjustment covered all 792 directional tests. Partial-year historical counts were truncated to the 2026 day of year. We report ratios, standardized scores, empirical upper-tail probabilities, Poisson or negative-binomial tails, and FDR adjustments. Conclusions were repeated across magnitude thresholds, spatial masks, and declustering methods.

The validation design deliberately links each numerical summary to a specific uncertainty. Subsampling evaluates whether a fitted partition is reproducible when the event set changes; temporal permutation evaluates whether close space–time pairs exceed a catalog-preserving null; circular shifts evaluate regional lag correlations without treating daily observations as independent; and matched partial-year windows avoid comparing an incomplete 2026 interval with complete historical years. No one procedure substitutes for the others. Their agreement or disagreement becomes part of the result, which is why effect sizes, empirical probabilities, model-based probabilities, and multiplicity adjustments are reported together.

The final inferential step is therefore hierarchical. Method performance is interpreted first because it determines what kind of cluster each algorithm can support; catalog variants are compared second because sequence removal changes the population being counted; and the partial-2026 assessment is evaluated last because it must inherit the prespecified threshold, exposure window, spatial mask, and multiplicity family. A result advances to the Discussion only when its denominator and conditioning choices remain visible in the accompanying table or figure. This structure provides a continuous bridge from the validation criteria in Section 4.3 to the multidimensional comparisons reported next.

5 Results

5.1 Method disagreement

The methods did not recover a common partition. K-Means selected $K = 10$ and assigned every event (silhouette 0.345). ST-SCKM A produced eight complete clusters (0.355), and SOM produced nine metaclusters

(0.303). DBSCAN retained most events but had weak separation (16 clusters, 6.3% noise, silhouette -0.018). HDBSCAN produced 180 compact retained-event clusters and 54.8% noise (silhouette 0.564; mean spatial dispersion 159 km; temporal MAD 197 days). ST-DBSCAN produced 759 short-lived structures and 51.7% noise (silhouette 0.365; 144 km; 16.9 days). Its cohesion is conditional on rejecting more than half the sample and substantial fragmentation. Table 8 keeps all of these dimensions together so that a high silhouette cannot be read without its noise and dispersion costs.

Mean subsample adjusted Rand indices were 0.981 for ST-SCKM, 0.866 for ST-DBSCAN, 0.860 for K-Means, 0.637 for SOM, 0.545 for DBSCAN, and 0.471 for HDBSCAN (Table 9). HDBSCAN’s compact retained clusters were less stable, while ST-SCKM’s complete partition was stable but broad. No method dominates all dimensions.

Read jointly, Tables 8 and 9 show that apparent method superiority changes with the question being asked. HDBSCAN provides the strongest retained-point separation, but its high noise fraction and lower resampling agreement limit any claim that it reveals a unique global partition. ST-DBSCAN provides the shortest temporal dispersion and strong local coherence, but its 759 clusters describe a sequence-scale decomposition rather than a compact global summary. K-Means, SOM, and ST-SCKM occupy the opposite end of that spectrum: they are easier to reproduce as complete partitions, yet their clusters are broad in both space and time. Fig. 7 translates these connected trade-offs into a common visual scale, while Fig. 9 retains the uncertainty across subsamples.

5.2 ST-SCKM attribute sensitivity

Space–time specification A had silhouette 0.355. Adding depth lowered it to 0.231; magnitude lowered it to 0.071; and both produced 0.014. Spatial dispersion increased from 1,864 km for A to 3,894 km for D, while temporal MAD increased from 1,126 to 1,622 days. Magnitude and depth therefore created separation less compatible with space–time cohesion in this representation.

The sensitivity surface in Fig. 8 clarifies that this deterioration is not simply a consequence of one unfortunate parameter combination. Across the evaluated configurations, positive spatial penalties did not create a consistent gain large enough to outweigh the resulting changes in geometry. The attribute experiment leads to the same methodological lesson from another direction: adding scientifically meaningful variables does not automatically improve an unsupervised representation. Their value depends on how distance is defined and which notion of coherence is being optimized. The result therefore supports a parsimonious space–time benchmark while leaving depth and magnitude available for post-cluster interpretation.

5.3 Raw and declustered structure

Gardner–Knopoff-type declustering retained 35,389 of 106,766 $M \geq 4.6$ events, whereas the nearest-neighbor alternative retained 68,006. On matched 5,000-event sam-

ples, broad K-Means metrics were similar across raw and declustered variants. The $M \geq 6$ historical Knox statistic observed 1,330 pairs within 100 km and 7 days compared with a temporal-permutation mean of 42.3 ($p = 0.002$). This establishes local catalog clustering, not distant triggering.

The difference between the two retained counts is

itself informative because it measures the sensitivity of the background concept to the declustering rule. Gardner–Knopoff windows remove events according to magnitude-dependent envelopes, whereas the nearest-neighbor model uses a fitted space–time–magnitude proximity distribution. Their disagreement cautions against treating either reduced catalog as an observed truth.

Table 6: Prespecified hyperparameter search spaces.

Method	Parameter	Search values	Selection rule
K-Means	K	8;10;12;14	multi-metric quality and stability; no single-score winner
DBSCAN	epsilon_scaled	0.30;0.40;0.60	quality subject to nondegenerate noise and cluster counts
DBSCAN	MinPts	5;8;12;20	parameter perturbation stability
HDBSCAN	min_cluster_size	20;30;50;100	persistence; noise; perturbation stability
HDBSCAN	min_samples	5;8;12;20	persistence; noise; perturbation stability
ST-DBSCAN	spatial_epsilon_km	50;100;200;300;500	separate spatial and temporal cohesion with stability
ST-DBSCAN	temporal_epsilon_days	1;3;7;14;30;60	separate spatial and temporal cohesion with stability
ST-DBSCAN	MinPts	5;8;12;20	noise and fragmentation constraints
ST-SCKM	K	8;10;12;14	multi-dimensional quality and stability
ST-SCKM	spatial_weight; temporal_weight	0.5/0.5;0.7/0.3;0.9/0.1	sensitivity; interpretability
ST-SCKM	KNN_k	5;10;20	spatial cohesion and fragmentation
ST-SCKM	lambda	0;0.05;0.20	constraint benefit versus oversmoothing
SOM	grid	4x4;5x5;6x6	quantization/topographic error and metacluster stability
SOM	learning_schedule	0.5 to 0.05; sigma grid/2 to 1	quantization/topographic error
Nearest-neighbor	b;fractal_dimension	estimated b; 1.4;1.6;1.8	mixture separation and agreement with independent declustering

Table 7: Selected development-only parameter configurations.

Method	Selected parameters	Clusters	Noise	Silhouette	Selection score
DBSCAN	eps=0.2; MinPts=20	29	0.313	0.137	0.053
HDBSCAN	min cluster=20; min samples=12	46	0.482	0.385	0.255
K-Means	K=10	10	0	0.372	0.37
SOM	grid=6x6; epochs=12; metaclusters=10	9	0	0.349	0.347
ST-DBSCAN	eps_s=300 km; eps_t=30 d; MinPts=5	119	0.83	0.824	0.593
ST-SCKM	K=8; ws=0.7; wt=0.3; kNN=10; lambda=0	8	0	0.366	0.365

Note: Selection used historical data only. The score penalizes noise and fragmentation and is not proposed as a new validity index.

Table 8: Multidimensional clustering performance on the common historical sample.

Method	Clusters	Noise	Sil.	DB	Spatial km	Temporal d	ST coherence
K-Means	10	0	0.345	0.956	1,680	1,010	0.14
DBSCAN	16	0.063	-0.018	0.893	2,891	2,156	0.079
HDBSCAN	180	0.548	0.564	0.553	159.035	196.781	0.538
SOM	9	0	0.303	1.072	1,782	1,181	0.128
ST-DBSCAN	759	0.517	0.365	1.118	144.029	16.869	0.749
ST-SCKM (A)	8	0	0.355	0.998	1,864	1,126	0.128
ST-SCKM (B)	8	0	0.231	2.919	2,566	1,232	0.105
ST-SCKM (C)	8	0	0.071	9.558	3,467	1,383	0.085
ST-SCKM (D)	8	0	0.014	8.448	3,894	1,622	0.076

Note: Metrics for density methods are conditional on non-noise events and must be read with the noise fraction.

Table 9: Stability under ten 80% subsamples.

Method	Replicates	ARI mean	ARI SD	ARI min	ARI max
K-Means	10	0.86	0.153	0.674	0.993
DBSCAN	10	0.545	0.029	0.495	0.588
HDBSCAN	10	0.471	0.304	0.183	0.831
ST-DBSCAN	10	0.866	0.016	0.839	0.891
ST-SCKM	10	0.981	0.007	0.965	0.99
SOM	10	0.637	0.067	0.575	0.774

Method comparison exposes trade-offs rather than a single winner

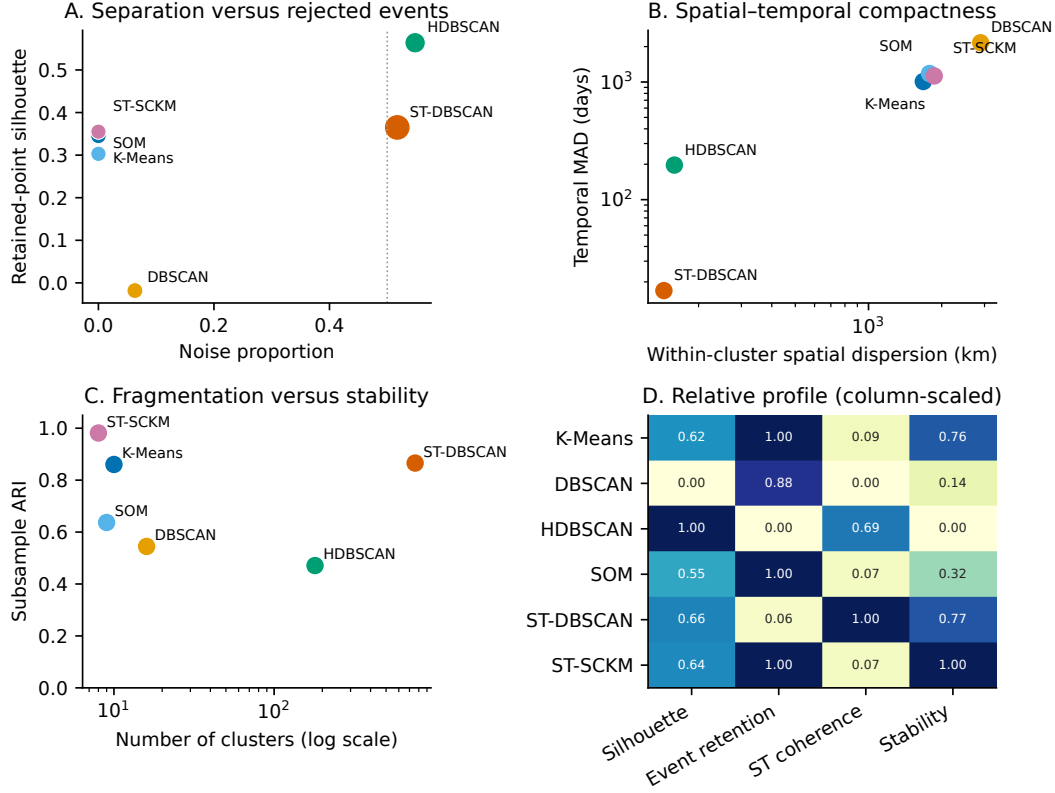


Fig. 7: Four-dimensional method comparison on the common historical sample: (A) retained-point separation versus rejected events, (B) spatial versus temporal compactness, (C) fragmentation versus subsample stability, and (D) the column-scaled diagnostic profile. The panels expose trade-offs that cannot be represented by one ranking score.

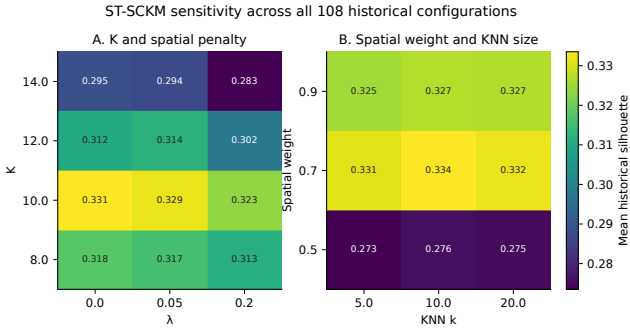


Fig. 8: ST-SCKM parameter sensitivity on historical development data: (A) mean silhouette across cluster count and spatial penalty, and (B) mean silhouette across spatial weight and KNN neighborhood size.

Nevertheless, both provide a useful stress test: a conclusion that persists in the raw catalog and under both reductions is much less dependent on one particular sequence definition. Fig. 10 shows the annual consequences of these choices before the partial-2026 comparison is interpreted.

5.4 Partial 2026 activity

The primary raw catalog contains 2,690 $M \geq 4.6$ events in the partial-2026 window. The matched historical mean was 2,526.5 (95% empirical interval 1,695.8–4,079.8), giv-

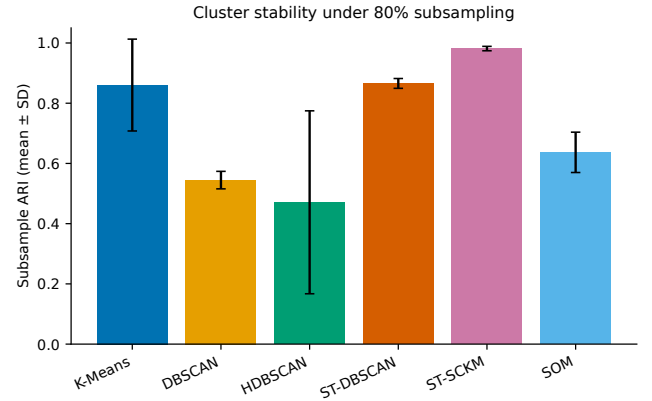


Fig. 9: Stability under 80% subsampling.

ing a rate ratio of 1.065, standardized score of 0.230, and empirical $p = 0.370$. Gardner–Knopoff gave 916 versus 828.3 (ratio 1.106; score 1.074; $p = 0.148$), and nearest-neighbor declustering gave 1,643 versus 1,591.9 (ratio 1.032; score 0.258; $p = 0.407$). None indicates a global anomaly at the working completeness threshold; Table 11 places the global and Philippines raw/declustered results side by side.

The raw Philippines count was 329 versus 115.4 (ratio 2.851; score 6.779). A negative-binomial model gave FDR-adjusted $p = 0.000107$, while the discrete empirical $p = 0.037$ became 0.610 across the planned family. More

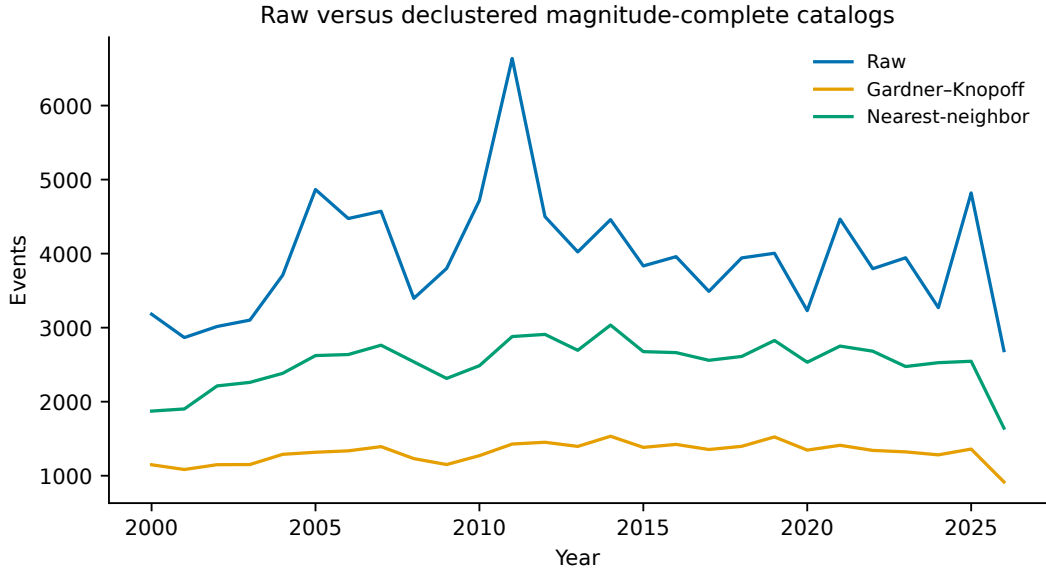


Fig. 10: Annual counts in the raw and two declustered magnitude-complete catalogs.

importantly, Gardner–Knopoff gave 48 versus 43.8 (ratio 1.095; $p = 0.333$) and nearest-neighbor gave 105 versus 82.2 (ratio 1.278; $p = 0.111$). The raw elevation is sequence-sensitive rather than robust background-rate evidence.

The high-magnitude tail gives a different result. There were 21 $M \geq 6.5$ events and 10 $M \geq 7.2$ events through the cutoff; all ten large events are identified in Table 10. Under the primary 500-km mask, the $M \geq 7.2$ historical matched-window mean was 4.81 and the 97.5th percentile was 9.5; the observed count of 10 therefore had a rate ratio of 2.08 and standardized score of 2.33. Because one of only 26 historical years also reached or exceeded 10, the add-one empirical upper-tail probability was 0.074. The exceedance was robust to the 300-, 500-, and 700-km masks, with standardized scores from 2.31 to 2.75. This is evidence for an elevated large-event tail, not for an elevated total earthquake rate.

Tables 10 and 11 deliberately place event-level and count-level evidence next to one another. The event inventory prevents the high-threshold result from becoming an abstract count by identifying when and where each $M \geq 7.2$ earthquake occurred. The anomaly table then shows that this tail behavior does not generalize to the complete-catalog threshold. In other words, the same partial year can be ordinary in its overall M4.6 rate, sequence-sensitive in one regional raw count, and elevated in its largest magnitudes. Those statements concern different parts of the distribution and should remain separate rather than being combined into a single label for 2026.

The distinction also explains the visual sequence that follows. Fig. 11 contrasts raw and declustered regional standardized scores, Fig. 12 places the regional counts in historical context, and Fig. 13 links the large-event tail to its geographic and temporal distribution. None of these displays alone establishes a mechanism. Together they show which features are widespread, which are concentrated in a productive sequence, and which depend on a high magnitude threshold.

5.5 Lagged association, null models, and robustness

None of 792 directional region–lag correlations survived FDR correction; the minimum adjusted p was 0.396. Table 12 displays the ten smallest unadjusted probabilities rather than hiding the strongest candidates, and Table 13 contrasts the completeness-threshold result with the $M \geq 7.2$ tail across all three boundary buffers. The lag result gives no statistical support for a systematic sequence moving among the defined Pacific regions at the tested lags. Temporal-permutation Knox tests, by contrast, found strong local space–time clustering, as expected in a catalog containing sequences. Local dependence and cross-region lag evidence are distinct.

The multiplicity correction matters because hundreds of directional comparisons almost guarantee a few small unadjusted probabilities even when the global pattern is compatible with the null. The ten displayed candidates in Table 12 are therefore diagnostics, not discoveries. Their adjusted probabilities remain large, their signs are mixed, and their selected lags do not form a coherent progression around the Pacific margin. Fig. 14 makes that absence of an ordered directional pattern visible. By contrast, the boundary-buffer results in Table 13 show a consistent threshold pattern: the M4.6 count remains close to its historical mean, whereas the $M \geq 7.2$ count exceeds the historical 97.5th percentile for every tested corridor width. The temporal-permutation distribution and threshold-by-buffer sensitivity in Fig. 15 then separate strong local clustering from the threshold-specific global tail result.

5.6 Integrated synthesis of the empirical results

The empirical results form a sequence of linked comparisons rather than a collection of independent rankings. First, the common-sample benchmark establishes that method choice changes the type of structure recovered. Complete partitioning methods summarize broad regimes, while density and explicitly spatiotempo-

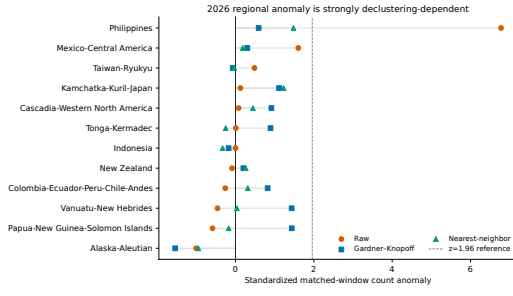


Fig. 11: Regional partial-2026 standardized scores for raw and two declustered catalogs. The Philippines raw-catalog elevation collapses after sequence removal.

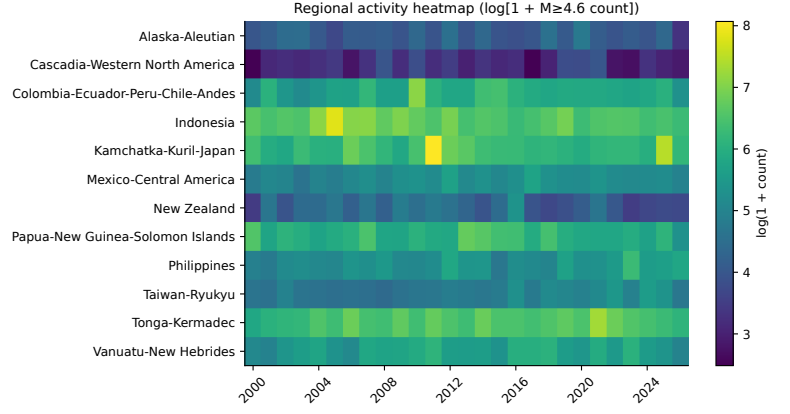


Fig. 12: Regional magnitude-complete activity through time.

ral methods isolate smaller structures by rejecting or fragmenting observations. Second, the stability experiment

shows that repeatability does not resolve this conceptual difference: a broad partition can be highly stable without

Table 10: All $M \geq 7.2$ earthquakes in the operational mask through 18 August 2026 at 15:56 UTC.

Date	M	Depth km	Tsunami	Region	USGS place
2026-03-24	7.5	234	0	Tonga-Kermadec	158 km W of Neiafu, Tonga
2026-03-30	7.3	120.628	0	Vanuatu-New Hebrides	51 km ENE of Luganville, Vanuatu
2026-04-01	7.4	35	0	Indonesia	129 km ESE of Bitung, Indonesia
2026-04-20	7.4	25	1	Kamchatka-Kuril-Japan	102 km ENE of Miyako, Japan
2026-06-07	7.8	56.991	0	Philippines	25 km SW of Kablalan, Philippines
2026-06-24	7.2	10	0	Colombia-Ecuador-Peru-Chile-Andes	20 km E of San Felipe, Venezuela
2026-06-24	7.5	10	0	Colombia-Ecuador-Peru-Chile-Andes	20 km W of Catia La Mar, Venezuela
2026-07-17	7.3	22	0	Mexico-Central America	52 km W of Puerto Madero, Mexico
2026-08-10	7.4	110.285	0	Mexico-Central America	5 km S of San José del Palmar, Colombia
2026-08-14	7.7	10	0	Indonesia	68 km NNW of Ende, Indonesia

Note: Tsunami is the official USGS GeoJSON flag (1=yes, 0=no), not an impact measure. The complete 21-event $M \geq 6.5$ inventory is included in table_10_2026_major_earthquakes.csv.

Table 11: Primary global and Philippines anomaly results under raw and declustered catalogs.

Catalog	Region	2026	Hist. mean	Ratio	z	Emp. p	Model FDR
Raw	GLOBAL	2690	2,526	1.065	0.23	0.37	0.823
Raw	Philippines	329	115.385	2.851	6.779	0.037	1.07e-04
GK	GLOBAL	916	828.308	1.106	1.074	0.148	0.823
GK	Philippines	48	43.846	1.095	0.591	0.333	0.823
NND	GLOBAL	1643	1,592	1.032	0.258	0.407	0.823
NND	Philippines	105	82.154	1.278	1.48	0.111	0.732

Note: The complete regional table is supplied as table_11_regional_anomalies.csv.

Table 12: Ten smallest unadjusted regional lag-test probabilities.

Source	Target	Lag d	Corr.	p	FDR p	FDR _i 0.05
TON	NZ	1	0.094	0.002	0.396	False
CAS	VAN	7	0.043	0.002	0.396	False
TON	NZ	3	0.074	0.002	0.396	False
NZ	PHL	60	0.038	0.002	0.396	False
KJ	VAN	60	0.131	0.004	0.396	False
NZ	TON	1	0.103	0.004	0.396	False
KJ	MCA	1	0.058	0.004	0.396	False
MCA	ALA	7	0.047	0.004	0.396	False
TON	VAN	7	0.056	0.006	0.413	False
NZ	IND	3	0.068	0.006	0.413	False

Note: Region codes: ALA Alaska-Aleutian; CAS Cascadia; AND Andes; IND Indonesia; KJ Kamchatka-Kuril-Japan; MCA Mexico-Central America; NZ New Zealand; PNG Papua New Guinea-Solomon Islands; PHL Philippines; TWR Taiwan-Ryukyu; TON Tonga-Kermadec; VAN Vanuatu-New Hebrides. None of all 792 directional tests survives FDR correction; the complete table is included as CSV.

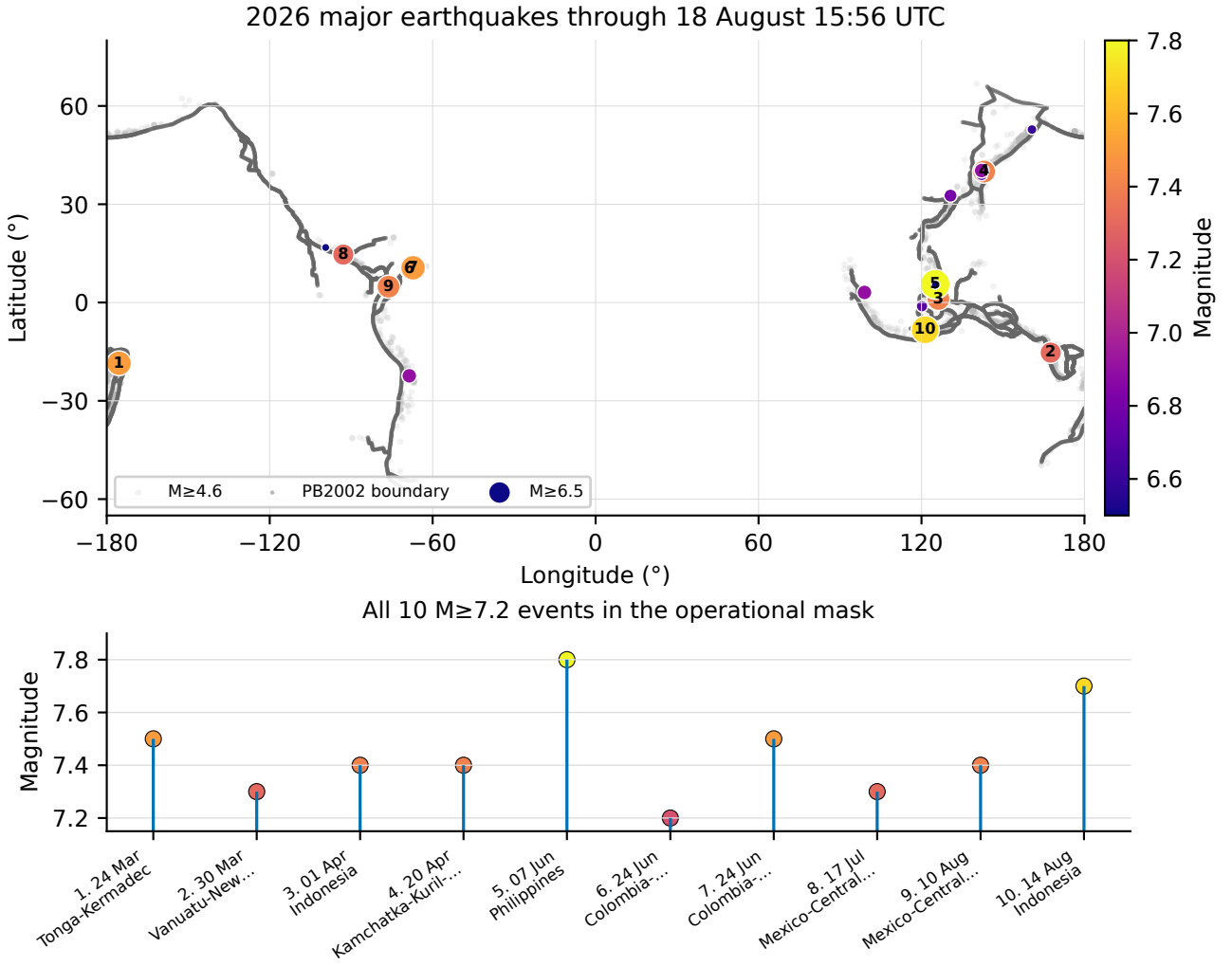


Fig. 13: Map of all $M \geq 6.5$ events (upper panel) and numbered timeline of the ten $M \geq 7.2$ earthquakes (lower panel) through the exact retrieval cutoff. Marker color encodes magnitude; numbers link mapped events to their chronological positions, and the complete USGS event table is supplied with the manuscript.

identifying local sequences, and a compact retained-event structure can be scientifically interpretable while remaining sensitive to resampling. The benchmark therefore narrows the claims that can be made from each method instead of selecting one algorithm for every purpose.

The ST-SCKM results reinforce that interpretation. Its strong adjusted Rand agreement shows that the selected complete partition is reproducible on the benchmark, yet the zero spatial penalty means that the pro-

posed neighborhood regularization did not provide the expected improvement. The method remains useful as a weighted space-time partition, but the empirical support concerns stability rather than a demonstrated advantage from spatial smoothing. Similarly, the depth and magnitude sensitivities show that additional attributes can rotate the fitted geometry away from the original target. These negative findings improve the comparison because they identify which components of the method

Table 13: Robustness across boundary buffers and magnitude thresholds.

Buffer km	M threshold	2026	Hist. mean	Ratio	z	$\hat{\epsilon}_{97.5\text{th pct.}}$
300	4.6	2543	2,406	1.057	0.202	False
300	6.5	21	20.115	1.044	0.135	False
300	7	10	6.615	1.512	1.156	False
300	7.2	10	4.346	2.301	2.747	True
500	4.6	2690	2,526	1.065	0.23	False
500	6.5	21	21.692	0.968	-0.096	False
500	7	10	7.269	1.376	0.848	False
500	7.2	10	4.808	2.08	2.331	True
700	4.6	2741	2,579	1.063	0.226	False
700	6.5	23	22.231	1.035	0.105	False
700	7	11	7.577	1.452	1.057	False
700	7.2	10	5.038	1.985	2.314	True

Note: The displayed thresholds contrast the complete-catalog result with the large-event tail. The full 30-specification grid is included as CSV.

are supported by the data and which are not.

The catalog-treatment results then connect algorithmic structure to seismic interpretation. Strong Knox statistics confirm that nearby earthquakes occur close together in time far more often than expected after temporal permutation. That result is consistent with after-shock sequences and swarms already known to be present in raw catalogs. Declustering removes different proportions of those dependent events, but the broad global rate conclusion remains similar under both alternatives. The regional Philippines elevation behaves differently: its attenuation after both declustering procedures indicates that the raw excess is dominated by sequence productivity. Thus, the analysis distinguishes the existence of local clustering from evidence for a changed independent background rate.

The 2026 comparison adds a second distinction, this time across magnitude thresholds. At $M4.6$, the observed partial-year count lies within broad historical variation and its effect size is small. At $M \geq 7.2$, the count is more than twice the matched historical mean and is above the historical 97.5th percentile for all three boundary buffers. The finite historical record makes the empirical tail probability coarse, so the effect size and percentile exceedance should be read alongside $p = 0.074$. This combination supports describing the large-event tail as elevated while avoiding a stronger claim that the entire earthquake rate is anomalous.

Spatial and temporal displays further constrain that conclusion. The large events are distributed among several Pacific-margin regions rather than forming one simple geographic chain, and the regional heatmap shows different historical baselines across strata. The lag analysis finds no directional relationship that survives correction across the planned family of region pairs and delays. Consequently, the descriptive impression of successive distant earthquakes is not matched by statistical evidence for systematic propagation at the tested scales. The appropriate synthesis is a dispersed high-magnitude tail embedded in a catalog that still contains strong local sequence dependence.

Finally, the robustness analysis identifies what would change the interpretation. A lower magnitude threshold removes the global anomaly because the much larger catalog is close to its matched historical expectation. Changing the boundary buffer does not remove the $M \geq 7.2$ percentile exceedance, so the tail result is not an artifact of the primary 500-km corridor. Declustering weakens the Philippines count elevation, showing dependence on sequence definition, while the global $M4.6$ conclusion remains stable. These contrasts make the result conditional but not arbitrary: each conclusion is tied to a clearly stated threshold, spatial definition, and catalog treatment.

6 Discussion

The analysis yields a coherent hierarchy of conclusions. Algorithmic disagreement is strongest when methods target different objects, catalog treatment matters most for regional raw-count elevations, and magnitude threshold matters most for the global 2026 assessment. These

patterns are connected: each demonstrates that an earthquake catalog does not have one method-independent clustering or anomaly label. The discussion therefore moves from method estimands, through the ST-SCKM adaptation, to the paired 2026 conclusions and their implications for reproducible reporting.

That hierarchy also prevents evidence from being transferred across analytical levels. A compact cluster does not by itself identify a tectonic source, a raw regional excess does not by itself imply a higher independent background rate, and a large-event-tail elevation does not imply that the full magnitude-complete catalog is anomalous. Each inference requires a different denominator and a different comparison set. The benchmark metrics describe fitted geometry on a common sample; the declustering comparisons describe sensitivity to sequence membership; and the matched-window analysis describes the position of the partial-2026 count within historical variation. Keeping those levels separate is the central condition for interpreting them together.

The results can therefore be read as a sequence of conditional decisions. If the objective is a stable global summary, the complete partitioning methods are defensible but spatially broad. If the objective is local sequence extraction, density-based or explicitly spatiotemporal methods are more appropriate, provided that noise rejection and fragmentation are reported. If the objective is current-year characterization, the conclusion must specify magnitude threshold, region, exposure time, and declustering status. This decision structure connects the numerical results to practical use without converting a comparative benchmark into a universal recommendation.

The same logic changes how uncertainty should be communicated. Internal validation indices quantify geometric separation under a chosen representation, but they do not measure whether a cluster corresponds to a fault system or independent physical source. Resampling stability quantifies reproducibility under event removal, but it does not guarantee that the stable grouping is optimal for every downstream use. Empirical count probabilities compare 2026 with the limited historical record, while count-model probabilities gain resolution by adding distributional assumptions. Presenting these quantities together exposes where the evidence is observational, algorithmic, or model-dependent and prevents precision in one layer from being mistaken for certainty in another.

Scale is the unifying source of these dependencies. At the basin scale, a small number of complete clusters can summarize long-lived geographic and temporal regimes, but the resulting within-cluster distances are too broad for sequence attribution. At the sequence scale, restrictive neighborhoods recover compact structures, but their parameters and noise decisions become locally consequential. At the annual-count scale, a high-magnitude tail can depart from its historical baseline even when the much larger completeness-threshold catalog remains typical. The discussion that follows evaluates the findings at their intended scales and treats cross-scale agreement as a robustness check rather than an assumption.

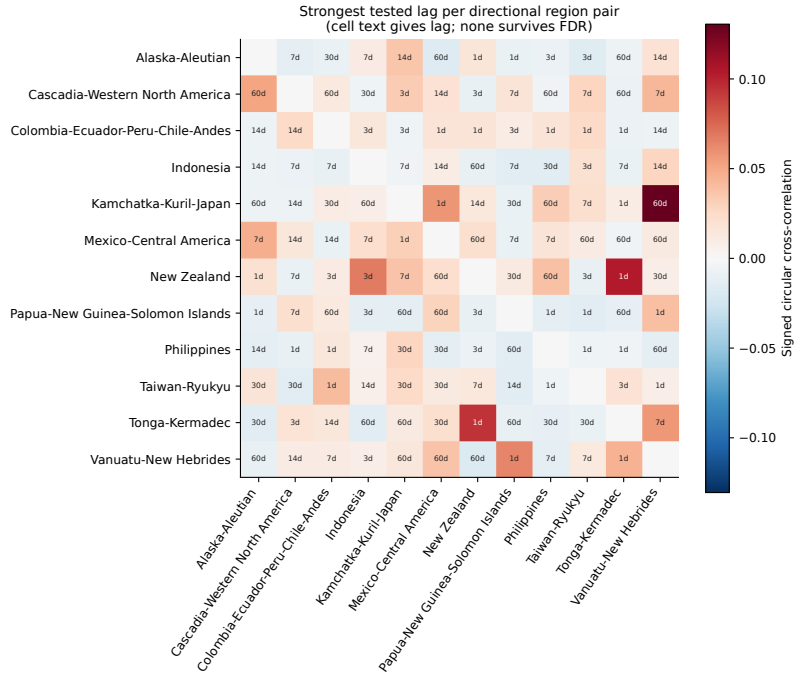


Fig. 14: Strongest tested lag for every directional region pair. Cell labels give lag in days; no direction or lag survives FDR correction.

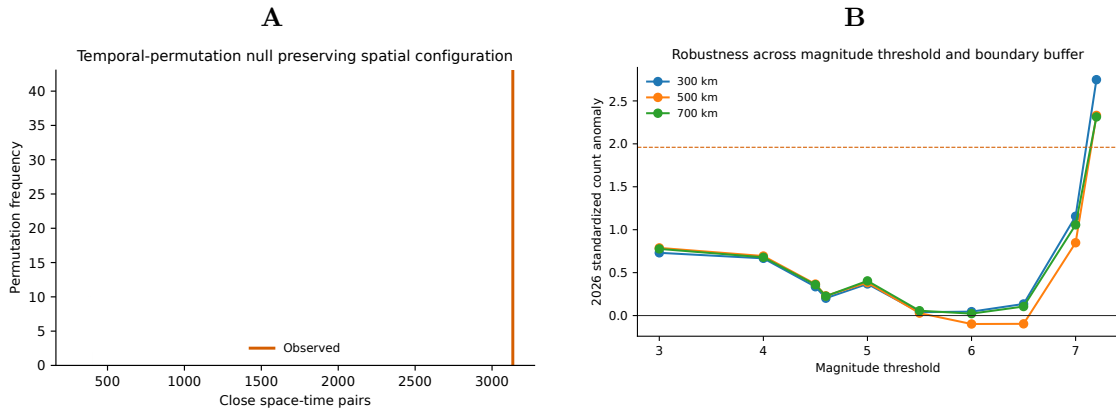


Fig. 15: Null-model and robustness diagnostics: (A) temporal-permutation distribution for the middle Knox specification and (B) global 2026 count sensitivity across magnitude thresholds and boundary buffers. Local space-time clustering is strong, while the global anomaly appears at $M \geq 7.2$ rather than at the completeness threshold.

6.1 Why the methods disagree

Method disagreement is the central benchmarking result, not a defect to be hidden by averaging ranks. K-Means, SOM, and ST-SCKM produced complete global partitions with eight to ten groups. Their high event retention makes them useful for broad catalog summaries, but it also guarantees that sparse background observations are attached to a centroid. HDBSCAN and ST-DBSCAN answered a narrower question: which events form sufficiently dense structures under their definitions? Their spatial and temporal cohesion improved precisely because 51.7–54.8% of the common sample was omitted as noise. A retained-point silhouette cannot be interpreted without that denominator.

The density methods also illustrate two different types of complexity. HDBSCAN adapted to variable density and obtained the highest retained-point silhouette, yet its 180 clusters were relatively unstable under subsampling. ST-DBSCAN’s separate kilometer and day thresholds

were physically legible, and its 16.9-day median temporal dispersion was far smaller than that of the complete partitions, but it fragmented the common sample into 759 clusters. A monitoring analyst seeking discrete local sequences might accept this fragmentation; an analyst summarizing basin-scale regimes probably would not. Method selection must follow the estimand.

Subsample stability provides a complementary perspective. ST-SCKM and K-Means were much more reproducible than HDBSCAN, but their within-cluster spatial dispersions exceeded 1,600 km and temporal median absolute deviations exceeded 1,000 days. Stability therefore establishes repeatability of a representation, not physical validity of every assigned group. The four-panel comparison in Fig. 7 makes these trade-offs visible rather than declaring a universal winner from a single index.

6.2 Interpretation of ST-SCKM

ST-SCKM produced the most stable complete partition, but its selected spatial disagreement penalty was $\lambda = 0$. That negative result is scientifically useful: under the common representation, sample, and search grid, the explicit nearest-neighbor regularizer did not improve the selection objective. The model effectively reduced to a weighted space-time centroid partition, so the study does not claim that spatial regularization was beneficial for this catalog.

Adding depth, magnitude, or both progressively reduced silhouette and spatiotemporal coherence while increasing geographic and temporal dispersion. This does not mean those variables are unimportant to earthquake physics. It means that appending them to this standardized unsupervised distance changed the geometry in a way that was less compatible with the chosen space-time objective. Depth and magnitude are better retained as cluster attributes here; a focal-mechanism or slab-structure study could justifiably define a different representation and validation target.

6.3 Two simultaneous 2026 conclusions

The 2026 results require two statements that are compatible rather than contradictory. At the M4.6 completeness threshold, the total count is only 6.5% above the matched historical mean and well inside historical variability. This conclusion is stable under the two declustering procedures. The catalog therefore does not show a broad increase in magnitude-complete earthquake occurrence through the cutoff.

At $M \geq 7.2$, however, ten events are more than twice the historical mean and exceed the historical 97.5th percentile under every tested boundary width. This directly addresses the descriptive concern that 2026 contains an unusually visible concentration of large earthquakes. The empirical $p = 0.074$ is not below 0.05 because the reference set contains only 26 comparable years and one historical year also had at least ten events. Reporting the exceedance, effect size, standardized score, and discrete empirical probability together is more informative than reducing the result to a binary “significant/not significant” label.

The threshold-specific pattern also prevents overgeneralization. A large-event tail elevation does not imply that smaller events are globally elevated, that the events share one causal chain, or that future large earthquakes are predictable. Conversely, describing the M4.6 rate as typical must not erase the unusual number of $M \geq 7.2$ events. Fig. 13 and Table 10 make the latter result explicit rather than burying it in an $M \geq 6.5$ inventory.

6.4 Declustering, regional anomalies, and lag evidence

The Philippines example shows why raw count anomalies cannot automatically be interpreted as background-rate changes. Its raw M4.6 count was 2.85 times the historical mean with a model-based FDR-adjusted probability of 0.000107, yet both declustering procedures greatly attenuated the elevation. Gardner–Knopoff reduced the stan-

dardized score from 6.78 to 0.59, while nearest-neighbor declustering reduced it to 1.48. The robust interpretation is a productive sequence in the raw catalog, not clear evidence that independent Philippine background activity changed.

The Knox and lag analyses answer different questions. Temporal permutation yielded far more close space-time $M \geq 6$ pairs than expected, confirming strong local sequence dependence. By contrast, none of 792 directional region-lag associations survived false-discovery-rate correction. Local aftershock clustering is therefore present, while the tested data provide no reliable evidence of a systematic sequence propagating across the Pacific regions. The absence of a corrected regional lag does not rule out individual dynamic-triggering cases; it limits the broader pattern claim tested here.

6.5 Implications for reproducible seismicity analysis

Three reporting practices follow from the results. First, every clustering metric should be paired with retained-event fraction and cluster count. Second, anomaly statements should specify magnitude threshold, spatial mask, time exposure, and declustering status. Third, a current-year analysis should use an exact retrieval timestamp because the catalog is still growing and reviewed solutions can change. The package implements these practices through cached query chunks, GeoJSON detail enrichment for major events, explicit configuration, full tuning grids, and machine-readable tables.

The analysis is best viewed as a reproducible statistical baseline for subsequent tectonic interpretation. Researchers can replace the global mask with fault-specific geometry, introduce focal mechanisms or plate-relative coordinates, fit regional ETAS models, or extend the historical reference period as new data arrive. Those extensions may change the scientific estimand, but they can be compared transparently because the present results expose rather than conceal their conditioning choices.

7 Limitations

The first group of limitations concerns observation. ComCat is authoritative but heterogeneous in network coverage, detection threshold, processing history, and magnitude type. We did not homogenize every magnitude to M_w , and the global M4.6 threshold remains a compromise because the implemented goodness-of-fit target was not attained. The agreement of the corrected maximum-curvature estimate across periods supports consistent processing, but it cannot establish identical completeness in every region. Consequently, small differences in low-threshold counts should not be interpreted more precisely than the catalog supports.

The second group concerns spatial and sequence definitions. PB2002 buffers approximate a diffuse Ring of Fire rather than delineating every active source, and the twelve regional strata simplify complex tectonic segmentation. Declustering is also model-dependent: Gardner–Knopoff windows and nearest-neighbor mixture separation estimate different background populations. The

nearest-neighbor implementation uses a scalable candidate approximation instead of exhaustive all-pairs search, while the clustering benchmark sample cannot preserve every small local sequence. Robustness across these choices strengthens a conclusion, but agreement does not turn an operational definition into a unique physical partition.

The third group concerns the current-year comparison. The 2026 interval ends at an exact August retrieval time, so every count is provisional and ComCat solutions may be revised. The empirical tail test has only 26 historical comparison years, making its attainable probabilities discrete and its upper quantiles sensitive to individual historical years. Count-model probabilities add resolution but introduce distributional assumptions. For this reason, the manuscript emphasizes effect sizes, matched-window percentiles, and consistency across masks rather than relying on one probability threshold.

Finally, the event metadata and statistical associations have limited physical scope. Tsunami flags were enriched from current USGS GeoJSON details for the major-event table, but a binary flag is not a tsunami-impact analysis. Regional lag correlations operate on aggregated daily counts and cannot identify a rupture-to-rupture stress-transfer pathway. No result identifies a fault, a physical triggering path, a precursor, or a future earthquake location. These limitations define the boundary between reproducible catalog characterization and claims that would require event-specific geophysical modeling.

8 Conclusions

Six clustering paradigms characterize the same Pacific catalog differently because they target different forms of structure. Density and explicit spatiotemporal methods recover compact retained-event sequences but depend strongly on noise rejection, fragmentation, and scale thresholds. Complete partitioning methods provide more stable global summaries while producing clusters that are broader in space and time. ST-SCKM belongs to the latter group: it yields the most stable complete partition in the benchmark, yet neither a positive spatial penalty nor the addition of magnitude and depth improves the selected space-time representation.

The partial-2026 results must likewise remain threshold-specific. $M \geq 4.6$ activity is consistent with matched historical variability, whereas the ten $M \geq 7.2$ events form an elevated high-magnitude tail that exceeds the historical 97.5th percentile across all tested boundary buffers. The raw Philippines elevation is not robust to either declustering procedure, and no regional lag survives multiplicity correction. Thus, the catalog supports a dispersed large-event-tail elevation without a catalog-wide rate anomaly or evidence for systematic cross-Pacific propagation.

Taken together, the analysis demonstrates the value of carrying catalog construction, method estimand, noise fraction, temporal exposure, declustering status, and multiplicity correction into the final interpretation. The resulting conclusions are narrower than a prediction claim but more reproducible and scientifically informative: they show exactly which features of the 2026 catalog are un-

usual, which are sequence-dependent, and which remain typical under the tested definitions.

Declarations

Ethics approval and consent to participate

Not applicable. This study uses publicly available earthquake catalog data and does not involve human participants, human data, or animal subjects.

Consent for publication

Not applicable.

Availability of data and materials

The earthquake data used in this study were obtained from the publicly available USGS Earthquake Catalog. The reproducibility package contains the analysis code, configuration files, processed datasets, tuning results, figures, tables, and experiment records required to reproduce the reported analyses. The package also documents the data retrieval cutoff, query parameters, random seeds, and relevant computational settings.

Competing interests

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

Not applicable.

Acknowledgements

Not applicable.

Use of artificial intelligence tools

Artificial intelligence tools were used for language proof-reading and technical assistance with LaTeX formatting. The authors take full responsibility for the content and integrity of the manuscript.

Authors' information

Not applicable.

References

- Jinmeng Bi, Cheng Song, and Yong Ma. Clustering of Earthquake Sequence and Its Effect on b Value in North China. *Pure and Applied Geophysics*, 182(3):1163–1176, 2024. doi: [10.1007/s00024-024-03588-9](https://doi.org/10.1007/s00024-024-03588-9).
- Derya Birant and Alp Kut. ST-DBSCAN: An algorithm for clustering spatial-temporal data. *Data & Knowledge Engineering*, 60(1):208–221, 2007. doi: [10.1016/j.datak.2006.01.013](https://doi.org/10.1016/j.datak.2006.01.013).
- Peter Bird. An updated digital model of plate boundaries. *Geochemistry, Geophysics, Geosystems*, 4(3):2001GC000252, 2003. doi: [10.1029/2001gc000252](https://doi.org/10.1029/2001gc000252).
- P. Bountzis, E. Lippiello, S. Baccari, and G. Petrillo. Automatic Earthquake Declustering Using the Nearest-Neighbor Distance. *Earth and Space Science*, 13(1):e2025EA004539, 2026. doi: [10.1029/2025ea004539](https://doi.org/10.1029/2025ea004539).
- Ricardo J. G. B. Campello, Davoud Moulavi, and Joerg Sander. Density-Based Clustering Based on Hierarchical Density Estimates. *Lecture Notes in Computer Science Advances in Knowledge Discovery and Data Mining*, pages 160–172, 2013. doi: [10.1007/978-3-642-37456-2_14](https://doi.org/10.1007/978-3-642-37456-2_14).
- Gabriel Candia and Jorge Macedo. A Homogenized and Declustered Earthquake Catalog for Continental Chile. *Seismological Research Letters*, 2026. doi: [10.1785/0220250449](https://doi.org/10.1785/0220250449).
- Paola Corrado, Marcus Herrmann, and Warner Marzocchi.

- Earthquake Magnitude Correlations Expose Short-Term Catalog Incompleteness. *Seismological Research Letters*, 96(2A): 909–919, 2024. doi: [10.1785/0220240277](https://doi.org/10.1785/0220240277).
- Luis Delgado, Billy Peralta, Orietta Nicolis, and Mailiu Díaz. Integrating spatio-temporal density-based clustering and neural networks for earthquake classification. *Expert Systems with Applications*, 277:127186, 2025. doi: [10.1016/j.eswa.2025.127186](https://doi.org/10.1016/j.eswa.2025.127186).
- David Essing and Piero Poli. Unraveling Earthquake Clusters Composing the 2014 Alto Tiberina Earthquake Swarm via Unsupervised Learning. *Journal of Geophysical Research: Solid Earth*, 129(1):e2022JB026237, 2024. doi: [10.1029/2022jb026237](https://doi.org/10.1029/2022jb026237).
- Stefania Gentili, Piero Brondi, Giuliana Rossi, Monica Sugan, Giuseppe Petrillo, Jiancang Zhuang, and Stefano Campanella. Seismic clusters and fluids diffusion: a lesson from the 2018 Molise (Southern Italy) earthquake sequence. *Earth, Planets and Space*, 76(1):157, 2024. doi: [10.1186/s40623-024-02096-3](https://doi.org/10.1186/s40623-024-02096-3).
- Derreck Gossett Gossett, Michael Brudzinski, Xiong Qiquan, and Jesse Hampton. Pattern of Earthquake Magnitude Clustering Based on Interevent Distance and Time. *Seismica*, 3(2), 2024. doi: [10.26443/seismica.v3i2.1094](https://doi.org/10.26443/seismica.v3i2.1094).
- B. GUTENBERG and C. F. RICHTER. Magnitude and Energy of Earthquakes. *Nature*, 176(4486):795–795, 1955. doi: [10.1038/176795a0](https://doi.org/10.1038/176795a0).
- J. A. Hartigan and M. A. Wong. Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, 28(1):100, 1979. doi: [10.2307/2346830](https://doi.org/10.2307/2346830).
- Muh Akbar Idris, Muhammad Nur Aidi, and Anik Djuraidah. Performance evaluation of ST-DBSCAN and spatio-temporal spatially constrained K-means (ST-SCKM) for wildfire risk zoning and resilience analysis. *Journal of Safety Science and Resilience*, page 100357, 2026. doi: [10.1016/j.jnlssr.2026.100357](https://doi.org/10.1016/j.jnlssr.2026.100357).
- T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464–1480, 1990. doi: [10.1109/5.58325](https://doi.org/10.1109/5.58325).
- Hisahiko Kubo, Makoto Naoi, and Masayuki Kano. Recent advances in earthquake seismology using machine learning. *Earth, Planets and Space*, 76(1):36, 2024. doi: [10.1186/s40623-024-01982-0](https://doi.org/10.1186/s40623-024-01982-0).
- Chen Li, Ziyi Li, Mengqiao Duan, and Lianqing Zhou. Spatiotemporal Distribution of the Magnitude of Completeness and b-Values in Mainland China Based on a Fused Multi-Source Earthquake Catalog. *Entropy*, 27(11):1137, 2025. doi: [10.3390/e27111137](https://doi.org/10.3390/e27111137).
- Leila Mizrahi, Shyam Nandan, and Stefan Wiemer. The Effect of Declustering on the Size Distribution of Mainshocks. *Seismological Research Letters*, 92(4):2333–2342, 2021. doi: [10.1785/0220200231](https://doi.org/10.1785/0220200231).
- Christian Molkenhain, Gert Zöller, Sebastian Hainzl, and Matthias Holschneider. Bayesian Earthquake Forecasting Using Gaussian Process Modeling: GP-ETAS Applications. *Seismological Research Letters*, 95(6):3532–3544, 2024. doi: [10.1785/0220240170](https://doi.org/10.1785/0220240170).
- S. Mostafa Mousavi and Gregory C. Beroza. Machine Learning in Earthquake Seismology. *Annual Review of Earth and Planetary Sciences*, 51(1):105–129, 2023. doi: [10.1146/annurev-earth-071822-100323](https://doi.org/10.1146/annurev-earth-071822-100323).
- Orietta Nicolis, Luis Delgado, Billy Peralta, Mailiu Díaz, and Marcello Chiodi. Space-time clustering of seismic events in Chile using ST-DBSCAN-EV algorithm. *Environmental and Ecological Statistics*, 31(2):509–536, 2024. doi: [10.1007/s10651-023-00594-3](https://doi.org/10.1007/s10651-023-00594-3).
- Yoshihiko Ogata. Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical Association*, 83(401):9–27, 1988. doi: [10.1080/01621459.1988.10478560](https://doi.org/10.1080/01621459.1988.10478560).
- Muthiara Hazimah Putri and Devni Prima Sari. Earthquake Point Clustering in Sumatra Island using Spatio-Temporal Density-Based Spatial Clustering Application with Noise (ST-DBSCAN) Algorithm. *Mathematical Journal of Modelling and Forecasting*, 3(1):8–15, 2025. doi: [10.24036/mjmf.v3i1.33](https://doi.org/10.24036/mjmf.v3i1.33).
- Robert Shcherbakov and Sidhanth Kothari. Earthquake Declustering Using Supervised Machine Learning. *Bulletin of the Seismological Society of America*, 115(4):1627–1641, 2025. doi: [10.1785/0120250004](https://doi.org/10.1785/0120250004).
- Dwi Agustin Nuriani Sirodj, Muhammad Nur Aidi, Bagus Sartono, Utami Dyah Syafitri, and Bayu Pranata. Identification of Earthquake Prone Zones in Sumatra using Density Based Spatial Clustering of Applications with Noise. *CAUCHY: Jurnal Matematika Murni dan Aplikasi*, 10(2): 916–928, 2025. doi: [10.18860/cauchy.v10i2.36120](https://doi.org/10.18860/cauchy.v10i2.36120).
- Samuel Stockman, Daniel J. Lawson, and Maximilian J. Werner. SB-ETAS: using simulation based inference for scalable, likelihood-free inference for the ETAS model of earthquake occurrences. *Statistics and Computing*, 34(5):174, 2024. doi: [10.1007/s11222-024-10486-6](https://doi.org/10.1007/s11222-024-10486-6).
- Mohammad Talebi, Mehdi Zare, and Antonella Peresan. Quantifying the features of earthquake clusters in north-central Iran, based on nearest-neighbor distances and network analysis. *Physics of the Earth and Planetary Interiors*, 353:107215, 2024. doi: [10.1016/j.pepi.2024.107215](https://doi.org/10.1016/j.pepi.2024.107215).
- Matteo Taroni and Aybige Akinci. Good practices in PSHA: declustering, b-value estimation, foreshocks and aftershocks inclusion; a case study in Italy. *Geophysical Journal International*, 224(2):1174–1187, 2020. doi: [10.1093/gji/ggaa462](https://doi.org/10.1093/gji/ggaa462).
- Daniel T. Trugman. A High-Precision Earthquake Catalog for Nevada. *Seismological Research Letters*, 95(6):3737–3745, 2024. doi: [10.1785/0220240106](https://doi.org/10.1785/0220240106).
- Rahul Kumar Vijay, Satyasai Jagannath Nanda, and Ashish Sharma. A review on clustering algorithms for spatiotemporal seismicity analysis. *Artificial Intelligence Review*, 58(8):231, 2025. doi: [10.1007/s10462-025-11229-3](https://doi.org/10.1007/s10462-025-11229-3).
- Xinyi Wang, Jiawei Li, Ao Feng, and Didier Sornette. Estimating Magnitude Completeness in Earthquake Catalogs: A Comparative Study of Catalog-Based Methods. *Journal of Geophysical Research: Solid Earth*, 130(9):e2025JB031441, 2025. doi: [10.1029/2025jb031441](https://doi.org/10.1029/2025jb031441).
- J. Woessner. 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. doi: [10.1785/0120040007](https://doi.org/10.1785/0120040007).
- Ilya Zaliapin and Yehuda Ben-Zion. Earthquake Declustering Using the Nearest-Neighbor Approach in Space-Time-Magnitude Domain. *Journal of Geophysical Research: Solid Earth*, 125(4):e2018JB017120, 2020. doi: [10.1029/2018jb017120](https://doi.org/10.1029/2018jb017120).
- Ilya Zaliapin, Andrei Gabrielov, Vladimir Keilis-Borok, and Henry Wong. Clustering Analysis of Seismicity and Aftershock Identification. *Physical Review Letters*, 101(1):018501, 2008. doi: [10.1103/physrevlett.101.018501](https://doi.org/10.1103/physrevlett.101.018501).
- Chengxiang Zhan, Jiancang Zhuang, and Stephen Wu. Auto-BETA: Fully Automated Earthquake Catalog Replenishment Tool. *Seismological Research Letters*, 2026. doi: [10.1785/0220250253](https://doi.org/10.1785/0220250253).